



UNIVERSITAT_{DE}
BARCELONA

Development of DNA methylation biomarkers to study epigenetic inheritance and integration of environmental stimuli

Núria Sánchez Baizán



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – SenseObraDerivada 4.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – SinObraDerivada 4.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-NoDerivs 4.0. Spain License.**

Development of DNA methylation biomarkers to study epigenetic inheritance and integration of environmental stimuli

Núria Sánchez Baizán

Ph.D. Thesis 2023



Barcelona, 2023

*A la meva família, amb especial
estima als meus pares i avis*

Design of the main and back covers: Joan Mir Arguimbau

Universitat de Barcelona

Tesi Doctoral

Development of DNA methylation biomarkers to study epigenetic inheritance and integration of environmental stimuli

Memòria presentada per Núria Sánchez Baizán
per optar al grau de Doctora per la Universitat de Barcelona

Programa de Doctorat “Genètica”

Tesi desenvolupada al Grup de Biologia de la Reproducció (GBR), Departament
de Recursos Marins Renovables, Institut de Ciències del Mar (ICM-CSIC),
Barcelona

Director: Dr. Francesc Piferrer Circuns

Doctoranda: Núria Sánchez Baizán

Tutor: Dr. Florenci Serras Rigalt

Barcelona, juliol de 2023

Preface

This Ph.D. thesis was carried out at the Group of Biology of Reproduction (GBR) of the Department of Renewable Marine Resources, Institute of Marine Sciences (ICM-CSIC) in Barcelona, under the supervision of Dr. Francesc Piferrer and at the Ph.D. Genetics Program of the Universitat de Barcelona during the years 2018 to 2023. The aim of this thesis was to contribute to the understanding of the genetic and environmental factors on sex development in sea bass (*Dicentrarchus labrax*).

The thesis is structured as follows:

Block A: Transcriptomics during sex differentiation

Chapter 1.

Improved biomarker discovery through a plot twist in transcriptomic data analysis

Chapter 2.

Extraordinary variability in gene expression dynamics during gonadal sex differentiation across vertebrates

Block B: Epigenetic biomarkers and the effect of sex and temperature

Chapter 3.

Development of epigenetic biomarkers with diagnostic and prognostic value to assess the lasting effects of early temperature changes in farmed fish

Chapter 4.

A cost-effective method for complementing genetic selection with epigenetic selection in fish production

Chapter 5.

Multigenerational temperature additive effects: the epigenetic regulation of masculinization and its inheritance in the European sea bass

Each chapter corresponds to an article either published, submitted or soon to be submitted:

Sánchez-Baizán, N., Ribas, L. and Piferrer, F., 2022. Improved biomarker discovery through a plot twist in transcriptomic data analysis. *BMC biology*, 20(1), pp.1-26. doi.org/10.1186/s12915-022-01398-w

Sánchez-Baizán, N., Jarne-Sanz, I., Rocco, A., Scharl, M., Piferrer, F., 2023. Extraordinary variability in gene expression dynamics during gonadal sex differentiation across vertebrates. *Genes and Development*. Submitted.

Valdivieso, A., Sánchez-Baizán, N., Mitrizakis, N., Papandroulakis, N. and Piferrer, F., 2023. Development of epigenetic biomarkers with diagnostic and prognostic value to assess the lasting effects of early temperature changes in farmed fish. *Aquaculture*, 563, p.738918. doi.org/10.1016/j.aquaculture.2022.738918.

Sánchez-Baizán, N., Herlin, M., López, M. and Piferrer F. A cost-effective method for epigenetic biomarker development towards integration of genetic and epigenetic selection in fish production. *In preparation*.

Sánchez-Baizán, N., Allal, F., Denkena, J., Colomé-Tatché, M., Vandeputte, M. and Piferrer, F. Multigenerational temperature additive effects: epigenetic regulation of masculinization and its inheritance in the European sea bass. *In preparation*.

During the realization of this thesis, Núria Sánchez Baizán also contributed to the following articles:

Ribas, L., Crespo, B., Sánchez-Baizán, N., Xavier, D., Kuhl, H., Rodriguez, J.M., Diaz, N., Boltana, S., MacKenzie, S., Moran, F. and Zanuy, S., 2019. Characterization of the European sea bass (*Dicentrarchus labrax*) gonadal transcriptome during sexual development. *Marine Biotechnology*, 21, pp.359-373. doi.org/10.1007/s10126-019-09886-x. Note: this paper is related to chapter 1 and 2.

Piferrer, F., Anastasiadi, D., Valdivieso, A., Sánchez-Baizán, N., Moraleda-Prados, J., Ribas, L., 2019. The Model of the Conserved Epigenetic Regulation of Sex. *Frontiers in Genetics*, 10, 857. doi:10.3389/fgene.2019.00857.

Geffroy, B., Besson, M., Sánchez-Baizán, N., Clota, F., Goikoetxea, A., Sadoul, B., Ruelle, F., Blanc, M.O., Parrinello, H., Hermet, S., Blondeau-Bidet, E., Pratlong, M., Piferrer, F., Vandeputte, M., Allal, F. 2021. Unraveling the genotype by environment interaction in a thermosensitive fish with a polygenic sex determination system. *Proceedings of the National Academy of Sciences*, 118(50), e2112660118. doi.org/10.1073/pnas.2112660118.

Note: this paper is related to chapter 4 and 5

The results obtained in this thesis were presented at the following scientific meetings:

Sánchez-Baizán, N. and Piferrer, F. *Effects of sex and temperature in the European sea bass epigenome*. 12th International Symposium on Reproductive Physiology of Fish. Crete, Greece. May 2023.

Sánchez-Baizán, N., Beato, S., François, A., Vandeputte, M., Piferrer, F. *Methodological considerations to evaluate epigenetic heritability in the European sea bass epigenome*. Society for Molecular Biology and Evolution. GS9 on Epigenomics in Evolution. Online. February 2023.

Sánchez-Baizán, N., Denkena, J., Colomé-Tatché, M., Vandeputte, M., Allal, F., Piferrer, F. *Identification of European sea bass neomales through epigenomic analysis*. EPIgenetics in MARine biology congress (EPIMAR). Massachusetts, USA, October 2022.

Sánchez-Baizán, N., Ribas, L., Piferrer, F. *Increased biomarker discovery through a twist on the analysis of transcriptomic data during sexual development in the European sea bass, mouse, and humans*. 30th CECE & 9th ISFE Joint Conference, Faro, Portugal. September 2022.

Sánchez-Baizán, N., Denkena, J., Allal, F., Colomé-Tatché, M., Vandeputte, M., Piferrer, F. *Development of epigenetic biomarkers for the European sea bass (*Dicentrarchus labrax*)*. 6th Genomics in Aquaculture Symposium (GIA2022), Granada, Spain. May 2022.

Sánchez-Baizán, N., Jarne-Sanz, I., Piferrer, F. *Gonadal transcriptome dynamics during sex differentiation across vertebrates*. XIII Congress of the Iberian Association of Comparative Endocrinology (AIEC). Online. September 2021.

Sánchez-Baizán, N., Vandeputte, M., Allal, F., Piferrer, F. *Epigenetic Multigenerational epigenetic effect of masculinizing temperatures in the European sea bass (*Dicentrarchus labrax*)*. 6th Genomics in Aquaculture Symposium (GIA2020). Online. May 2021.

Sánchez-Baizán, N., Anastasiadi, D., Vandeputte, M., Allal, F., Piferrer, F. *Epigenetic inheritance and integration of masculinizing temperatures of the European sea bass (*D. labrax*)*. Epigenetic Inheritance: Impact for Biology and Society. University of Zurich, August 2019. Zurich, Switzerland.

Sánchez-Baizán, N., Ribas, L., and Piferrer, F. *Weighted correlation network analysis and transcriptome dynamics during early gonadal differentiation of the European sea bass (*Dicentrarchus labrax*)*. NGS Conference. International Society for Computational Biology and Bioinformatics. Barcelona, Spain. April 2018.

The research presented here was supported by AGL2016-78710-R (EPIMARK) and PID2019-108888RB-I00 (EPIPURE) grants from the Ministerio de Ciencia e Innovación, both awarded to Dr. Francesc Piferrer. Also with funding from the Spanish government through the ‘Severo Ochoa Centre of Excellence’ accreditation (CEX2019-000928-S).

Núria Sánchez Baizán was supported by a scholarship (BES-2017-079744) from the Government of Spain (August 2018–December 2022) and a contract linked to EPIPURE project (PID2019-108888RB-I00). (March 2023 – present).

During the course of this research, Núria Sánchez did a stay of three months during 2022 at the Institute of Computational Biology, Helmholtz Zentrum München from the German Research Center for Environmental Health and at the Biomedical Center (BMC) from the Ludwig Maximilian University of Munich, Munich, Germany, under the supervision of Dr. Maria Colomé-Tatché. Travelling and living expenses were supported by the Spanish grant: BES-2017-07974.

Agraïments / Acknowledgments

El meu somni sempre ha estat poder dedicar-me a la ciència. Aquesta tesis representa el primer pas per poder-ho fer i això no hauria estat possible sense el suport, l'ajuda i la confiança de moltes persones a les quals els estaré sempre molt i molt agraïda.

En primer lloc, vull agrair profundament i posar en valor tota la feina i suport del meu director Dr. Francesc Piferrer. Sense ell aquesta tesis no hauria estat possible, no només per la seva tasca essencial de direcció i suport econòmics sinó, sobretot, per la seva confiança en mi, des del primer dia. Una confiança que a vegades fins i tot va ser més que la meua pròpia. Li estic agraïda per transmetre la seva passió per la nostra feina, per emprendre els reptes amb energia i il·lusió i per ensenyar-nos a no rendir-nos mai. La seva curiositat incansable es contagia i per això agraeixo també que m'hagi ensenyat a qüestionar-m'ho tot per poder-ho entendre. Per tot això, malgrat que en alguns moments hagi estat difícil, aquests han estat superats en escriure pels moments en que hem gaudit junts fent la feina que més ens agrada. D'ell m'emporto per sempre molt records, eines, consells, trucs i maneres de fer. Moltíssimes gràcies de tot cor!

En segundo lugar, esta tesis también ha sido posible gracias a miembros presentes y pasados del “Grup de Biologia de la Reproducció”. Me gustaría hacer especial mención de Dafni Anastasiadi, no solo por todo lo que me enseñó durante el tiempo que compartimos juntas en el laboratorio sino también porque su excelente e inspirador trabajo. Gracias por tu valentía en atreverte a trabajar en ámbitos como la bioinformática y el aprendizaje automático y hacernos ver que no son mundos inaccesibles y por lo tanto que me atreviera a adentrarme en ellos. También quiero agradecer a Alejandro Valdivieso por las horas de discusión científica sobre resultados y figuras, por compartir su conocimiento en el laboratorio y en el análisis. No podría dejar de agradecer a Susanna Pla, quien me acogió desde el principio y con quien nos apoyamos mutuamente en los momentos complicados y de quien jamás olvidaré su ejemplo de fuerza de voluntad. Finalmente, mi querida amiga y compañera Silvia Beato, siempre dispuesta a ayudar, con quien hemos compartido horas de laboratorio, despacho y viajes. Muchísimas gracias por tu apoyo en todo. Also thank you very much to Umar Farouk. I appreciate that he is always there to show the positive side of things. Además, me gustaría agradecer a todos los estudiantes de grado y master a quienes he tenido el placer de enseñar pero que también me han enseñado muchísimo: Emmanouil Diakos, Ignasi Jarne, Emma Feliu y Maria Ortega. Voldria agrair a la Silvia Joly per ensenyar-me els protocols de laboratori, pel seu suport i per la seva feina com a responsable del laboratori de biologia molecular.

I would like to thank Maria Colomé-Tatché and all the members of the Computational Epigenetics group for welcoming me and teaching me a whole new approach to see research in biology from a computational perspective. Thanks a lot Maria for having me in your group. It was a very nice experience even though the circumstances of the Covid-19 pandemic. Particularly, I would like to thank Anna Danese and Johana Denkena for their patience in introducing me to coding and for inviting me to their home to do home-office while covid was still an issue and for the amazing adventure that it was skiing in the Alps with you! I will always remember my stay in Munich.

Also important contributors to this thesis were Alicia Felip and Ana Gómez from Torre de la Sal (IATS-CSIC), Marc Vandeputte and François Allal from Ifremer in Montpellier, France. Ramji Bhandari from the University of Greensboro, North Carolina, USA and Tapas Chackraborty from the National University Corportaion Kyushu University, Japan. I express deep gratitude for sharing their expertise and for the nice discussion and support.

Vull agrair a tots els companys del departament de Recursos Marins Renovables del ICM per fer que l'ambient de treball sigui tant agradable i que sigui un espai on conèixer persones amb inquietuds i interessos relacionats amb el mar d'una manera tan diversa. Gràcies a: Alba, Alberto, Amalia, Anna, Anabel, Batis, Cristina, Elena Ll., Elena F., Elvira, Eve, Fernando, Gemma, Guiomar, Joan M., Joan N., Joan S., Jordi, Laia, Marc, Maria, Marta B., Marta C., Marta P., Mariona, Mireia, Morane, Mercedes, Montse, Nixon, Noelia, Ricardo, Roger, Tosca, Valerio, Vanesa, Víctor i Xènia.

No podria deixar de mencionar que estic immensament agraïda a totes les amistats que des de prop o des de la distància sempre heu estat allà. Gràcies Maria, per compartir aquesta experiència de primera mà, per escoltar-me amb paciència i donar-me bons consells. A les amistats de Granollers de tota la vida, en especial a la Marina i la Gemma. També a les amistats i al director del Cor de Cambra de Granollers que m'han acompanyat en aquesta llarga aventura. To Sandra Rybicki, my dearest best friend who is always there no matter what country we are at, she always finds the way to be close and supportive. Thank you for being the pillar of strength and support that I can always rely on.

Finalment, vull agrair amb tot el meu cor a la meva gran família, que ho és en número però també per tot allò que m'han donat. I és que sento que tot allò que m'atreveixo a fer a la vida, com ara aquesta tesis, és gràcies a tots i cada un d'ells. Per proporcionar-me una xarxa de seguretat que em fa ser més forta i valenta del que seria sense ells. Als meus avis i als meus pares, en especial, perquè m'han transmès la cultura de l'esforç, la curiositat per la ciència i els viatges. A la meva germana Laura que és i serà sempre una roca on agafar-me. I per acabar, moltíssimes gràcies al David el meu company de viatge per motivar-me i animar-me, sobretot quan més ho necessitava. El seu suport incondicional i confiança han estat un far durant aquest trajecte. Gràcies per creure sempre en mi, per ajudar-me a gestionar els nervis, per comprendre'm i acompanyar-me, sempre.

Voldria recalcar que em sento molt afortunada de tenir aquest ventall de persones al meu voltant i que això és el que ha fet possible que m'atrevis, continués i acabés aquesta tesis. Tot i que sóc conscient i ho valoro en el meu dia a dia, durant el desenvolupament d'aquesta tesi, enmig de la pandèmia i l'aïllament que aquesta va implicar, això es va fer més evident. Vam experimentar com els congressos i les reunions on-line no donaven els mateixos fruits. Per tant, com a aprenentatge de l'experiència que ens va tocar viure, m'emporto el recordatori de que el contacte social amb les persones és essencial tant per als éssers humans com per a la ciència. Moltíssimes gràcies a tots els que ho heu fet possible!

Resum

Els mecanismes epigenètics juguen un paper vital en la regulació de l'expressió gènica i en la determinació del fenotip. Aquesta tesi tenia com a objectiu investigar el transcriptoma, el metiloma i la influència de factors intrínsecs (com el sexe) i factors ambientals (com la temperatura) en el llobarro (*Dicentrarchus labrax*). En primer lloc, ens vam centrar en comprendre l'expressió gènica de les gònades durant els primers estadis de la diferenciació sexual. Mitjançant això, proposem una nova aproximació metodològica que combina l'anàlisi de xarxes de co-expressió gènica ponderada (WGCNA) amb l'anàlisi de l'expressió gènica diferencial (DEGs) per a una millor anàlisi transcriptòmica. A més, vam explorar la diferenciació sexual gonadal en diferents espècies de vertebrats, identificant gens clau i possibles marcadors nous. A continuació, ens vam centrar en el desenvolupament de biomarcadors epigenètics. Com a resultat, es van identificar marques de metilació del DNA com a indicadors de l'entorn tèrmic experimentat durant el desenvolupament primerenc. A més, es va desenvolupar un mètode amb baix cost i alta efectivitat per al desenvolupament de biomarcadors epigenètics i la seva integració amb la selecció genètica en l'aqüicultura. Finalment, vam explorar l'impacte del sexe i la temperatura en el metiloma, revelant efectes multigeneracionals i marcadors per a femelles genèticament resistents a la masculinització induïda per la temperatura. Aquests resultats tenen implicacions per a l'aqüicultura i contribueixen a la nostra comprensió dels mecanismes adaptatius i de la herència epigenètica.

Summary

Epigenetic mechanisms play a vital role in gene expression regulation and phenotype determination. This thesis aimed to comprehensively investigate the transcriptome, the methylome and the influence of intrinsic (such as sex) and environmental factors (such as temperature) on the European sea bass (*Dicentrarchus labrax*). First, we focused on understanding the genetic expression of the gonads during early sex differentiation. By doing so, we proposed a novel approach combining weighted gene co-expression network analysis (WGCNA) with differential gene expression analysis (DEGs) for improved transcriptomic analysis. Further, we also explored gonadal sex differentiation across different vertebrate species, identifying key genes and potential novel markers. Then, we focused on the development of epigenetic biomarkers. As a result, DNA methylation marks were identified as indicators of the thermal environment experienced during early development. Additionally, a cost-effective method for epigenetic biomarker development and its integration with genetic selection in aquaculture was developed. Lastly, the impact of sex and temperature on the methylome was explored, revealing multigenerational effects and markers for genetic resistant females to temperature-induced masculinization. These findings have implications for aquaculture and contribute to our understanding of adaptive mechanisms and epigenetic inheritance.

Table of contents

Preface	iii
Agraïments / Acknowledgments	vii
Resum / Summary	ix
General introduction	1
1. Gonadal development in vertebrates	3
1.1 Sex determination	3
1.1.1 Sex determining genes	5
1.1.2. Polygenic sex determination	6
1.1.3 Environmental sex determination	7
1.2 Sex differentiation	8
1.3 Sex development: from a hierarchical cascade to networks	10
1.4 Activation and repression programs	11
2. Epigenetics, a bridge between genomic and environmental information	12
2.1 Epigenetic mechanisms	12
2.1.1 DNA methylation and methods to assess DNA methylation	13
2.2 Epigenetics in cell identity acquisition	14
2.3 Environmental influence and epigenetics	16
2.4 Epigenetic inheritance	17
2.4.1 Types of epigenetic inheritance	17
2.4.2 Epigenetic reprogramming	18
3. Epigenetic biomarkers	19
3.1 Development of epigenetic biomarkers for aquatic organisms	19
3.2 Examples of epigenetic biomarkers	21
4. The European sea bass model	21
4.1 Sea bass biology	21
4.2 Sexual development of the sea bass	21
4.3 Environmental factors influencing sea bass sexual phenotype	22
4.4 Problems when reared in captivity	22
Aims	25
Report of the supervisor	29
Results	33

Block 1: Transcriptomics during sex differentiation 35

Chapter 1: Improved biomarker discovery through a plot twist in transcriptomic data analysis 37

Abstract	39
1. Introduction	40
2. Results	44
3. Discussion	56
4. Conclusions	60
5. Materials and Methods	60
Supplementary Information.....	66
Acknowledgements	79

Chapter 2: Extraordinary variability in gene activation and repression programs during gonadal sex differentiation across vertebrates 81

Abstract	83
1. Introduction	84
2. Materials and methods	85
3. Results	88
4. Discussion	97
Supplementary Information.....	102
Acknowledgements	107

Block 2: Epigenetic biomarkers and the effect of sex and temperature 109

Chapter 3: Development of epigenetic biomarkers with diagnostic and prognostic value to assess the lasting effects of early temperature changes in farmed fish 111

Abstract	113
1. Introduction	114
2. Materials and methods	115
3. Results	120
4. Discussion	132
Supplementary Information.....	137
Acknowledgements	144

Chapter 4: A cost-effective method for epigenetic biomarker development and their integration with genetic selection in finfish aquaculture 147

Abstract	149
1. Introduction	150

2. Materials and methods	152
3. Results	158
4. Discussion	172
Supplementary Information.....	177
Acknowledgements	184
 <i>Chapter 5: Multigenerational temperature additive effects: the epigenetic regulation of masculinization and its inheritance in the European sea bass</i>	185
Abstract	187
1. Introduction	188
2. Materials and Methods	190
3. Results	194
4. Discussion	207
Supplementary Information.....	215
Acknowledgements	232
 <i>General discussion</i>	233
 <i>Conclusions</i>	247
 <i>References</i>	251
 <i>Annex</i>	287

General introduction

1. Gonadal development in vertebrates

Sexual reproduction is the predominant mode in vertebrates, although some exceptions exist and can occur through various types of sexual systems, ranging from gonochorism, where individuals have fixed and separate sexes, to simultaneous hermaphroditism, where an individual can produce both male and female gametes at the same time (Ah-King and Nylin, 2010; Pla, 2019). To produce different sexes, sexual development is the complex process encompassing sex determination and sex differentiation (Piferrer and Guiguen, 2008).

It could be said that sexual development starts during early embryonic development (blastula), when some of the epiblasts (first pluripotent primary lineage that differentiates after fertilization) differentiate into primordial germ cells (PGCs). PGCs are the early developmental predecessors of eggs and sperm, which carry important genetic and epigenetic information to future generations (Johnson and Alberio, 2015). Once PGCs are differentiated, they migrate to the region where the gonadal crests will form. When the gonadal crests region is colonized, the PGCs give rise to the undifferentiated gonad (Hansen and Pelegri, 2021). Then, the germ cells along with interstitial and supporting cells start male or female differentiation based on the genetic and environmental inputs (DeFalco and Capel, 2009). PGCs develop into spermatogonia in males and oogonia in females. Thus, testis is composed of three main cell types: Sertoli, Leydig cells and germ cells at different stages of maturation, from spermatogonia to spermatozoa. The Sertoli cells and spermatogonia reside in seminiferous tubules where spermatogenesis occurs. Conversely, the ovary is composed of germ and supporting cells. Those germ cells are oogonia that become oocytes and are released as eggs. Such oocytes are surrounded by granulosa and theca cells in follicles (Lubzens *et al.*, 2010; Nagahama, 1983).

The underlying molecular mechanisms by which sex differences are achieved have been studied during at least the last three decades. And, as result, multitude of genes, factors and mechanisms have been identified that play crucial roles in the gonads, reflecting the complexity of these developmental processes in vertebrates. However, the conundrum of genes controlling sex determination and differentiation are still not fully understood. In the following sections, we will provide a summary on the variability of mechanisms of sex determination and differentiation found across the five major groups of vertebrates.

1.1 Sex determination

Sex determination (SD) is the biological process that establishes an individual's sex (Bull, 1983). Among vertebrates, a diverse array of SD mechanisms exists, ranging from genetic SD (GSD) to environmental SD (ESD) (Devlin and Nagahama, 2002). GSD is the mechanism by which sex is established based on genetic factors at fertilization or thereafter. The genetic factors that determine sex can range from a single nucleotide polymorphism (SNP), a duplicated gene, a master gene or a major factor combined with one or more secondary factors (known as a multifactorial system), or a group of genes that act together (polygenic system, PSD). GSD can be further divided into Chromosomal SD (CSD) and PSD if the genetic factors are located in sex chromosomes or sex chromosomes are not present and thus they are distributed in the autosomes, respectively. In CSD the sex is determined based on the combination of heterogametic sexual chromosomes inherited from the gametes of the parents. In PSD, multiple sex-determining loci are found on different autosomal chromosomes (Moore and Roberts, 2013). In vertebrates, GSD is the dominant mechanism, ranging from fish to mammals (**Figure 1**), and typically results in an equal proportion of males and females (1:1), as outlined by Fisher's principle, due to natural selection pressures (Fisher, 1930).

Most mammals have XX/XY sex determining system, where the presence of the heteromorphic sex chromosome Y determines the development of male gonads (Koopman *et al.*, 1991). This chromosome contains the sex determining region (*Sry*) gene, which begins the process of male development. The XY system evolved between 166 and 148 million years ago and has remained stable in most mammals, although some exceptions are known, such as monotremes, where sex determination is not completely understood (Capel, 2017). Similarly, birds have a ZZ/ZW chromosomal system, however in this system, the female is the heterogametic sex (*i.e.*, females have two different sex chromosomes, ZW, and males have two of the same sex chromosomes, ZZ). In this group, the establishment of the male fate is controlled by a gene on the Z chromosome; the Doublesex and mab-3 related transcription factor 1 (*DMRT1*) (Smith *et al.*, 2009).

In ESD, sex is established by external factors (Bull, 1983) during early sensitive periods after fertilization (Penman and Piferrer, 2008). This mechanism is not present in birds or mammals, but it is found in some reptiles and gonochoristic fish, providing them with high sexual plasticity (Mank *et al.*, 2006). The factors influencing sex determination are abiotic (Piferrer, 2018a).

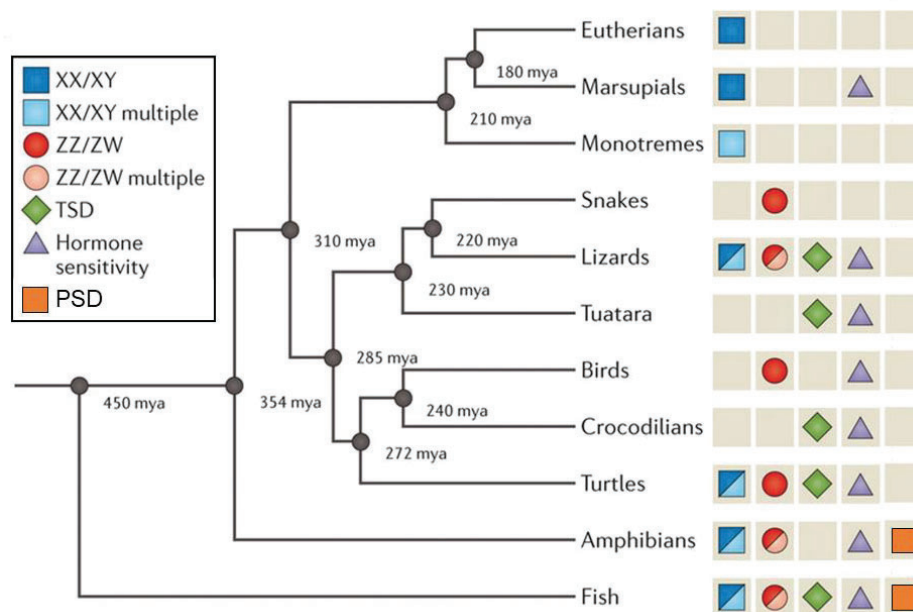


Figure 1. Phylogenetic tree showing the diversity of sex determination systems in vertebrates. Abbreviations: TSD: Temperature sex determination. PSD: Polygenic sex determination. Figure modified from Capel (2017).

Those environmental factors affecting the sexual phenotype are basically temperature and pH (Baroiller *et al.*, 2009; Ospina-Alvarez and Piferrer, 2008), with temperature sex determination (TSD) being the most common factor in ESD, although other factors such as photoperiod (Villamizar *et al.*, 2012) or density (Ribas *et al.*, 2017a) have also been revealed. In TSD, the temperature during embryonic development determines the sex of the offspring. Both in reptiles and amphibian examples of GSD and ESD can be found, however, fish are by far the group with more variable sex determining systems, including GSD with CSD and PSD, as well as ESD, notably with TSD. The SD systems can vary even among closely related species, as in medaka species like *Oryzias latipes*, which has an XX/XY sex determining system and *O. hubbsi*, which has a ZZ/ZW sex determining system (Takehana *et al.*, 2008, 2007).

In the classical view on sex determination, GSD and ESD were seen as mutually exclusive mechanisms. However, the current understanding recognizes them as two endpoints of a spectrum, where various combinations of GSD and ESD are feasible (Sarre *et al.*, 2004). This means that in some species, sex determination is dependent on a combination of genetic and environmental factors. Additionally, it has been proposed that three sources of phenotypic variation may significantly contribute to sex determination, instead of just two: genes, environment, and random factors (Perrin, 2016).

1.1.1 Sex determining genes

As described above, in CSD, a master sex-determining loci can be responsible to switch on the machinery towards male or female development. To date, more than 20 SD genes have been identified in vertebrates (**Table 1**). Such SD genes primarily arise through one of two evolutionarily conserved processes. The first process involves gene duplication and insertion (Feron *et al.*, 2020; Nanda *et al.*, 2002). The second process involves allelic diversification of an already existing locus. Allelic diversification, consists in an autosomal gene that acquires a male or female sex-promoting function. This mechanism ensures the continued presence and dominance of the male-specifying allele, establishing it as a male-specific and dominant sex-determining gene (Pan *et al.*, 2016). Such diversification includes the variation of the whole gene, a fragment or a single nucleotide polymorphism (SNP) (Kamiya *et al.*, 2012). For example, a SNP found in the anti-Müllerian hormone receptor type 2 (*amhr2*) in Japanese pufferfish (*Takifugu rubripes*) (Kamiya *et al.*, 2012) and in the Hydroxysteroid 17-Beta Dehydrogenase 1 (*hsd17b1*) gene in the Yellowtail amberjack (*Seriola dorsalis*) (Koyama *et al.*, 2019).

Other than variation within genes, fragments or SNPs, recently, it was found that SD genes can also arise from variation induced by regulatory elements such as transposable elements (so-called “jumping genes”). In the sablefish, for example, allelic diversification of the Gonadal soma-derived factor (*gsdf*) gene was found to occur through transposable element exaptation (Herpin *et al.*, 2021). Transposable element exaptation refers to the process by which a transposable element, a segment of DNA capable of moving or copying itself within a genome, acquires a new functional role contributing to the evolution of existing functions.

However, most of the SD genes evolved from genes that already had essential functions in sexual development (Nagahama *et al.*, 2021), such as transcription factors (*Dmrt1* or SRY-Box Transcription Factor 3, *Sox3*) or genes from the TGF- β signaling pathway (Anti-Müllerian Hormone, *Amh*; Anti-Müllerian Hormone Receptor Type 2, *Amhr2*; Bone Morphogenetic Protein Receptor Type 1B, *Bmpr1b*; *Gsdf* and Growth Differentiation Factor 6, *Gdf6*) (Pan *et al.*, 2021b). The only exception so far is the conserved salmonid sex-determining gene, sexually dimorphic on the Y-chromosome (*sdY*), which evolved from a gene involved in regulating the immune response. *sdY* is part of a jumping sex locus that has been transposed into different ancestral autosomes during the evolution of salmonids (Bertho *et al.*, 2021). Nowadays, with the development of new long-read sequencing technologies, such as Nanopore and PacBio, allowing to produce high-contiguity chromosome-level genome assemblies, the discovery of SD genes has and is expected to continue increasing in the future (Zheng *et al.*, 2022). The discovery of SD genes has helped to understand the plasticity of sexual reproduction, found under evolutionary pressures through rapid turnover in sex chromosomes and SD genes to ensure survival (Nagahama *et al.*, 2021). For example, (Pan *et al.*, 2021a), suggested that sex reversal (*i.e.*, which occurs when the developmental trajectory of the gonad undergoes a change, leading to a shift in the sexual phenotype of the organism towards the opposite sex) (Weber and Capel, 2018), and genetic bottlenecks could be a non-adaptive explanation for evolution of sex determination mechanisms, after the study of several sex determining transitions between different species and populations.

Table 1. Diversity of master sex determining genes across vertebrate species. Table modified and updated from Nagahama *et al.*, (2021). The SD genes in teleosts are highlighted in grey. Abbreviations: GF, growth factor; H, anti-Müllerian hormone; HR, hormone receptor; IRF, interferon regulatory factor; SE, steroidogenic enzyme; SR, steroid receptor; TF, transcription factor; CAS, Crk-associated substrate

Species	GSD type	SD gene	Protein type	Reference
Mammals	XY/XX	<i>SRY/Sry</i>	TF	Sinclair <i>et al.</i> , 1990
Monotremes	X ₁ X ₁ X ₂ X ₂ X ₃ X ₄ X ₅ X ₅ / X ₁ Y ₁ X ₂ Y ₂ X ₃ Y ₃ X ₄ Y ₄ X ₅ Y ₅	<i>AMH</i>	H	Grafodatskaya <i>et al.</i> , 2007
Chicken, <i>Gallus gallus domesticus</i>	ZZ/ZW	<i>DMRT1</i>	TF	Smith <i>et al.</i> , 2009
African clawed frog, <i>Xenopus laevis</i>	ZZ/ZW	<i>DM-W</i>	TF	Yoshimoto <i>et al.</i> , 2008
Japanese wrinkled frog, <i>Glandirana rugosa</i>	ZZ/ZW	<i>Z-AR</i>	HR	Oike <i>et al.</i> , 2017
Medaka, <i>Oryzias latipes</i>	XY/XX	<i>dmy</i>	TF	Matsuda <i>et al.</i> , 2002
Medaka, <i>Oryzias latipes</i>	XY/XX	<i>dmrt1by</i>	TF	Nanda <i>et al.</i> , 2002
Patagonian pejerrey, <i>Odontesthes hatcheri</i>	XY/XX	<i>amhy</i>	H	Hattori <i>et al.</i> , 2012
Luzon medaka, <i>Oryzias luzonensis</i>	XY/XX	<i>gsdf^Y</i>	GF	Myosho <i>et al.</i> , 2012
Japanese pufferfish, <i>Takifugu rubripes</i>	XY/XX	<i>amhr2</i>	HR	Kamiya <i>et al.</i> , 2012
Rainbow trout, <i>Oncorhynchus mykiss</i>	XY/XX	<i>sdY</i>	IRF	Yano <i>et al.</i> , 2012
Indian medaka, <i>Oryzias dancena</i>	XY/XX	<i>sox3^Y</i>	TF	Takehana <i>et al.</i> , 2014
Nile tilapia, <i>Oreochromis niloticus</i>	XY/XX	<i>amhy</i>	H	Li <i>et al.</i> , 2015
Turquoise killifish, <i>Nothobranchius furzeri</i>	XY/XX	<i>gsdf6Y</i>	GF	Reichwald <i>et al.</i> , 2015
Chinese tongue sole, <i>Cynoglossus semilaevis</i>	ZZ/ZW	<i>dmrt1</i>	TF	Cui <i>et al.</i> , 2017
Seriola fishes	ZZ/ZW	<i>hsd17b1</i>	SE	Koyama <i>et al.</i> , 2019
Northern pike, <i>Esox lucius</i>	XY/XX	<i>amhbY</i>	H	Pan <i>et al.</i> , 2019
Turbot, <i>Scophthalmus maximus</i>	ZW/ZZ	<i>sox2-W</i>	TF	Martínez <i>et al.</i> , 2021
Channel catfish, <i>Ictalurus punctatus</i>	XX/XY	<i>bcar1-Y</i>	CAS	Bao <i>et al.</i> , 2019
Cod, <i>Gadus morhua</i>	XX/XY	<i>zkY</i>	ZK	Kirubakaran <i>et al.</i> , 2019
Atlantic herring, <i>Clupea harengus</i>	XX/XY	<i>bmpr1bbY</i>	HR	Pettersson <i>et al.</i> , 2019
Sablefish, <i>Anoplopoma fimbria</i>	XX/XY	<i>gsdf</i>	GF	Herpin <i>et al.</i> , 2021
Sebastes rockfish, <i>Sebastes schlegelii</i>	XX/XY	<i>amhy</i>	H	Song <i>et al.</i> , 2021
Ayu, <i>Plecoglossus altivelis</i>	XX/XY	<i>amhr2bY</i>	HR	Nakamoto <i>et al.</i> , 2021
Atlantic Halibut, <i>Hippoglossus hippoglossus</i>	XX/XY	<i>gsdf</i>	GF	Edvardsen <i>et al.</i> , 2022
Southern catfish, <i>Silurus meridionalis</i>	XX/XY	<i>amhr2y</i>	HR	Zheng <i>et al.</i> , 2022
Spotbanded scat, <i>Selenotoca multifasciata</i>	XX/XY	<i>dmrt1Y</i>	TF	Jiang <i>et al.</i> , 2022
Senegalese sole, <i>Solea senegalensis</i>	XX/XY	<i>fshr</i>	HR	de la Herrán <i>et al.</i> , 2023

1.1.2. Polygenic sex determination

In PSD, multiple genetic factors independently inherited act together to bring about the sexual phenotype within a species (Kosswig, 1964; Moore and Roberts, 2013). PSD systems can arise from modifications of autosomal loci in the genome that create new inputs for the regulation of the gonad development. PSD is likely to have played a role in the evolution of other sex-determining systems as they provide a model for evolutionary transitions from GSD system to another SD system, where a novel SD gene would co-exist and interact with an ancestral SD gene within a population (Ezaz *et al.*, 2006).

PSD, was discovered almost 60 years ago in platyfishes (Kosswig, 1964). In the platyfish *Xiphophorus maculatus*, where the sex of an individual is determined by the interaction of three alleles (X, Y, and W) at the same sex determination locus (Volf and Scharl, 2001). In some East African cichlids, XY and ZW sex determination systems are located on two separate chromosomes, and their interaction and

dominance determine sex (Roberts *et al.*, 2016). The European sea bass (*Dicentrarchus labrax*) has PSD, with more than three Quantitative Trait Locus (QTLs) associated with sex ratio (Vandeputte *et al.*, 2007).

Although PSD may be inherently unstable, it can remain stable if different polymorphisms result in alternative fitness benefits according to the environmental context (Moore and Roberts, 2013). Meisel, (2021), used forward population genetic simulations to identify selection pressures that could maintain PSD. He found that maintenance of PSD depends on the dominance of fitness effects (*i.e.*, when the presence of one copy of the gene is enough to produce the same phenotype as having two copies) and that over-dominant fitness effects of male-determining proto-Y chromosomes tend to have sexually antagonistic effects (male-beneficial and female-detrimental) which are more likely to maintain PSD than dominant, recessive, or additive fitness effects.

Nowadays it is thought that truly polygenic systems are uncommon in nature, and represent transitions between systems or found in hybrid species where the sex determination system is disrupted (Schartl *et al.*, 2023). The truly attribution of PSD is challenging because loci which contribute only partially and have minor effects on the sexual phenotype are less apparent and difficult to characterize in analyses of genetic crosses. Schartl *et al.*, (2023), have recently revisited the actual definition of PSD and highlighted that a species should not be attributed to have PSD when: *i*) they have polyallelic systems at the same locus or when several steps in a biochemical pathway are driven by a single gene, *ii*) these genes do not act in the same individual, but instead work as polymorphic within the population and *iii*) when these multiple factors co-exist due to turnover stage when one system is taking over from another. Thus, it is important to distinguish between actual PSD system in which multiple independent genes are combined together from other SD mechanisms, such as monogenic SD with multiple alleles, digenic with two copies of the allele or digenic with neo-chromosome turnover (**Figure 2**).

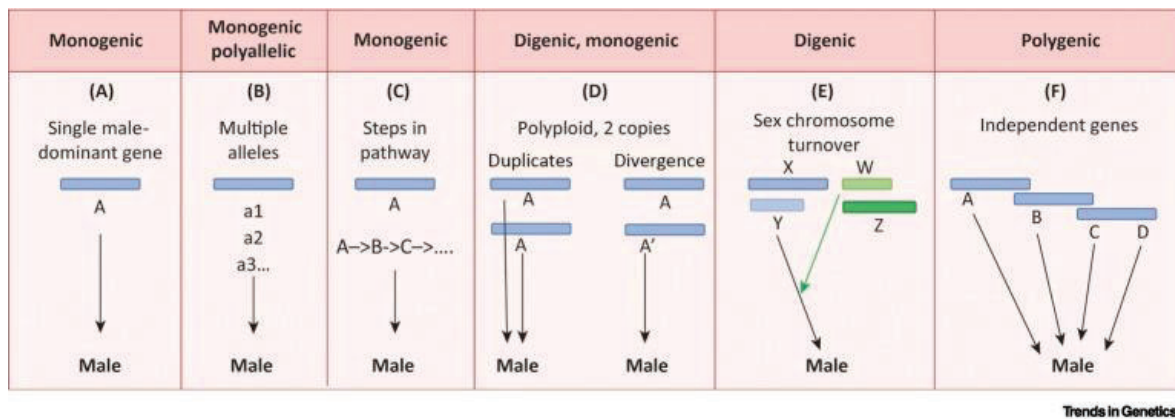


Figure 2. Diagram of monogenic, digenic, and polygenic male sex determination. Figure taken from Schartl *et al.*, 2023.

1.1.3 Environmental sex determination

ESD is expected to be advantageous in offspring inhabiting fluctuating environments, as it may benefit one sex over the other (Charnov and Bull, 1977). TSD is the most common ESD mechanism in vertebrates, where the temperature at which embryos develop determines gonadal development. TSD occurs in wild amniotes such as reptiles, crocodiles, tuatara, most turtles (including all sea turtles), and some lizards (Sarre *et al.*, 2011). In those reptiles, egg incubation temperature is critical for sexual fate (Deeming and Ferguson, 1988). Previous research suggested that TSD and GSD in reptiles were independent physiological events, but a study on live-bearing lizards (The snow skink, *Niveoscincus*

ocellatus) found that divergent SD mechanisms were closely linked. The study concluded that high altitudes' fluctuating climatic conditions caused snow skinks in highlands to adopt GSD to avoid reduced fitness and imbalanced sex ratios (Pen *et al.*, 2010).

In reptiles, there are three patterns of sex ratio response to temperature during the thermo-sensitive period: 1) species where low temperatures produce female-biased sex ratios and high temperatures produce male-biased sex ratios, 2) the opposite pattern where low temperatures produce male-biased sex ratios and high temperatures produce female-biased sex ratios and 3) species in which balanced sex ratios are obtained at intermediate temperatures and females are produced at extreme temperatures (Valenzuela and Lance, 2004). An example of a reptile with TSD is the red-eared turtle (*Trachemys scripta*) (Wibbels *et al.*, 1998), species that when incubated at 26°C develops males and when incubated at 31°C degrees, all individuals develop as females (Czerwinski *et al.*, 2016). However, in fish, there is one pattern of response to temperature in which high temperature produces more males (Ospina-Alvarez and Piferrer, 2008). An example of a fish with TSD is the Atlantic silverside (*Menidia menidia*) (Conover and Van Voorhees, 1990) and the Argentinian silverside or pejerrey (*Odontesthes bonariensis*), which also has Anti Müllerian Hormone (*amhY*) as a sex determining gene (Strüssmann *et al.*, 1997).

Temperature is an abiotic factor which can have important implications in vertebrates at multiple levels of the organism, but one of the most important effects related to sexual phenotype is that it can also alter sex ratios in PSD species such as the European sea bass (Piferrer *et al.*, 2005) and even in GSD species such as the Nile tilapia (*Oreochromis niloticus*) (Bezault *et al.*, 2007). In fact, the classification between species with TSD and species with GSD and temperature effects (GSD+TE) need critical consideration because most studies have been performed under laboratory conditions and therefore, the significance of temperature in terms of ecology or adaptation is not necessarily demonstrated (Ospina-Alvarez and Piferrer, 2008). An example of how related are these two systems is the central bearded dragon (*Pogona vitticeps*), which has ZZ males and ZW female sex chromosomes whose influence can be overridden in genetic males and by high temperatures, causing male-to-female sex reversal. In this species, Whiteley *et al.*, (2021), found that the genes involved in early gonadal development differ dramatically between genetically driven and thermally driven females, converging to produce a common phenotype: ovaries. Another example of how unclear is the limit between GSD and ESD is the case described by Gessler *et al.*, (2023). Through the study of the transcriptomic developmental time series of two reptiles, *Apalone spinifera* (with GSD) and *Chrysemys picta* (with TSD), they found that significant transcriptional plasticity in the developing gonads can persist for more than 145 million years, even after the establishment of SD through the evolution of sex chromosomes. They also observed that certain genes exhibited thermal plasticity. This finding suggests that GSD species possess an overlooked evolutionary potential in terms of standing thermosensitivity, which could be utilized during future shifts in developmental programming, such as a reversal from GSD to TSD, if ecological conditions favor such changes (Gessler *et al.*, 2023).

1.2 Sex differentiation

Sex differentiation is the process by which an undifferentiated gonad acquires the sexual phenotype and is transformed into a testis or an ovary through molecular, genetic and physiological mechanisms (Piferrer and Guiguen, 2008; Takahashi, 1977). Although the genes involved in converting the gonads into ovaries or testis were found relatively conserved during sex differentiation, recently, it became apparent that the temporal and relative expression, are different across vertebrates. In some instances, there are genes whose function is even related to the opposite sex (Nagahama *et al.*, 2021). Additionally, the level of variability of genes involved in sex differentiation and their transcriptomic dynamics are

unknown. Nevertheless, in this section we will summarize the cellular and molecular mechanisms known so far to be underlying gonadal sex differentiation among the five major groups of vertebrates, from fish to mammals.

Within the group of fish, teleosts are the most diverse group (Ravi and Venkatesh, 2018). During the emergence of teleosts, a significant event known as the teleost-specific whole-genome duplication took place in their ancestral lineage around 300 Ma (Stöck *et al.*, 2021). Currently, more than 35,000 species of fish have been registered representing more than half of all known vertebrate species (Fricke *et al.*, 2019). Gonadal sex differentiation, in teleosts depends on the relative ratio between androgens and estrogens, which is determined by the activity of the steroidogenic enzymes. One of the key downstream target genes during gonad differentiation is aromatase, an enzyme responsible of the conversion of androgens into estrogens encoded by the cytochrome P450, family 19, subfamily A, polypeptide 1 (*cyp19a1*) (Simpson *et al.*, 1994). In teleosts, aromatase is coded by two genes, the gonad-specific *cyp19a1a* and brain-specific *cyp19a1b* (Chang *et al.*, 1997). High levels of estrogen, particularly estradiol-17 β (E₂) are essential for ovarian development in fish and their suppression or downregulation leads to the development of a permanent and functional testis (Guiguen *et al.*, 2010; Piferrer *et al.*, 1994). On the other hand, 11-Ketotestosterone (11-KT) is the major androgen hormone in fish involved in testis development and in spermatogenesis (Miura *et al.*, 1991). In some fish species in which sex differentiation is most well-known, such as medaka and Nile tilapia, sex differentiation of males is driven by the expression of *gsdf*, *dmrt1* and *amh*. In opposite, ovarian differentiation in fishes is initiated by the expression of R-Spondin 1 (*rspo1*), Forkhead Box L2 (*foxl2*) and *cyp19a1a* (Kobayashi *et al.*, 2004; Nagahama *et al.*, 2021).

Amphibians diverged from Amniotes 335 Ma, leading to Anura (frogs and toads) and Urodela (newts and salamanders) (Stöck *et al.*, 2021). Most amphibian exhibit GSD with either male (XY/XX) or female (ZZ/ZW) heterogamety. Although, environmental factors can influence artificially sex ratio as sex reversal, there is no evidence that environmental factors can influence the GSD in amphibians under natural conditions (Perrin, 2021; Piferrer and Anastasiadi, 2021). In *Ranidae*, it was shown that the location of the SD genes is not fixed, as they were associated with different linkage groups across different species, and even between populations of the same species. This was shown in the Japanese wrinkled frog (*Glandirana rugose*), where both XX/XY and ZW/ZZ populations were found (Perrin, 2021). Nevertheless, the only SD gene characterized so far in anurans is the W-linked DM-domain gene (*DM-W*). The W-linked DM-domain gene (*DM-W*) is female-specific and was described in the African clawed frog (*Xenopus laevis*) with ZZ/ZW heterogamety. However, this gene did not show a relevant role during differentiation (Nagahama *et al.*, 2021; Stöck *et al.*, 2021). Instead, key genes involved in ovarian differentiation of *X. laevis* are *foxl2*, *cyp19a1*, and Follistatin (*fst*). The development of testis occurs through the expression of *amh*, Cytochrome P450 Family 17 Subfamily A Member 1 (*cyp17a1*), *dmrt1*, Follicle Stimulating Hormone Receptor (*fshr*), Prostaglandin D2 Synthase (*ptgds*), and SRY-Box Transcription Factor 9 (*sox9*) (Piprek *et al.*, 2018, 2017).

Turtles evolved 270–250 Ma from the Archosaurian lineage (birds and crocodiles) (Stöck *et al.*, 2021). Most squamates present GSD, while ESD is predominant in all crocodiles, most turtles, and some lizards. In turtles, testis differentiation begins through activation of *dmrt1* which leads to upregulation of *amh* and *sox9*. The late expression of *dmrt1* in the painted turtle, *Chrysemys picta*, is an example in which this gene has an important role during sex differentiation and not during sex determination. In this species, Mizoguchi and Valenzuela, (2020) revealed an isoform of this gene indicating that alternative splicing may be a common feature of *dmrt1* across vertebrates, as isoforms are also found in crocodilians, birds, mammals and fish. These findings highlight the divergent regulation of *Dmrt1* and

sexual development across vertebrates. At female promoting temperatures (FPT), *cyp19a1* and *foxl2* are the key genes leading to ovarian differentiation (Czerwinski *et al.*, 2016; Nagahama *et al.*, 2021). Another example of the work done in sex differentiation in reptiles is the gonadal transcriptome analysis during differentiation in central bearded dragon, a species with GSD. In this species, the male pathway genes *dmrt1* and *amh* and the female pathway gene *foxl2* were involved in early sex differentiation. However, *sox9* (the central player of the mammalian male differentiation pathway) was not differentially expressed. The authors suggested that in this species, a default male trajectory could occur if not repressed by a W-linked master gene tilting the balance of gene expression towards the female pathway (Wagner *et al.*, 2023).

Birds which evolved from the crocodiles about 240 Ma (Stöck *et al.*, 2021) have female-heterogametic sex chromosomes, and unlike amphibians, the female sex-determining gene has not been discovered. Testis development is initiated by the activity of the double doses of *dmrt1* in the primordial germ cells, followed by the coordinated activation of Hemogen (*HEMGN*), *AMH*, and *SOX9*. In contrast, ovarian development is dominated by estrogen production regulated through the genes such as *CYP19A1*, *FOXL2*, *RSPOL*, and *WNT4* expressed during ovarian development (Nagahama *et al.*, 2021). Additionally, Ayers *et al.*, (2015) described three genes relevant for ovarian differentiation: cal-157pain-5 (*CAPN5*), G-protein coupled receptor 56 (*GPR56*), and fibroblast growth factor receptor 3 (*FGFR3*). In the most primitive mammals, monotremes (which evolved 130 Ma), the sex chromosomes are not homologous to those of other mammals but rather to the ZW chromosomes of birds, which originated through duplications (Kratochvíl *et al.*, 2021; Wallis *et al.*, 2008). For example, the platypus (*Ornithorhynchus anatinus*) has five pairs of X and Y chromosomes that fuse during meiosis to form a chain. Despite having the gene *Dmrt1* on one of the X chromosomes, these mammals lack the *Sry* gene, and *Dmrt1* is not involved in sex determination. On the other hand, most marsupials and placental mammals share the same pair of sex chromosomes, X and Y, with the *Sry* gene playing a role in sex determination. Therefore, since the sex chromosomes of marsupials and placental mammals are homologous to autosomes in other vertebrates, it is believed that the *Sry* gene evolved after the divergence of monotremes from marsupials and placental (eutherian) mammals about 166 Ma (Beukeboom and Perrin, 2014; Keating *et al.*, 2022). Placental mammals have GSD with heteromorphic sexual chromosomes (XX/XY) system. The mouse is the model organism within vertebrates for which sexual development is most well-known. Up to the embryonic stage 11.2 days post coitum (dpc), the genital ridges are at the bipotential stage. The *Sry* determining gene is found in the Y chromosome and initiates the development of testes. Between 10.0 and 11.2 dpc the animals having *Sry* gene in the Y chromosome begin testes development. *Sry*, activates *Sox9* with a positive feedback loop existing between *Pdg2/Sox9* and Fibroblast Growth Factor 9 (*Fgf9*)/*Sox9*. In the XX gonads, genes such as *Rspo1*, Wnt Family Member 4 (*Wnt4*), and *Foxl2* are expressed during early ovarian development (Nagahama *et al.*, 2021).

1.3 Sex development: from a hierarchical cascade to networks

From the point of view of developmental biology, the complex process of vertebrate sexual development has given rise to a number of theories. In 1995, Adam Wilkins proposed the so-called bottom-up theory according to which purifying selection acts only at the bottom of a hierarchical cascade (Wilkins, 1995). This idea was supported by Graham *et al.*, 2003, who referred to the genes at the top as “masters” selected through positive selection, and those at the bottom, as “slaves”. Accordingly, while masters can vary from one species to another, the slaves maintain structural and functional conservation across species (Graham *et al.*, 2003; Williams and Carroll, 2009). In 2006, the term “threshold dichotomy” to sex determination was proposed instead of viewing the SD gene as a simple switch on/off option (Mittwoch, 2006). This term is referred to opposing phenotypes that arise from multiple genes with

quantitative effects and are separated by a physiological threshold (Martínez *et al.*, 2014) (**Figure 3A**). Mittwoch, suggested that in mammals, sex determination is a combination of genetic and environmental influences. In the absence of the Y chromosome, the system favors female development, while in its presence, the effect of genetic and environmental factors is minimized in XY embryos, making sex inheritance resemble the effects of a single gene (Mittwoch, 2006). The threshold nature of sex determination is applicable regardless of the mechanism, even in species with GSD that are exposed to extreme temperatures.

In 2011, using medaka fish, Herpin and Scharf (2015), found that transcriptional regulation of *foxl2* was independent from aromatase expression, revealing that the linear downstream regulation of sex determination could be less conserved than previously thought. These and other findings lead to questioning the straightforward linear and hierarchical cascade, and instead, the picture of a complex network of antagonizing pathways emerged (Heule *et al.*, 2014). These studies also revealed that the genes actually participating in the regulation of gonadal sex differentiation could be diverse even between closely related species with similar gonadal development at the morphological, histological, and cellular levels (Herpin and Scharf, 2015; Pan *et al.*, 2016). Therefore, nowadays it is accepted that, despite the diversity of sex determining mechanisms, gonadal sex differentiation involves a complex network of multiple regulatory interactions of genes associated with male and female pathways (**Figure 3B**). Those are expressed simultaneously as opposing forces, until a factor or combination of them tilts the balance towards one fate or the other (Capel, 2017).

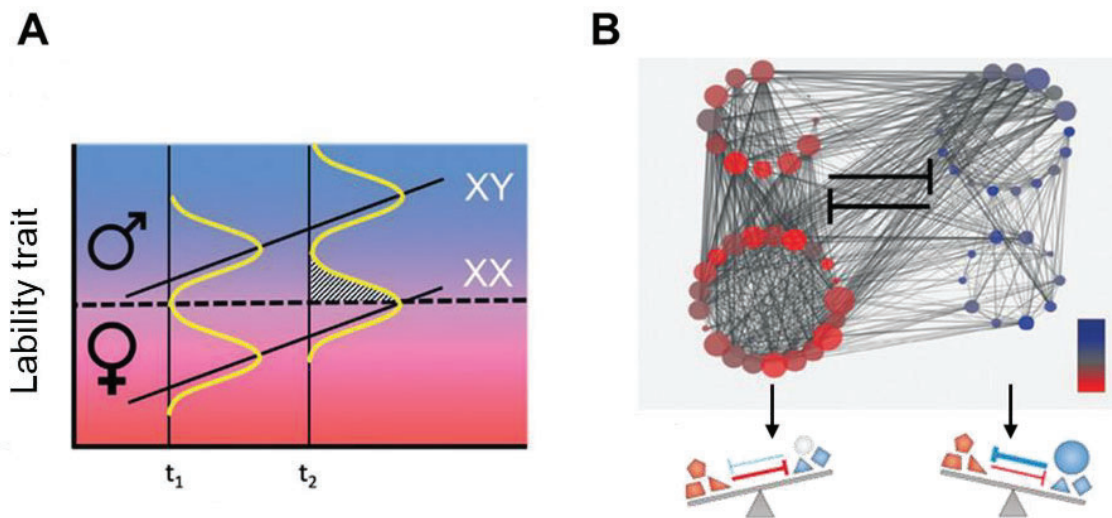


Figure 3. A) Diagram representing the threshold dichotomy nature of sex. Regardless of the sex determining system, even in a species with, for example, a XX/XY genetic sex determination, if the temperature is extreme, it can cause some XX individuals to become males (dashed area) if the threshold value of a lability trait exceeds a certain threshold (dotted line). **B)** Cassettes of genes associated with male (blue) and female (red) pathways are simultaneously expressed in two opposing sub-networks until a threshold is reached and the system tilts the balance towards male or female phenotype. Figure from Piferrer, 2018b.

1.4 Activation and repression programs

The genes involved in both male and female developmental programs are distributed throughout the genome as they are also involved in other developmental processes of the organism besides sexual development proper. Thus, in such context, a model based solely on the activation of either the male or female program is too simplistic (Poulat, 2021). In 2013, Munger *et al.*, (2013), analyzed the fine time course of gene expression for each sex separately during sex differentiation in the mouse. This study

allowed to identify cascades of activation and also cascades of repression. Between embryonic stages E.11.0 and E.12.0, in the XY gonad of the mouse, male pathway genes were actively upregulated while approximately half of the genes associated with the female pathway became sexually dimorphic through downregulation in the XY gonad rather than through upregulation in the XX gonad. Thus, repression of the alternative phenotype is critical for development and maintenance of the testis or ovary programs. However, the mechanisms of such repression are still not completely understood. Dupont and Capel (2021) reviewed several studies showing the importance of the chromatin state and epigenetic mechanisms regulating such patterns of activation and repression of gene expression and cell fate commitment, with histone modifications and DNA accessibility having a direct role in gene regulation. Nevertheless, it remains unknown how widespread and to what proportion these mechanisms of gene activation and repression are found across vertebrates. Thus, fine-tuning of the spatiotemporal expression of genes involved in sexual development is critical.

2. Epigenetics, a bridge between genomic and environmental information

The term of epigenetics was coined by Conrad Waddington in 1942 (Waddington, 1942). In its origins, he linked the term of epigenetics to development. He defined epigenetics as “to the study of the way genes and their products bring the phenotype into being”. Thus, Waddington draw the developmental system as a landscape in which bifurcating and deepening valleys run from a plateau (**Figure 4**). Such epigenetic landscape was modulated by underlying interactions and those interactions were caused by both genetic and environmental factors (Jablonka and Lamb, 2002). Today, epigenetics is seen as the branch of biology that contributes to the study of mechanisms responsible for changes in gene expression that do not involve changes to the DNA sequence, and which can be inherited during cell division (mitosis) as well as across generations (meiosis), affecting gene expression and therefore the phenotype or set of observable characteristics in an individual (Dupont *et al.*, 2009; Russo *et al.*, 1996). Epigenetic modifications have been shown to play a crucial role in regulating a wide range of biological processes from embryonic development, epigenetic activation and silencing of genes, to disease onset and progression (Heyn *et al.*, 2012).

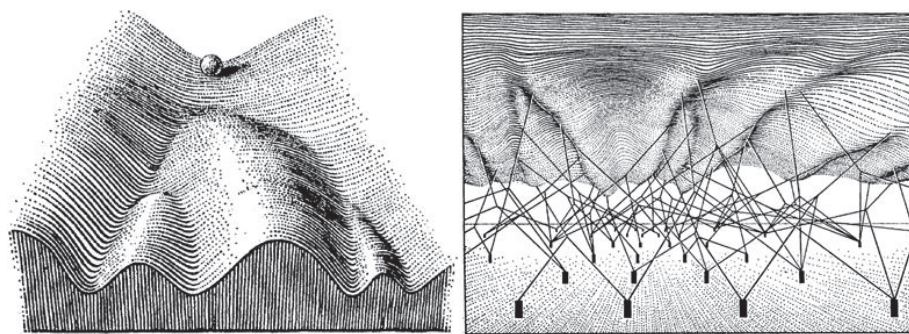


Figure 4. A) Epigenetic landscape drawn by Conrad Waddington. B) Underlying interaction of the epigenetic landscape with genetic and environmental influences. Figures from Waddington, (1957).

2.1 Epigenetic mechanisms

Three primary molecular mechanisms are widely recognized as key epigenetic processes: DNA methylation, histone variants and modifications, and regulatory non-coding RNAs. In this thesis, we will provide a brief overview of these mechanisms, while focusing primarily on DNA methylation. DNA methylation is the epigenetic mechanism that involves the covalent transfer of a methyl-group from S-

adenosyl-L-methionine (SAM) to the fifth position of a cytosine, creating 5-methylcytosine (^5mC) or to the sixth position of an adenine, creating N^6 -methyladenine (Ratel *et al.*, 2006; Wu *et al.*, 2016). ^5mC is the most widely studied DNA modification and is present in most species (Chen *et al.*, 2003). DNA methylation is catalyzed by DNA methyltransferases (DNMTs), the “writers”, in vertebrates: *dnmt1* and *dnmt3*. *Dnmt1* adds a methyl group to hemi-methylated CpGs, maintaining DNA methylation patterns across cell divisions, while *dnmt3* is responsible for de novo methylation and is involved in maintenance of DNA methylation. On the other hand, the “erasers” are from the Ten–eleven translocation (TET) family of enzymes (TET1, TET2 and TET3), which catalyze the oxidation of 5-methylcytosine (^5mC) to 5-hydroxymethylcytosine (^5hmC), 5-formylcytosine (^5fC) and 5-carboxylcytosine (^5caC), key intermediates in the DNA demethylation processes (Gong and Zhu, 2011).

Histone modifications are post-translational modifications that occur on the proteins known as histones, which package DNA into nucleosomes forming the chromatin. These modifications regulate transcription, and participate in basic DNA functions, such as replication, repair and recombination. The interplay of histone modifications with DNA methylation are important regulators of gene expression (Garzón-Porras *et al.*, 2023; Ito, 2007; Park *et al.*, 2022). The enzymes that modify chromatin are sensitive to changes in environmental and metabolic cues, therefore act as sensors through which environmental conditions may influence gene expression (Carlberg *et al.*, 2014). The most common histone modification types are methylation (mostly associated with gene silencing), acetylation, phosphorylation (both associated with gene activation), and ubiquitination (associated with silencing and activation) (Turner, 2005).

Non-coding RNAs (ncRNAs) are a group of RNA molecules that do not encode functional proteins, and were previously thought to only regulate gene expression post-transcriptionally. However, it was shown latter that non-coding RNA (ncRNAs), such as micro (miRNAs), PIWI-Interacting (piRNAs), endogenous Small interfering (siRNAs), and long ncRNAs, are considered a crucial epigenetic mechanism that regulates gene expression (Castro-Muñoz *et al.*, 2023; Zaratiegui *et al.*, 2007). ncRNAs have been involved in several epigenetic phenomena, such as transcriptional and transposon silencing, parental imprinting, formation of heterochromatin, and regulation of sex chromosomes for dosage compensation and developmental regulation (Mattick, 2007; Ponting *et al.*, 2009). The mechanisms by which ncRNAs act include preventing DNA-binding factors from binding, acting as scaffolds for protein interactions, and recruiting chromatin modifying enzymes to the DNA (Carlberg *et al.*, 2014; Zhou *et al.*, 2023).

2.1.1 DNA methylation and methods to assess DNA methylation

DNA methylation can be induced by genetic variation (Richards, 2006), spontaneous epimutations (Becker *et al.*, 2011), and environmental factors (Weaver *et al.*, 2004), although the latter depends on the interaction with genetic variation as well. DNA methylation is a mechanism by which the phenotype can be fine-tuned under the environmental influence, providing the organism the opportunity to develop an adaptive phenotype, thus, contributing to phenotypic plasticity (Jablonka and Lamb, 2007).

DNA methylation is a key epigenetic mechanism that regulates gene expression in eukaryotes. Gene promoter methylation is typically linked to transcriptional silencing, and to explain the relationship between DNA methylation and gene expression in the regulation of sexual phenotype emergence, the Conserved Epigenetic Regulation of Sex (CERS) model was proposed by (Piferrer *et al.*, 2019). The CERS model postulates that, in females, DNA methylation of pro-female genes will be low while its expression is high and, in contrast, DNA methylation of pro-male genes is high associated to its low expression. In turn, in males, the opposite occurs. Such model applies to sex differentiation in

gonochoristic species and sex change in hermaphroditic species regardless of the underlying SD mechanism. However, despite the generalized dogma that DNA methylation is associated with gene silencing, some studies in human indicate that methylation downstream of the transcription start site is also associated to gene expression. Additionally, a negative correlation between DNA methylation in the first intron and gene expression across various tissues and species was found (Anastasiadi *et al.*, 2018a). Thus, all the nuances regarding the interplay between DNA methylation and the genome regulating gene expression remains unclear.

Currently, there are several methods for assessment of DNA methylation which consist in distinguishing unmethylated cytosines from methylated cytosines (^5mC) in the DNA sequence (Feng and Lou, 2019). Such methods can be classified within three levels: 1) Detection strategy (*e.g.*, enzyme based, bisulfite conversion), 2) Coverage level (targeted approaches, *i.e.*, locus specific, genome wide, whole genome) and 3) Resolution level (*e.g.*, global, single nucleotide resolution). Within the first level (detection), regarding the technology to distinguish between ^5mC and Cs there are four main principles: (i) digestion of DNA with methylation-sensitive restriction enzymes, (ii) enrichment of methylated genomic DNA fragments using antimethylcytosine antibody or methyl-binding domain (MBD) proteins, (iii) sequencing of bisulfite-converted DNA and (iv) direct measurement.

The second level has to do with the extent of the genome (where) being interrogated, which can be: (i) locus-specific, in which the marks are measured in target regions, (ii) genome-wide, in which the marks are measured across a representative portion of the genome. (iii) whole-genome, where the entire genome is assessed (Anastasiadi, 2016). Finally, the third level refers to the level of resolution, which can be grouped into: (i) low resolution, provides information on the mean global ^5mCs content of the entire sample, however, the exact genomic locations of the ^5mCs remain unknown. (ii) medium resolution, when the ^5mCs can be localized within an approximate context, and (iii) single nucleotide resolution, when the exact locus of ^5mCs and Cs can be determined. In **table 2**, there is an updated summary of the most used methods to measure DNA methylation and their corresponding levels of classification.

Nowadays, sequencing of bisulfite-converted DNA is the gold standard of ^5mC detection techniques. This technique is based on the treatment of genomic DNA with sodium bisulfite, which results in the deamination of unmethylated cytosines into uracil and leaving methylated cytosines intact (Frommer *et al.*, 1992). Converted DNA is PCR-amplified to produce sufficient template for analysis, causing uracil to convert to thymine. The methylation of the DNA can be read using traditional sequencing methods like Sanger or Illumina, providing accurate and quantitative data at the individual base-pair resolution level (Gouil and Keniry, 2019).

2.2 Epigenetics in cell identity acquisition

As mentioned above, epigenetic modifications are involved in many normal cellular processes, an important one being its contribution to cellular identity. Thus, cells with identical genome information acquire specific gene expression patterns through epigenetic mechanisms, giving rise to the different cell types. In the human, for example, more than 200 different cell types, result from one genome (Heyn *et al.*, 2012). DNA methylation plays a crucial role in cellular differentiation and maintenance of the tissue-specific phenotypes (Taylor, 2012). Regions that are uniquely unmethylated in a particular cell type are often found in transcriptional enhancers with binding sites for tissue-specific transcriptional regulators. Interestingly, DNA methylation profiles between individuals are more similar than inter-tissue profiles of the same individual, suggesting strong tissue-specific signatures of DNA methylation

Table 2. Classification of methods to analyze DNA methylation according to the type of detection method, coverage and resolution level. Table modified from Ortega-Recalde and Hore, 2023; the methods added are shown in blue

Detection method	Genome coverage			Resolution level	References
	Gene- or locus-specific	Genome-wide	Whole genome		
Enzymatic		LUMA, MSAP, RLGS		Low	Ando and Hayashizaki, 2006; Guevara <i>et al.</i> , 2017; Mohsen <i>et al.</i> , 2006
		MSAP-seq, MRE-seq, methyl-seq		Medium	Brunner <i>et al.</i> , 2009; Chwialkowska <i>et al.</i> , 2017; Maunakea <i>et al.</i> , 2010
	<i>HpaIII/MspI</i> -PCR		EM-seq, TAPS, Methyl-MAPS	High	Ficz <i>et al.</i> , 2011; Ito <i>et al.</i> , 2010; Li <i>et al.</i> , 2010; Liu <i>et al.</i> , 2019; Mellén <i>et al.</i> , 2012; Shi <i>et al.</i> , 2017; Vaisvila <i>et al.</i> , 2021; Wang <i>et al.</i> , 2013
Affinity			ELISA, HPLC	Low	Armstrong <i>et al.</i> , 2011; de Oliveira <i>et al.</i> , 2020
	MeDIP-PCR, MBD-PCR	MeDIP-chip, MeDIP-seq, MBD-seq	MethylCap-seq	Medium	Brinkman <i>et al.</i> , 2010; Mohn <i>et al.</i> , 2009
	MSP		hME-Seal	High	Song <i>et al.</i> , 2017
Bisulfite conversion				Low	Ramalho-Carvalho <i>et al.</i> , 2018
			Low-coverage WGBS	Medium	Ortega-Recalde <i>et al.</i> , 2021
	BSP, COBRA, MBS, Ms-SNuPE,	RRBS, bsRAD-seq, TIME-seq, TAB-seq, microarrays	WGBS, PBAT, T-WGBS, oxBS-seq	High	Anastasiadi <i>et al.</i> , 2018b; Bibikova and Fan, 2009; Booth <i>et al.</i> , 2013; Darst <i>et al.</i> , 2010; Eads and Laird, 2002; Gonzalgo and Liang, 2007; Grehl <i>et al.</i> , 2018; Griffin <i>et al.</i> , 2021; Meissner <i>et al.</i> , 2005; Nakabayashi, 2020; Schumacher <i>et al.</i> , 2006; Trucchi <i>et al.</i> , 2016; Yu <i>et al.</i> , 2018
Direct detection			HPLC-UV, LC-MS/MS	Low	Sidoli <i>et al.</i> , 2016
			SMRT-seq, nanopore	High	Kaya-Okur <i>et al.</i> , 2019; Kim and Dekker, 2018; Migeot and Hermand, 2018; Nakato and Sakata, 2021; Skene and Henikoff, 2017

Abbreviations: BSP: direct bisulfite sequencing; bsRAD-seq: Bisulfite Restriction site Associated DNA; COBRA: Combined bisulfite restriction analysis; ELISA: enzyme-linked immunosorbent assay; EM-seq - Enzymatic Methyl-seq; HPLC - High Performance Liquid Chromatography; hME-Seal - selective chemical labeling and capture; LC-MS/MS - liquid chromatography coupled with mass spectrometry; LUMA: Luminometric Methylation Assay; MeDIP: Methylated DNA ImmunoPrecipitation; MBD: Methyl-CpG-Binding Domain; Methyl-MAPS: Methylation Mapping Analysis by Paired-end Sequencing; MethylCap-seq: Methylated DNA Capture sequencing; MRE: Methyl-sensitive Restriction Enzyme; MSAP: Methylation Sensitive Amplified Polymorphism; MSP: Methylation Specific PCR; MBS: Multiplex Bisulfite Sequencing; oxBS-seq: Oxidative bisulfite sequencing; PBAT-seq: post-bisulfite adapter tagging; RLGS: Restriction Landmark Genomic Scanning; RRBS: Reduced Representation Bisulfite Sequencing; SMRT-seq: Single Molecule, Real-Time sequencing; TAB: Tet-Assisted Bisulfite Sequencing; TAPS: TET-assisted pyridine borane sequencing; TIME-seq: Tagmentation-based Indexing for Methylation Sequencing; T-WGBS: Tagmentation-based whole-genome bisulfite sequencing; WGBS: Whole Genome Bisulfite Sequencing.

(De Bustos *et al.*, 2009). During cell type differentiation there is a general trend of hypermethylation (Suelves *et al.*, 2016), and furthermore, the distinct epigenome of a given cell type changes over time (Heyn *et al.*, 2012). Recently, an updated human methylome atlas has been released with the identification of thousands markers for 39 cell types. Another example is the landmark paper by Klughammer *et al.*, (2023), in which an atlas of DNA methylation from 580 species reveals links between genomes and epigenomes throughout vertebrate evolution. This type of reference information is crucial for investigating gene regulation, disease-associated genetic variations, and discovering potential tissue-specific biomarkers (Loyfer *et al.*, 2023). Histone modifications also play a role during cell fate acquisition, however it was found that regulation by transcription factors could have a comparatively more important regulatory role than histone modifications, as unmarked chromatin would permit rapid gene activation and deactivation during development (Pérez-Lluch *et al.*, 2015).

2.3 Environmental influence and epigenetics

For a given genotype, the epigenetic variation can be originated by the underlying genetics, by environmental variation, by the genotype x environmental interactions, and as a result of stochastic variation. The importance of epigenetics in environmental responses is well-established in plants and in microorganisms (Cavalli and Heard, 2019; Vogt, 2021), when compared to that of vertebrates. Plants have shown considerable phenotypic variation because they are sessile and cannot avoid unfavorable environmental conditions. Moreover, the underlying genetic variation is smaller, since asexual reproduction is common in plants (Kroon and Groenendaal, 1997; Vogt, 2021). In the thale cress (*Arabidopsis thaliana*), the process of vernalization involves gradual epigenetic silencing, proportional to the length of exposure to cold, causing downregulation of a gene responsible for floral repression (Song *et al.*, 2012).

To date, the mechanisms by which the environmental cues are integrated in the form of epigenetic marks is not understood. In 2009, Berger *et al.*, proposed an operational model on how environmental information may be integrated and maintained into the genome through epigenetic mechanisms. The model consists in three parts: 1) the epigenator, an environmental signal that enters the nucleus, causing subsequent cascade of events, 2) the epigenetic initiator, a cellular signal that establishes specific chromatin contexts in response to epigenator signals, relying on sequence recognition to identify precise chromatin coordinates and, 3) the epigenetic maintainer, a signal that perpetuates chromatin changes without initiating them, preserving the epigenetic state in subsequent generations (*e.g.*, DNA methylation, histone modifications, and histone variants) (Berger *et al.*, 2009; Piferrer, 2013). In particular, for DNA methylation, one proposed mechanism to integrate environmental information is through the activation of signaling pathways that ultimately result in changes to the activity of DNMTs (Campbell and Wood, 2019; Samudio-Ruiz and Hudson, 2012). Another proposed pathway is through changes in the availability of methyl donor molecules, such as S-adenosylmethionine, which are required for DNMT activity (Campbell and Wood, 2019; Mahmoud and Ali, 2019; Niculescu and Zeisel, 2002; Pizzorusso and Tognini, 2020).

Some of the environmental factors influencing the phenotype of vertebrates, but not limited to, are nutrition, stress, pollutants, exposure to toxins, social environment, temperature, pH, domestication, etc. (Anastasiadi and Piferrer, 2019; Emam *et al.*, 2020; Klingenberg, 2019; Podgorniak *et al.*, 2019). Some of the most well-known examples of environmental influence through epigenetic mechanisms involve nutrition. For example, it was found that in part, what differs between a queen honeybee and the female workers (with identical genome) was that more than 550 genes were differently methylated between the two castes (Lyko *et al.*, 2010). These genes were involved in crucial cellular functions such as RNA synthesis and metabolism, which explains differences in behavior and reproductive ability. Other studies

showed that the methylation profile of genes linked to stress and metabolism can be modified in rats as a result of the maternal diet with a persistent effect in the long term (Lillicrop *et al.*, 2005; Szyf *et al.*, 2005). In humans, The Dutch Hunger Winter, which occurred during World War II when the German blockade led to famine in the Netherlands from 1944-1945 constitutes a prime example of the long-term effects arising from environmentally- (diet in this case) induced epigenetic changes. Exposure to famine during pregnancy caused persistent effects on health, including increased risk of obesity, diabetes, and cardiovascular disease (Heijmans *et al.*, 2008).

In species with TSD, one of the known mechanisms that integrates temperature information is the differential DNA methylation of the aromatase promoter, the enzyme that converts androgens into estrogens, which occurs between males and females. It was first discovered in the European sea bass (Navarro-Martín *et al.*, 2011), subsequently in turtles (Matsumoto *et al.*, 2013) and alligators (Parrott *et al.*, 2014). The differential DNA methylation between males and females has been observed in several other fish species, such as the Japanese flounder (*Paralichthys olivaceus*) (Wen *et al.*, 2014) and the Nile tilapia (Sun *et al.*, 2016), among others. The half-smooth tongue sole (*Cynoglossus semilaevis*) and the Japanese pufferfish can even undergo sex-reversal due to high and low temperature, respectively, causing changes in DNA methylation, with potential population consequences (Shao *et al.*, 2014; H. Zhou *et al.*, 2019). A recent study by Liu *et al.*, (2022), investigated DNA methylation dynamics during gonadal development in the yellow pond turtle (*Mauremys mutica*), a reptile with TSD. The study highlights the involvement of DNA methylation in integrating temperature information in this species.

2.4 Epigenetic inheritance

The inheritance of epigenetic marks occurs in two situations: one during cell divisions (mitotic), such as the epigenetic marks involved in cell identity being transmitted to daughter cells (Probst *et al.*, 2009). The other situation of epigenetic inheritance (meiotic) occurs when epigenetic marks are passed to subsequent generations, a form of inheritance with critical ecological and evolutionary implications (Jablonka and Lamb, 2006). Here, we will briefly describe some of key concepts regarding epigenetic inheritance according to the nomenclature proposed by Piferrer, (2023) in turn in accordance to what is most commonly used in mammals.

2.4.1 Types of epigenetic inheritance

According to the number of generations in which the effects induced by a stimulus can be observed, we can distinguish the following categories: **intragenerational** and **intergenerational**. Intragenerational inheritance occurs when the effects induced by a stimulus are observed within the exposed generation, *i.e.*, only the F0 generation. In intergenerational inheritance, the effects induced by a stimulus are identified in two or more generations, including the F0 generation. Within intergenerational inheritance there are two subtypes: multigenerational, when the marks induced by a stimulus are found ‘in all the directly exposed generations, and transgenerational, where the epigenetic marks acquired due to direct exposure are found in one or more subsequent unexposed generations. To demonstrate multigenerational inheritance in a pregnant mammal it is necessary to find the marks in F0, F1, and F2. In a non-pregnant female mammal, in a male or in fish, at F0 and F1. To demonstrate transgenerational epigenetic inheritance (TEI) in a pregnant mammal is necessary to find the marks in F0, F1, F2, as well as the F3 (unexposed generation) at least; in a non-pregnant female or male mammal and in fish, the F0, F1 (directly exposed) or more generations. Of note, not all intragenerational or multigenerational effects must have an epigenetic nature, however, transgenerational effects can only be epigenetic (Piferrer, 2023).

Examples of epigenetic variation inheritance across generations has been found in plants (Bender and Fink, 1995; Martienssen and Colot, 2001), prokaryotes (Lim and van Oudenaarden, 2007), eukaryotes (Colot and Rossignol, 1999) and vertebrates (Meehan and Stancheva, 2001) including fish (Shao *et al.*, 2014) and mammals (Takahashi *et al.*, 2023). In fish there are several examples: it was demonstrated that hypoxia causes transgenerational impairments of spermatogenesis in fish (Wang *et al.*, 2016). In zebrafish, there is evidence of TEI of epigenetic marks influenced by the effect of cadmium ancestral exposure, leading to feminization and even overriding temperature-induced masculinization of subsequent generations (Pierron *et al.*, 2021). TEI in mammals has been a matter of debate, however, Takahashi *et al.*, (2023) recently demonstrated that DNA methylation of promoter-associated CpG islands (CGIs) can be transmitted from parents to their offspring in mice. After edition of methylation profile in targeted regions, the acquired methylation and the phenotypic traits were maintained and transmitted across multiple generations. The results showed that the edited methylation underwent reprogramming in parental PGCs and was subsequently reestablished in the next generation at post-implantation stages. Nevertheless, the mechanisms and the extension of generations in mammals and other organism remain unknown. A comprehensive review on epigenetic inheritance and the reproductive mode of both plants and animals showed that multigenerational persistence of epigenomic patterns is common in both plants and animals and that the reproductive mode and germline development can influence the prevalence and persistence of such epigenetic inheritance (Anastasiadi *et al.*, 2021b).

2.4.2 Epigenetic reprogramming

Epigenetic reprogramming is the process by which epigenetic marks derived from the parents are changed during early embryogenesis and gametogenesis, contributing to the development of the new individual (Jiang *et al.*, 2013). These changes include the erasure and re-establishment of parental genomic imprints in germ cells, the erasure of epimutations, and the generation of totipotent or multipotent cells (Hajkova *et al.*, 2002). In mammals, reprogramming is a well-studied phenomenon (Du *et al.*, 2022; Wang *et al.*, 2014). In both humans and mouse, there are two waves of DNA demethylation (**Figure 5**): one occurs during the early stages of embryo development, and the other during germ cell specification. In the embryo, methylation reprogramming occurs between fertilization and formation of the blastocyst. At fertilization, a rapid paternal-specific asymmetric loss of methylation is observed (Mayer *et al.*, 2000). During germ cell specification, hypermethylated PGCs undergo rapid genome-wide demethylation (Reik *et al.*, 2003) coinciding with erasure of parent-of-origin marks (Santos and Dean, 2004; Tucker *et al.*, 1996). In amphibians, such as the Western clawed frog (*Xenopus tropicalis*), the genome is heavily methylated at both intergenic and intragenic regions during gastrulation, global demethylation does not occur (Bogdanović *et al.*, 2011; Du *et al.*, 2022).

In fish, only a few species have been studied. In zebrafish maternal methylome gradually changes to resemble sperm, allowing zygotic genome activation. In some fish species, epigenetic marks required for germ cell specification are inherited through the germplasm, allowing demethylation unnecessary in the zebrafish (Jiang *et al.*, 2013; Potok *et al.*, 2013). In medaka, however, a strategy similar to that in the mammals was found. Methylation reprogramming occurs in both early embryos and PGCs, raising the question on whether this species inherit epigenetic marks from the germ plasm or not (Wang and Bhandari, 2019). The mangrove killifish (*Kryptolebias marmoratus*), a self-fertilizing hermaphrodite fish, epigenetic reprogramming occurs at a later stage, is more extensive, and lasts longer compared to other vertebrates (Fellous *et al.*, 2018). Overall, these findings suggest that epigenetic reprogramming in fish may be species-specific, which raises questions about the extent and mechanism of inheritance of epigenetic marks to subsequent generations.

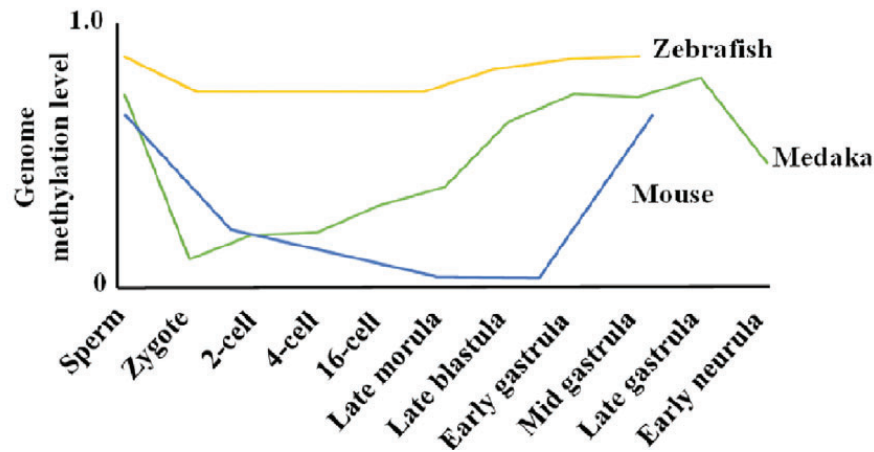


Figure 5. Comparison of DNA methylation profile in sperm prior to fertilization and in the zygote during early embryonic development in zebrafish (*Danio rerio*), medaka and mouse obtained from Jiang *et al.*, (2013), Potok *et al.*, (2013), Wang and Bhandari, (2019), and Wang *et al.*, (2014). Figure from Zhang *et al.*, (2023).

3. Epigenetic biomarkers

Epigenetic biomarkers (EBs) are induced changes, such as DNA methylation at specific loci, that can be used to detect phenotypic traits, environmental changes, pollution, treatments or diseases (U.S Food and Drug Administration; FDA; Anastasiadi and Beemelmans, 2023). These loci are sensitive to changes caused by the environment (internal or external) and such changes are relatively stable over time. Another relevant characteristic of EBs is that they can have diagnostic and prognostic value (Costa-Pinheiro *et al.*, 2015). This is because epigenetic changes in response to environmental cues can appear before the manifestation of phenotypic changes with potential consequences for performance. Thus, epigenetic DNA methylation-based biomarkers hold great promise in various fields such as medicine and disease diagnosis, as well as breeding programs for both plant and animal production (Turcotte *et al.*, 2022).

3.1 Development of epigenetic biomarkers for aquatic organisms

One of the most advanced fields involved in the development of EBs is that of cancer research. DNA methylation biomarkers of cancer are suitable for early detection across tissues and areas relevant to identification and cancer treatment. During the last 20 years, there has been a continuously increasing number of publications related to cancer EBs (Locke *et al.*, 2019). In this regard, the advances in high-throughput sequencing technologies and the large amounts of resources have allowed generating methylome databases from thousands of patients (Kamińska *et al.*, 2019; Locke *et al.*, 2019) such as The Cancer Genome Atlas (TCGA), MethHC 2.0 (Huang *et al.*, 2021), MethyCancer (He *et al.*, 2008), DNMIVD (Ding *et al.*, 2020), etc. Given the large amount of information being produced it is possible to use powerful machine learning (ML) algorithms for the classification of healthy or tumor cells and to classify the different subtypes (Cesaroni *et al.*, 2014; Guo *et al.*, 2015, 2021; Hosseini *et al.*, 2023; Huang *et al.*, 2018; Modhukur *et al.*, 2021; Wei *et al.*, 2006). However, the number of patients used in publications relevant to robust ML methods are in the order of hundreds or thousands (Arslan *et al.*, 2021; Bowler *et al.*, 2022; Rauschert *et al.*, 2020). Most of these studies rely on whole genome bisulfite sequencing of large number of samples, which can become prohibitive in other fields of research such as ecology and conservation biology, agronomy, etc. Alternatively, they rely on the use of high-throughput DNA methylation microarrays which, to date, have only been designed for the human

genome. Consequently, the advantages of employing powerful tools like machine learning algorithms have been rarely explored due to these limitations. Nevertheless, it remains unclear whether, in the context of production, such strategies could be adapted to implement epigenetic biomarkers that integrate with genetic breeding programs. Further research is needed to determine the feasibility and potential benefits of incorporating epigenetic biomarkers in previously existing breeding programs.

Recently, Anastasiadi and Beemelmanns (2023), described an approach for Epigenetic Biomarker Development Procedure (EBDP) with a focus on aquatic organisms (**Figure 6**). Briefly, the procedure consists in four phases: 1) Genome screening for candidate features, 2) Identification of strong candidate features, 3) Prediction of the outcome and 4) Validation of the methodology with external samples. Furthermore, during EBDP important considerations should be taken into account. Those are that EBs can depend on multiple factors causing biases (*e.g.*, cell, tissue, genetics, sex, age). Moreover, the epigenetic marks can change during tissue development, age and be affected by environmental factors. Finally, the number of samples to be used and the costs associated to sequencing technologies should be considered.

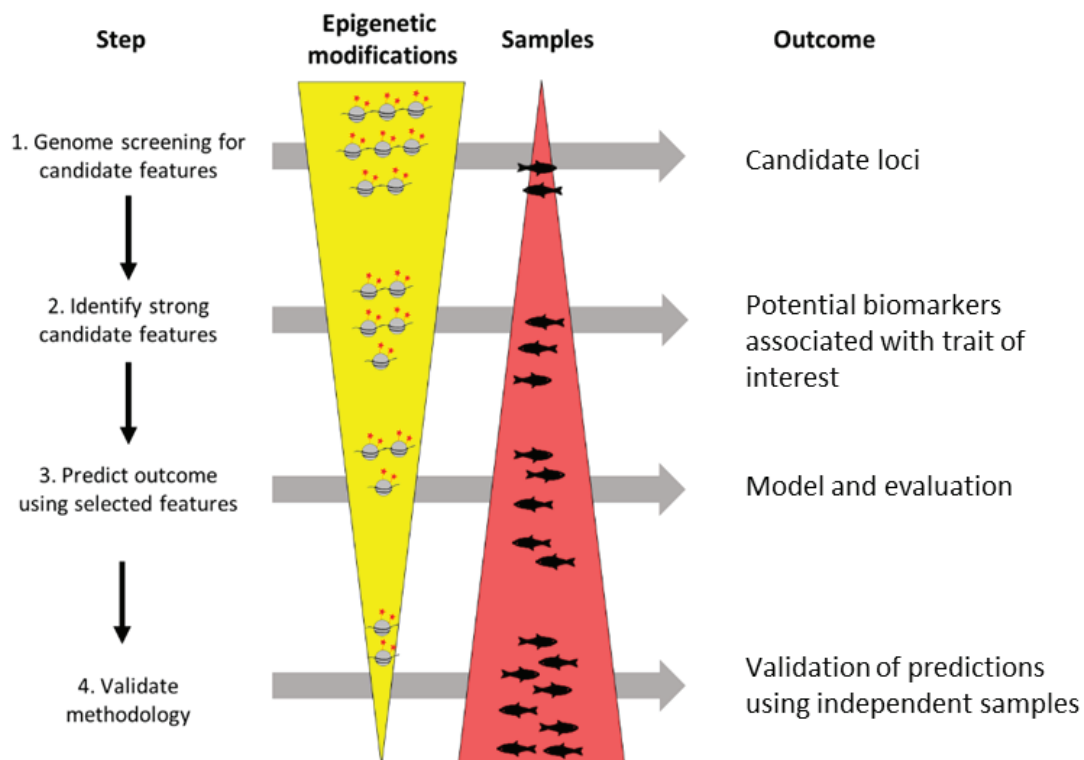


Figure 6. Diagram of the epigenetic biomarker development procedure (EBDP). The procedure is designed so that by targeting a smaller number of epigenetic loci and increasing the number of samples, optimization of resources can be achieved. The outcome expected from each step is described on the right panel. Modified from Anastasiadi and Beemelmanns, (2023).

3.2 Examples of epigenetic biomarkers

One of the applications of EBs is the development of epigenetic clocks. EBs have been used to build epigenetic clocks, a tool to accurately determine the age in humans (Horvath, 2013) and in other vertebrates, such as humpback whales (Polanowski *et al.*, 2014), birds (De Paoli-Iseppi *et al.*, 2019), reptiles (Mayne *et al.*, 2022), as well as several fishes like in the sea bass (Anastasiadi and Piferrer, 2020), Australian lungfish (Mayne *et al.*, 2021), medaka (Bertucci *et al.*, 2021) and the northern

red snapper (Weber *et al.*, 2022), among others (Anastasiadi and Piferrer, 2023; Piferrer and Anastasiadi, 2023).

In the European sea bass, several EBs have been developed. For example, to predict sex (Anastasiadi *et al.*, 2018b), to identify early stages of domestication (Anastasiadi and Piferrer, 2019), determine biological age (Anastasiadi and Piferrer, 2020), characterize the effect of heat waves (Anastasiadi *et al.*, 2021a) and to assess the stress associated with rearing density (Krick *et al.*, 2021).

4. The European sea bass model

4.1 Sea bass biology

The European sea bass is a marine perciform teleost (family *Moronidae*) in (Nelson *et al.*, 2016). The species' distribution extends from the Northeastern Atlantic, encompassing the area from Senegal to Norway, and includes the Mediterranean and the Black Seas (Kottelat and Freyhof, 2007). It is a eurythermal species with a broad range of temperature tolerance (8°C-27°C) (Pickett and Pawson, 1994). Its main environment is pelagic, but its habitat can also include coastal lagoons, estuaries and coastal waters. Additionally, it can be found in brackish and fresh waters. Juveniles can be found at shallow depth (~10 m), while adults inhabit coastal waters at a maximum of ~100 m depth (Frimodt, 1995). Adult sea bass has predator behavior. They mostly feed on small pelagic fish as sardines, sprats and sand smelts, crustaceans and squids, through a wide range of hunting techniques (Nelson *et al.*, 2016). This species exhibits female-biased sexual size dimorphism (SSD) during early development (Blázquez *et al.*, 1999; Díaz *et al.*, 2013; Saillant *et al.*, 2001; Vandeputte *et al.*, 2007). Semi-natural enclosures showed female-biased populations (Arias, 1980; Vandeputte *et al.*, 2012). Studies on 13 wild European sea bass populations showed a skew towards females overall, but balanced sex ratios at young ages. However, the results could be influenced by differential longevity (which ranges between 15 to 25 years), sampling bias, or differential fishing pressure (Freyhof and Kottelat, 2007; Vandeputte *et al.*, 2012).

Temperature has a profound effect on the European sea bass ontogenetic plasticity (Koumoundouros *et al.*, 2001). In the Mediterranean Sea, spawning takes place annually in the pelagic open sea, during winter, at temperatures above 9°C (Freyhof and Kottelat, 2007). Spawning occurs in batches, producing between 250,000 to 500,000 eggs per kilogram (Barnabe, 1974). Once the eggs are released to the water, early development occurs at a temperature range between 13°C and 17°C in the wild (Finn and Kristoffersen, 2007). Sea bass eggs and larvae are pelagic and planktonic, respectively (Finn and Kristoffersen, 2007). They float in the upper layers of water and experience temperature variation. During larval stages, it is known that even small temperature increases significantly changed global DNA methylation and the expression of genes related to DNA methylation, stress response, muscle and organ formation (Anastasiadi *et al.*, 2017). However, the long-term effects of temperature variations during embryo development on the transcriptome and methylome in later stages of life have not been evaluated. Juveniles move to brackish water and stay inshore until their second summer. Depending on the region, sexual maturity is reached at 2-4 years in the Mediterranean and 4-8 years in the Atlantic. In the wild, larvae hatch within 4-9 days post fertilization (dpf), depending on water temperature (Pickett and Pawson, 1994).

4.2 Sexual development of the sea bass

The sea bass is a gonochoristic species with PSD resulting from continuous genetic and environmental influences (Geffroy *et al.*, 2021; Piferrer *et al.*, 2005; Vandeputte *et al.*, 2007; M. Vandeputte and

Piferrer, 2019). The first 60 days after hatching are a crucial period called the thermos-sensitive period (TSP) during which temperature plays a critical role in determining the phenotypic sex of the fish (Navarro-Martín *et al.*, 2009a). The sex differentiation was reported to occur between 128 and 250 dpf (Saillant *et al.*, 2003a), however, the first signs of gonadal differentiation at the molecular level are found at 110 dpf, where the expression of several genes becomes sexually dimorphic: for example the markers: Hydroxysteroid 17-Beta Dehydrogenase 10 (*hsd17β10*), upregulated in females (Ribas *et al.*, 2019). Nevertheless, the first marker identified of ovarian differentiation is gonadal aromatase, *cyp19a1a*, expressed higher in females around 120 dpf (Blázquez *et al.*, 2008) and the process of sex differentiation is complete by 200 dpf. Currently there are no known biomarkers for early testis differentiation. Additionally, previous studies on the sea bass transcriptome only focused on identifying differentially expressed genes, and other approaches such as network construction have not been explored. In the sea bass Faggion *et al.*, (2021), found that female-biased sexual size dimorphism (SSD) was strongly influenced by early growth differences between sexes, coinciding with the standard length window and before gonadal sex differentiation started. This highlights the complex interlink between sex and growth traits in sea bass and raises the question of whether early growth is a cause or a consequence of sex differentiation in this species. In this species, differences in growth were found before molecular markers of sex (*i.e.*, *cyp19a1a*) and that growth rates during the sex differentiation period did not affect sex ratios (Díaz *et al.*, 2013). However, Geffroy *et al.*, (2021) found that energetic pathways are involved in SD, which could explain why females tend to exhibit higher energy levels and improved growth compared to males. Further research is needed to unravel the complex mechanisms underlying sex development and its relationship with growth in sea bass.

4.3 Environmental factors influencing sea bass sexual phenotype

As mentioned above, sea bass has high plasticity in response to environmental temperature. For instance, early-life temperature affects early embryonic development (Cucchi *et al.*, 2012), morphometric characteristics (Georgakopoulou *et al.*, 2007; Koumoundouros *et al.*, 2001), swimming capacity (Claireaux *et al.*, 2006), growth rate and metabolism (Koumoundouros *et al.*, 2002). These effects can be mediated through long-lasting changes in gene expression (Díaz and Piferrer, 2015).

4.4 Problems when reared in captivity

The European sea bass, ranks as the third most significant species for finfish aquaculture in Europe in terms of volume. In the Mediterranean region (including Turkey, Greece, Spain, Italy, Croatia, France, Cyprus and Portugal), it holds even greater importance as the most prominent species, with a production of over 255,019 tons in 2022 (FEAP, 2022, <https://feap.info/index.php/data/>). In aquaculture conditions, several traits are altered: behavior (Benhaïm *et al.*, 2012; Millot *et al.*, 2011), acute stress tolerance (Millot *et al.*, 2011), skeletal deformities (Barahona-Fernandes, 1982), morphology, organoleptic characteristics and other muscle properties (Arechavala-Lopez *et al.*, 2013; Valente *et al.*, 2013). Regarding reproduction in captivity, sea bass males mature earlier than females (Roblin and Bruslé, 1983) and earlier than in the wild (male mature around the second year, while females mature during the third year; precocious males and females can mature even one year earlier). Since keeping the water at low temperature increases the costs of production, the favored families for production are those that produce higher number of individuals with higher weight and resistant to the temperature effect. At high temperature (21°C) a proportion of the genetic females can become masculinized leading to male-biased sex ratios (Geffroy *et al.*, 2021; Navarro-Martín *et al.*, 2011, 2009; Piferrer *et al.*, 2005). During the thermo-sensitive period, high temperature is associated to increased methylation of the promoter of *cyp19a1a* in the gonads. This is linked with less expression of this gene, which leads fish that would develop as females to differentiate as males (Navarro-Martín *et al.*, 2011). Nonetheless, it remains

elusive whether the methylation profile of other genes is also affected by temperature and whether DNA methylation mechanism plays a crucial role underlying sex reversal on the sea bass or not. In summary, in the context of aquaculture, families with maximum survival, growth and increased number of females when reared at high temperature, resistance to disease and without deformities are preferred. Currently, genetic breeding programs have been established to selectively breed for the growth trait (Garcia de Leon *et al.*, 1998; Vandeputte *et al.*, 2019). Nevertheless, the rest of problems in captivity remain a challenge to which the study of epigenetics and the development of epigenetic biomarkers could hold the potential to their solution.

Aims

Aims

The overall objective of this thesis was to study the influence of intrinsic factors, such as genetic background and sex, as well as environmental factors, such as temperature, on the methylome of the European sea bass. We set out also to understand how DNA methylation marks are transmitted to offspring and how they are exhibited at different developmental stages. On a broader scale, we ultimately attempted to better understand the association of genotype with phenotype and environmental effects.

To achieve the overall objective, this thesis was organized into five research chapters with the following specific objectives:

1. To evaluate whether using only differentially expressed genes for weighted co-expression network analysis or combining the entire dataset with a subsequent filter of differentially expressed genes yields more biologically meaningful results. This could aid in selecting the optimal approach for analyzing the gonadal transcriptome of the European sea bass during early sex differentiation.
2. To compare the gonadal transcriptome dynamics during sex differentiation of six vertebrate species, ranging from fish to mammals. Instead of comparing females to males, our analysis aimed to identify genes involved in sex differentiation of each species and to identify conserved potential novel markers of early sex differentiation by examining gene expression changes between two developmental stages within each sex separately.
3. To propose and assess an approach for the development of epigenetic biomarkers with both diagnostic and prognostic value for finfish production. Specifically, we tested whether DNA methylation changes can be detected before transcriptomic changes become evident in latter stages of life. To do so we investigated the effects on the methylome and the transcriptome in fish exposed to different thermal regimes during early development in sea bass.
4. To demonstrate the feasibility of utilizing machine learning for epigenetic biomarker development despite the limitation of sequencing hundreds of samples in the context of animal production. We tested if by implementing genome screening in a carefully selected small batch of samples combined with multiple filtering, evaluation and validation steps could lead to biomarkers indicators of maximum biomass and the desired sex, thereby integrating selection through epigenetic biomarkers alongside classical genetic selection.
5. To investigate the sea bass gonadal methylome, exploring both intrinsic (sex) and extrinsic (temperature) effects. Additionally, we sought to unveil the epigenetic signature of sex reversal in a species with polygenic sex determination system and to quantify the proportion of temperature-induced epimarks transmitted from the parental (in the sperm) to the first filial (in the gonads) generation.

Report of the supervisor

Report of the supervisor

Francesc Piferrer, supervisor of the Ph.D. thesis entitled “*Development of DNA methylation biomarkers to study epigenetic inheritance and integration of environmental stimuli*” by Núria Sánchez Baizán, certifies that the results obtained have been or will be submitted to peer-reviewed international journals. The entire thesis comprises five papers: two of them already published and three in preparation. The details of the papers, journal and their impact factor (Journal Citation Reports) are detailed below:

Chapter 1

“Improved biomarker discovery through a plot twist in transcriptomic data analysis”. Sánchez-Baizán, N., Ribas, L. and Piferrer, F. *BMC biology*, 20(1), pp.1-26. doi.org/10.1186/s12915-022-01398-w.

Impact factor (2022): 7.364

Journal Citation Reports:

Area: Biology

Rank: 10/94

Quartile: Q1

The Ph.D. candidate contributed in the design of the study, conducted the data analysis did all the graphs, participated in the interpretation of the results and in writing the manuscript.

Chapter 2

“Extraordinary variability in gene expression dynamics during gonadal sex differentiation across vertebrates”. Sánchez-Baizán, N., Jarne-Sanz I., Roco, A. S., Scharl, M., Piferrer F. *In preparation*.

The Ph.D. candidate participated in the design of the study, searched and selected the datasets to be analyzed. She also performed the analysis of the data and performed some of the graphs. She also participated in the discussion of the results, discussion and writing of the manuscript.

Chapter 3

“Development of epigenetic biomarkers with diagnostic and prognostic value to assess the lasting effects of early temperature changes in farmed fish”. Valdivieso, A., Sánchez Baizán, N., Mitrizakis, N., Papandroulakis, N., Piferrer, F. *Aquaculture*, 563:738918. doi.org/10.1016/j.aquaculture.2022.738918

Impact factor (2023): 5.135

Journal Citation Reports:

Area: Marine & Freshwater Biology

Rank: 8/113

Quartile: Q1

The Ph.D. candidate participated in data analysis, did some of the graphs and contributed to the interpretation of the results and in writing the manuscript.

Chapter 4

“A cost-effective method for complementing genetic selection with epigenetic selection in fish production”. Sánchez-Baizán, N., Herlin, M., López, M., Piferrer F. *In preparation*.

The Ph.D. participated in the sampling of the experiment. She also prepared the libraries for reduced representation bisulfite sequencing and multiplex bisulfite sequencing and designed the primers for bisulfite-converted PCR. She did the bioinformatic analysis of the data, all the graphs and diagrams, participated in the interpretation of results, discussion and writing of the manuscript.

Chapter 5

“Multigenerational temperature additive effects: the epigenetic regulation of masculinization and its inheritance in the European sea bass”. Sánchez-Baizán, N., Allal, F., Denkena, J., Colomé-Tatché, M., Vandeputte M., Piferrer F. *In preparation*.

The Ph.D. candidate participated in the sampling of the experiment. She also prepared the libraries for reduced representation bisulfite sequencing. She did the bioinformatic analysis of the data, all the graphs, participated in the interpretation of results, discussion and writing of the manuscript.

The director also informs, to the best of his knowledge, that colleagues that appear as co-authors in some of the papers included in this have not used, implicitly or explicitly, the data or these articles for the elaboration of other Ph.D. thesis.

Barcelona, July 27, 2023.

Institut de Ciències del Mar (ICM-CSIC)

Dr. Francesc Piferrer
Director

Results

Block 1

*Transcriptomics during sex
differentiation*

Chapter 1

*Improved biomarker discovery through a plot
twist in transcriptomic data analysis*

BMC Biology (2022)

20(1), pp.1-26

doi.org/10.1186/s12915-022-01398-w

Improved biomarker discovery through a plot twist in transcriptomic data analysis

Núria Sánchez Baizán, Laia Ribas and Francesc Piferrer

Institut de Ciències del Mar (ICM), Spanish National Research Council (CSIC), Barcelona, Spain

Abstract

Background: Transcriptomic analysis is crucial for understanding the functional elements of the genome, with the classic method consisting of screening transcriptomics datasets for differentially expressed genes (DEGs). Additionally, since 2005, weighted gene co-expression network analysis (WGCNA) has emerged as a powerful method to explore relationships between genes. However, an approach combining both methods, *i.e.*, filtering the transcriptome dataset by DEGs or other criteria, followed by WGCNA (DEGs + WGCNA), has become common. This is of concern because such approach can affect the resulting underlying architecture of the network under analysis and lead to wrong conclusions. Here, we explore a plot twist to transcriptome data analysis: applying WGCNA to exploit entire datasets without affecting the topology of the network, followed with the strength and relative simplicity of DEG analysis (WGCNA + DEGs). We tested WGCNA + DEGs against DEGs + WGCNA to publicly available transcriptomics data in one of the most transcriptomically complex tissues and delicate processes: vertebrate gonads undergoing sex differentiation. We further validate the general applicability of our approach through analysis of datasets from three distinct model systems: European sea bass, mouse, and human.

Results: In all cases, WGCNA + DEGs clearly outperformed DEGs + WGCNA. First, the network model fit and node connectivity measures and other network statistics improved. The gene lists filtered by each method were different, the number of modules associated with the trait of interest and key genes retained increased, and GO terms of biological processes provided a more nuanced representation of the biological question under consideration. Lastly, WGCNA + DEGs facilitated biomarker discovery.

Conclusions: We propose that building a co-expression network from an entire dataset, and only thereafter filtering by DEGs, should be the method to use in transcriptomic studies, regardless of biological system, species, or question being considered.

Keywords: Gene expression analysis, Gene networks, Weighted gene co-expression network analysis (WGCNA), Sex determination and differentiation, Gonadal development, Biomarker discovery

1. Introduction

The transcriptome is defined as the entire set of messenger RNA (mRNA or transcripts) expressed by a cell or tissue type of an organism of a given genotype under certain internal and external influences. The total amount and types of transcripts vary depending on multiple factors such as stage of development, physiological state, or environmental conditions (Lowe *et al.*, 2017). Transcriptomics, the analysis of the whole transcriptome, started using complementary DNA (cDNA) clones to generate expressed sequence tags (ESTs), and the development of the first microarrays in the early 90s (Adams *et al.*, 1991; Piétu *et al.*, 1999; Velculescu *et al.*, 1997). An accurate comprehension of the transcriptome is essential to reveal the molecular constituents of cells and tissues, to understand the functional elements of the genome, and also to gain a better understanding of development and disease (Wang *et al.*, 2009). With the advent of high-throughput technologies to analyze the transcriptome, several bioinformatic challenges appeared in storing, pre-processing, and analyzing the large datasets produced. As a consequence, a variety of bioinformatic pipelines arose to overcome such challenges (Parmigiani *et al.*, 2003; Voshall and Moriyama, 2018). To date, the most commonly used method in various fields of biology and medicine is screening for differentially expressed genes (DEGs), which compares the mean expression levels for individual genes between two or more groups of samples (Ayers *et al.*, 2015; Clement *et al.*, 2007; Czerwinski *et al.*, 2016; Dergunova *et al.*, 2018; Lata *et al.*, 2010; Men *et al.*, 2017; Ribas *et al.*, 2016; Yu *et al.*, 2016).

In 2005, Zhang and Horvath proposed a bioinformatic application called weighted gene co-expression network analysis (WGCNA) to gain a deeper understanding of the transcriptome and to elucidate the underlying cellular processes based on the coordinated co-expression of genes encoding the interacting proteins (Zhang and Horvath, 2005). In contrast to DEG analysis, WGCNA is a gene screening method that takes advantage of the inherent variability in gene expression among different biological samples to illuminate higher-order relationships among genes. WGCNA generates clusters (called modules) of strongly correlated genes based on the Pearson correlation as a measure of their functional relatedness and assigns a different color to each module for easy identification purposes. WGCNA is a powerful method and has many advantages over DEG analysis since it also allows evaluating the association of modules with phenotypic sample traits using network properties. Furthermore, WGCNA facilitates the identification of candidate biomarkers and hub genes relevant to the process under study. The weighted co-expression network consists of an adjacency matrix reporting the connection strength between gene pairs (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). Since its development, its use has become exponentially widespread and allows integrating network parameters with genetic information from microarray datasets and, more recently, from RNA sequencing experiments (Duan *et al.*, 2015; Iancu *et al.*, 2015; Liu *et al.*, 2021). It should be noted that variations or methods other than WGCNA have also been developed, such as Differential Co-expression Analysis or metaDCN (Zhu *et al.*, 2017), THD-Module Extractor (Kakati *et al.*, 2016), Diffcoex (Tesson *et al.*, 2010), and module differential analysis for weighted gene co-expression network (MODA) (Li *et al.*, 2022). However, WGCNA remains by far the most commonly used in numerous research fields.

In reviewing the literature extensively, we found three main strategies to analyze the transcriptome (**Figure 1**) that can be broadly classified as (1) screening for differentially expressed genes between two conditions (from now on referred to as method #1 or simply as DEGs), (2) WGCNA and derivatives (from now on referred to as method #2 or simply as WGCNA), and (3) a method that has become also very popular that can be viewed as a combination of the two former and consists in filtering the transcriptome dataset by DEGs or other criteria, *e.g.*, considering only the most expressed genes, or the top 25% genes with more expression variance, etc., and only then applying WGCNA to the filtered dataset (from now on referred to as method #3 or DEGs + WGCNA). Another strategy that we have found in reviewing

the literature is the independent use of methods #1 (DEGs) and #2 (WGCNA) in the same study but then using only one of the two for the subsequent downstream analysis on the data (Ribas *et al.*, 2016; Zhao *et al.*, 2018; Gupta *et al.*, 2014; Deist *et al.*, 2017; Amiri *et al.*, 2018; Chen *et al.*, 2019; Wang *et al.*, 2019; Niemira *et al.*, 2020; Tu *et al.*, 2021). Some examples of studies following each one of the three broad methods defined above are shown in **Table 1**. Of note, regardless of the method used, usually, these analyses are followed by data visualization and functional analysis (gene ontology and/or pathways enrichment analysis).

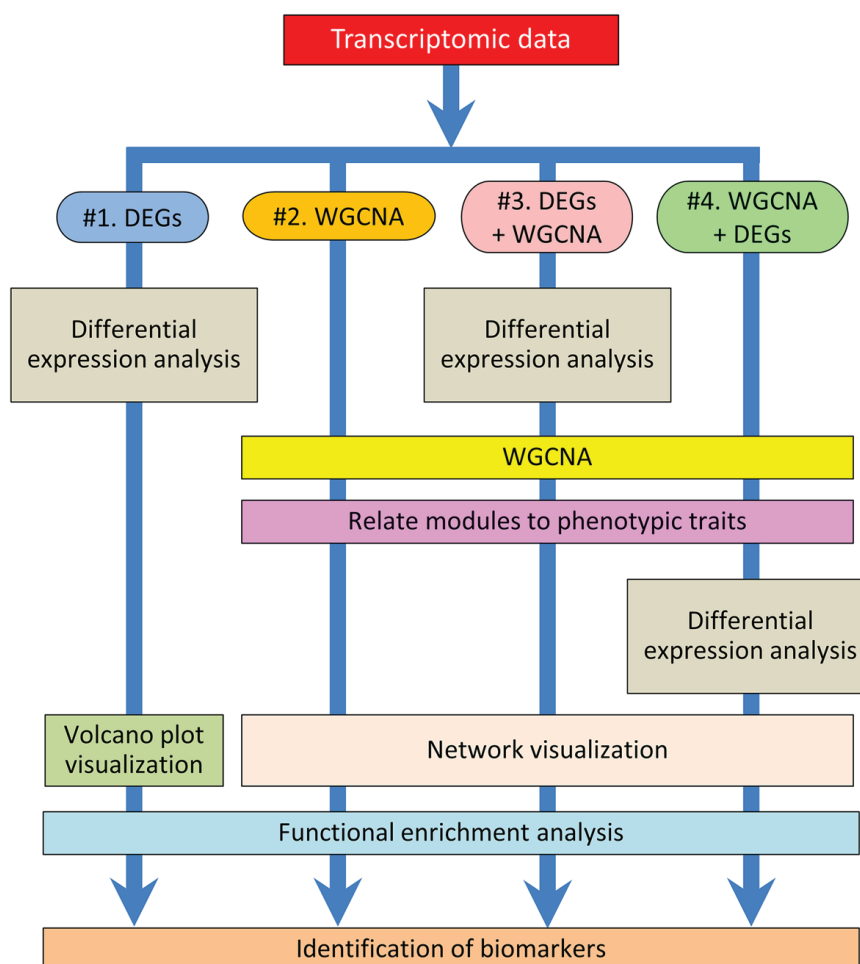


Figure 1. Flow diagram of the three most used methods in the literature to analyze transcriptomic data plus one method (#4) proposed. Method #1 Analysis of the differentially expressed genes (DEGs) between two conditions followed by volcano plot visualization, functional enrichment analysis (FEA), and biomarker identification. #2 Construction of a gene co-expression network by using the WGCNA package (Langfelder y Horvath 2008) in R, following the general WGCNA guidelines to perform FEA, identify modules and genes related to a trait of interest (Zhang and Horvath 2005). #3 To filter the transcriptome by DEGs and conduct WGCNA using the filtered dataset and continue with FEA. In this study, we propose method #4 to use WGCNA and filter the output of the selected modules by DEGs to follow with FEA and biomarker discovery.

The emergence of method #3 (DEGs + WGCNA) apparently seems a logical forward step to take: by first filtering DEGs, the analysis is limited to a set of previously selected genes, based on a statistical or quantitative threshold and, thus, it requires less computational power than the required by WGCNA of the entire dataset. However, and this is very important for the proper analysis of gene interactions, in a scale-free network (*i.e.*, a network whose characteristics are independent of the size or number of nodes making up the network) such as the gene networks, the fraction of nodes with degree k , where k is

connectivity (*i.e.*, the sum of connections of a node), follows a power-law $k-\alpha$ distribution, where α is some exponential. In this manner, the network topology is dominated by a few highly connected nodes, called hubs, linked to the rest of less connected nodes (Zhang and Horvath, 2005). Therefore, applying a filtering step such as DEGs previous to WGCNA, as done in method #3, might eliminate many of the less connected genes that, if not filtered, would contribute to some nodes being identified as hubs. This is of great concern because it thus could affect the core architecture of the network and lead to biased results and interpretations. In contrast, construction of a network with the unfiltered dataset allows drawing the complete map of the network.

As we have just seen, method #3 has the main advantage that one ends up dealing with a much-reduced gene list, albeit all genes are DEGs, but it has serious dangers that can affect the architecture of the networks and lead to false conclusions. Furthermore, the validity of method #3 was never established and, in fact, the developers of method #2 (WGCNA) recommend against the application of such filter before WGCNA (Langfelder and Horvath, 2017). Thus, to combine the benefits from WGCNA, taking advantage of the whole dataset and not affecting the topology of the network, with the strength and relative simplicity of DEG analysis, in the present study we propose a sort of plot twist in transcriptome data analysis. We propose a new strategy (referred to as method #4): to perform WGCNA using the entire transcriptome dataset and, only subsequently, filtering by DEGs (**Figure 1**). To validate this new method, we compared it with the performance of method #3, which is, as mentioned above, nowadays much spread in the literature.

For the purpose of this study, we selected one of the most transcriptomically complex tissues, the gonads (Shao *et al.*, 2014) and one of the most complicated developmental processes that involve many genes interacting in a delicate spatio-temporal and orchestrated manner: vertebrate gonadal sex differentiation. At the bi-potential stage, the vertebrate gonad expresses genes from the pro-male (leading to testis differentiation) and pro-female (leading to ovarian differentiation) pathways with opposing effects until the time when, depending on the species, a combination of genetic and/or environmental influences switch the balance towards the differentiation of one sexual phenotype and the repression of the other (Capel, 2017). Thus, vertebrate gonadal sex differentiation provides an excellent opportunity to compare different methods to analyze large amounts of gene expression data. Furthermore, we used data from the European sea bass, *Dicentrarchus labrax* (from now on sea bass), a modern teleost with a polygenic sex-determining (PSD) mechanism without sex chromosomes (Vandeputte and Piferrer, 2019), the mouse model, where gonadal sex differentiation has been extensively studied, and the human, a species where despite the limited availability of normal fetal gonads, it has been extensively studied for better understanding of the complex pathways related to disorders of sexual development (DSDs) (Hughes *et al.*, 2006; Lecluze *et al.*, 2020). As with most mammals, mouse and human have an XX/XY sex determining system where SRY is the master male sex determining gene (Lovell-Badge and Robertson, 1990). These three species with different sex-determining mechanism and differentiation dynamics were used to show that our approach was not only working on a complex process but also to show that the results are consistent regardless of the underlying genetic architecture. In all cases, we show that method #4 provides a more realistic and nuanced picture of the complex gene interactions taking place over time during gonadal sex differentiation and propose this method for future transcriptomic studies regardless of the biological system or question being considered.

Table 1. Examples of some transcriptomic studies classified according to the three different methods described (#1, DEGs; #2, WGCNA; # 3, DEGs + WGCNA; # 1+ 2 indicates the use of both methods, without combining them, as in #3) to analyze transcriptomic data since the development of WGCNA in 2005

Method #	Species	Cell type/ Tissue	Technology	Reference
1	<i>Rattus norvegicus</i>	gonads	Microarray	Clement <i>et al.</i> , 2007
1	<i>Setaria italica</i> L.	plant seeds	SSH	Lata <i>et al.</i> , 2010
1	<i>Gallus gallus</i>	gonads	RNA-seq	Ayers <i>et al.</i> , 2015
1	<i>Trachemys scripta</i>	embryos	RNA-seq	Czerwinski <i>et al.</i> , 2016
1	<i>Triticum aestivum</i> L.	plant seeds	Microarray	Yu <i>et al.</i> , 2016
1	<i>Homo sapiens</i>	breast	RNA-seq	Men <i>et al.</i> , 2017
1	<i>Xenopus laevis</i>	gonads	Microarray	Piprek <i>et al.</i> , 2018
1	<i>Rattus norvegicus</i>	brain	RNA-seq	Dergunova <i>et al.</i> , 2018
1	<i>Dicentrarchus labrax</i>	gonads	Microarray	Ribas <i>et al.</i> , 2019
2	<i>Homo sapiens</i>	brain	Microarray	Oldham <i>et al.</i> , 2008
2	<i>Homo sapiens</i>	brain	Microarray	Miller <i>et al.</i> , 2010
2	<i>Homo sapiens</i>	brain	RNA-seq	Iancu <i>et al.</i> , 2015
2	<i>Homo sapiens</i>	bone	Microarray	Chen <i>et al.</i> , 2016
2	<i>Homo sapiens</i>	liver	Microarray	Yin <i>et al.</i> , 2018
2	<i>Mus musculus</i>	placenta	Microarray	Cox <i>et al.</i> , 2019
2	<i>Homo sapiens</i>	brain	Microarray	Feltrin <i>et al.</i> , 2019
2	<i>Oplegnathus fasciatus</i>	spleen	RNA-seq	Kim <i>et al.</i> , 2020
2	<i>Homo sapiens</i>	umbilical cord, blood	RNA-seq	Wang <i>et al.</i> , 2020
2	<i>Homo sapiens</i>	lung	Microarray	Guillotin <i>et al.</i> , 2021
2	<i>Homo sapiens</i>	bladder	RNA-seq	Liu <i>et al.</i> , 2021
3	<i>Homo sapiens</i>	bladder	Microarray	Deng <i>et al.</i> , 2015
3	<i>Homo sapiens</i>	Colon mucose	Microarray	Lin <i>et al.</i> , 2018
3	<i>Homo sapiens</i>	podocyte cells	Microarray	Zuo <i>et al.</i> , 2018
3	<i>Bubalus bubalis</i>	mammary gland	RNA-seq	Deng <i>et al.</i> , 2019
3	<i>Homo sapiens</i>	bladder	Microarray	Di <i>et al.</i> , 2019
3	<i>Homo sapiens</i>	lung	Microarray	Ding <i>et al.</i> , 2019
3	<i>Homo sapiens</i>	bladder	RNA-seq	Pan <i>et al.</i> , 2019
3	<i>Bubalus bubalis</i>	blood	RNA-seq	Liu <i>et al.</i> , 2019
3	<i>Homo sapiens</i>	brain	RNA -seq	Kakati <i>et al.</i> , 2019
3	<i>Homo sapiens</i>	liver	RNA-seq	Bai <i>et al.</i> , 2020
3	<i>Scophthalmus maximus</i>	gill	RNA-seq	Cui and Ma, 2020
3	<i>Scophthalmus maximus</i>	kidney	RNA-seq	Huang <i>et al.</i> , 2020
3	<i>Homo sapiens</i>	breast	Microarray	Zhang <i>et al.</i> , 2020
3	<i>Rattus norvegicus</i>	Spinal cord	RNA-seq	Jiang <i>et al.</i> , 2020
3	<i>Homo sapiens</i>	stem cells	RNA-seq	Zhu <i>et al.</i> , 2021
3	<i>Bos taurus</i>	blood	RNA-seq	Zhao <i>et al.</i> , 2021
1+2	<i>Scophthalmus maximus</i>	gonads	Microarray	Ribas <i>et al.</i> , 2016
1 + 2	<i>Paralichthys olivaceus</i>	embryo	RNA-seq	Zhao <i>et al.</i> , 2018
1+2	<i>Homo sapiens</i>	brain	RNA-seq	Gupta <i>et al.</i> , 2014
1+2	<i>Gallus gallus</i>	lung	RNA-seq	Deist <i>et al.</i> , 2017
1+2	<i>Homo sapiens</i>	organoids	RNA-seq	Amiri <i>et al.</i> , 2018
1+2	<i>Homo sapiens</i>	placenta	Microarray	Chen <i>et al.</i> , 2019
1+2	<i>Mus musculus</i>	gonads	RNA-seq	Wang <i>et al.</i> , 2019
1+2	<i>Homo sapiens</i>	lung	RNA-seq	Niemira <i>et al.</i> , 2020
1+2	<i>Mus musculus</i>	pancreas	Microarray	Tu <i>et al.</i> , 2021

2. Results

2.1 Construction of WGCNA using different approaches

WGCNA was conducted on male and female samples at two developmental stages (sea bass and mouse) or three developmental stages (human) to investigate the gene network operating during gonadal development in the sea bass, mouse, and human (**Supplementary Figure 1**).

We applied method #3 to the sea bass transcriptome, composed of a total of 20,978 genes, and identified 8,434 DEGs between males and females at 250 days post fertilization (dpf) ($P < 0.05$) (**Supplementary Figure 2**). The normalized intensities of the 8434 DEGs (**Additional file 1**) were then used for WGCNA to build the network. The selected soft threshold used for the adjacency function closest to meet scale-free topology criterion and additional considerations was $\beta = 9$, leading to signed $R^2 = 0.7$ (**Figure 2A**). Regarding connectivity, the second parameter considered for the selection of the soft threshold indices, the mean value was < 1000 (**Figure 2B**) while the slope of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$ was -0.47 ($R^2 = 0.73$) (**Figure 2C**). In contrast, when the entire transcriptome was used (methods #2. WGCNA and #4. WGCNA + DEGs), the selected soft threshold obtained was $\beta = 5$, which resulted in a better fit, signed $R^2 = 0.87$ (**Figure 2D**), and more than doubled the mean connectivity to ~ 2000 (**Figure 2E**) while also improving the slope $-\gamma = -0.51$ ($R^2 = 0.82$) (**Figure 2F**).

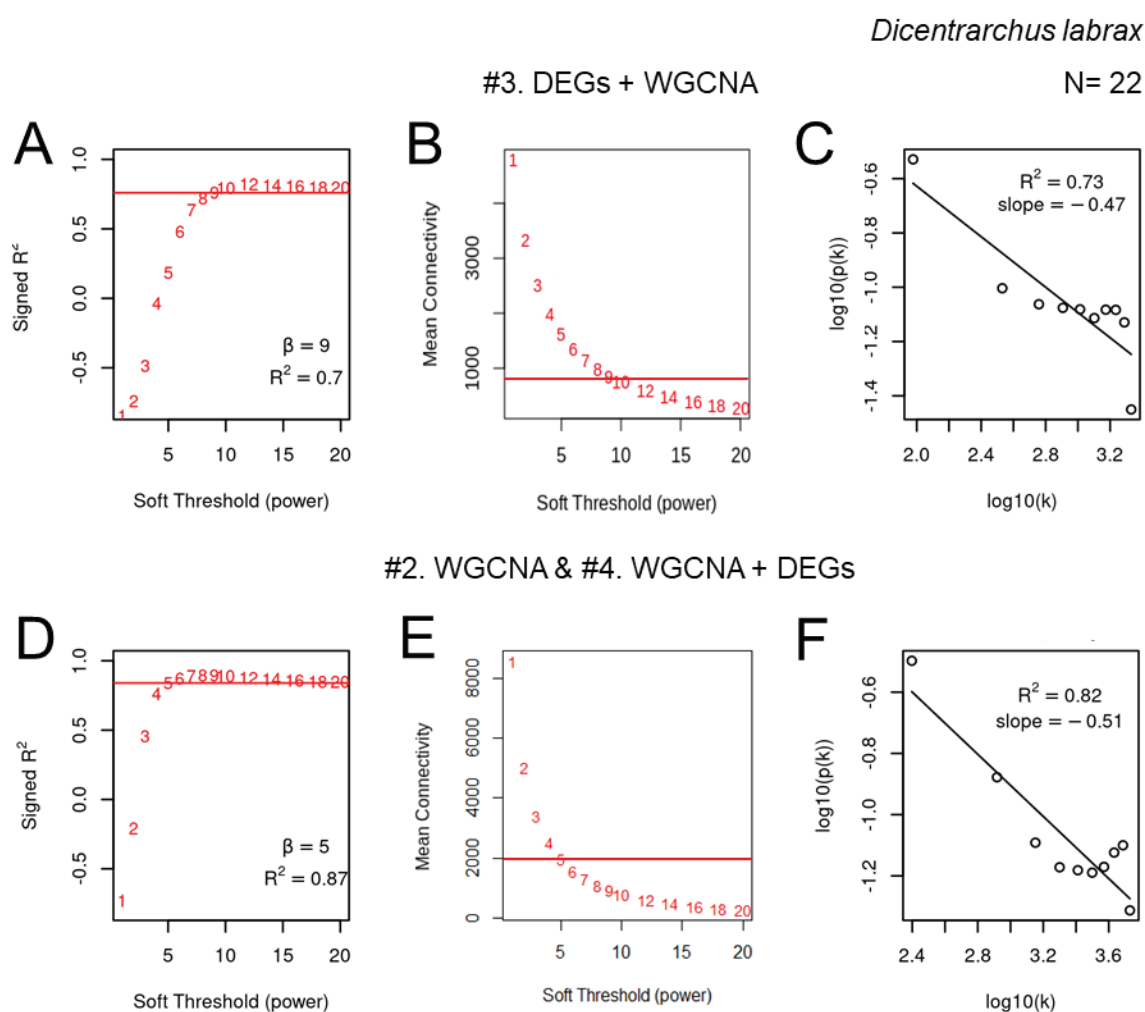


Figure 2. WGCNA model fit and soft threshold determination. Soft thresholding power analysis was used to obtain the scale-free fit index (ranging from 1 to 20) of network topology, for method #3 (A) and #2 and #4 (D), which until that step of the working pipeline are the same. Mean connectivity when using method #3 (B) and method #4

(E). The plot of $\log_{10}(p(k))$ vs. $\log_{10}(k)$ indicates that by using the chosen β value, C method #3 ($\beta = 9$) and F method #4 ($\beta = 5$), the network is close to a scale-free network because it is approximately following a straight line. k is the whole network connectivity and $p(k)$ is the corresponding frequency distribution.

The same difference between the two methods was observed in the mouse. Analysis of the gonadal transcriptome between males and females at 16.5 days post coitum (dpc) ($P < 0.05$) identified 8,109 DEGs (**Additional file 2**). Using method #3, the selected soft threshold was $\beta = 14$ which resulted in a signed $R^2 = 0.55$ (**Supplementary Figure 3A**), a mean connectivity value < 500 (**Supplementary Figure 3B**), and slope $-\gamma = -0.79$ ($R^2 = 0.51$) (**Supplementary Figure 3C**). However, when the entire gene expression dataset (14,088 genes) were used for the construction of the network (methods #2. WGCNA and #4. WGCNA + DEGs), the soft threshold was $\beta = 12$, resulting in signed $R^2 = 0.7$ (**Supplementary Figure 3D**), the mean connectivity ~ 500 (**Supplementary Figure 3E**), and the slope improved to $-\gamma = -0.88$ ($R^2 = 0.55$) (**Supplementary Figure 3F**).

The discrepancies between the methods were replicated using human data as well, where using method #3 the selected soft threshold was $\beta = 8$, resulting in signed $R^2 = 0.67$ (**Supplementary Figure 4A**), a mean connectivity value < 50 (**Supplementary Figure 4B**), and slope $-\gamma = -1.02$ ($R^2 = 0.63$) (**Supplementary Figure 4C**). With methods #2 WGCNA and #4 WGCNA + DEGs, the soft threshold was $\beta = 8$, resulting in signed $R^2 = 0.89$ (**Supplementary Figure 4D**), the mean connectivity > 200 (**Supplementary Figure 4E**), and the slope improved to $-\gamma = -1.38$ ($R^2 = 0.87$) (**Supplementary Figure 4F**).

To ensure that the improvement in the model fit were not caused by the selection of a particular soft threshold, results were calculated for a range of thresholds (**Table 2**). In the three species, method #4 outperformed method #3 when it referred to scale-free topology model fit, mean connectivity, and the slope of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$ being closer to -1 , regardless of the soft threshold selected.

2.2 Selection of modules related to sex

Because method #3 uses a smaller dataset than method #4, a different number of modules were produced using the average linkage hierarchical clustering algorithm in WGCNA. Thus, in the sea bass the 8,434 DEGs retained in method #3 were grouped into eight modules, of which two were strongly associated with the trait of interest, sex (red module: $R^2 = 0.93$, $P = 7e - 10$; blue module: $R^2 = -0.84$, $P = 1e - 06$) (**Figure 3A, C**). On the other hand, the 20,978 genes of methods #2 and #4 were grouped into 29 modules, of which four were strongly associated with sex (pink, $R^2 = 0.95$, $P = 7e - 12$; green, $R^2 = -0.79$, $P = 1e - 05$; blue sky, $R^2 = -0.8$, $P = 7e - 06$; magenta $R^2 = -0.72$, $P = 2e - 04$) (**Figure 3B, D**).

In the mouse, the 8,109 DEGs retained in method #3 were clustered into six modules, of which two were associated with sex (blue: $R^2 = 0.69$, $P = 0.01$; turquoise: $R^2 = -0.8$, $P = 0.002$). Using the entire transcriptome in methods #2 and #4, a total of 21 modules were obtained, of which two were associated with sex (red: $R^2 = 0.98$, $P = 1e - 07$; brown: $R^2 = -0.75$, $P = 0.008$) (**Supplementary Figure 5**).

The 1,479 differentially expressed transcripts at 6 postconceptional weeks (PCW) in human were grouped into eight modules, among which three of them were positively associated with sex (red: $R^2 = 0.64$, $P = 9e - 05$; blue: $R^2 = 0.8$, $P = 3e - 08$; yellow: $R^2 = 0.87$, $P = 2e - 10$), while one module was negatively associated with sex (turquoise: $R^2 = -0.91$, $P = 8e - 13$). In contrast, the analysis of 35,194 transcripts using method #4 resulted in 36 modules, of which seven modules were associated with sex

Table 2. Comparison of model fit parameters under different soft thresholds using methods #3 (DEGs+WGCNA) and #4 (WGCNA+DEGs) on the gonadal transcriptome during early sex differentiation in three different species, sea bass, mouse and human. The selected threshold and results obtained with the selected threshold are shown in bold

	Method	Threshold	Signed R ² (model fit)	Mean connectivity	slope (scale free topology)	R ²
<i>Dicentrarchus labrax</i>	#3 DEGs+WGCNA	5	0.19	1580	-0.12	0.08
		7	0.63	1100	-0.33	0.59
		9	0.7	800	-0.47	0.73
		11	0.8	650	-0.56	0.79
		13	0.8	450	-0.64	0.81
	#4 WGCNA+DEGs	5	0.89	2050	-0.51	0.82
		7	0.9	1400	-0.63	0.87
		9	0.91	900	-0.71	0.88
		11	0.9	700	-0.78	0.87
		13	0.9	500	-0.83	0.86
<i>Mus musculus</i>	#3 DEGs+WGCNA	8	0.13	540	0.02	-0.3
		10	0.33	375	-0.5	0.24
		12	0.47	300	-0.65	0.4
		14	0.55	230	-0.79	0.51
		16	0.62	180	-0.89	0.57
	#4 WGCNA+DEGs	8	0.62	495	-0.62	0.4
		10	0.67	325	-0.77	0.49
		12	0.7	220	-0.88	0.55
		14	0.7	200	-0.96	0.59
		16	0.73	160	-1.02	0.62
<i>Homo sapiens</i>	#3 DEGs+WGCNA	4	0.45	105	-0.74	0.38
		6	0.61	50	-0.93	0.56
		8	0.67	25	-1.02	0.63
		10	0.68	19	-1.1	0.64
		12	0.69	15	-1.17	0.65
	#4 WGCNA+DEGs	4	0.7	950	-1.3	0.66
		6	0.79	410	-1.36	0.77
		8	0.89	210	-1.38	0.87
		10	0.93	100	-1.42	0.92
		12	0.95	85	-1.45	0.95

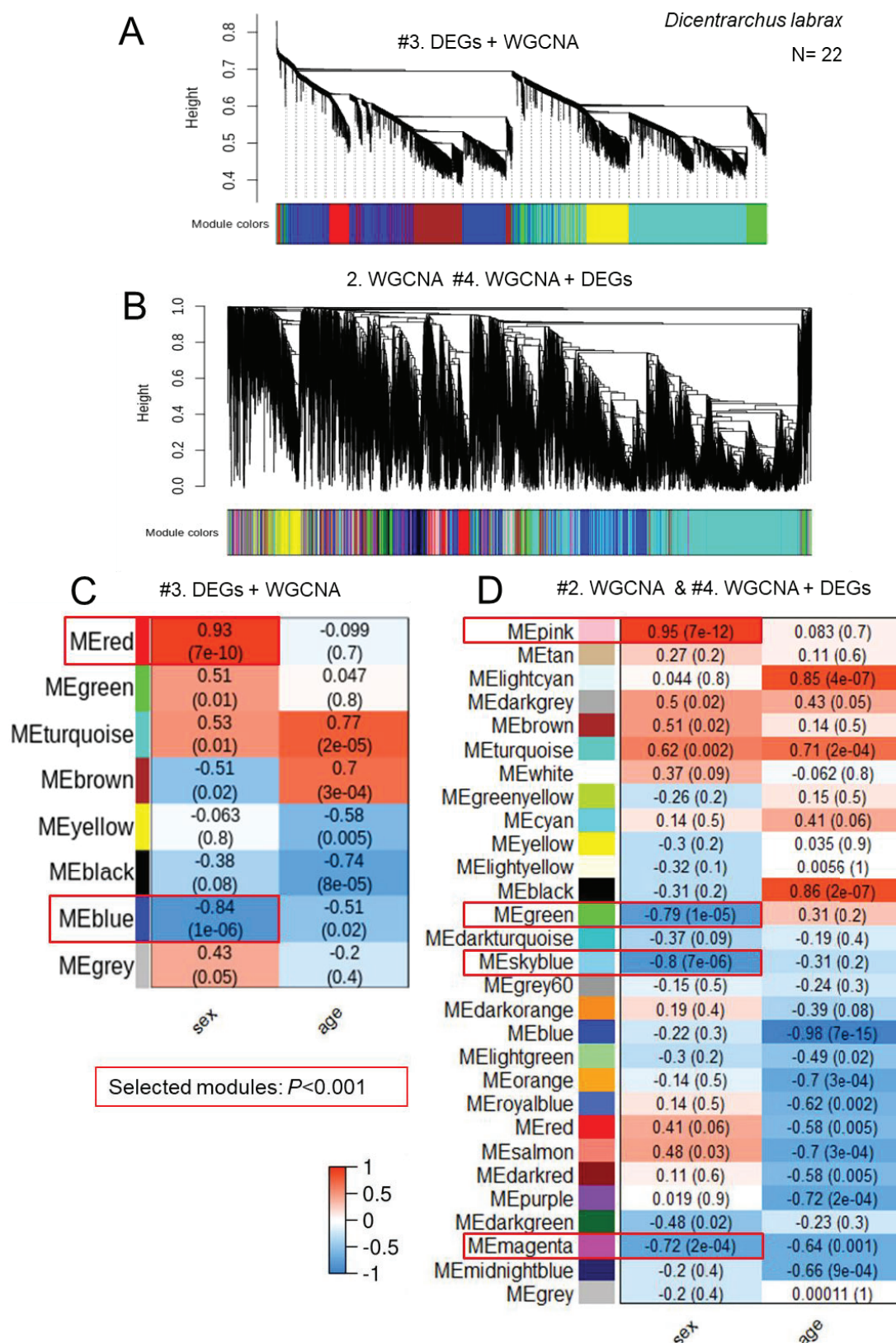


Figure 3. Identification of gene modules associated with sex. Gene hierarchical cluster analysis using method #3 (A) and #4 (B) using the sea bass gonadal transcriptome. Heat map of the correlation of sex with module eigengene distances using method #3 (C) and #4 (D). Each color represents a module in the constructed gene co-expression network by WGCNA. The heat map is colored from red (1, positive) to blue (-1, negative) to indicate the level of correlation of each module with the trait of interest. The red boxes highlight the selected modules for further analysis associated with sex ($P < 0.01$) (color in the online version). Genes not assigned to any of the previous modules are included in the grey module.

(light-yellow: $R^2 = -0.6$, $P = 3e-04$; black: $R^2 = -0.85$, $P = 5e-10$; dark olive-green: $R^2 = -0.57$, $P = 6e-04$; red: $R^2 = -0.72$, $P = 4e-06$; dark red: $R^2 = -0.66$, $P = 3e-05$; green: $R^2 = 0.72$, $P = 4e-06$; green-yellow: $R^2 = 0.78$, $P = 2e-07$) (**Supplementary Figure 6**).

Next, considering only the genes from the modules that significantly associated with sex, we selected those that, in addition, showed a significant correlation between modular membership and the gene significance for sex. All the modules associated with sex in the sea bass using methods #3 (**Figure 4A**), #2, and #4 (**Figure 4B**) showed a significant positive correlation and were kept for further analysis. However, in the weighted network obtained from the mouse data using method #3 the blue module was discarded because it did not pass the established criteria ($\text{cor} = 0.25$, $P = 1.8e-28$), while the turquoise module was kept for further analysis ($\text{cor} = 0.64$, $P < 1e-200$) (**Supplementary Figure 7A**). Also the two modules associated with sex obtained from mouse data using the methods #2 (WGCNA) and #4 (WGCNA + DEGs) were kept for further analysis given their positive significant correlation (red: 0.96, $P < 1e-200$; brown: 0.67, $P < 1e-200$) (**Supplementary Figure 7B**). In human, four modules were retained using method #3 (red: 0.75, $P = 5.3e-99$; blue: 0.81, $P = 4.6e-120$; yellow: 0.87, $P = 6.6e-27$; turquoise: 0.84, $P = 1.7e-145$) (**Supplementary Figure 8A**), and seven modules were kept for further analysis using method #4 (light-yellow: 0.74, $P < 1e-200$; black: 0.86, $P < 1e-200$; dark green: 0.8, $P = 1.4e-13$; red: 0.77, $P < 1e-200$; dark red: 0.81, $P = 1.3e-56$; green: 0.74, $P < 1e-200$; yellow-green: 0.82, $P < 1e-200$) (**Supplementary Figure 8B**). Thus, a total of 5,876 genes from the four selected modules in sea bass, 3,069 genes obtained from the three chosen modules in the mouse, and 8,452 transcripts within the seven selected modules from human were kept for further analysis with method #4 (**Supplementary Figure 2**).

2.3 Co-expressed DEGs involved in gonadal differentiation

After the selection of interesting modules, we further filtered the genes within those modules by its gene significance (GS) for sex ($|GS| > 0.2$) and intra-modular membership ($|k_{IM}| > 0.8$). Furthermore, to implement the method #4 (WGCNA + DEGs), we further filtered the gene lists with the DEGs between males and females at the second stage of development, *i.e.*, 250 dpf in the sea bass, 16.5 dpc in the mouse, and 6 PCW in the human, to identify the most relevant genes of each module. The number of genes co-expressed and, further, that are DEGs at 250 dpf was 3,782 in sea bass (pink modules, $n = 365$; green module, $n = 621$; magenta module, $n = 930$; sky blue module, $n = 1,866$ genes) (**Additional file 3** and **Supplementary Figure 2**). In the mouse, there were 2,205 co-expressed DEGs at 16.5 dpc (red module, $n = 544$; brown module, $n = 1,661$) (**Additional file 4**), and in the human there were a total of 1,023 differentially co-expressed transcripts (**Supplementary Figure 2**).

2.4 Comparison of methods using network properties and statistics

We calculated several topological parameters to compare the performance of methods #2 (WGCNA), #3 (DEGs + WGCNA), and #4 (WGCNA + DEGs). Some of the most relevant parameters are shown in **Table 3**, including (1) the number of nodes, (2) number of edges, (3) characteristic path length, (4) average number of neighbors, (5) heterogeneity, (6) betweenness centrality, (7) mean degree, and (8) maximum degree. The detailed definitions of topological parameters were published by the developer of the NetworkAnalyzer (Doncheva *et al.*, 2012). In the three species, method #4 showed a lower number of nodes than method #2 since only co-DEGs were retained; however, method #4 resulted in a much higher number of nodes and edges than method #3 in the sea bass and human. The average number of neighbors, which indicates the average connectivity of a node, showed higher (better) results when using method #4 than method #3, in proportion to the number of nodes obtained from each method (average number of nodes/number of nodes) $\times 100$, in sea bass: method #2 = 75%, method #3 = 14%, method #4

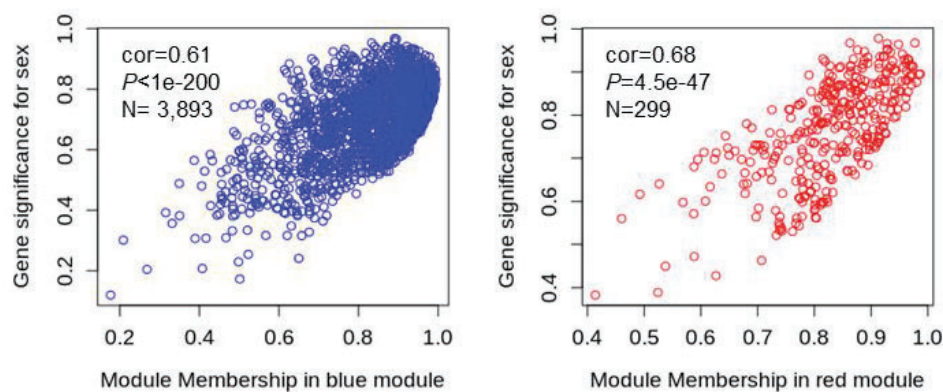
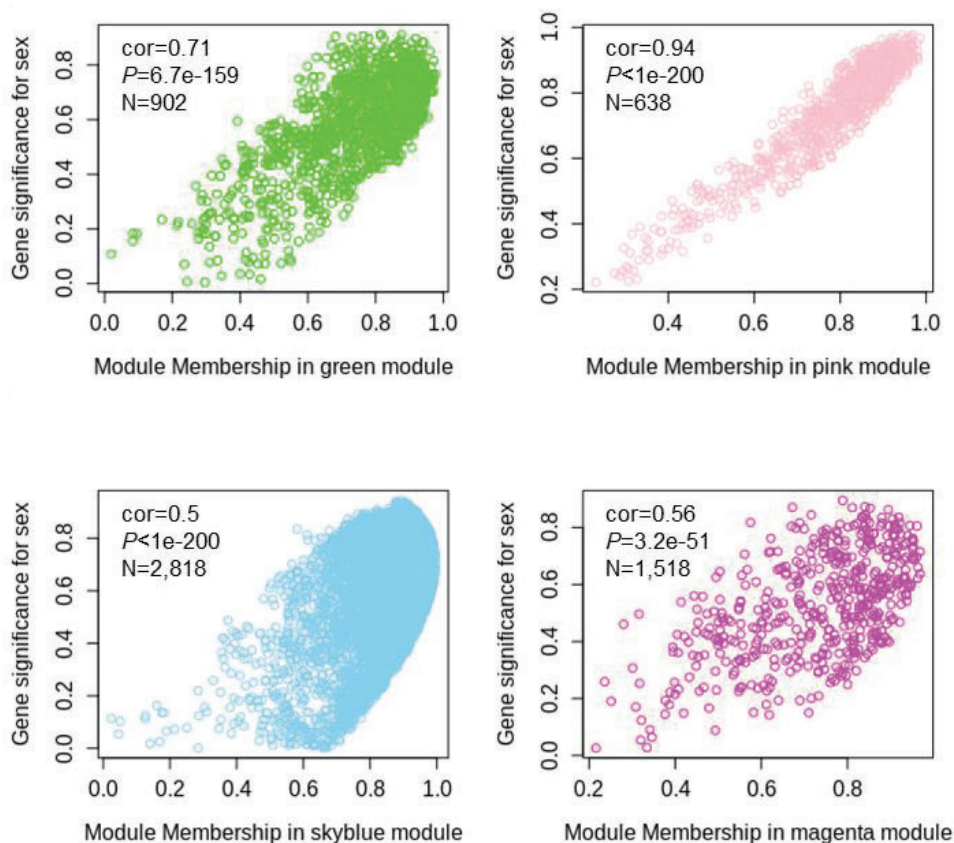
A**#3. DEGs + WGCNA****B****#2. WGCNA & #4. WGCNA + DEGs**

Figure 4. Filtering of genes based on network properties in the sea bass. Scatterplots of correlation between GS vs. module membership of each module which were used for determination of interesting modules for sex trait in the gonadal transcriptome of sea bass using method #3 (A) and #2 and/or #4 (B). The modules were selected when GS was positively and significantly correlated to modular membership. Gene significance is defined as $-\log$ of the p -value of association of the gene with the trait. Module membership (k_{ME}) measures how correlated each gene is to a particular module eigengene (Langfelder and Horvath, 2008).

= 72%, and mouse: method #2 = 40%, method #3 = 10%, method #4 = 58%, but not in human: method #2 = 3.3%, method #3 = 37%, method #4 = 4.2%. Regarding the characteristic path length, the expected distance between two connected nodes (the shortest, the more compacted is a network), method #4 showed better results than method #3, except in human where the network of the module from method #3 was much smaller and, hence, resulted in shorter characteristic path length. Network heterogeneity is an important parameter for biomarker discovery as it reflects the tendency of a network to contain hub nodes (Horvath and Dong, 2008). Such parameter was higher when using method #4 in the sea bass and the human. The maximum betweenness centrality was obtained from method #3 in the sea bass and the mouse, as well as from the human using method #4. The maximum degree was higher in the three species when using method #4. Taken together, the topological parameters of the different networks produced using the three different methods showed that the most robust networks were achieved using methods #2 and #4. Besides, method #4 allows further filtering of the interesting genes without altering main network properties and without removing too many genes that could be importantly related to the trait of interest. Of note, when method #3 resulted in better parameter values, this was always in the mouse, incidentally the species in which the minimum number of samples recommended for network construction was not reached ($n < 15$), leading to less robust conclusions. Considering this and that the construction of the network in the mouse required the highest soft threshold ($\beta = 12$ in the mouse vs. $\beta = 5$ in the sea bass and $\beta = 8$ in the human) while achieving the poorer model fit using method #4, the results support the original recommendation of at least 15 samples to construct networks with method #2 (WGCNA) and #4 (WGCNA + DEGs).

Table 3. Network statistics results when using methods #2, #3, and #4 in sea bass, mouse, and human transcriptomic data

Network statistics		Method #2	Method #3	Method #4
		DEGs	DEGs+WGCNA	WGCNA+DEGs
<i>Dicentrarchus labrax</i>	N° of nodes	638	291	365
	N° of edges	180,759	6,059	98,783
	Average n° of neighbors	480.7	42.3	263.42
	Characteristic path length	1.37	2.35	1.66
	Network heterogeneity	0.34	0.68	0.77
	Mean degree	480.7	41.6	263.42
	Max degree	720	135	349
	Max betweenness	0.003	0.07	0.03
<i>Mus musculus</i>	N° of nodes	2,257	2,697	1,661
	N° of edges	1,016,156	377,169	798,757
	Average n° of neighbors	903,648	280.21	964.1
	Characteristic path length	1.66	2.39	1.43
	Network heterogeneity	0.563	0.865	0.365
	Mean degree	900.4	105.2	964.1
	Max degree	1,867	327	1,556
	Max betweenness	0.003	0.02	0.002
<i>Homo sapiens</i>	N° of nodes	558	48	481
	N° of edges	5632	382	3375
	Average n° of neighbors	20.2	15.9	14.0
	Characteristic path length	2.4	1.7	2.2
	Network heterogeneity	1.88	0.72	2.3
	Mean degree	20.2	15.9	14.03
	Max degree	309	42	309
	Max betweenness	0.25	0.09	0.39

2.5 Proportion of genes related to sexual development identified by the different approaches in the sea bass and the mouse

We found that the proportion of retained genes and, within those, of key genes for gonadal sex differentiation among the four methods tested were similar in the sea bass and mouse (compare **Figure 5A** and **Supplementary Figure 9A**). When using method #1, filtering the transcriptome by DEGs, a large proportion of the transcriptome was kept. Up to 40.2% in sea bass (**Figure 5A**) and 57.6% in mouse (**Supplementary Figure 9A**) of the transcriptomes were differentially expressed. Among them, 78.6% of the key genes previously known to be involved in sexual development were found in the sea bass, and 78% of them were identified in the mouse when using method #1.

When using method #2 for the same datasets, we obtained smaller lists of genes being related to sex. In sea bass, we found that 28% of the transcriptome was involved, of which 64% of the key genes were identified. Similarly, in mouse, 22% of the transcriptome was co-expressed and associated with sex and 76% of the key control genes were detected. The third method, the application of DEG filter previous to network construction, leads to slightly smaller gene lists: 19.9% in the sea bass, and 20% of the mouse transcriptome. However, the proportion of key genes detected by this method was much smaller in the sea bass (34%, 1 gene in the red module and 8 genes in the blue module, **Figure 5B**) and in the mouse (50%, 25 genes, **Supplementary Figure 9B**). Finally, when using the proposed method #4, to perform WGCNA first and to apply the DEG filter afterwards, it provided us with the smaller proportions of potential novel genes related to sex trait (18% in sea bass and 15.6% of the transcriptome in the mouse) without harming the capability to detect the key genes previously known to be involved in sex development. In sea bass, 64% of the key genes were identified (**Figure 5A, C**), and in mouse, 58% of the key genes were captured (**Supplementary Figure 9A and C**).

It is important to note that the lists of genes produced by methods #3 and #4 differ. Method #3 yielded a total of 4,192 co-expressed genes in sea bass, 1,870 of which were uniquely detected using this method. On the other hand, method #4 produced a total of 3,782 genes, of which only 877 were unique to that method (**Figure 6A**). In the mouse dataset, method #3 detected a total of 2,835, of which 1,023 were unique to that method, while method #4 detected 2,205, of which 393 genes were uniquely detected by this method (**Supplementary Figure 9D**). Furthermore, the Jaccard index revealed low similarity between the gene lists obtained from methods #3 and #4 (45.02% in sea bass, and 56.13% in mouse).

2.6 Network visualization and gene expression profiles of the selected genes

The data presented so far indicates that our proposed approach (method #4) gives more meaningful results of the biological process being studied than the currently used method #3. To further explore its advantages, we used the gene lists obtained with method #4 in sea bass for further network visualization and plotted the mean gene expression profiles of the chosen modules. Interestingly, the average gene expression profile of the genes making up the four significant modules showed clear differences in their temporal dynamics during sea bass sex differentiation between 110 and 250 dpf. Thus, co-expressed genes of the green module were upregulated in males and downregulated in females, reaching very disparate expression levels (**Figure 6B**). Genes of the sky blue module did not change expression in males but were actively downregulated in females (**Figure 6C**) while genes of the magenta modules were downregulated in both sexes (**Figure 6D**). Finally, genes of the pink module started upregulated in females but were strongly downregulated by 250 dpf, while in the males their expression increased slightly (**Figure 6E**).

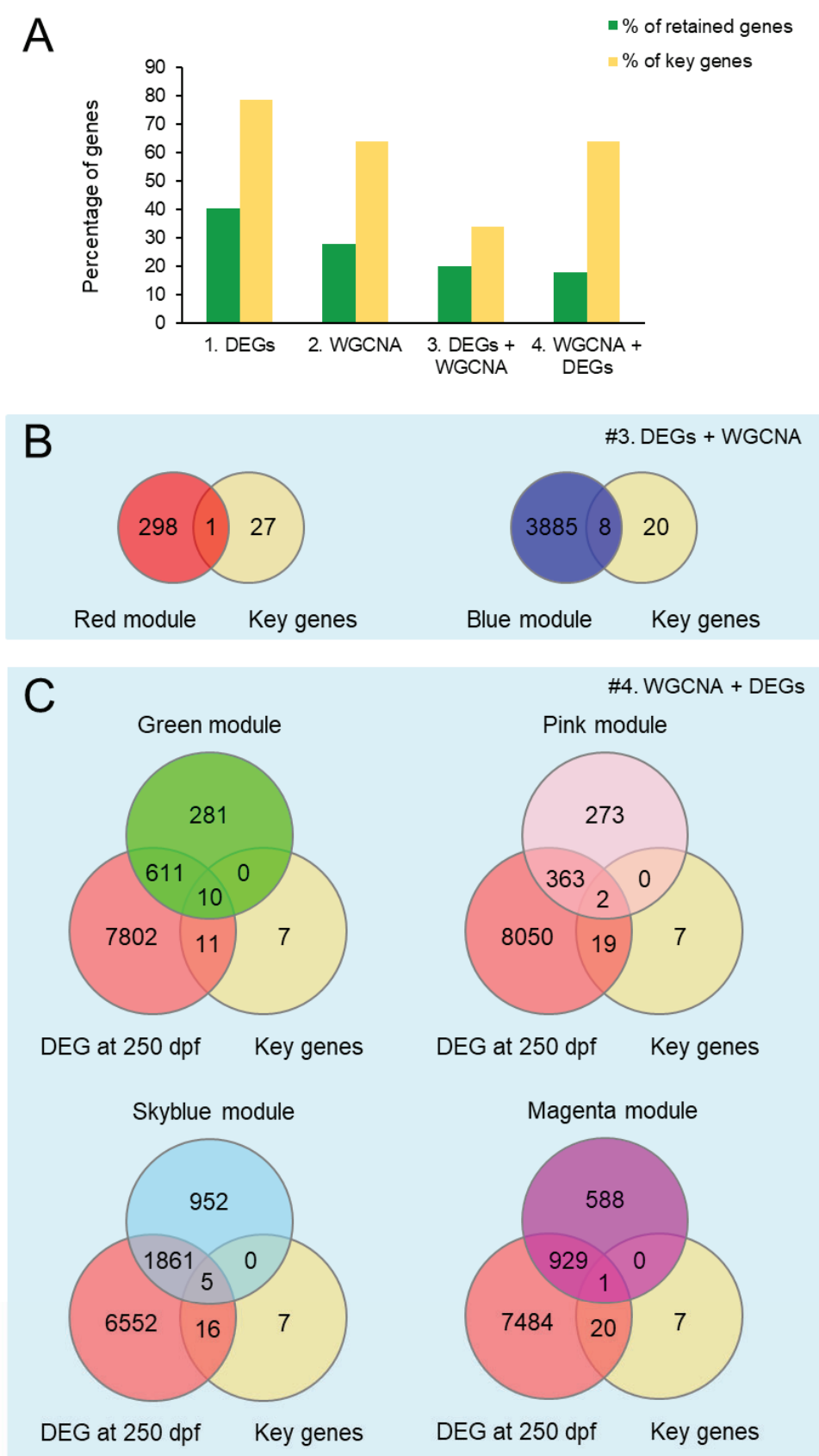


Figure 5. Identification of key genes involved in sex differentiation in the sea bass when using the different methods to analyze transcriptomic data. A Percentage of genes and key genes retained according to the methods mentioned in this study. B Venn diagram of the key genes found in each module using method #3 and C in method #4.

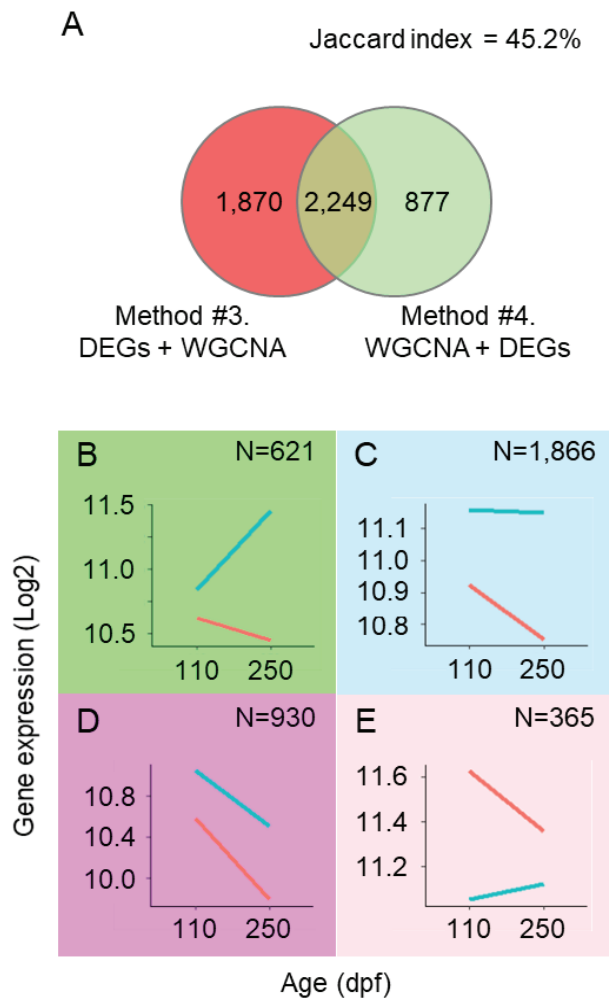


Figure 6. A. Comparison of the gene lists obtained from the sea bass dataset when using method #3 and #4. The Jaccard index indicates low similarity between gene lists. Gene expression profile of the co-expressed DEGs from the four modules associated with sex and filtered by method #4. The background color of the plots indicates the color of the modules they belong to: **B.** green, **C.** sky blue, **D.** magenta, and **E.** pink module.

Genes making up the green module ($n = 621$) were those whose expression levels differed the most over time. For representation purposes, **Figure 7** shows only a subset of these genes ($n = 102$) having the highest number of connections (width > 0.3). From the inside to the outside, the four concentric circles show the genes with a higher degree (number of connections), including Cytochrome C Oxidase Copper Chaperone COX17 (*cox17*), Cyclin A2 (*ccna*), Cytochrome P450 Family 11 Subfamily A Member 1 (*cyp11a1*), Ajuba LIM Protein (*ajuba*), and Proteasome 20S Subunit Alpha 6 (*psma6b*). Importantly, 18 of the 28 key genes selected for their role in sea bass gonadal sex differentiation were represented in the different modules. Nine of them were in the green module with the highest degree (from higher to lower): Anti-Müllerian Hormone (*amh*), Cytochrome P450 Family 17 Subfamily A Member 1 (*cyp17a1*), Follicle Stimulating Hormone Receptor (*fshr*), Steroidogenic Acute Regulatory Protein (*star*), Androgen Receptor (*ar*), Gonadal soma derived factor (*gsdf*), Heat Shock Protein Family a Member (*hsp70*), Cytochrome P450 Family 11 Subfamily B Member 1 (*cyp11b1*), and Luteinizing Hormone/Choriogonadotropin Receptor (*lhr*).

In the magenta module, the key gene found was Hydroxysteroid 17-Beta Dehydrogenase 10 (*hsd17b10*), and some of the genes with the highest degree were Vimentin (*vim*), Protein Kinase C and Casein Kinase Substrate in Neurons 3 (*pacsin3*), and Hephaestin Like 1 (*hephl1*) (**Supplementary Figure 10**).

In the pink module, two of the key genes were found among the selected genes and with the highest degree of connectivity: Estrogen Receptor beta 2 (*erb2*), and Hydroxy-Delta-5-Steroid Dehydrogenase, 3 Beta- And Steroid Delta-Isomerase 1 (*hsd3b*). Similarly, other genes included were as follows: Aquaporin 10 (*aqp10*), Zinc Finger AN1-Type Containing 3 (*zfand3*), POP4 Homolog, Ribonuclease P/MRP Subunit (*pop4*), Tubulin Alpha 8 (*tuba8*), Cytochrome C Oxidase Subunit 5A (*cox5a*), and Acyl-CoA Oxidase 3, Pristanoyl (*acox3*) (**Supplementary Figure 11**). Lastly, with the chosen degree threshold for better network visualization, fewer genes were shown in the sky blue network, including Ubiquitin-Specific Peptidase 5 (*usp5*), Keratin 8 (*krt8*), and Schwannomin Interacting Protein 1 (*schip1*) (**Supplementary Figure 12**).

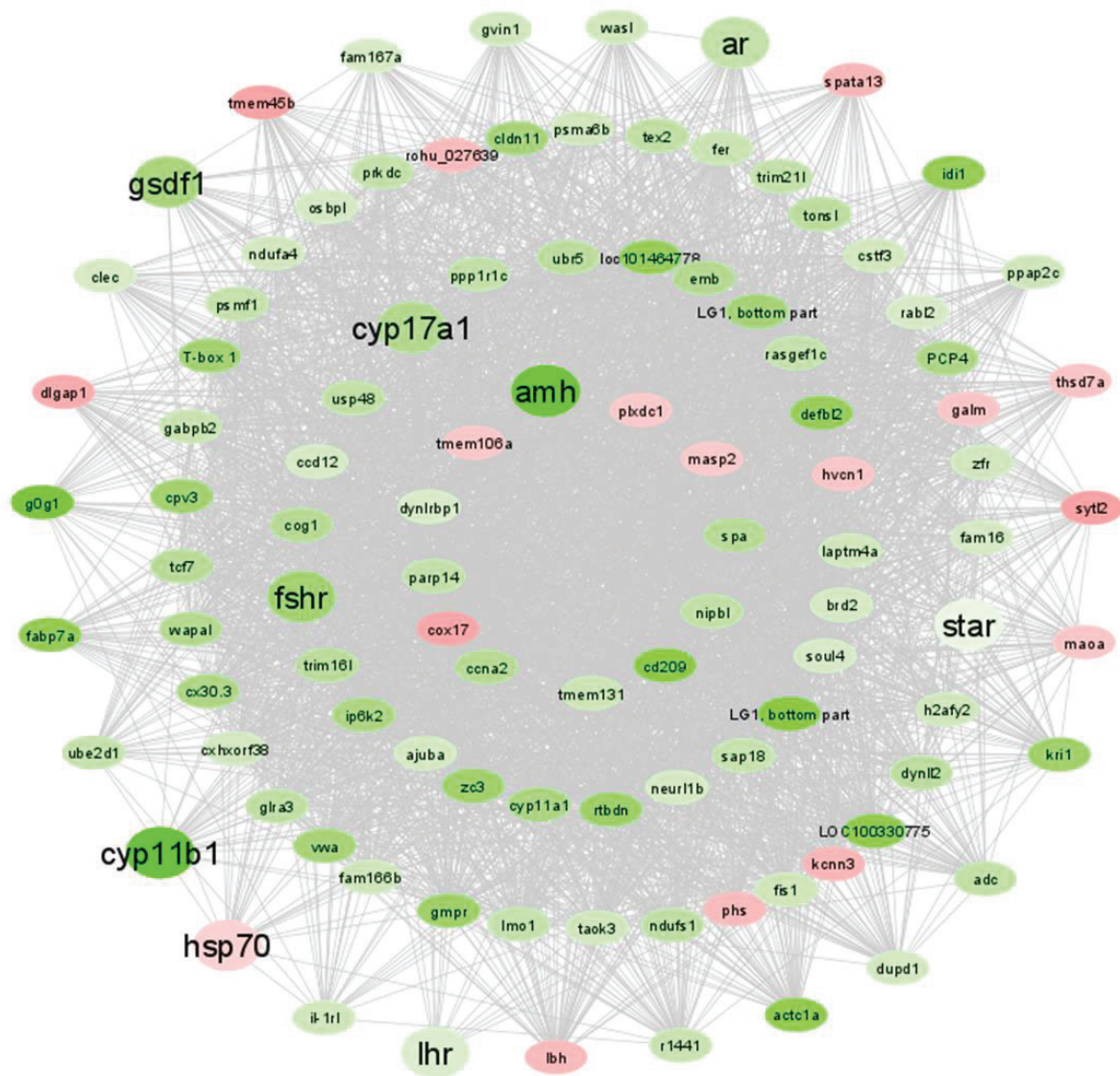


Figure 7. DEGs co-expression network of the green module using method #4 in the sea bass. Downregulated genes at 250 days post fertilization in females are shown in red and upregulated genes in males are indicated in green (in fold change, FC). The position of the nodes or genes indicates the degree range of each gene, which is higher to lower from inside to outside of the net. Key genes (N = 9) are shown in bigger nodes and written in bold.

2.7 Prediction of biomarkers of early gonadal differentiation in the European sea bass

Since the sea bass has PSD, and thus there are no genetic markers for sex, we wanted to focus on genes involved in the process of gonadal sex differentiation to see if we could at least identify robust markers of the early stages of this process. To do so, we used the gene lists obtained from WGCNA built from the entire transcriptome and filtered the output with DEGs only at 110 dpf. We obtained a total of 2,089 genes which were strongly correlated to sex. Among these genes, two of the key genes found were *hsd17b10*, in the magenta module, and the ortholog of SRY-Box Transcription Factor 9 a (*sox9a*), which was detected in the sky blue module. Furthermore, hierarchical cluster analysis of the 28 key genes during the sea bass gonadal sex differentiation showed three main gene expression patterns: (1) genes upregulated in males at 250 dpf, (2) genes upregulated in females at 250 dpf, and (3) genes upregulated in males or females already at 110 that increase their expression by 250 dpf (**Figure 8**). *sox9a* was

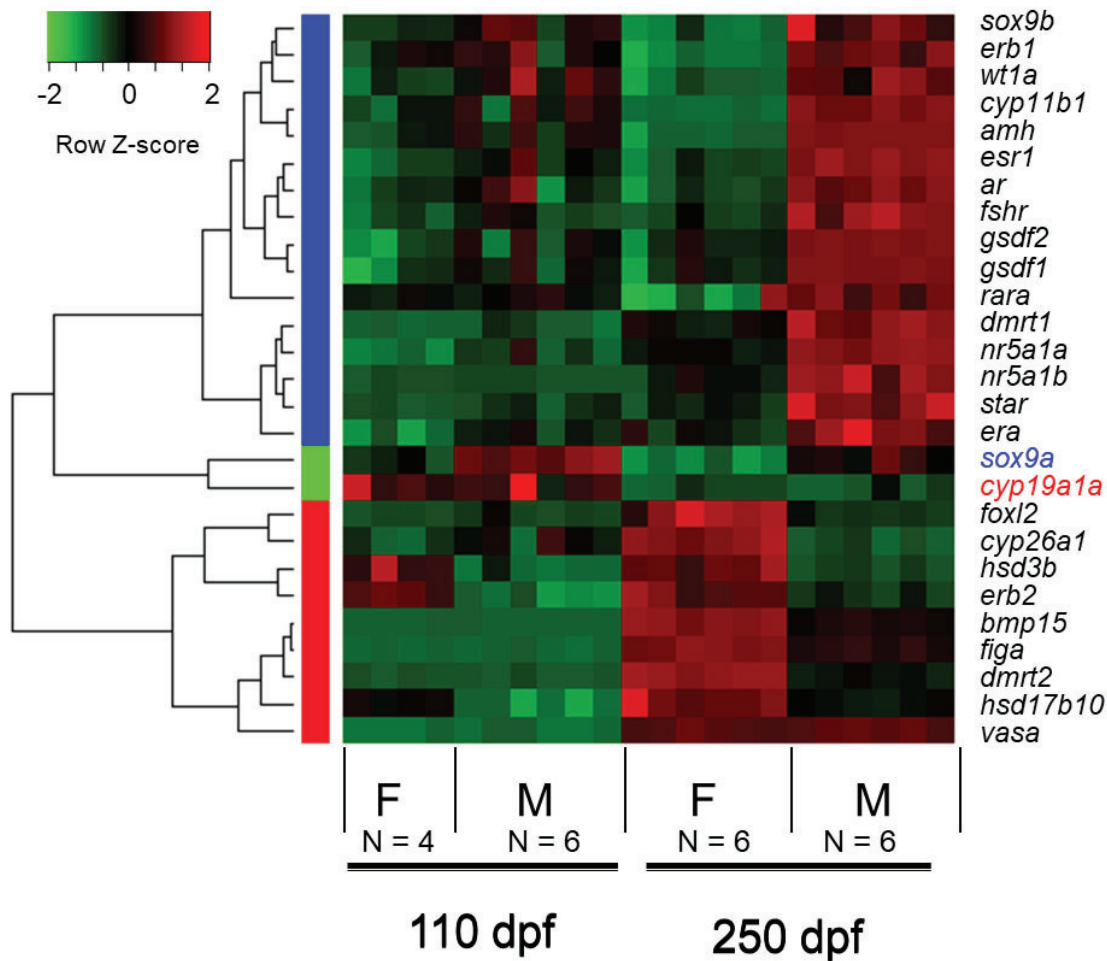


Figure 8. Hierarchical clustering analysis and heat map of the key genes in the sea bass. Upregulation is indicated in red and downregulation is shown in green. The left bars indicate the three main gene expression patterns found. The blue bar includes the genes upregulated in males at 250 dpf, the red bar includes the genes upregulated in females at 250 dpf while the two genes included in the green bar are those identified as early biomarkers, which show a different gene expression pattern already at 110 dpf between males and females.

clustered together with *cyp19a1*, the already described biomarker in previous studies for ovarian differentiation. They were clustered together according to the third gene expression pattern mentioned above. Altogether, this suggested that *sox9a* expression could be used as early marker of male sex differentiation in the sea bass.

2.8 Prediction of biomarkers of early gonadal differentiation in the mouse

In the case of the mouse, there is no need for biomarkers for sex because they have genetic sex determination (GSD) with sexual chromosomes, which allow identifying the sex as early as the zygote stage. However, we used methods #2 and #4 to test their performance in biomarker identification using a well-known model. For this objective, method #1 was not used because it generated lists of several thousand candidate genes from which it was impossible to choose a few candidates. Similarly, method #3 was discarded because of the observed bias in the results. However, when using method #2 in the brown module, for example, we found a total of eight key genes. When using method #4, the number of potential biomarkers of sex at the first stage was five genes: Wnt Family Member 4 (*wnt4*), Fibroblast Growth Factor Receptor 2 (*fgfr2*), Bone Morphogenetic Protein 2 (*bmp2*), Follistatin (*fst*), and GATA Binding Protein 6 (*gata6*). The three genes filtered out by method #4 were not DEG at the E11.0–12.0 stage between males and females but were associated with sex by WGCNA. Such genes are relevant for

sex development in the mouse but not among the first to show differential expression between males and females.

Therefore, the filter by DEGs in method #4 is essential to increase the probability of finding genes that could show differential gene expression when compared to method #2. This allows the prediction of the sex on other individuals based on the gene expression measurement of a biomarker. Hence, method #4 was the approach that allowed obtaining the shorter lists of candidates meeting all the criteria from WGCNA and, on top, they were among the first to show differential expression at the stage of interest.

2.9 Comparison of the functional enrichment analysis from two different methods in the mouse dataset

Next, we performed gene ontology (GO) enrichment analysis of the filtered gene lists produced with methods #3 and #4 in mouse (the system we used that had the most curated annotation). The results showed that GO terms directly related to biological processes associated with gonadal development were more enriched and fell under a more robust significance threshold when using method #4 than when using method #3 (**Additional file 5**). Secondary alcohol biosynthetic process (GO:1,902,653), cholesterol biosynthetic process (GO:0,006,695), sterol biosynthetic process (GO:0,016,126), and steroid biosynthetic process are a few examples (GO:0,006,694). Furthermore, only method #4 captured the term GO cell morphogenesis involved in differentiation (GO:0,000,904).

3. Discussion

Studies based on transcriptomics data can be broadly classified into three different major methods: #1 DEGs, #2 WGCNA, and #3 filtering by DEGs (or other filters such as keeping the 20% with more gene expression variance, or the top 5,000 DEGs) before WGCNA (Bai *et al.*, 2020; Deng *et al.*, 2019; Liu *et al.*, 2021; Zhao *et al.*, 2021). To combine the power from WGCNA with the filter from the traditional and robust statistical analysis without introducing bias, we propose a new pipeline: (4) WGCNA of the entire transcriptome filtered by DEGs only after network construction.

In the present study, we compared the performance of filtering by DEGs before (method #3) or after (method #4) WGCNA in the analysis of one of the most transcriptomically complex organs, the gonads, during the critical period of sex differentiation. And to do so, we used three different vertebrates, a modern fish, the mouse model, and human. The fish has a very different sex-determining system than the mouse and the human that, in contrast, have a very similar sex-determining system. In all species, method #4 (WGCNA + DEGs) clearly and consistently outperformed method #3 (DEGs + WGCNA) (**Figure 9**). WGCNA establishes an adjacency matrix of correlation values which is transformed using a power function. The value used to power the matrix is chosen based on parameters that ensure a scale-free topology (in order of importance): signed $R^2 \geq 0.8$, or the β value which maximizes a scale-free independence, high mean connectivity, and the slope ($-\gamma$) of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$ was around -1 . In sea bass, the best scale-free topology model fit using method #3 resulted in index $\beta = 9$ leading to $R^2 = 0.7$, low mean connectivity, and the $-\gamma$ slope of -0.47 . However, using method #4, for the same dataset, a much better scale-free topology model fit ($R^2 = 0.87$) was obtained with a lower power index, $\beta = 5$. Also, higher mean connectivity and $-\gamma$ was slightly closer to -1 . In mouse, using method #3, even a high soft threshold index of $\beta = 14$ lead to a weak model fit $R^2 = 0.55$, low mean connectivity, and the slope of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$ of -0.79 . Using method #4 with the same dataset improved all the parameters during co-expression network construction. The R^2 increased from 0.55 to 0.7, and the mean connectivity increased from hundreds to

a few thousands. Furthermore, the $-\gamma$ slope decreased from -0.79 to -0.88 . The improvement of model fit was also observed in human using method #4 ($\beta = 8$, $R^2 = 0.89$, mean connectivity > 200 , and $-\gamma = -1.38$ ($R^2 = 0.87$) when compared to method #3 (method #3: $\beta = 8$, $R^2 = 0.67$, mean connectivity value < 50 , and $-\gamma = -1.02$ ($R^2 = 0.63$)). The results from six analyses (using two methods in three species) under different soft threshold values showed that the network model fit is much better when using method #4, objectively.

Another important outcome of our comparison related to the network construction is the number of modules created by each method. Thus, method #3 created 8 modules while method #4 created 29 in sea bass. In the mouse, these figures were 6 and 21, and in human 8 and 36, respectively. These results consistently show that the number of modules associated with the trait of interest is considerably smaller when using method #3, clearly showing that potentially interesting genes and connections are lost already before WGCNA is performed, diminishing its power. These results were consistent with values found in the literature regardless of vertebrate species or question being answered, where method #3 produces fewer modules (5 modules in Lin *et al.*, (2018); 12 in Zuo *et al.*, (2018); and 6 in Deng *et al.*, (2019), than studies using WGCNA without previous filtering (17 modules in Cox *et al.*, (2019); 23 in Oldham *et al.*, (2008); and 14 in Guillotin *et al.*, (2021)).

Several strategies exist to compare gene co-expression networks (Kakati *et al.*, 2019; Ovens *et al.*, 2021a). We calculated topological parameters for comparison of the network module most associated with the trait of interest using methods #2, #3, and #4. All the network parameters from method #4 outperform method #3 in two or all the three species studied. One of the instances was the characteristic path length in human, which showed a shorter path length in the network built using method #3. This could be because of the much smaller number of nodes obtained with such a method, which could provide a shorter length despite being a less compacted network, as noticed elsewhere (Doncheva *et al.*, 2012). Some of the other parameters that did not clearly improve using method #4 compared to method #3 resulted from the mouse. Such networks were built with 11 samples (less than the minimum sample size recommended by the developers of WGCNA) (Langfelder and Horvath, 2017). Therefore, results should be considered with precaution. We observed in the literature that out of 20 studies in **Table 1** using WGCNA or DEGs + WGCNA, some used $n < 15$ ($n = 13$, Guillotin *et al.*, 2021; $n = 10$, Wang *et al.*, 2020; $n = 8$, Deng *et al.*, 2019; $n = 14$, ; and $n = 8$ Liu *et al.*, 2019), although the majority used > 15 samples to construct robust WGCNA ($n = 160$, Oldham *et al.*, 2008; $n = 183$, Cox *et al.*, 2019; $n = 25$, Kim *et al.*, 2020, $n = 430$, Liu *et al.*, 2021, $n = 45$ Jiang *et al.*, 2020, $n = 36$ Tu *et al.*, 2021, $n = 228$ Niemira *et al.*, 2020, $n = 48$ Deist *et al.*, 2017. Considering the developers' recommendation and our results, we take the opportunity to further emphasize that at least 15 samples in total are required to perform method #2 and extend this recommendation to method #4.

Regarding the results of biological relevance, which was to identify genes related with sex and therefore with reproduction, higher proportions of genes previously known to be involved in sex differentiation were retained for further analysis when screening with method #4. Of note, not only the number of genes captured was larger in both species, but also essential key genes in a given process (here used as key controls), described in previous works, were also better represented with method #4. For example, key genes found in sea bass were as follows: *amh* (Rocha *et al.*, 2016), *gsdf* (Crespo *et al.*, 2013), *ar* (Blázquez and Piferrer, 2005), and *erb* (Rodríguez *et al.*, 2003). Similarly, some of the key genes found in mouse were as follows: *dmrt1* (Raymond *et al.*, 1998), *amh* (Staack *et al.*, 2003), *ar* (Matsushita *et al.*, 2018), *foxl2* (Ottolenghi *et al.*, 2005), *bmp6* (Wang *et al.*, 2015), and *wnt4* (Vainio *et al.*, 1999). These genes were retained with method #4 but not with method #3. Thus, our approach results in a more meaningful representation of the biological question under consideration.

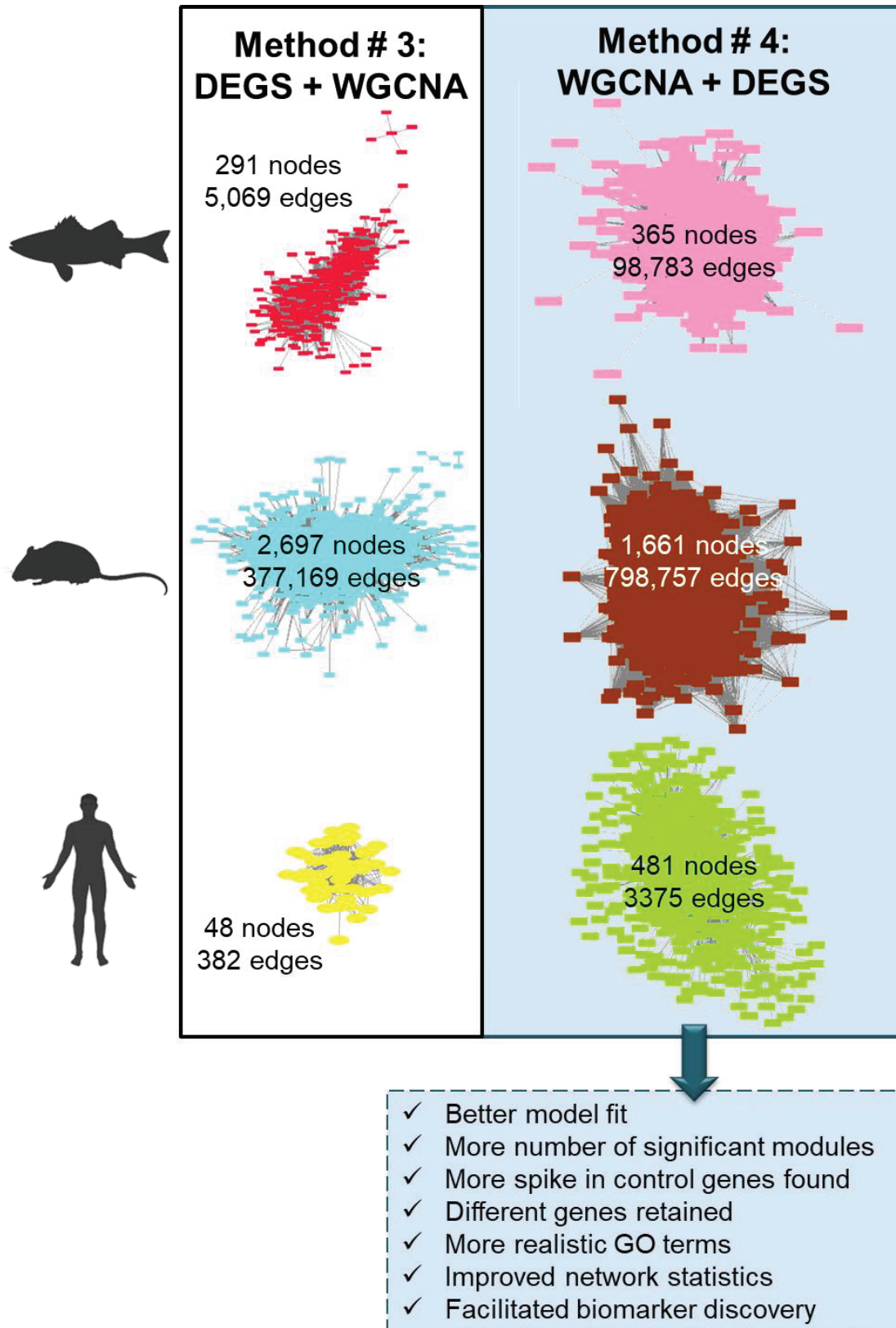


Figure 9. Comparison of module networks built using method #3 (DEGs + WGCNA) and method #4 (WGCNA+ DEGs). The selected modules for network statistics comparison and visualization were those that showed the strongest association with sex on each method and species. Hence, we compared the modules: red vs. pink in the sea bass, turquoise vs. the brown in the mouse, and yellow vs. the green-yellow in human. For visualization, we filtered the networks by edge weight > 0.2 threshold.

Another important insight of this study is that we clearly show that methods #3 and #4 generate different gene lists. As a result, this has an impact on the downstream functional analysis and can affect the resulting GO biological processes, as well as the interpretation of the results (**Additional file 5**). First, the overrepresented GO biological processes from the gonadal transcriptome of the mouse at 16.5 dpc slightly varied depending on the filtering strategy and the order in using such filters leading to different results. First, we found that GO terms from the category of biological processes related to gonad development functions were more enriched and had a higher significance threshold when using method #4. Second, only method 4 yielded some terms related to the gonadal sex differentiation process. Altogether, these results provide evidence that more accurate gene lists related to the target phenotypic trait were achieved using method #4 (WGCNA + DEGs) than when using the DEG filter or gene lists from pre-filtered WGCNA as in method #3.

We are aware that the results from gene ontology enrichment analysis should not be considered proof of biological validity in the analysis of high-throughput data considering the multiple sources of bias existing in functional enrichment analysis tools (Timmons *et al.*, 2015). However, to date, this is one of the main methods used to summarize information from high-throughput experiments and here we showed how the previous filtering of DEGs can affect the results of WGCNA.

In search of early sex differentiation markers in the sea bass, several genes have been identified so far in the literature. Blázquez *et al.*, (2009) defined *cyp19a1a* as a suitable molecular marker of ovarian differentiation, which showed the first significant gene expression difference between sexes at 120 dph. Ribas *et al.*, (2019) identified the *hsd17b10* gene as an early marker for ovarian differentiation at 110 dpf. At this stage, *cyp19a1a* expression levels were higher in females but not significantly different from males. In the present study, among the key genes involved in early gonadal development, three genes stood out from the combined filter of WGCNA and the subsequent DEG filter at 110 dpf. Among these three genes, *hsd17b10* was found, hence, supporting the results from Ribas *et al.*, (2019). The other gene identified as marker of testis early differentiation was *sox9a*. Noteworthy, this gene could not be identified in previous studies using method #1 (Ribas *et al.*, 2019).

sox9 is a multifunctional transcription factor found in different tissues and plays crucial roles in vertebrate development, including cell proliferation and differentiation (Vining *et al.*, 2021). This gene was defined as a “hub” gene of testis differentiation after sex determination in vertebrates. As reviewed elsewhere (Vining *et al.*, 2021), *sox9* has a conserved role in male gonadal development and a highly conserved protein sequence. In fishes, two orthologs exist due to gene duplication: *sox9a* and *sox9b* (Voldoire *et al.*, 2017). Although the two orthologs are retained, they are related to different functions and tissues depending on the fish species. For example, *sox9a*, is expressed in the ovary of the medaka (*Oryzias latipes*) and the platy fish (*Xiphophorus maculatus*), while in the zebrafish (*Danio rerio*), *sox9a* is expressed only in the testis (Voldoire *et al.*, 2017), like the current results found in the sea bass. In this species, *sox9a* was recently found to be involved in the epigenetic regulation of the temperature induced sex ratio during sex differentiation (Geffroy *et al.*, 2021).

Based on the parameters obtained in the model fit during network construction, network statistics, and the biological relevance of the results obtained, we highlight produced by WGCNA without removing the most important (based on gene expression) genes. WGCNA is a powerful method still being used (Guillotin *et al.*, 2021; Kim *et al.*, 2020; Liu *et al.*, 2021; G. Wang *et al.*, 2020), with a large proportion of the studies using specifically the inappropriate method #3 (Bai *et al.*, 2020; Cui and Ma, 2020; Huang *et al.*, 2020; Jiang *et al.*, 2020; Zhang *et al.*, 2020; Zhao *et al.*, 2021; Zhu *et al.*, 2021). More recent alternative methods to WGCNA to identify modules of differentially co-expressed genes have been

developed: THD-Module Extractor (Kakati *et al.*, 2016), DiffCoEx (Tesson *et al.*, 2010), and MODA (Li *et al.*, 2022). And, although in the comparison of these methods with WGCNA, the THD-Module Extractor method was claimed more effective in finding modules with higher functional relevance and biological significance than WGCNA, this result could be due to the previous filter of the dataset. Precisely, as stated by the authors, the use of DEGs before WGCNA lead them to the soft threshold of power fit to fail (Kakati *et al.*, 2019). Therefore, a comparison of THD-Module Extractor method with the method #4 (WGCNA + DEGs) described in this study remains to be done.

4. Conclusions

In conclusion, WGCNA is a robust and systematic approach commonly used in transcriptomics. With time, several strategies have appeared and became quite common, *i.e.*, method #3, to use the parameters and filters provided by this tool but somehow running against its original philosophy. We evaluated the model fit and the biological relevance of the results obtained using method #3 (DEGs + WGCNA) and our proposed method #4 (WGCNA + DEGs) with the gonadal transcriptome of three different vertebrates, sea bass, mouse, and human, obtained from two high-throughput technologies: microarray and RNA-seq. The results showed that method #4 is more efficient in filtering for smaller gene lists that contain the genes most related to the trait of interest. Thus, we propose to analyze transcriptomic data using WGCNA to build a co-expression network from the entire dataset and, only subsequently, filter by DEGs. Such a strategy combines the powerful method of producing a network with the filter from the traditional and robust statistical analysis without introducing bias. Additionally, we produced new lists and network visualization of genes related to early sex differentiation in the sea bass with the corresponding functional enrichment analysis and identifying a novel biomarker for testis early differentiation (*sox9a*). Last but not least, the filter by DEGs in method #4 increases the probability of robust biomarker discovery when compared to method #2. We propose that method #4 should be the method to use in future transcriptomic studies regardless of biological system, species, or question being considered.

5. Materials and Methods

After WGCNA development, numerous studies started to use this method in combination with the previous filtering of the dataset (method #3). Combining WGCNA with other filtering methods like DEGs can be a strategy to further filter for the target genes associated with a trait of interest. However, whether this filtering step before WGCNA affects the results and to what extent has never been shown yet. In this study, we aim to compare method #3 with the here proposed new method #4: applying the filtering step after construction of the co-expression network. To determine which one is more efficient, we used both methods to study the gonadal transcriptome of three vertebrates: sea bass, mouse, and human. Hence, a total of six analyses were performed. Then, we compared the model fit, connectivity, and other network statistics to compare both methods for each of the three species. Additionally, we used existing information on key genes for the sexual development and reproduction of sea bass and mouse to measure how many key genes were found with each method. Finally, the enriched genes by GO term enrichment analysis obtained from the gene lists produced by methods #3 and #4 were compared in the mouse. The methods are implemented in R software and all source code has been made publicly available on GitHub as part of the WGCNA_DEGs project at: https://github.com/Nsbaizan/WGCNA_DEGs.

5.1. Literature review

Before any comparison of methods #3 and #4 was performed, we wanted to know if the use of method #3 was exceptional or if we could find several studies where this method was used. We used Scopus and Web of Science databases to find published studies using each method. To search a few examples of method #1, we used the following keywords in the topic field: “((DEG) OR (“differentially expressed genes”)) AND ((Microarray) OR (RNA sequencing))”. To find studies using methods #2 or #3, we used the following keywords in the topic field: “((WGCNA) OR (“weighted gene co-expression network analysis”)) AND ((Microarray) OR (RNA sequencing))”. After reading detailed information on the methods section of each paper, we could classify between methods depending on whether WGCNA was used without (method #2) or after filtering by DEG (method #3). We found that the use of the WGCNA is widespread in transcriptomic data studies and that method #3 is not exceptional but, rather, it is of common use even in recent publications.

5.2. Species and processes studied: the sea bass

In sea bass, sex depends on the combination of several pro-male and pro-female autosomal genes plus environmental (temperature) influences (Vandeputte and Piferrer, 2019). Sex determination is thought to occur between 60 and 100 dpf (Díaz and Piferrer, 2015), around 120 dpf the first molecular signs of gonadal sex differentiation appear in the form of differences in the expression of gonadal aromatase, *cyp19a1a* (Blázquez *et al.*, 2008), and the first histological differences appear around 150 dpf when fish are around 8 cm long. Sex differentiation proceeds earlier in females than males, and the process is completed by 250 dpf in females and around 350 dpf in males when females are 12.7 ± 5.7 cm (mean $\pm$ SD), and 11.2 ± 0.6 cm of SD in males, and 16 ± 1.3 in females and 14.8 ± 1.1 cm of SD in males, respectively (Ribas *et al.*, 2019). Gene expression during gonadal sex differentiation in the sea bass has been extensively studied, first by targeted approaches (Cerdá-Reverter *et al.*, 2000; Gornati *et al.*, 2004; Mateos *et al.*, 2002; Socorro *et al.*, 2007) and more recently by using a homologous and validated microarray containing 43,803 probes. In the latter case, fish before (110 dpf), during (250 dpf), and after (350 dpf) gonadal sex differentiation were analyzed (Ribas *et al.*, 2019), and DEGs and enriched signaling pathways were identified. Additionally, *hsd17 β 10* was identified as a marker of early ovarian differentiation.

5.3. Species and processes studied: the mouse

Sex differentiation in the mouse is well characterized (Nagahama *et al.*, 2021) and thus less information will be provided. Briefly, key genes during mouse gonadal sex differentiation include sex-determining region Y (*sry*), *amh*, and *sox9*, while pro-female genes include *wnt4*, R-spondin 1 (*rspo1*), a member of the R-spondin family, and β catenin (*ctnnb1*) (Nagahama *et al.*, 2021). In mouse, sex differentiation starts between embryonic day (E) 11.0–E12.0 and ends at E16.5 (Menke *et al.*, 2003; Munger *et al.*, 2009).

5.4. Species and processes studied: the human

Sex differentiation in the human has been studied extensively despite the limited availability of normal fetal human gonads (Lecluze *et al.*, 2020). In brief, the gonadal primordium arises from the coelomic epithelium around the 4th PCW and differentiation reaches the end towards the 17th PCW. Testis differentiation is activated by the expression of the Y-linked transcription factor *sry* during the 6th PCW in supporting cells leading to the expression of *sox9* and *amh* genes. In the absence of *sry*, *rspo1/Wnt4*/ β catenin pathway, *foxl2* activates the transcriptional cascade required for ovarian differentiation.

5.5. Datasets

Transcriptomic data during sea bass sex differentiation was previously obtained using a homologous microarray (Ribas *et al.*, 2019) and can be downloaded from the Gene Expression Omnibus (GEO) database with the accession number GSE115841 (Wang *et al.*, 2019). For the aim of the present study, we selected data available from fish at two key developmental stages: four females and seven males at the beginning of sex differentiation at 110 dpf with a length of 5.2 ± 0.5 cm (mean $\pm$ SD). At this time, gonads are still morphologically undifferentiated but fish can be sexed measuring *cyp19a1a* expression levels (Blázquez *et al.*, 2008a). We also used six females (12.7 ± 5.7 cm) and six males (11.2 ± 0.6 cm) in the middle of sex differentiation period at 250 dpf. Thus, we used 23 fish in total for analysis. The original downloaded file consists of 43,801 probe copies representing 20,978 transcripts with normalized expression values, corrected for batch effect. Microarray intensity values were directly used for the determination of differentially expressed genes and/or network construction.

Transcriptomic data during mouse sex differentiation was obtained using RNA-seq (Wang *et al.*, 2019) and can be downloaded from the GEO database under the accession number GSE117590. For the present study, we selected transcriptomic data from twelve samples, three males and three females at two embryonic stages: 12.5 dpc, corresponding to the beginning of gonadal sex differentiation, and 16.5 dpc, at the end of sex differentiation. After trimming the raw reads, the alignment to reference genome (v. GRCm39 GCA_000001635.9) (2.1.0) (Kim *et al.*, 2015) was performed to obtain a dataset with 55,416 genomic features using *featureCounts* (v2.0.0) (Yang Liao *et al.*, 2014). Counts were processed using *edgeR* package (v3.34.0) (Robinson *et al.*, 2010)—limma workflow which includes counts pre-processing and exploratory data analysis before obtaining lists of DEGs as described in (Law *et al.*, 2016).

Transcriptomic data during human sex differentiation was obtained using RNA-seq (Lecluze *et al.*, 2020) and can be found at NCBI GEO under accession number GSE116278. We selected transcriptomic data from a total of 32 samples which were grouped into three stages as follows: four males and four females at 6 PCW, corresponding to the beginning of gonadal sex differentiation (first stage); 8 males and 8 females at 7 PCW (second stage); as well as two males and two females at 13–14 PCW, and two males and two females at 17 PCW, corresponding to the end of sex differentiation (third stage). The total of 35,194 finely annotated transcripts in (Lecluze *et al.*, 2020) were used to test methods #3 and #4.

5.6. Transcriptomic data analysis overview

The gonadal transcriptome of the sea bass, mouse, and human during sex differentiation were analyzed with two different methods: #3 (DEGs + WGCNA) and #4 (WGCNA + DEGs). In method #3, we first created a list of the DEGs between males and females at 250 dpf in sea bass and at 16.5 dpc in mouse and then we used this list for network construction. For the human, we used exactly the same list of differentially expressed transcripts between males and females at 6 PCW previously published (Lecluze *et al.*, 2020). In method #4, we first carried out WGCNA and then used the same list of DEGs as before to filter the dataset just after network generation but prior to network visualization. Therefore, we carried out a total of six analyses (two methods $\times$ three species). In the following sections, we describe in detail how we determined the DEGs and how we carried out WGCNA in each case.

5.7. Weighted gene co-expression network analysis

We implemented the WGCNA using the R package (v.1.51) according to the authors' recommendations (Langfelder and Horvath, 2012, 2008) in R statistical software (R Core Team, 2021; RStudio Team,

2021). We first checked for the presence of outliers using the hierarchical clustering of samples with Euclidean distance (Langfelder and Horvath, 2008). In the sea bass dataset, one male at 110 dpf grouped with the male samples at 250 dpf at a height > 15 , distant from the rest of samples and thus removed from further analysis. The rest of the samples were used to generate the Pearson correlation matrices. Among the mouse samples, one was identified as an outlier using the hierarchical clustering of samples. One of the females at 16.5 dpc was at a height > 30 far from the rest of samples and hence, removed from further analysis. Thus, in total we analyzed 22, 11, and 32 samples from sea bass, mouse, and human, respectively.

To build unsigned weighted networks, the adjacency matrix was calculated as $a_{mn} = |\text{cor}_{mn}|^\beta$, where a_{mn} is the adjacency between gene m and gene n , cor_{mn} is the Pearson correlation, and β is the soft-power threshold. Unsigned networks allow the connection of genes that are both positively and negatively correlated; the absolute value of the Pearson correlation is used as a co-expression similarity measure (Dai *et al.*, 2019; Mason *et al.*, 2009). In contrast, a signed network would not include the connections between strong negatively correlated genes (Dai *et al.*, 2019; Mason *et al.*, 2009). Sexual development in vertebrates is known to be orchestrated by mutually antagonizing male and female pathways (Capel, 2017) in which not only upregulation of a set of genes tilts the balance towards the development of one sex or the other but, importantly, the concomitant mechanism of active downregulation of numerous genes in the opposite sex is also required (Munger *et al.*, 2013, 2009). Therefore, using the unsigned network allowed drawing the connections between genes or nodes that were either positively or negatively correlated, since no difference was made between gene inhibition and activation patterns.

To fit the scale-free topology model, we tested several soft threshold powers (ranging from 1 to 20) to which co-expression similarity is raised. To quantify how well a network met the scale-free topology criterion, the model fit was measured as the signed linear regression model fitting index R^2 . After selection of the value leading to the best fit to the scale-free topology model (signed R^2 closest to 1), additional considerations described by the developers were taken into account: high mean connectivity (k), and the slope of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$, closest to -1 (Langfelder and Horvath, 2012, 2008). Subsequently, the adjacency matrix was transformed into a topological overlap matrix, and gene modules were detected by hierarchical average linkage clustering analysis for the gene dendrogram, setting the parameters as default (minimal gene module size = 30, and the threshold to group similar modules was set to 0.25).

5.8. Selection of modules associated with sex

After the modules were defined, the module eigengene (ME) distances were calculated to elucidate potential relationships of modules with two phenotypic traits: sex and age of the samples. For simplicity, we focused on the results related to sex development in the present study, although the study of other traits could be considered following the same workflow. Hence, we chose the significant modules that met the following thresholds regarding sex trait: absolute $R^2 > 0.5$ and $P < 0.001$ in the sea bass and human datasets or $P < 0.01$ for the mouse transcriptome. To assess the correlation strength, we calculated the module significance (MS), the average absolute GS of all the genes involved in the module. The key modules kept for further analysis were those with the highest MS score among all modules produced.

For the genes within the selected modules, we calculated intra-modular membership (k_{IM}) for each gene to determine how well-assigned is a gene within a module and its relationship with the trait of interest as described by Langfelder and Horvath (2008). Additionally, we calculated GS for the sex trait of each gene within the modules. The gene significance is defined as $-\log$ of the p -value of association of the gene with the trait, in our case, sex.

To identify the modules with most interesting genes associated to the trait of interest, we selected the modules with a significant correlation ($|\text{cor}| \geq 0.5$, $P < 0.001$ in sea bass and human; $|\text{cor}| \geq 0.5$, $P < 0.01$ in mouse). To identify the genes most interesting within those modules, we filtered by K_{IM} and GS, where the higher the absolute value of GS of a gene the more biologically significant it is for the trait of interest. Thus, we kept for further analysis all the genes within the selected modules with an absolute intra-modular membership higher than 0.8 (Langfelder and Horvath 2008).

5.9. Analysis of differentially expressed genes

The DEGs were determined by fitting a linear model, using the same empirical Bayesian statistics in both datasets. For the sea bass data, we used the Quantile method in the Linear Models for Microarray Analysis (Limma) (v. 3.44.3) R package (Ritchie *et al.*, 2015). We compared sea bass female vs. male gonads at 110 and 250 dpf. The same statistical test was applied to the mouse transcriptome dataset using *edgeR* (v.3.30.3) (Robinson *et al.*, 2010) and *GLimma* package (v.2.2.0) (Su *et al.*, 2017) in R software (v. 3.4.1) (R Core Team, 2021; RStudio Team, 2021). In this case, the comparison of female vs. male at 16.5 dpc was used. For all the comparisons, genes with a false discovery rate (FDR) based on the Benjamini–Hochberg method were defined as differentially expressed (adjusted $P < 0.05$). We preprocessed the two transcriptomes using different packages because the dataset from the sea bass was obtained from a microarray experiment (normalized intensity values matrix), and the mouse data were obtained from an RNA sequencing experiment (normalized expression counts matrix). While data pre-processing steps were necessarily different because of the technologies used to obtain the data, the mathematics behind DEG determination were essentially the same.

5.10. Key genes with reproduction-related functions

To determine which methodology would unveil more genes relevant for the target process, we generated lists of genes previously known to be involved with gonadal sex differentiation for each of the two species studied. For the sea bass, we selected a total of 28 genes from the literature related to the sea bass or other well-studied fish species like zebrafish (Nagahama *et al.*, 2021; Ribas *et al.*, 2017b, 2019) (**Additional file 6**). In addition, we produced a heat map with gene hierarchical analysis of the selected genes in sea bass using *gplots* (v3.1.1) (Warnes *et al.*, 2020) and *ggplot2* (v3.3.5) (Wickham *et al.*, 2021) packages. For the mouse, where gonadal sex differentiation has been thoroughly studied, we selected 50 genes compiled from recent and comprehensive reviews (Nagahama *et al.*, 2021; J. Wang *et al.*, 2019b; Yildirim *et al.*, 2020) (**Additional file 7**).

Up to this step, all analyses were performed in sea bass and mouse species to test whether the comparison between methods #3 and #4 was reproducible with transcriptomes of different species and obtained with different technologies (microarray and RNA-seq). From this point onwards, only the results from the sea bass were further explored by network visualization and identification of early gene expression markers of sex differentiation.

5.11. Network visualization and statistics

In order to compare the three methods involving network construction, we selected the most associated module with sex (positively or negatively) to calculate network parameters. The modules were as follows: red (DEGs + WGCNA) vs. pink (WGCNA + DEGs) in the sea bass, turquoise (DEGs + WGCNA) vs. brown (WGCNA + DEGs) in the mouse, and yellow (DEGS + WGCNA) vs. yellow-green (WGCNA + DEGs) in the human. We exported network results to Cytoscape software (version 3.5.1) (Shannon *et al.*, 2003) using the function *exportNetworkToCytoscape* from WGCNA package by including the adjacency matrix of each module as input and defining the adjacency threshold to 0.2.

Cytoscape was then used to visualize and to analyze the networks using NetworkAnalyzer (Doncheva *et al.*, 2012).

5.12. Identification of biomarkers of early gonadal sex differentiation

To find genes involved in the early stages of gonadal sex differentiation in the sea bass, we used the gene lists of the four modules associated with sex, filtered by gene significance and intra-modular membership as described above, *i.e.*, absolute gene significance higher than 0.2 and an absolute intra-modular membership higher than 0.8. Additionally, for those genes that met the criteria, we checked which ones were also DEGs between males and females at 110 dpf. Among the selected genes, we checked for the presence of genes previously known to be involved in gonadal sex differentiation in this species.

5.13. Functional enrichment analysis of the mouse dataset

We chose to investigate potential differences in GO term analysis caused by different methods using the mouse, a model species with a much more curated and up to date annotation. The co-regulated DEGs associated with sex at 16.5 dpc were enriched for Gene Ontology analysis using the GO Enrichment Analysis bioinformatic PANTHER tool (Mi *et al.*, 2017), with the list of genes captured by the RNA-seq experiment ($n = 14,088$) serving as the background reference list.

Improved biomarker discovery through a plot twist in transcriptomic data analysis

Núria Sánchez Baizán, Laia Ribas and Francesc Piferrer

Supplementary Information

The online version contains supplementary material available at:

<https://doi.org/10.1186/s12915-022-01398-w>

Additional file 1. Table with DEGs in the differentiating gonads of the European sea bass between males and females at 250 dpf. File format in .xls.

Additional file 2. Table with DEGs in the differentiating gonads of the mouse between males and females at 16.5 dpc. File format in .xls.

Additional file 3. Table of co-expressed DEGs in the differentiating gonads of the European sea bass between males and females at 250 dpf. File format in .xls.

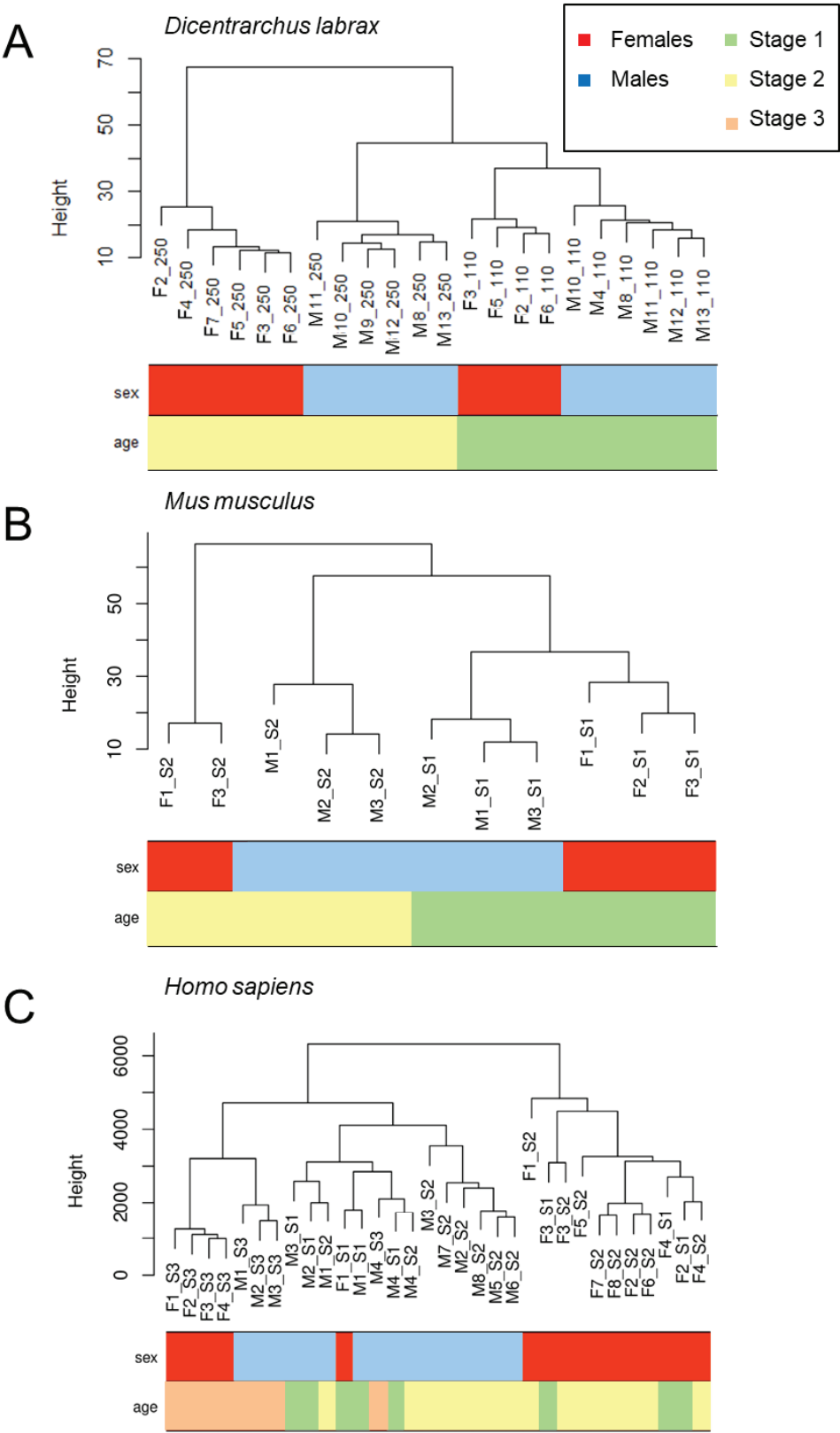
Additional file 4. Table of co-expressed DEGs in the differentiating gonads of the mouse between males and females at 16.5 dpc. File format in .xls.

Additional file 5. Significant GO terms enriched from the co-expressed DEGs in the differentiating gonads of the mouse between males and females at 16.5dpc using both methods #3 and #4. File format in .xls.

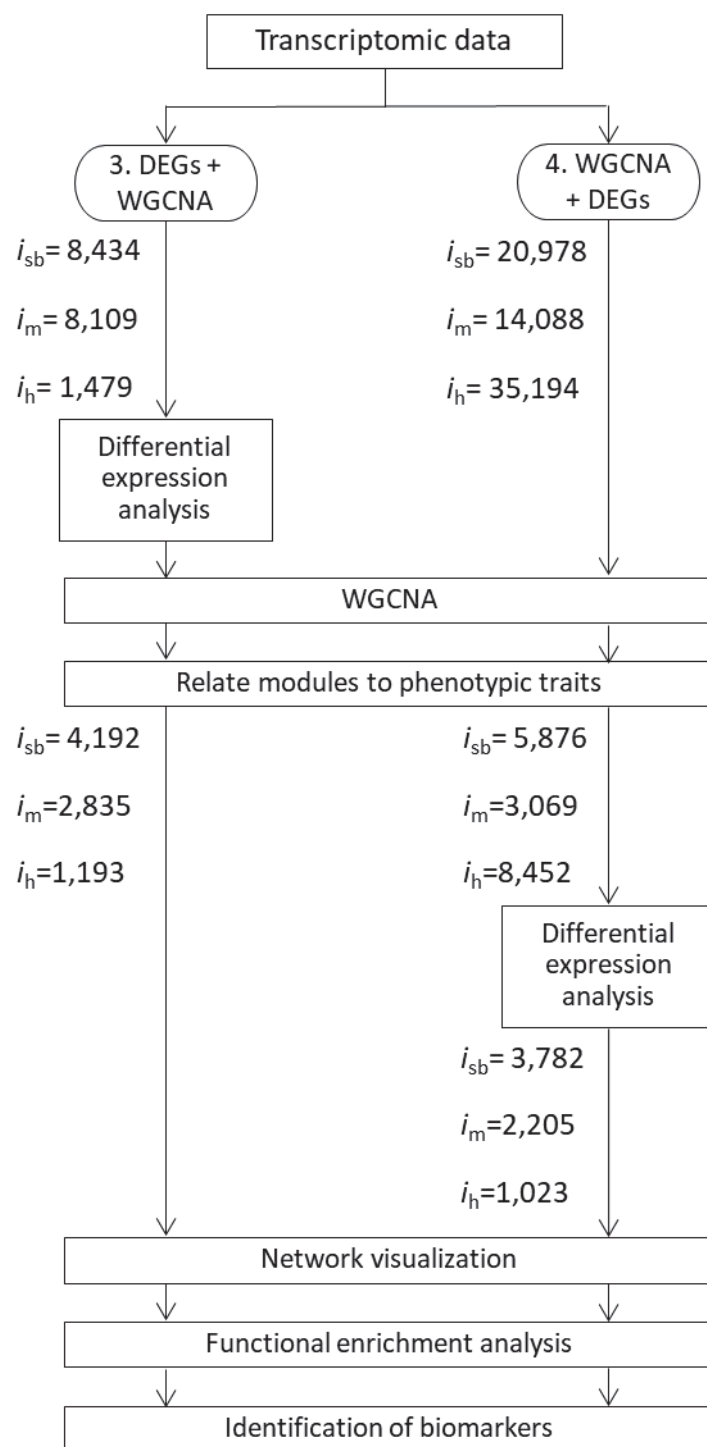
Additional file 6. Key genes involved in the gonadal development of the European sea bass and other teleosts. Data compiled from (Nagahama *et al.*, 2021; Ribas *et al.*, 2019, 2017). File format in .xls.

Additional file 7. Key genes involved in the gonadal development of the mouse. Data compiled from (Nagahama *et al.*, 2021; J. Wang *et al.*, 2019b; Yildirim *et al.*, 2020). File format in .xls.

Supplementary Figures



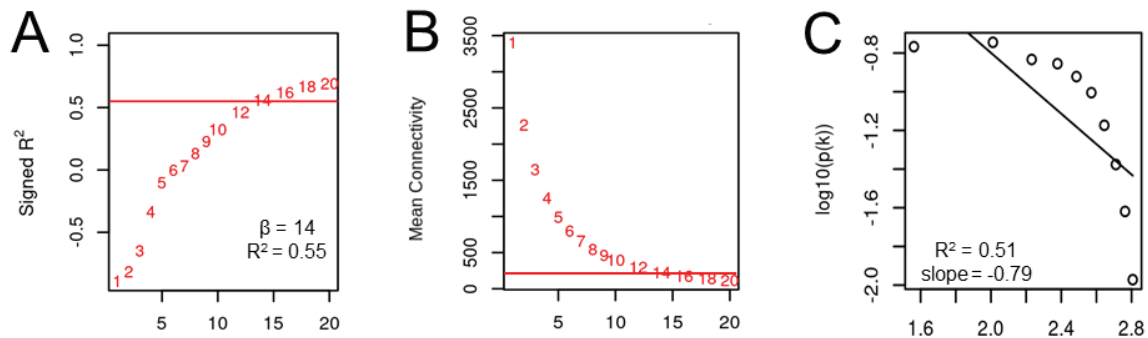
Supplementary figure 1. Sample dendrogram and trait information on sex and age using hierarchical clustering analysis.



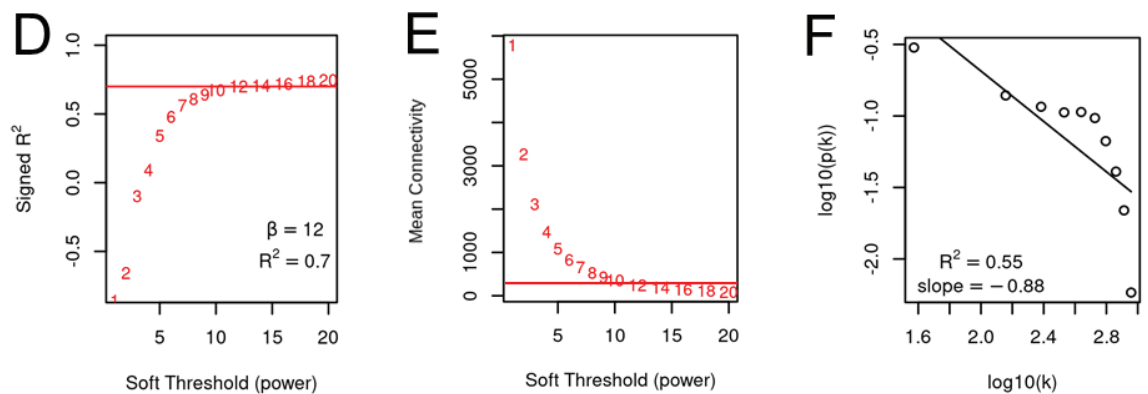
Supplementary figure 2. Flow diagram and number of genes of methods #3 and #4.

Mus musculus

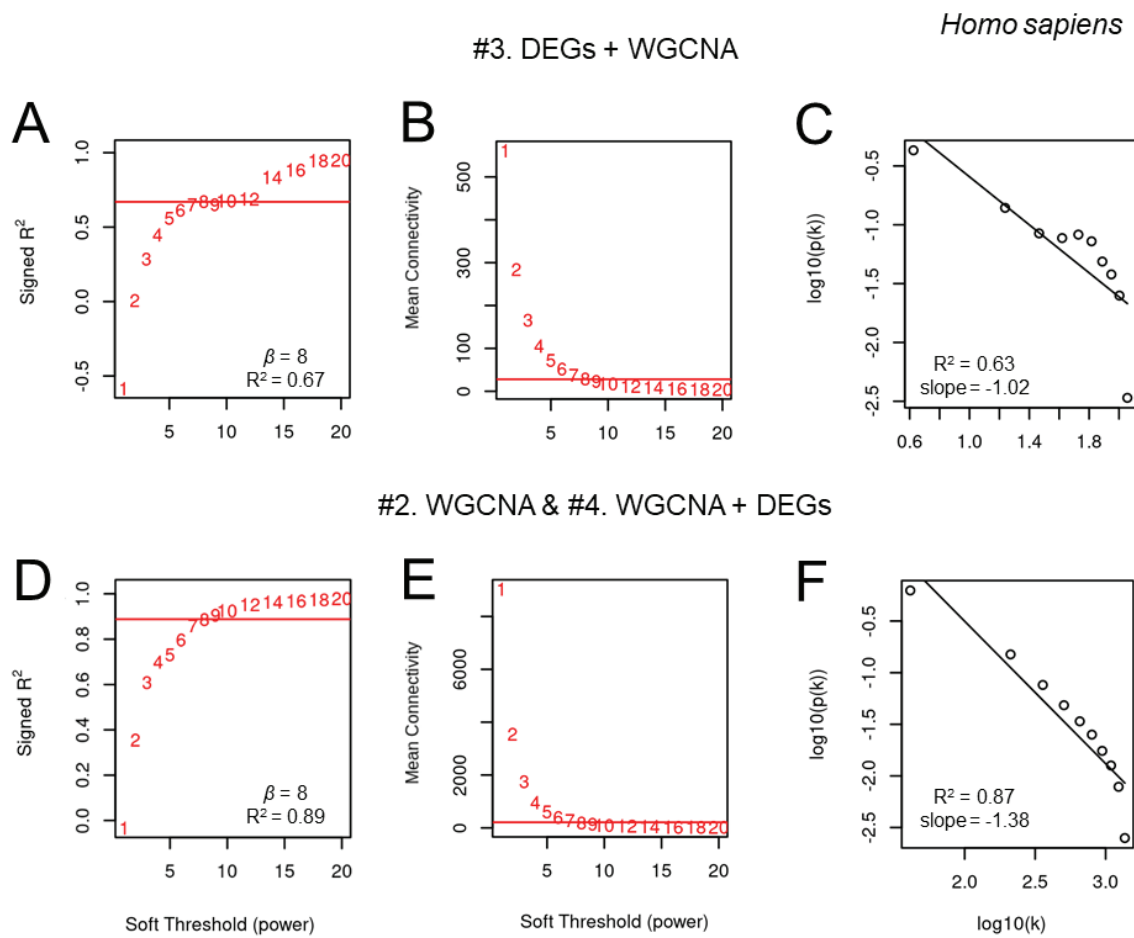
#3. DEGs + WGCNA



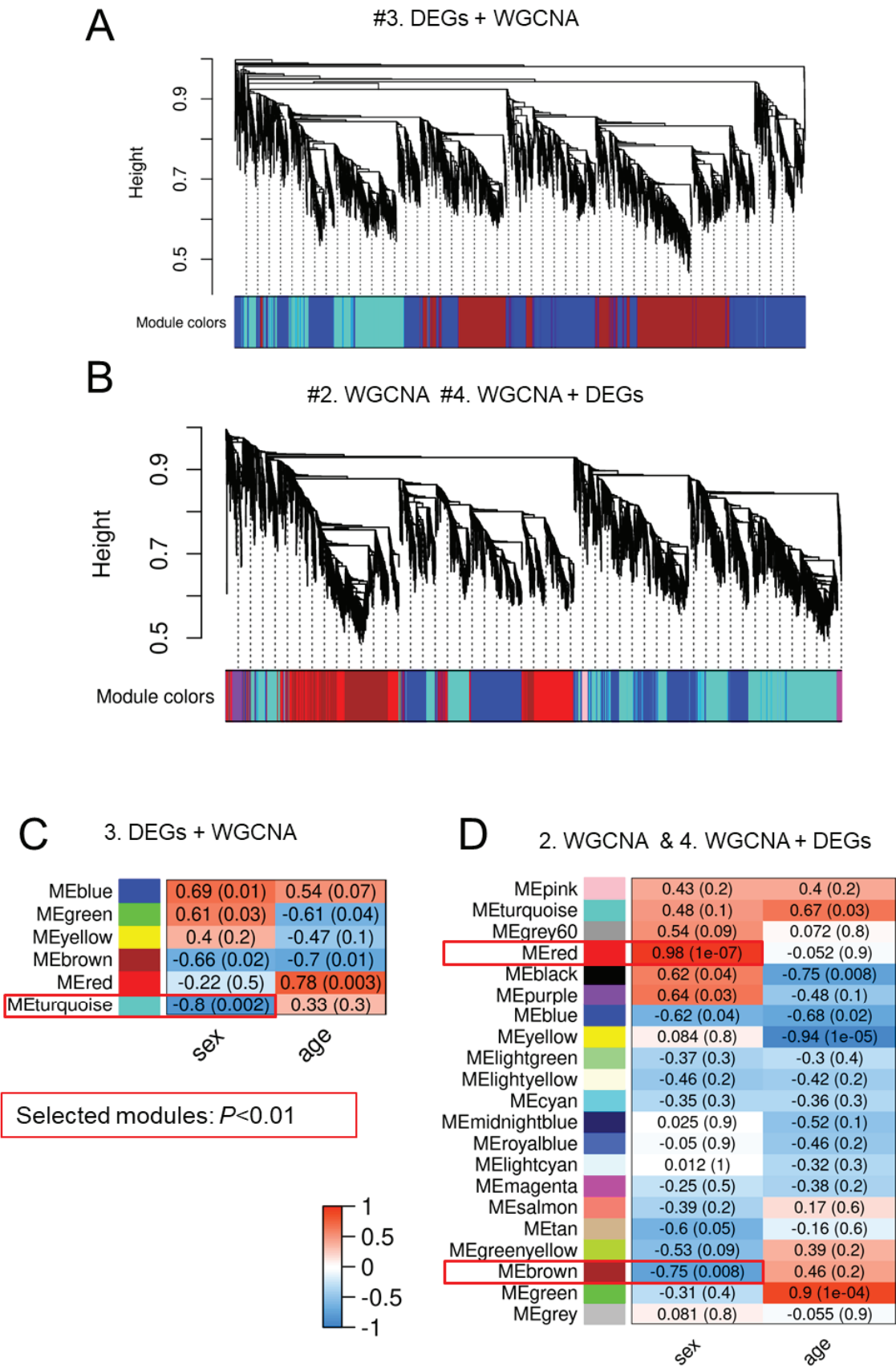
#2. WGCNA & #4. WGCNA + DEGs



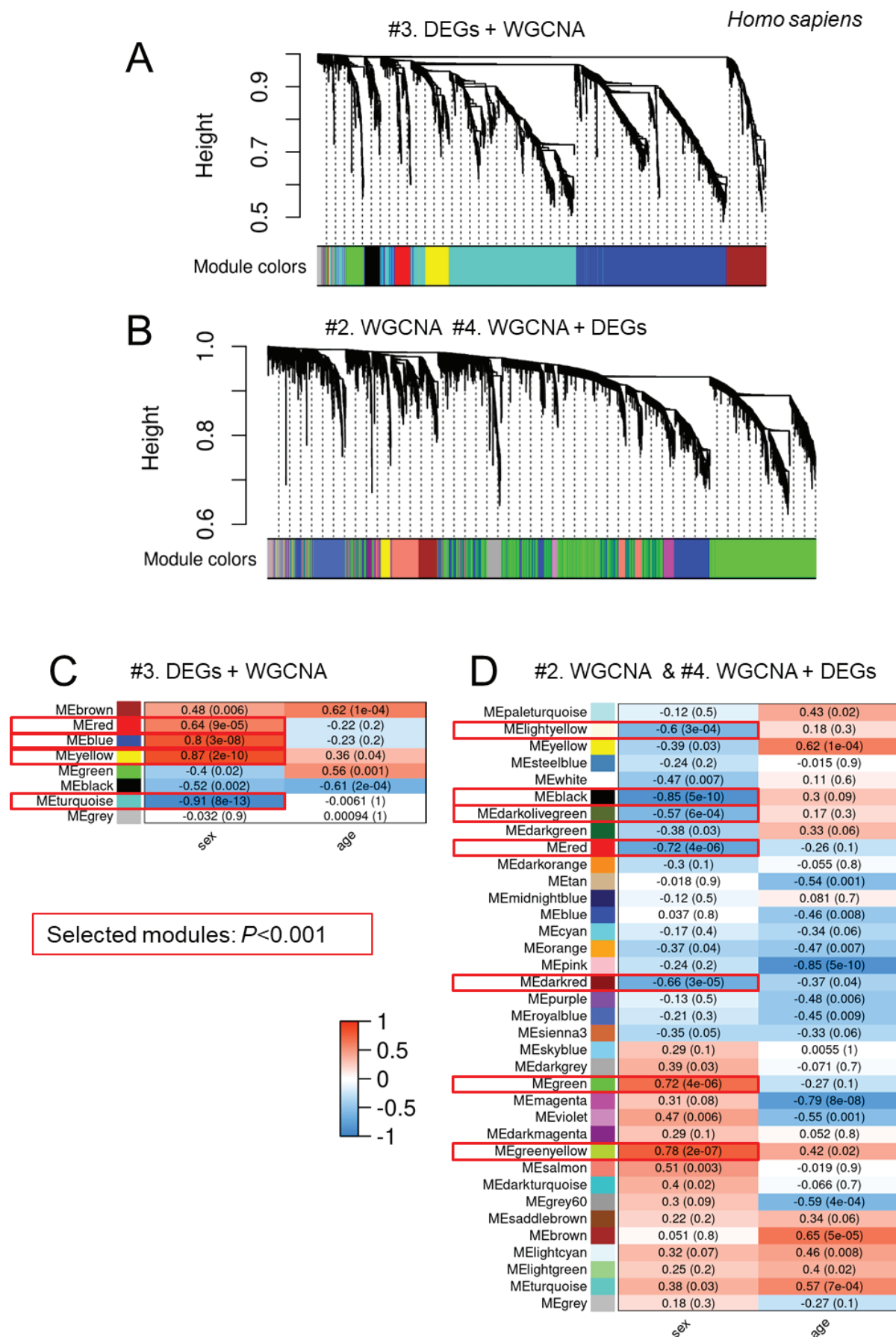
Supplementary figure 3. Determination of soft threshold power in the WGCNA using the mouse gonadal transcriptome.



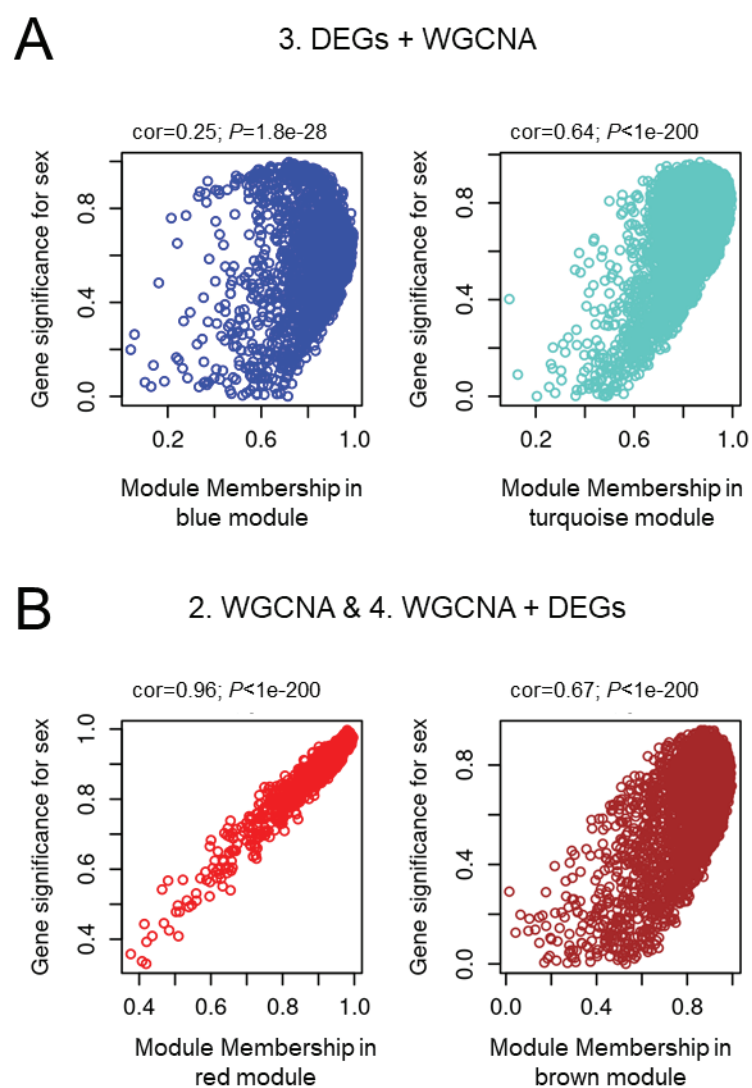
Supplementary figure 4. Determination of soft-threshold power in the WGCNA using the human gonadal transcriptome.



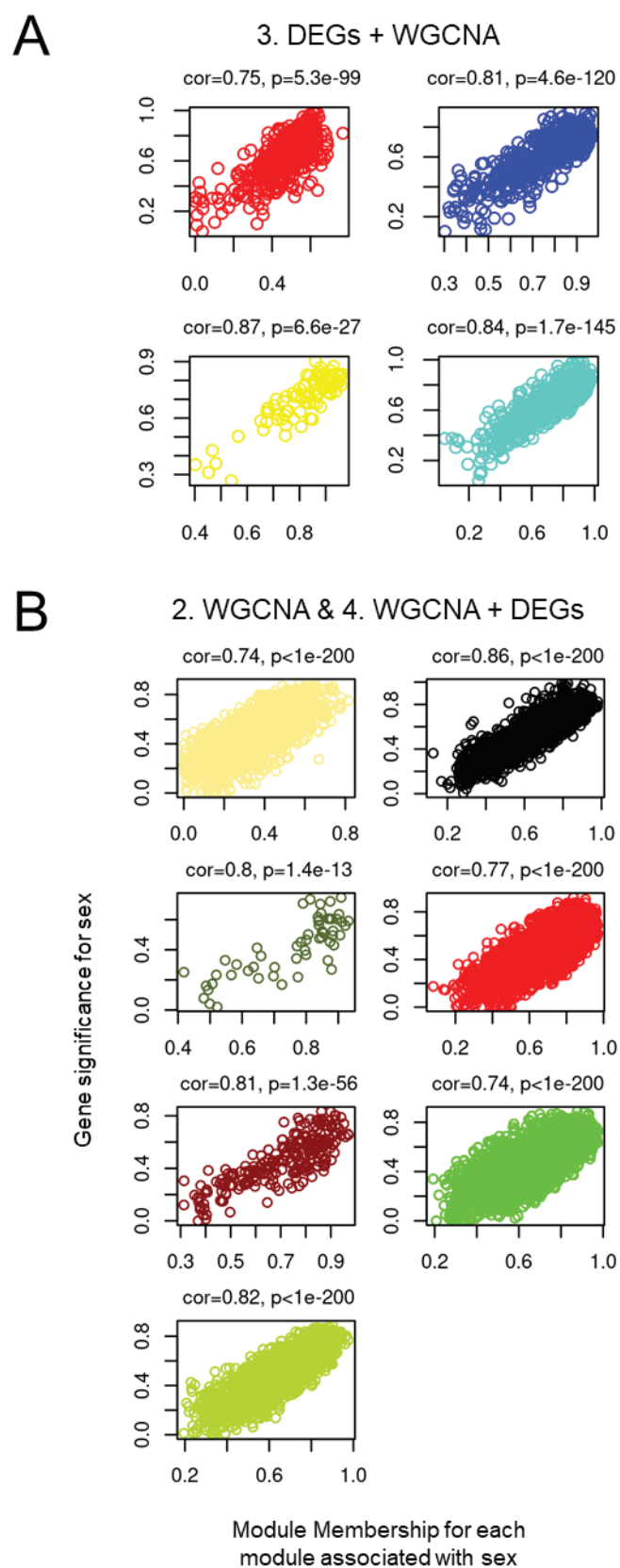
Supplementary figure 5. Identification of gene modules associated with sex in the mouse.



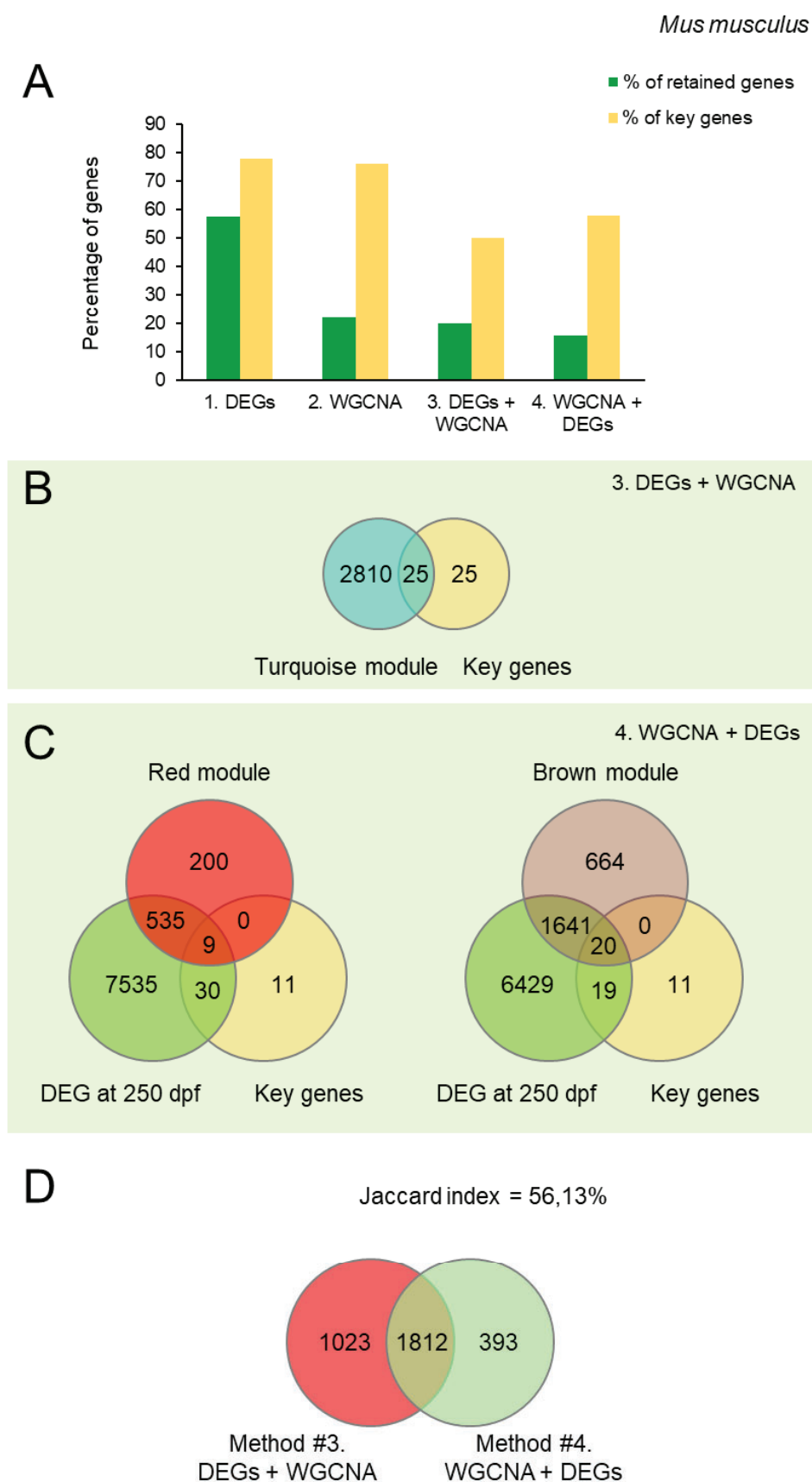
Supplementary figure 6. Identification of gene modules associated with sex in the human.

Mus musculus

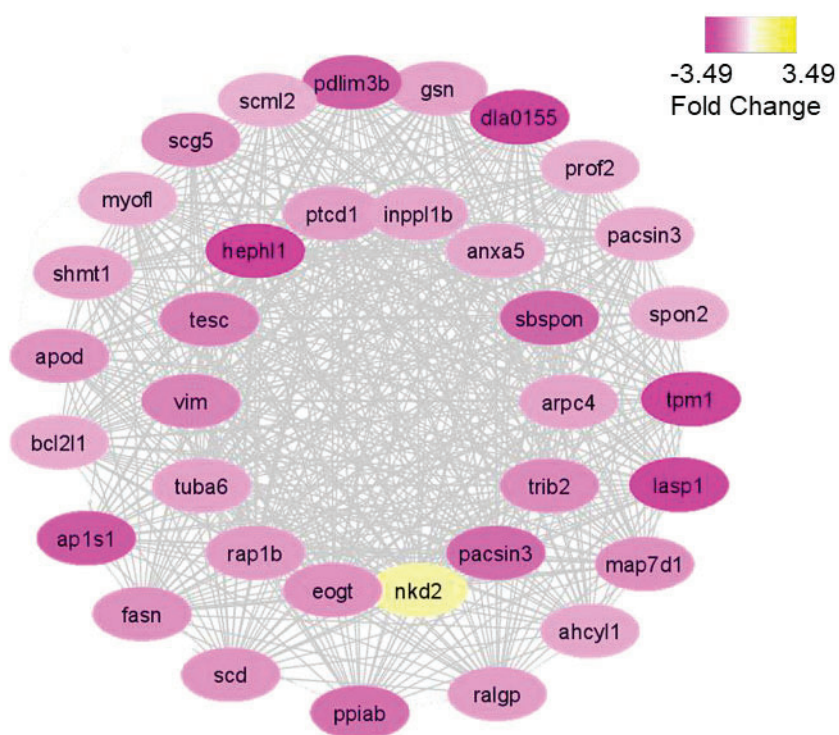
Supplementary figure 7. Filtering of genes based on network properties in the mouse.

Homo sapiens

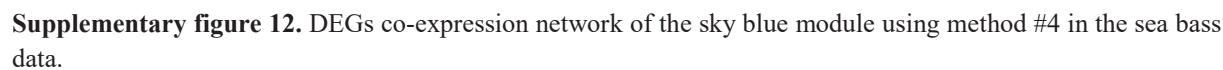
Supplementary figure 8. Filtering of genes based on network properties in the human.



Supplementary figure 9. Identification of key genes involved in sex differentiation of mouse.



Supplementary figure 10. DEGs co-expression network of the magenta module using method #4 in sea bass data.



Acknowledgements

We thank Dr. Steve Horvath for his expert assistance and Dr. Frédéric Chalmel for his help in obtaining the data from their previous publication in the required format.

Authors contributions

NS, LR, and FP designed research; NS collected and analyzed the data. NS wrote the first draft of the manuscript. NS and FP wrote the manuscript. All authors revised and approved the final version of the manuscript.

Funding

NS was supported by a Spanish Ministry of Science (SMS) predoctoral scholarship (BES-2017–079744). Research was supported SMS grant no. PID2019-108888RB-I00 to FP. Also with funding from the Spanish government through the ‘Severo Ochoa Centre of Excellence’ accreditation (CEX2019-000928-S).

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories. The datasets analyzed are available at the GEO database with the accession numbers:

GSE115841 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115841>),
GSE117590 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>),
GSE116278 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116278>).

Competing interests

The authors declare that they have no competing interests

Chapter 2

*Extraordinary variability in gene activation and
repression programs during gonadal sex
differentiation across vertebrates*

Submitted

Extraordinary variability in gene activation and repression programs during gonadal sex differentiation across vertebrates

Núria Sánchez-Baizán¹, Ignasi Jarne-Sanz¹, Alvaro S. Roco^{2,3}, Manfred Schartl^{3,4}, Francesc Piferrer¹

1. Institut de Ciències del Mar (ICM), Spanish National Research Council (CSIC), Barcelona, Spain
2. Developmental Biochemistry, Biocenter, University of Wuerzburg, 97074 Wuerzburg, Germany
3. Department of Experimental Biology, Faculty of Experimental Sciences, University of Jaén, Campus Las Lagunillas S/N, 23071 Jaén, Spain
4. Xiphophorus Genetic Stock Center, Texas State University, San Marcos, TX 78666, USA

Abstract

Gonadal sex differentiation is a complex process involving the dimorphic expression of many genes, initially thought to be achieved mainly by upregulation in one sex but nowadays known to result of the action of both activation and repression programs, at least in mouse. However, the situation in the rest of vertebrates is unknown. Here, we characterized gene expression dynamics and quantified the proportion of genes activated and repressed in each sex of six vertebrate species, from fish to mammals. Our results unambiguously show an extraordinary variability across species, with activation/repression involving between 8% and 92% of the male- and female-enriched genes in one sex or the other depending on the species but without relationship with position in the vertebrate phylogenetic tree. Furthermore, key genes such as *dmrt1*, *sox9* and *amh*, if enriched, were consistently enriched only in males, whereas key genes such as *foxl2*, *cyp19a1a* and *fst*, if enriched, were only enriched in females. Previously, *sox9* and *amh* had been shown to be hub genes. In contrast, the top 10% most expressed genes were not enriched always in the same sex and included 16 potential novel markers of early sex differentiation. We propose that gonadal sex differentiation in vertebrates can be seen as a hierarchical network with hub genes such *sox9* and *amh* and less connected nodes. In such scenario, evolutionary pressures may affect genes depending on the number of connections.

Keywords Gonadal development, gene activation, gene repression, gene expression networks, transcriptomics.

1. Introduction

Sexual systems in vertebrates are diverse and include gonochorism (or separate sexes), hermaphroditism and unisexuality (Pla *et al.*, 2022). During development, the fate of the bi-potential gonad is first established by the process of sex determination, followed by the process of gonadal sex differentiation, in which it is transformed into ovaries or testis. Sex-determining systems in vertebrates range from environmental (ESD) to genotypic (GSD), depending on the factors or combination of them that trigger the decision. GSD and ESD are seen as two ends of a continuum rather than two mutually exclusive options (Sarre *et al.*, 2004).

The genes involved in sex differentiation are fairly conserved from fish to mammals (Capel, 2017; Hirst *et al.*, 2018; Nagahama *et al.*, 2021; Piferrer and Guiguen, 2008; Rhen and Schroeder, 2010; Ruiz-García *et al.*, 2021; Stöck *et al.*, 2021; Yildirim *et al.*, 2020). However, most transcriptomic studies carried out so far have focused on sex-specific differences in gene expression (and less frequently with the associated gene networks) at a particular developmental stage (**Supplementary figure 1A**). Using this approach, the consensus emerged that genes involved were relatively conserved during vertebrate sex differentiation. Yet, recently, it has become apparent that the temporal and relative expression of some key genes, as well as the relative position of these genes within the network, are different across vertebrates. In some instances, there are genes whose function is even related to the opposite sex (Nagahama *et al.*, 2021). For example, in the mouse, testis development is determined by the expression of the Sex Determining Region Y (*Sry*) gene, which activates SRY-Box Transcription Factor 9 (*Sox9*) (Rahmoun *et al.*, 2017). In turn, *Sox9* activates a positive feedback loop between fibroblast growth factor 9 (*Fgf9*) and lipocalin type prostaglandin D2 synthase (*Ptgds*) (Colvin *et al.*, 2001; Moniot *et al.*, 2009). Anti-Müllerian hormone (*Amh*) and Doublesex and Mab-3 Related Transcription Factor 1 (*Dmrt1*) is another target of *Sox9*, and its expression in Sertoli cells causes the regression of Müllerian ducts between 13.5 and 14.5 days post coitum (dpc) (Behringer *et al.*, 1994; Nagahama *et al.*, 2021). Nevertheless, in the chicken, testis differentiation is initiated by *DMRT1* and activates *AMH* and Hemogen (*HEMGN*). In turn, those genes lead to the expression of *SOX9* (instead of *Sox9* followed by *Amh* and *Dmrt1* in the mouse) (Nagahama *et al.*, 2021). This, therefore, constitutes an example of genes being expressed in a different order.

One approach to investigate sexual differentiation in a manner different from the male vs. female approach mentioned above is to analyze gene expression dynamics within each sex but at different developmental stages during sex differentiation (**Supplementary figure 1B**). Munger *et al.*, (2013) used this approach to study transcriptomic dynamics during gonadal sex differentiation in the mouse. Importantly, such approach revealed that sexually dimorphic gene expression patterns are established not only by activation but also by repression programs (Capel, 2017; Munger *et al.*, 2013). Moreover, they combined the fold change (FC) expression values between developmental stages for each sex (*i.e.*, two FC values) into a single data point using Cartesian coordinates. This allowed to quantify the proportion of genes that are regulated by each one of the following three types of mechanisms: 1) genes that are up-regulated (activated) in one sex with concomitant down-regulation (repression) in the opposite sex, 2) genes that are up-regulated in one sex with constant expression in the opposite sex, or 3) genes that exhibit constant expression in one sex with combined active down-regulation in the opposite sex.

Surprisingly, and to the best of our knowledge, the analysis of gene expression dynamics within each sex at different developmental stages during sex differentiation has never been used again in the 10 years elapsed since the seminal study of Munger *et al.*, (2013). Such a two-stage comparison approach

accurately characterized overall changes in gene expression and revealed for the first time that while testis development required the upregulation of most male-related genes with concomitant downregulation of a considerable amount of female-related genes, in contrast, ovarian differentiation involved the upregulation of a lower number of female-related genes and a very small amount of male-related genes. Thus, the type of gene expression dynamics that were described for the mouse using this approach remain essentially undescribed for the rest of the vertebrate species where sex differentiation has been studied. Analysis of gene expression dynamics is an important tool for studying position within the network, temporal and relative expression of genes involved in complex developmental processes (Batut *et al.*, 2022; Reis *et al.*, 2001).

Here, we used both published and unpublished data to further exploit this approach in other vertebrate species. We selected species for which the transcriptome was available at the beginning (here referred as T1) and towards the end (referred as T2) of the gonadal sex differentiation period, representing five of the major groups of vertebrates: the European sea bass (*Dicentrarchus labrax*) (Ribas *et al.*, 2019) the platyfish (*Xiphophorus maculatus*), the clawed frog (*Xenopus laevis*) (Piprek *et al.*, 2018) the red-eared slider turtle (*Trachemys scripta elegans*) (Czerwinski *et al.*, 2016) and the chicken (*Gallus gallus*) (Ayers *et al.*, 2013). Of note, we also included the mouse (*Mus musculus*) not only as a representative of mammals but also to replicate previous results (Munger *et al.*, 2009; 2013). First, genes that become enriched in males relative to females at T2 were named Male Enriched Genes (MEGs), while genes that become enriched in females relative to males at T2 were named Female Enriched Genes (FEGs). Second, we analyzed gene expression dynamics to calculate the expression FC between T1 and T2 for each sex separately. Next, we combined the two FC values into a single data point in a scatterplot to reveal novel expression dynamics in vertebrate species other than the mouse. Additionally, we identified not only which genes were relevant for gonadal sex differentiation, but also their dynamics in both sexes and compared them across vertebrates, *i.e.*, we quantified the number of genes that change their expression between T1 and T2 through each of the three mechanisms mentioned above. Finally, we identified the genes with the highest expression change between stages and determined whether those were conserved or not across species. This allowed the identification of new potential markers of early sex differentiation in vertebrates.

2. Materials and methods

2.1. Species selection and data collection

We performed a comprehensive review to select the most comparable and publicly available gonadal transcriptome datasets (microarray or RNA-seq) during sex differentiation of at least one species from each one of the five groups of vertebrates. The criteria considered were the sex determining system, the developmental stage and the number of samples available. For the analysis of gene expression dynamics within each sex two developmental stages were chosen: at the beginning of sexual differentiation (T1), when gonads were still morphologically undifferentiated; and towards the end of sex differentiation (T2). When data from several stages was available, careful examination of the development dynamics was made in order to determine the which ones were most equivalent (Roux *et al.*, 2015). A minimum of two samples per sex and stage were included (*i.e.*, eight or more samples per species) for statistical purposes. The species selected were European sea bass (*Dicentrarchus labrax*), platyfish (*Xiphophorus maculatus*), African clawed frog (*Xenopus laevis*), red-eared slider turtle (*Trachemys scripta elegans*), chicken (*Gallus gallus*) and mouse (*Mus musculus*). Details on selected species, developmental stages, source of data and data format are summarized in **Supplementary table 1**, and details on the background biology of the selected species can be found in the **Supplementary materials and methods** section.

2.1.1 Fishes

To represent fish, gonadal transcriptome data of the sea bass and the platyfish were used. To work with the gonadal transcriptome of the sea bass, we selected samples of four females and seven males at 110 days post fertilization (dpf) when the gonads are morphologically undifferentiated but can be sexed by gene expression of sex markers (*i.e.*, *cyp19a1a*, Blázquez *et al.*, 2009), and twelve gonads (six testes and six ovaries) of fish at 250 dpf, towards the end of differentiation (Ribas *et al.*, 2019). Data are publicly accessible in GEO (accession number GSE11584). The raw downloaded file consists of 43,801 probe copies representing 20,978 transcripts with quantile normalized expression values, and corrected for batch effect. To include another fish species with a different sex determining system we produced the gonadal transcriptome of the platyfish using RNA-sequencing technology.

2.1.2. RNA sequencing of gonadal tissue during sex differentiation of the platyfish

The first stage selected for the platyfish was embryonic stage 24 and day 7 according to (Tavolga, 1949), which correspond to 17 dpf and 30 dpf, respectively. At stage 24 (17 dpf) the gonads are still morphologically undifferentiated but show already differential expression of the earliest gonadal male (*dmrt1*) and female (*cyp19a1a*) markers. The second stage, day 7 after birth (30 dpf), which represents a later stage of testis and ovary development. Material from genotypic male and female individuals (10 for 17 dpf and 5 for 30 dpf) were pooled. RNA transcriptomes from two replicates of all samples were obtained from 100 bp paired-end Illumina (HiSeq 4000) reads (approx. 20 million reads per sample). These data are available on the NCBI Sequence Read Archive (BioProject PRJNA944639).

2.1.3 Amphibian, reptile, bird and mammal

To represent amphibians, we selected data from developing gonads of the African clawed frog published by Piprek *et al.*, (2018) under accession number GSE105103. We used three samples for each sex at Nieukoop and Faber (NF) 50 and NF53, which produced a matrix with 20,219 annotated transcripts. In reptiles, we obtained transcriptomic data from Czerwinski *et al.*, (2016), who studied the differentiating gonads of the red-eared slider turtle. For that species we used two replicates for developing testis and ovaries at stages 15 and 21. Data is available from the Sequence Read Archive (SRA) database under the code SRP079664. For birds, gonadal transcriptomic data of the chicken were obtained from Ayers *et al.*, (2013). For the chicken, two replicates per sex were available at E4.5 and E6.0 stages. These data were downloaded from SRA under the accession number SRP014719. Finally, the gonadal transcriptome data of the mouse were obtained from Munger *et al.*, (2009), and consisted of 25,697 annotated probes for three females and five males at E11.0 and three females and two males at E12.0. All the selected samples were from the strain 129S1, available from GEO with code GSE41948.

2.2 Data processing

The methods were implemented in R software and R studio (R Core Team, 2021; RStudio Team, 2021). The source code has been made publicly available on GitHub repository available from: https://github.com/Nsbaizan/Transcriptomic_dynamics.

On one hand, data of the European sea bass, the African clawed frog, and the mouse were obtained using microarrays. Therefore, normalized intensity datasets were downloaded from the National Center for Biotechnology Information (NCBI) database. Boxplots of gene expression confirmed the normal distribution of the data. On the other hand, data from the platyfish, the red-eared slider turtle and chicken were obtained with RNA-sequencing. Raw sequences of the red-eared slider turtle and chicken were downloaded from the European Nucleotide Archive (ENA) database. For the three datasets, quality

reports were produced using *FastQC* software (version 0.11.9) (Wingett and Andrews, 2018). Adapters and sequences with a quality score lower than 30 were trimmed with *Bbduk* software (version 38.90) (Bushnell, 2014). Reads were aligned against their respective reference genome using *Hisat2* software (version 2.2.1) (Kim *et al.*, 2015). To map the platyfish reads the newest annotation available was used, the X_maculatus-5.0-male from 2021 in Ensembl, with 24,209 annotated genes. The red-eared slider turtle reads were mapped to CAS_Tse_1.0 genome acquired from NCBI which contains 28,415 annotated genes. The reference genome used to map chicken reads was GRCg6a from Ensembl, with a total of 20,937 genes. *samtools* software (version 1.12) (Li *et al.*, 2009) was used to convert and format aligned files and expression count matrices were obtained using *featureCounts* software (version 2.0.1) (Yang Liao *et al.*, 2014). Normalization, low gene expressed genes filter, and determination of differential expression were performed using DESeq2 package (v.1.34.0) (Love *et al.*, 2022, 2014).

2.3 Determination of differentially expressed genes throughout sex differentiation

Determination of differentially expressed genes (DEGs) between T1 and T2 during sex differentiation was performed separately for each sex and each species. Specifically, the approach used to visualize the transcriptomic changes as described by Munger *et al.*, (2013) consisted in plotting the significant FC expression between two developmental stages of genes with $1.5 \log_2$ FC or higher expression in either sex between T1 and T2. The FC in the female gonad is plotted on the X-axis, and the FC in the male gonad is plotted on the Y-axis (**Supplementary figure 2**). Genes that became enriched in testes relative to ovaries at T2 are shown in blue and are referred to as Male Enriched Genes (MEGs). In contrast, genes that became enriched in ovaries relative to testes at T2 are shown in red and referred to as Female Enriched Genes (FEGs). Genes that were similarly up- or downregulated in both sexes are shown in grey.

We pre-processed the data according to what is best for each type of source data (microarray, RNA-seq) and then we compared the resulting FC, which was calculated using exactly the same functions to build a linear model to determine DEGs between T1 and T2. The $\log_2$ transformed expression values at T2 were used to perform ANOVA analysis against values at T1 using the *limma* R package (Ritchie *et al.*, 2015). The model was fit for all the probes using the *lmFit* function. The function *eBayes* was used to calculate moderated t-statistics and rank the statistical significance of each gene. Only significant genes (Benjamini-Hochberg adjusted $P < 0.05$) of FC higher than 1.5, or lower than -1.5, were kept for visualization of the results. Since RNA sequencing technology identifies more DEGs (Rao *et al.*, 2019), we used more stringent filtering steps to make more comparable analysis between datasets obtained from microarray and RNA sequencing technologies. First, we removed genes considered as outliers when their read count was greater than 19.17 in Cooks distance between samples of the same group. Then, an independent filtering was also performed to optimize the multiple testing results with an alpha value of 0.1. Both the outlier removal and the independent filtering strategies were applied using the function *results* from the DESeq2 R package (Love *et al.*, 2022, 2014).

We generated a separate scatter plot for each species studied. Also, we produced line plots to visualize the expression profiles of a few selected genes over time. Finally, we also created separated scatter plots of the FC for a panel of six key genes related to sex and reproduction: *sox9*, *dmrt1*, *amh*, Forkhead Box L2 (*foxl2*), gonadal aromatase (*cyp19a1*), and Follistatin (*fst*). Finally, we studied the mean FC and standard deviation of those genes considering the six species studied.

2.4 Gene and protein nomenclature

In this study we have used the existing and approved nomenclature which was established in vertebrates 30 years ago and reflects evolutionary relationships (McCarthy *et al.*, 2023). Thus, mammalian and avian gene symbols are in uppercase and italics (e.g. *AMH*), with the exception of rodents, which have gene symbols in lowercase and italics (e.g. *Amh*). Amphibian, reptile and fish gene symbols are in lower case and italics (e.g. *amh*). When referring to a gene in all vertebrates in general, gene symbols are in lower case and italics (e.g. *amh*). For proteins of vertebrates, we use the all uppercase gene symbol (e.g., *AMH*) as indicated in the International Protein Nomenclature Guidelines (International Protein Nomenclature Guidelines, 2020).

3. Results

3.1 Sexually dimorphic gene expression patterns during development for each species

The total number of MEGs and FEGs that were either upregulated or downregulated during ovarian differentiation and testicular differentiation varied greatly across the different vertebrate species analyzed (**Table 1**). Hence, sexually dimorphic gene expression in the gonad is achieved by activation and repression in different proportions across vertebrates (**Figures 1 and 2**). For example, in the European sea bass, there were a total of 4,129 DEGs out of the 20,978 annotated genes between 110 and 250 dpf in males and females (**Figure 1A; Supplementary table 2**). Testis differentiation occurred through upregulation of 333 genes and concomitant downregulation of 254 genes. However, ovarian differentiation involved the upregulation of 1,605 genes combined with the downregulation of 1,646 genes. The remaining 292 DEGs between stages were regulated in the same direction in both sexes. In the platyfish, out of 24,209 annotated genes, there were 6,242 DEGs between 17 and 30 dpf (**Figure 1B; Supplementary table 3**). Among the MEGs there were 1,652 genes upregulated in males and 198 downregulated in females while there were 340 FEGs upregulated in females and 1,039 FEGs downregulated in males.

Another example is the red-eared slider turtle, which showed 4,330 DEGs between stages 15 and 21 out of the 28,415 annotated genes (**Figure 2A; Supplementary table 5**). During testicular differentiation, 371 genes were upregulated in males while 876 genes were downregulated in females. Ovarian differentiation of the red-eared slider turtle involved the upregulation of 1,037 genes as well as the downregulation of 441 genes in males. A total of 1,605 genes appeared similarly regulated between stages 15 and 21 in males and females.

In the mouse (**Figure 2C**), we found 809 out of the 25,697 annotated genes that were differentially expressed between E.11.0 and E.12.0 in the 129S1 strain. Among the DEGs, 213 genes were upregulated in males and 20 were downregulated in females during testis differentiation. For development of the ovaries, 77 genes were upregulated in females combined with the downregulation of 131 genes in males. Similar expression dynamics in both sexes between E11.0 and E12.0 was observed for 368 DEGs (**Figure 2C; Supplementary table 7**). These are exactly the same results described by Munger *et al.*, (2013) when analyzing the mouse gonadal transcriptome. This is very important because: 1) it validates our analysis workflow and code to be used with the other species, and 2) allows meaningful comparisons across species since exactly the same method is applied. The total number of DEGs obtained for each species and sex is summarized in **Table 1**.

Table 1. Differentially upregulated (up) or downregulated (down) genes between the beginning (T1) and towards the end (T2) of sex differentiation in six species of vertebrates. The sex determining Mechanism (SDM), the method used to obtain the data (Microarray, M; RNA-sequencing, R) and the total number of DEG in each sex and the total number of coding genes present in each species is indicated

Group	Species	T1	T2	SDM	Method	Ref*	Males		Females		Both sexes	Total DEGs	Total genes
							UP	Down	UP	Down			
Fish	European sea bass	110 dpf	250 dpf	PSD	M	1	333	254	1,605	1,646	292	4,129	20,978
	Platyfish	17 dpf	30 dpf	XY/XX	R	2	1,652	1,039	340	198	3,013	6,242	24,209
Amphibian	Frog	NF50	NF53	ZZ/ZW	M	3	738	1,309	1,098	1,753	536	5,434	20,219
Reptile	Turtle	15 stage	21 stage	TSD	R	4	371	441	1,037	876	1,605	4,330	28,415
Bird	Chicken	E.4.5	E6.0	ZZ/ZW	R	5	432	266	361	308	2,358	3,725	20,937
Mammal	Mouse	E11.0	E12.0	XY/XX	M	6	213	131	77	20	368	809	25,697

References: 1) Ribas *et al.*, 2019; 2) unpublished; 3) Pipek *et al.*, 2018; 4) Czerwinski *et al.*, 2016; 5) Ayers *et al.*, 2015; 6) Munger *et al.*, 2013.

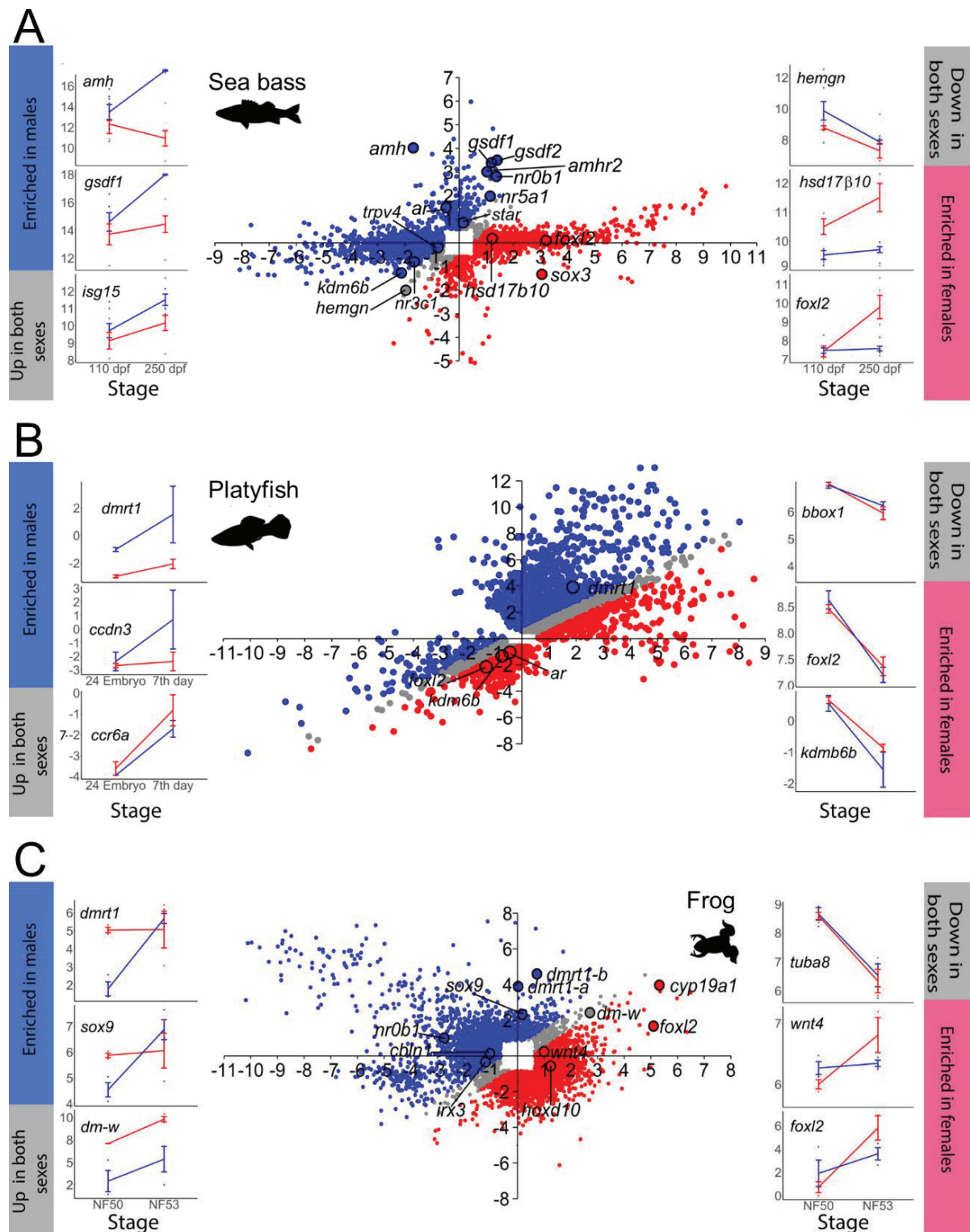


Figure 1. Gene expression changes at the beginning of sexual differentiation (T1), when gonads were still morphologically undifferentiated; and towards the end of sex differentiation (T2) in males and females. Scatterplot axes indicate $\log_2$ FC. Probes that exhibited ≥ 1.5 FC and $P < 0.05$ in at least one sex between T1 and T2 are plotted in the Y axis for males and in the X axis for females. Probes that become enriched in males relative to females at T2 are plotted in blue and probes that become enriched in females relative to males at T2 are plotted in red. Probes with similar expression in both sexes at T2 are plotted in grey in the upper right quadrant if upregulated or in the lower left quadrant if downregulated. Thus, enrichment in one sex relative to the other can be achieved by activation, repression, or a combination of both regulatory mechanisms. The line plots on the sides (blue lines = males; red line = females) show some examples of genes that are enriched in males (left, blue bar), that are upregulated in both sexes (left, grey bar), that are enriched in females (right, red bar), or that are downregulated in

both sexes (right, grey bar). The X-axis indicates the developmental stage and the Y-axis represents $\log_2$ normalized counts per million (CPM) for RNA-seq-derived data or $\log_2$ normalized intensities for microarray-derived data. **A.** sea bass between 110 dpf and 250 dpf, **B.** platyfish between 1 and 7 dpf, and **C.** frog between NF53 vs. NF50.

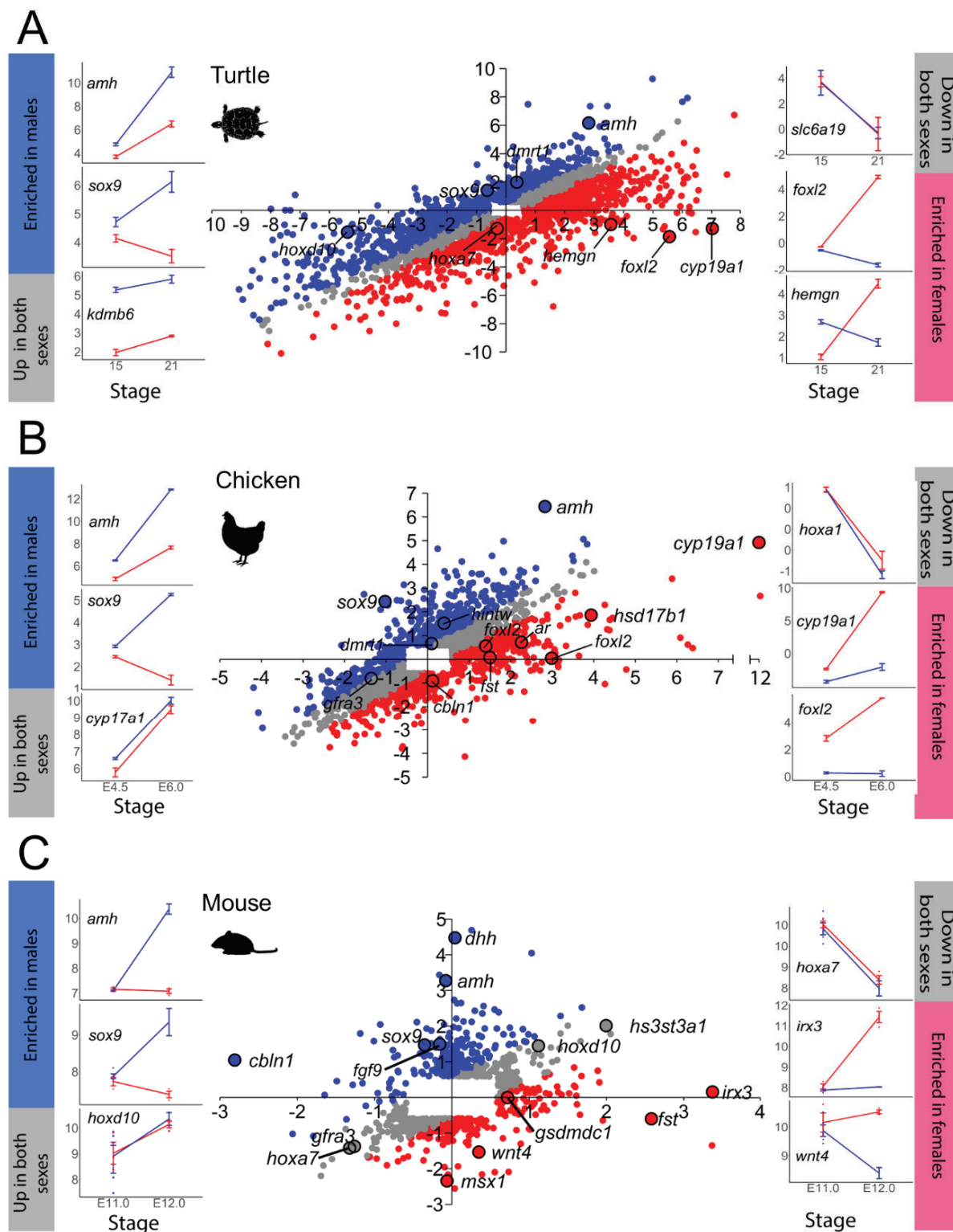


Figure 2. Same as in Fig. 1 for: **A.** turtle at 15 and 21 stages; **B.** chicken between E4.5 and E6.0 stages; and **C.** mouse between E11.0 vs. E12.0 stages.

3.2 Global transcriptomic dynamics comparison across vertebrates

Consistently across the species studied, higher expression in one sex was achieved not only through active upregulation in that sex but also by concomitant downregulation in the opposite sex or by a combination of both types of regulation. Therefore, gonadal differentiation towards a particular sex is more complex than the upregulation of the genes typically associated with that sex. Moreover, important differences among species were evident. This can be clearly seen by plotting the percentage of upregulated and downregulated DEGs per sex in each species (**Figure 3**). Overall, we found that the ratio of MEGs and FEGs that were either upregulated or downregulated during ovarian and testicular differentiation greatly varied across the different vertebrate species analyzed, including the two fish species. For example, in the sea bass, ovarian development required upregulation of about 85% of FEGs and downregulation of a similar percentage of MEGs. In contrast, testicular differentiation involved a much smaller percentage of upregulated MEGs and downregulated FEGs, around 15%. In the platyfish, these percentages were essentially reversed and thus it was testicular rather than ovarian development that involved the major number of DEGs, both up- and downregulated. Similar to the platyfish, in the mouse, ovarian differentiation required a smaller percentage of upregulated FEGs (37%) or downregulated MEGs (8%) than testis differentiation (63% FEGs and 92% MEGs). Thus, in the sea bass ovarian differentiation requires simultaneous upregulation of most FEGs and concomitant downregulation of most MEGs, while in the platyfish and the mouse active up- and downregulation takes place during testis rather than in ovarian development. Finally, in the turtle and the frog the situation tended to resemble more the situation in the sea bass, albeit with less pronounced differences between the two sexes. The chicken, however, displayed the more balanced situation since the number of MEGs and FEGs upregulated and downregulated in each sex were similar.

3.3 Comparison of key gene dynamics across species

To directly examine the gene expression dynamics across species and by sex, we concentrated on a panel of previously described key genes. By ‘key genes’ we mean genes shown to be essential for sex differentiation in vertebrates (Nagahama *et al.*, 2021; Ribas *et al.*, 2017b; Yildirim *et al.*, 2020), although this knowledge has been driven by studies in mammals. From those genes we selected six for which a significant change in expression was found ($FC \geq 1.5$ and $P < 0.05$) in at least three species and in at least one sex: *amh*, *sox9*, *dmrt1* in males and *cyp19a1a*, *foxl2*, and *fst* in females. We noted obvious differences across species (**Figure 4A**).

amh was clearly more enriched in males ($FC > 5$) with simultaneous upregulation in females ($FC \sim 2.5$) in the chicken and turtle. In contrast, in the mouse upregulation occurred exclusively in males (X value ~ 0). Yet, in the sea bass not only active upregulation in males but also concomitant downregulation in females was observed. In frog and the platyfish *amh* did not show any significant change, at least between the two developmental stages considered. The transcription factor *sox9* was consistently upregulated in males across species. However, it was downregulated during ovarian differentiation in the turtle, chicken and mouse but not in the frog and the sea bass. *dmrt1* was consistently upregulated in all but the mouse, either only in males (frog and chicken) or also with concomitant upregulation in females (particularly in the platyfish, but also in the sea bass and turtle) albeit at lower FC values (**Figure 4A**).

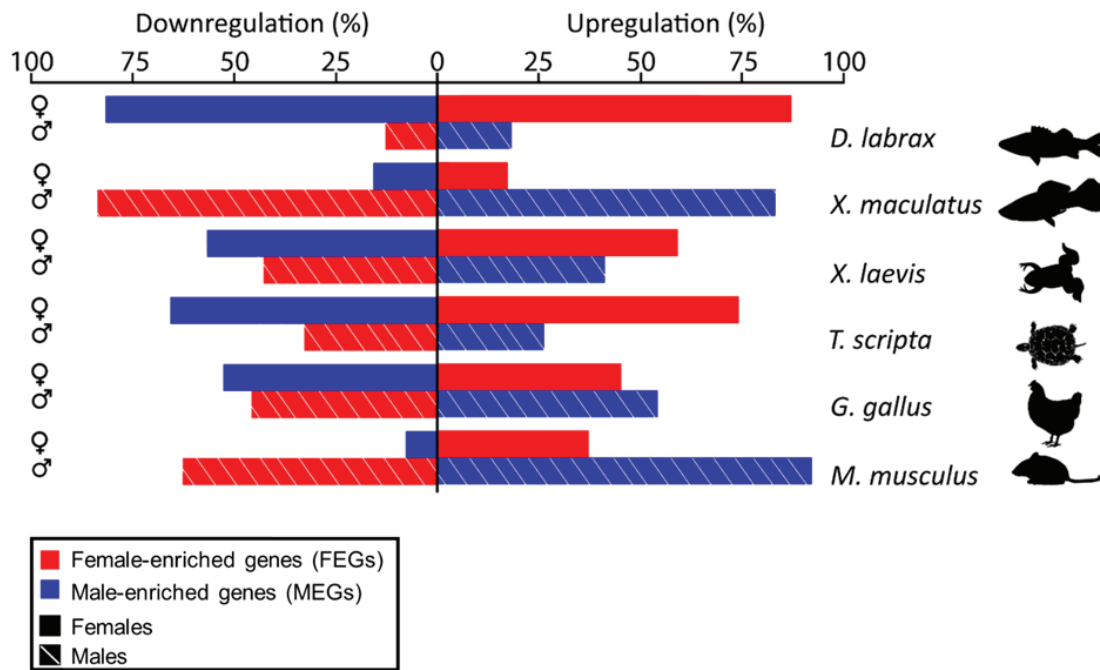


Figure 3. Bar plot showing the percent of male- (blue) or female-enriched (red) genes that are upregulated (right) or downregulated (left) during ovarian (top, solid lines) or testicular differentiation (bottom, striped lines) in the different vertebrate species analyzed and compared in this study. Notice clear species- and sex-specific differences in the amount of activation and repression of male- and female-enriched genes.

Regarding the key FEGs, *cyp19a1a* was upregulated in both sexes but with higher FC values in females. In the turtle there was active downregulation in males whereas in the platyfish downregulation in both sexes was observed. The transcription factor *foxl2* showed the most variable dynamics across species. In the frog it was upregulated in both sexes, with a higher increase in females; upregulated in females without significant change in the males of the sea bass and the chicken, downregulated in both sexes (although enriched in females at the end of the studied period) of the platyfish, and it was upregulated in females while downregulated in males of the turtle. Lastly, *fst* was upregulated in both sexes, with a higher increase in females of the frog, upregulated in females of the chicken (no change in males), and upregulated in females while simultaneously being downregulated in males of the mouse (**Figure 4A**).

A particular case was that of the platyfish, species in which only two of the six key genes showed significant changes between stages: *foxl2* achieved female enrichment through downregulation in the males and *dmrt1* became a MEG through upregulation in the males. Collectively, these findings showed that among the six key genes selected *amh*, *sox9* and *dmrt1* were consistently MEGs, had similar change in both sexes or exhibited no significant change, but they never were FEGs. Conversely, *cyp19a1a*, *foxl2* and *fst* were consistently FEGs, had similar change in both sexes or exhibited no significant change, but they never were MEGs. This applied across the six studied species regardless of their sex determining system

3.4 Variability of transcriptome dynamics of key genes of sexual differentiation across species

We also calculated the mean and standard deviation (SD) of the FC in males and females for each of the six key genes named above of the six studied species in order to better assess the magnitude and variability of the expression of these key genes according to sex. We observed higher mean FC and SD values in males for MEGs (*amh*, *sox9*, and *dmrt1*) and in females for FEGs (*cyp19a1*, *foxl2*, and *fst*)

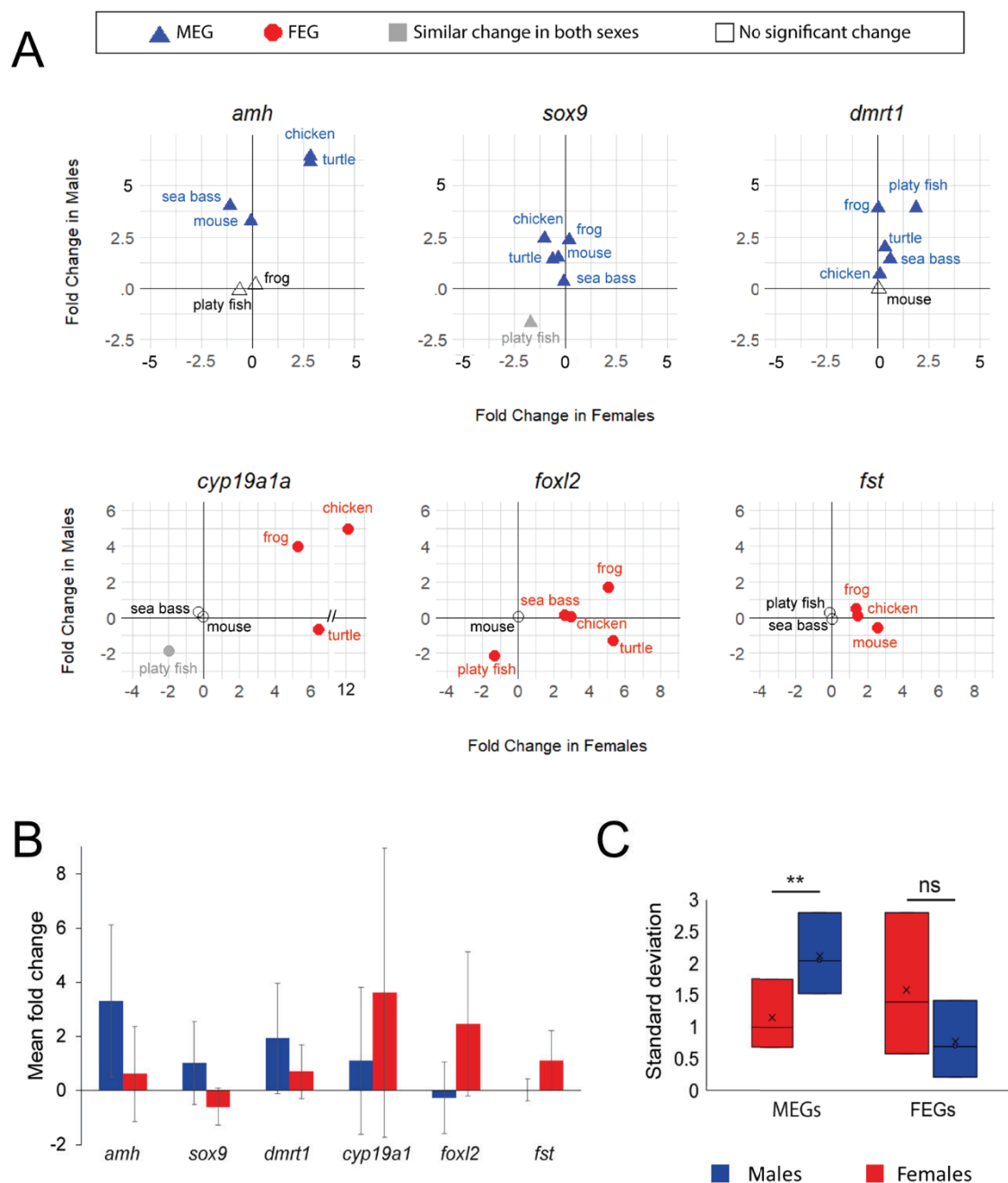


Figure 4. **A.** Changes in expression between T1 and T2 of six **key genes** involved in sexual development in vertebrates (in $\log_2$ FC): *amh*, *sox9*, *dmrt1*, *cyp19a1a*, *foxl2*, and *fst*. Male-enriched genes (MEGs) at T2 are shown in blue triangles, female-enriched genes (FEGs) at T2 are shown in red circles, genes with significant change between T1 and T2 but without differences between sexes are plotted in grey, and genes without significant change are plotted with open symbols. **B.** Bar plot of the mean FC of the six key genes mentioned above (blue for males; red for females). Bars are standard deviation and indicates the variability of FC across species. Notice higher variation and fold change values in males for MEGs and in females for FEGs, indicating that across species there is more variation in the degree of upregulation than of downregulation. **C.** Box plot of the standard deviation of the same six key genes by sex (blue in males; red in females). This shows higher sex-specific variability in male-enriched genes (MEGs) when compared to female-enriched genes (FEGs).

(**Figure 4B**). Furthermore, sex-related differences in SD values were significant ($P < 0.01$) for MEGs (**Figure 4C**). Together, these results suggest that, the six selected key genes involved in testicular differentiation are mostly driven by upregulation while more variable dynamics were found for key

FEGs, and that there is more gene expression variability across species in the sex for which their expression is most important (MEGs in males and FEGs in females).

3.5 Comparison of top gene dynamics across species

The approach followed in this study not only allows revealing important differences in the transcriptomic dynamics of key genes across species but also can help uncovering other important genes involved in sex differentiation of vertebrates previously not associated with this process. We reasoned that genes with the highest absolute FC values could be relevant. Thus, we filtered the datasets of the six species to retain only the top 10% of DEGs with highest absolute FC values regardless of sex and direction of change (from now on referred to as ‘top genes’). Then, we compared the six gene lists to find top genes shared among at least three or more species. There were 19 top genes including three key genes, *i.e.*, genes previously known to have essential roles in sex differentiation of vertebrates: *amh*, *cyp19a1* and *foxl2* (**Table 2**).

Notably, while the six key genes were consistently enriched in one sex across all species compared (expression was either significantly enriched in one sex, in both, or non-significant), less conserved enrichment was found among the top genes—those that were not previously classified as key genes—which were enriched in one sex in some species and in the opposite sex in other species. Some examples are shown in **Figure 5**. Thus, in contrast to what has been shown earlier for the key genes, the most expressed genes are not necessarily enriched always in the same sex towards the end of sex differentiation in different vertebrate species.

The gene that stand out first of this list besides the previously known key genes is *fshr*. This gene is the receptor for follicle stimulating hormone and functions in gonad development of many species but is not always considered a key gene for vertebrates. The results show that even though it was not a DEG between stages in the mouse, important gene expression changes occurred in the rest of the five species studied. In the frog, the sea bass and the platyfish *fshr* was classified as a MEG with relatively little change in the opposite sex, while in the turtle and the chicken, *fshr* was classified as a FEG, with moderate upregulation in males. This constitutes a clear example of genes enriched during sex differentiation in one sex or another depending on the species.

Some examples of other genes that met the strict filtering criteria to become top genes were: deleted in azoospermia like (*dazl*), an RNA binding protein classified as MEG in the platyfish and the frog but FEG in the turtle. The gene uncoupling protein 3 (*ucp3*) encodes for a mitochondrial uncoupling protein member of the mitochondrial anion carrier proteins family. It was classified as a MEG in the platyfish and the frog but as a FEG in the sea bass. Carboxypeptidase B1 (*cpb1*) is an enzyme classified as MEG in the frog and the sea bass but as FEG in the platyfish. Asparaginase (*aspg*) is a coding enzyme protein classified as MEG in the frog and the sea bass but as FEG in the turtle. Claudin 10 (*cldn10*), which encodes for a component of the cellular membrane that works as a tight junction strand, is classified as MEG in the mouse, the platyfish and the sea bass but as FEG in the turtle.

Table 2. List of the genes found in common in at least three of the six species of vertebrates considered in this study which were amongst the top 10%-fold change expression between stages in at least one sex, either up or downregulated. Numbers in parenthesis indicate the number of species. Among these 19 genes there are three (highlighted in bold) previously known to be involved in sex differentiation (named as key genes in this study). The information in the ‘Gene type’ column was obtained from GeneCards database (Stelzer *et al.*, 2016)

Species where found as top genes	Gene symbol	Gene name	Gene type
(4) sea bass, frog, turtle, chicken	<i>foxl2</i>	Forkhead Box L2	Transcription factor
(4) sea bass, frog, turtle and chicken	<i>fshr</i>	Follicle Stimulating Hormone Receptor	Receptor
(4) sea bass, turtle, chicken, mouse	<i>amh</i>	Anti-Müllerian Hormone	Protein hormone
(3) frog, turtle, chicken	<i>cyp19a1</i>	Cytochrome P450 Family 19 Subfamily A Member 1	Enzyme
	<i>asz1</i>	Ankyrin Repeat, SAM And Basic Leucine Zipper Domain Containing 1	Binding protein
(3) platyfish, sea bass, frog	<i>cpb1</i>	Carboxypeptidase B1	Enzyme
	<i>ucp3</i>	Uncoupling Protein 3	Mitochondrial uncoupling protein
(3) sea bass, frog, chicken	<i>myl2</i>	Myosin Light Chain 2	Sarcomeric protein
	<i>iyd</i>	Iodotyrosine Deiodinase	Enzyme
(3) frog, chicken, mouse	<i>spon2</i>	Spondin2	Binding protein
(3) sea bass, frog, mouse	<i>eno3</i>	Enolase3	Isoenzyme
(3) sea bass, frog, turtle	<i>hmg20a</i>	High Mobility Group 20A	Binding protein
	<i>aspg</i>	Asparaginase	Enzyme
(3) sea bass, turtle, chicken	<i>g0s2</i>	G0/G1 Switch 2	Mitochondrial protein
	<i>unc45b</i>	Unc-45 Myosin Chaperone B	Regulatory component of progesterone receptor
(3) platyfish, frog, chicken	<i>ccna1</i>	Cyclin A1	Binding protein
(3) platyfish, frog, turtle	<i>dazl</i>	Deleted In Azoospermia Like	RNA binding protein
(3) platyfish, turtle, mouse	<i>cdkn1a</i>	Cyclin Dependent Kinase Inhibitor 1A	Inhibitor enzyme
(3) platyfish, sea bass, turtle	<i>cldn10</i>	Claudin 10	Tight junction strand (integral membrane protein)

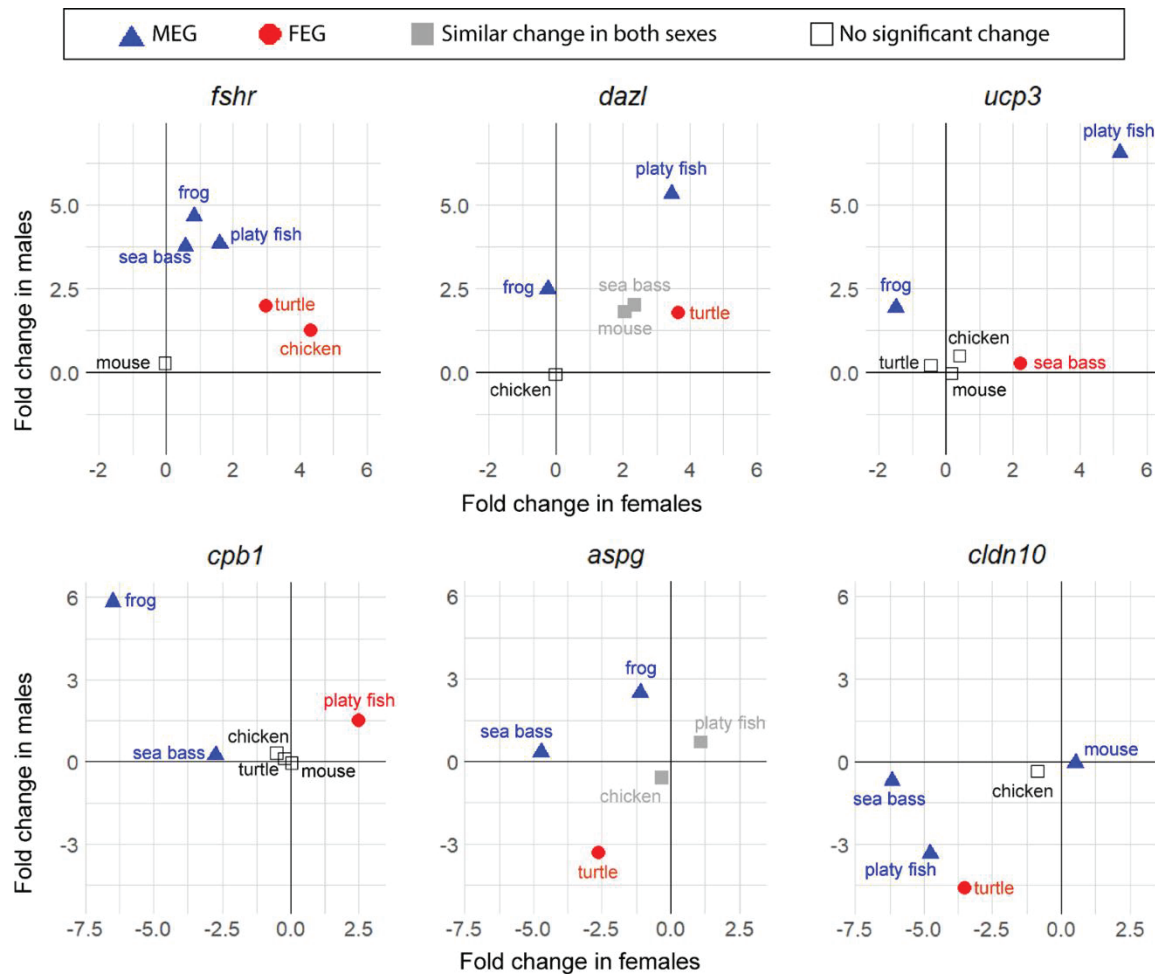


Figure 5. Changes in expression between T1 and T2 of six **top genes** involved in sexual development in vertebrates (in log₂ FC): *fshr*, *dazl*, *ucp3*, *cpb1*, *aspg*, and *cldn10*. Male-enriched genes (MEGs) at T2 are shown in blue triangles, female-enriched genes (FEGs) at T2 are shown in red circles, genes with significant change between T1 and T2 but without differences between sexes at T2 are plotted in grey, and genes without significant change between T1 and T2 are plotted with open symbols.

4. Discussion

Sex differentiation requires the orchestrated participation of many genes tightly regulated in a strict spatial and temporal fashion. The current consensus in the field is that key genes involved in gonadal differentiation are relatively conserved, but that “their relative positions and, in some cases, their functions in testicular and ovarian differentiation differ” (Nagahama *et al.*, 2021). In the present study, we aimed answering how actually conserved is the expression program of genes involved in gonadal sex differentiation between species during the beginning and towards the end of gonadal differentiation. By conserved we mean: 1) that they are expressed at comparable stages across species, 2) whether they participate in the differentiation of the same sex across species as inferred from a strong expression enrichment in that sex, and 3) what regulatory mechanisms (upregulation, downregulation or a combination of both) are used to achieve differential expression between the two sexes.

Traditionally, sex differentiation has been studied based on quantification of gene expression differences between one sex against the other, at one or more developmental stages. However, understanding this process also requires identifying mechanisms leading to gene enrichment in one sex relative to the other between two developmental stages, for each sex separately. To date, such different perspective has been

used on the mouse only (Munger *et al.*, 2013) but —and to the best of our knowledge— surprisingly in no other species. In this paper, we characterized expression dynamics in five other species, and also performed a systematic comparison across six vertebrates. We found that, even though the six selected key genes of sex differentiation (*amh*, *sox9*, *dmrt1*, *cyp19a1*, *foxl2* and *fst*) show a consistent sex-specific expression trajectory connected to their known functions across the studied species (Nagahama *et al.*, 2021), their dynamics can be variable. Further, the proportion of genes following different regulatory mechanisms also greatly varies across species. Finally, we identified 16 potential markers of early sex differentiation, including *fshr*, *dazl*, *cldn10* and *aspg*. These genes show the highest increase in expression between the two studied developmental stages in at least three of the six studied species. This warrants further studies to determine their role during gonadal sex differentiation. Together, these results unambiguously show that despite some players are fairly conserved, there are important overall differences in the expression dynamics between vertebrates. Hence, we encourage further research on more species for which the gonadal transcriptome might be available, while highlighting the need to create new data at comparable developmental stages.

Even though the selection of comparable developmental stages was made with particular care, it is crucial to consider them when interpreting the results, since another selection could lead to different number of DEGs between developmental stages. An additional challenge when comparing transcriptomic data from different species is related to the methods used to obtain the datasets. The comparison of data from microarray and RNA sequencing is possible, as high correlation between gene expression (FC) values was shown when relating the same samples in the two platforms (Black *et al.*, 2014; S. Zhao *et al.*, 2014). Nevertheless, several adjustments were made to ensure proper comparison of data analyzed in this study (see methods section). Furthermore, we validated the results using raw data files from the mouse to exactly replicate the results with our pipeline. The common limitation of all transcriptomic comparative studies is that these directly rely on the quality of reference genome annotation (Roux *et al.*, 2015). Hence, if one gene is well annotated in many species this increases the chances of that gene being found as conserved. However, if one gene is poorly annotated then it could be overlooked after filtering steps when looking for common gene name or symbol.

Despite these limitations and the need for caution when comparing between species, our results disclose new insights for the understanding of the complex process of sex differentiation in vertebrates. Overall, we found that a large proportion, between 3 to 27% (mean = 18%), of the known coding genes changed their expression between T1 and T2 in all studied species. This is in accordance with a comprehensive study on the transcriptional dynamics in human fetal testis and ovary development by (Lecluze *et al.*, 2020), which identified in the order of thousands coding and non-coding genes with early sexually dimorphic expression profiles associated with sex differentiation.

The effective manner to display gonadal transcriptomic data designed by Munger *et al.*, (2013) in a scatterplot allowed obtaining a global view of the mechanisms used to achieve dimorphic expression changes in the gonad of the mouse. Surprisingly, these proportions remained completely unknown until now for other vertebrates. The present study reveals that in other species these proportions are completely different. To date, the only study that globally compared transcriptomes during sexual differentiation was focused on studying the order in which groups of genes were expressed simultaneously or not between turtle and mouse (Czerwinski *et al.*, 2016). These authors found that even though antagonistic and strongly canalized forces were conserved, groups of shared genes between the mouse and the turtle showed heterochronic dynamics of genes which converged to upregulate or downregulate aromatase (Czerwinski *et al.*, 2016).

Using such approach of the scatterplot in the mouse, it was found that not only upregulation is involved; “downregulation in the opposite sex” also occurs to reach gene dimorphism. In mouse, most genes with higher expression in males achieve this through upregulation in the XY gonad and only a small group of genes through downregulation in XX gonads. However, approximately half of all FEGs were upregulated in the XX gonad and the other half were downregulated in the XY gonad (Capel, 2017; Munger *et al.*, 2013). Thus, in the mouse testis development involves changes (up- and downregulation) of a greater number of genes than ovarian differentiation. This observation aligns with the concept proposed by Jost (1947) regarding sex differentiation stating that females are the default sex, less dependent on gonadal hormones. Regarding the other species, we found for the first time that the same mechanism also occurs in the platyfish by a different proportion of genes (about 85% of the FEGs achieved enrichment through downregulation in the males). Surprisingly, the opposite strategy occurs in other species: a large proportion of genes reached higher expression in males through downregulation in females. This was found in the sea bass (82% of its MEGs) and, to a lesser extent, in the turtle (66%) and the frog (57%). Hence, particularly in the sea bass, the change of in expression of a higher number of genes was required to develop an ovary than a testis. Taken together, these results reveal that there is no relationship between how dimorphic expression in the gonad is achieved and position in the vertebrate phylogenetic tree.

The comparison of the gonadal transcriptome between species confirms that the key genes for sex differentiation that we studied are fairly conserved, as previously suggested (Capel *et al.*, 2017; Herpin and Scharl 2015; Nagahama *et al.*, 2021). For example, *amh*, *sox9* and *dmrt1* achieved sex-dimorphic expression through upregulation in males and relatively little or no change in females. Interestingly, the standard deviation of the mean FC of these genes between species was higher in the males than in the females, suggesting that evolutionary pressures could be acting in a sex-specific manner. On the other hand, key genes related to ovarian development (*cyp19a1*, *foxl2* and *fst*) showed non-significant differences between sexes on the standard deviation of FC. These genes showed more variable expression dynamics. For instance, *foxl2*, a transcription factor required for ovarian differentiation classified as FEG in five of the six studied species (except mouse) achieved higher expression in females at T2 through very different mechanisms depending on the species. Such type of information was, to the best of our knowledge, not shown in such a clear and comparative manner as the present study does.

In other cases, genes such as the androgen receptor (*ar*), which encodes for a protein that binds androgens to stimulate transcription of its target genes, showed less degree of conservation across species regarding its expression during gonadal differentiation. In fishes, the role of *ar* is variable. In medaka, treatment with an androgen receptor antagonist, cyproterone acetate, induced secondary feminization (Schreck *et al.*, 1974), suggesting a role of *ar* in testis development. We found that in the sea bass *ar* is a MEG through upregulation in the males and kept constant in the females, in agreement with a previous study showing an important role for testis differentiation of *ar* (Blázquez and Piferrer, 2005). However, in the platyfish this gene could be important for ovarian differentiation since it was a FEG through stronger downregulation in the males than in the females during the period considered. Thus, these results support the notion that important genes composing the network involved in differentiation can be substantially different even among closely-related species (Capel, 2017; Herpin and Scharl, 2015).

So far we have concentrated in differences in the regulatory dynamics of genes previously known to be involved in testicular and ovarian differentiation across species. We wondered whether the approach used in this study could also reveal genes that had not been considered as key genes in gonad sex differentiation. We reasoned that these genes (referred to as ‘top genes’) could exhibit important changes

in expression in various of the analyzed species in this study. Thus, we searched among the 10% genes with the highest FC between T1 and T2, in an analogous manner than in studies seeking gene markers (Khazaei *et al.*, 2018; Marshall *et al.*, 2013; Miller *et al.*, 2011). Among these top genes we found some without previously known relation to gonadal development, such as *myl2*, involved in embryonic heart muscle structure and function (Weterman *et al.*, 2013). Many of the top genes were known for having a function related to reproduction or expressed in the gonadal tissue at some point but they were not previously associated with gonadal differentiation. For instance, *dazl* was shown to be involved in the differentiation of germ cells (Xi and Li, 2009); *cldn10*, was identified as a biomarker for ovarian cancer in the chicken (Seo *et al.*, 2010); and *asz1*, was related with gametogenesis and piRNA metabolic process (Wang *et al.*, 2017; Yi *et al.*, 2014).

Among the top genes identified, *fshr* deserves attention. Although it was not enriched in the same sex for all species, high significant expression changes were found in sea bass, platyfish, and frog as MEG, and in turtle and chicken as FEGs. *fshr* is an important gene for female differentiation and maintenance in gonochoristic fishes like Japanese medaka (Murozumi *et al.*, 2014) and zebrafish (*Danio rerio*) (Zhang *et al.*, 2015), while in sea bass it plays an important role during testis differentiation. In the hermaphroditic honeycomb grouper (*Epinephelus merra*), this gene is directly involved in gonadal sex change (Kobayashi *et al.*, 2010). Other than in fish, *fshr* is important for gonadal development of amphibians (Nakamura, 2013) and birds (Akazome *et al.*, 2002). Thus, we suggest that this gene should be considered as a key gene involved in vertebrate gonadal sex differentiation.

The evolution of genes and mechanisms involved in vertebrate sexual development has given rise to various theories, including the bottom-up theory proposed by Adam Wilkins (Wilkins, 1995). This theory suggests that purifying selection acts only at the bottom of a hierarchical cascade. This theory was supported by (Graham *et al.*, 2003) who referred to the genes at the top as "masters" and those at the bottom as "slaves". Accordingly, while masters can vary from one species to another, the slaves maintain structural and functional conservation across species (Graham *et al.*, 2003; Williams and Carroll, 2009). Since then, subsequent studies revealed a more complex network of antagonizing pathways involved in gonadal sex differentiation and, importantly, with different genes participating in the regulation of male and female pathways even between closely related species (Herpin and Scharl, 2015; Heule *et al.*, 2014; Martinez-Bengochea *et al.*, 2022; Pan *et al.*, 2016; Stöck *et al.*, 2021). Nowadays it is accepted that, despite the diversity of sex determining mechanisms, gonadal sex differentiation involves a complex network of multiple regulatory interactions of genes associated with male and female pathways. Those are expressed simultaneously as opposing forces, until a factor or combination of them tilts the balance towards one fate or the other (Capel, 2017). Our results support this view and provide the first truly comprehensive and quantitative examination of gene expression differences during sex differentiation across vertebrates.

As shown in this study, the six selected key genes for gonadal sex differentiation when they were enriched it was always in the same sex, which supports Lynch hypothesis stating that only the final gene product produces a phenotype exposed to selection (Lynch, 2007). Therefore, developmental pressures can act upstream of the key genes without changing the final phenotype (Herpin and Scharl, 2015; Lynch, 2007). However, other genes such as *fshr* and *ar* were involved in testicular or ovarian differentiation depending on the species, as discussed above. This is in line with another previous hypothesis based on the link between a sex-determining gene to a sex differentiation gene (Herpin and Scharl, 2015; van Doorn and Kirkpatrick, 2007). Based on what has been discussed above and the behavior of key genes (if enriched, always in the same sex) and the genes with the highest FC (enriched in one sex or the other, depending on the species), it is tempting to propose that neither a hierarchical

cascade nor a network where all nodes are equally important to produce a male or female phenotype, but instead a hierarchical network with important hub genes and less connected nodes controlling the process. Hub genes would correspond to genes, including key genes, that are always enriched in the same sex while less connected genes would be those genes involved in testicular or ovarian differentiation depending on the species. Such networks with hub genes, some of them key genes such as *sox9* and *amh*, were characterized during gonadal sex differentiation of sea bass, mouse and humans (Sánchez-Baizán *et al.*, 2022). Thus, in such a scenario, one could speculate that evolutionary pressures may affect genes depending on the number of connections that they have in the network, with hub genes more robust to changes, which would explain why, if they are enriched, they are always enriched in the same sex at the end of gonadal differentiation across vertebrates. This is compatible with the basic underlying principle about the existence of antagonizing signals that ensure canalization of the male or female developmental pathway and that tolerate plasticity (Capel, 2017).

To conclude, we investigated the variability in gene expression dynamics during gonadal sex differentiation across six vertebrate species with different sex determining mechanisms. We found that the proportion of genes being up- or downregulated leading to sex-related differences in gene expression varied greatly across species. Interestingly, we found that a large proportion of genes acquired enrichment through downregulation in the opposite sex, not only in females of the mouse but also in the platyfish and males of sea bass, turtle, and frog. Also, the six studied key genes were consistently enriched in one sex only (if enriched) across the six studied species. Of note, among these key genes, those involved in testicular differentiation were mostly driven by upregulation, while slightly more variable dynamics were found for key FEGs. The key genes involved in ovarian differentiation were mostly enriched by upregulation in the females but in some instances also by different dynamics: downregulation in the males (such as *foxl2* in the platyfish), upregulation in both sexes with higher increase in the females (*cyp19a1* in the frog and the chicken) or a combination of upregulation and downregulation (*foxl2* in the turtle and *fst* in the mouse). Additionally, we identified 16 new markers of early sex differentiation in vertebrates. Overall, this study provides the first comprehensive and quantitative examination of gene expression differences during sex differentiation across vertebrates, shedding light on the mechanisms underlying gonadal sex differentiation and the variability across species. Based on the results of this study and previous literature we suggest that a hierarchical network with important hub genes and less connected nodes underlies the process of sex differentiation, where evolutionary pressures could affect genes depending on the number of connections. These findings could have important implications for understanding the evolution of sex determination and differentiation in vertebrates.

**Extraordinary variability in gene activation and repression
programs during gonadal sex differentiation
across vertebrates**

Núria Sánchez-Baizán, Ignasi Jarne-Sanz, Alvaro S. Roco, Manfred Scharl,
Francesc Piferrer

Supplementary Information

Background information of the selected species

Fishes

Teleosts are the vertebrates exhibiting the highest variety of sexual systems, where gonochorism, hermaphroditism (both simultaneous and sequential) and unisexuality are found. However, about 95% of teleosts are gonochoristic (Pla *et al.*, 2021). Also, they are the group with the most varied range of sex determination mechanisms from GSD to ESD, including chromosomal sex determination (CSD) and polygenic sex determination (PSD) (Mank *et al.*, 2006; Scharl *et al.*, 2023).

The sea bass (*Dicentrarchus labrax*) is a gonochoristic species with PSD influenced by the environment (Piferrer *et al.*, 2005; Vandeputte *et al.*, 2007; M. Vandeputte and Piferrer, 2019). The first signs of gonadal differentiation at the molecular level are found at 110 days post fertilization (dpf), where the expression of several genes becomes sexually dimorphic: for example the markers: Hydroxysteroid 17-Beta Dehydrogenase 10 (*hsd17 β 10*), upregulated in females (Ribas *et al.*, 2019) and *sox9a* upregulated in males (Sánchez-Baizán *et al.*, 2022). Nevertheless, the first marker identified of ovarian differentiation is gonadal aromatase, Cytochrome P450 Family 19 Subfamily A Member 1 (*cyp19a1a*), which is expressed higher in females around 120 dpf (Blázquez *et al.*, 2008). Sex differentiation is complete by 200 dpf. The gonads used in this study were from fish of 5.2 ± 0.5 cm of standard length (SD) at 110 dpf, females of 12.7 ± 5.7 cm of SD at 250 dpf and males of 11.2 ± 0.6 cm of SD at 250 dpf.

The platyfish (*Xiphophorus maculatus*), is a species of the livebearing (ovoviviparous) genus *Xiphophorus*. This genus is particularly diverse in sex-determining systems. The platyfish has a CSD with multiple sex chromosome mechanisms (WY, WX or XX females; XY or YY males) (Zhang *et al.*, 2011). Although the gonadal transcriptome has been studied to identify genes differentially expressed between male and female gonads in adults, the transcriptome has never been studied during the period of sex differentiation. RNA-seq transcriptomes were obtained from T1 = embryonic stage 24 (approx. 17 dpf, (Tavolga, 1949) which follows the sex determination stage, and T2 = 7 d old newborn fish (approx. 30 dpf), representing the later gonadal differentiation stage. The highly inbred platyfish strain Jp163A (WLC#6628) was used, which has XY sex determination with no environmental modification noticed under laboratory conditions. All animals were kept and sampled in accordance with the applicable EU and national German legislation governing animal experimentation. In particular, all experimental protocols were approved through an authorization (55.2532-2-215) of the Veterinary Office of the District Government of Lower Franconia, Germany, in accordance with the German Animal Protection Law (TierSchG) and in accordance with ARRIVE guidelines. All fish were cut into three pieces: head, abdomen with gut removed, and tail. The head and tail portion were used for PCR genotyping and the abdomen containing the developing gonad for RNA extraction.

Amphibians

Amphibians exhibit GSD with either male (XY/XX) or female (ZZ/ZW) heterogamety. The only sex-determining gene characterized in anurans so far is the W-linked DM-domain gene (*dm-w*). *dm-w* is a female-specific gene and was described in the African clawed frog (*Xenopus laevis*) with ZZ/ZW heterogamety (Yoshimoto *et al.*, 2008). However, this gene did not show a relevant role during gonad differentiation (Nagahama *et al.*, 2021; Stöck *et al.*, 2021). Instead, key genes involved in ovarian differentiation of *X. laevis* are *foxl2*, *cyp19a1*, and *fst*. The development of testis characterized by the expression of *amh*, Cytochrome P450 Family 17 Subfamily A Member 1 (*cyp17a1*), *dmrt1*, Follicle Stimulating Hormone Receptor (*fshr*), *ptgds*, and *sox9* (Piprek *et al.*, 2018). The individuals used in this study were staged according to

Nieuwkoop and Faber (1964) and the selected stages were NF50 and NF53. Morphologically, at NF50 the gonads are undifferentiated with few germ cells. The earliest visible signs of differentiation were described at NF53 (approximately 23 dpf) when there is visible difference in the location of the germ cells (Pipek *et al.*, 2017). However, at the molecular level there were clear changes of gene expression of sex differentiation key genes (Pipek *et al.*, 2018). We highlight that (Pipek *et al.*, 2017) defined NF50 as undifferentiated gonads and the NF53 stage as the onset of differentiation based on morphological differences. However, at the molecular level the differences appeared prior to morphological differentiation. We selected the chosen stages based on several comparisons of the transcriptome between NF50, NF53 and NF56 as well as the results from Pipek *et al.*, (2018) to finally consider NF50 as the beginning of molecular differentiation and NF53 towards the end because at NF56 the ovaries are completely recognizable.

Reptiles

Most squamates present GSD while ESD is predominant in all crocodiles, most turtles, and some lizards. To represent reptiles, we studied the gonadal transcriptome of a turtle (*Trachemys scripta elegans*). This species is a TSD species, whose sex differentiation starts at stage 15 and ends at stage 21. The male gonadal development begins with the activation of *dmrt1*, which leads to upregulation of *amh* and *sox9*. At female promoting temperatures (FPT), *cyp19a1* and *foxl2* are the key genes leading to ovarian differentiation (Czerwinski *et al.*, 2016; Nagahama *et al.*, 2021). The data used in this study was obtained from gonads sampled from embryos kept at male- or female-producing temperatures. The stages selected for the study were stage 15 and stage 21 based on the comprehensive time course analysis of gonadal development. Czerwinski *et al.*, (2016) found only a total of 10 DEGs between sexes at stage 12, and hence, considered the beginning of molecular sex differentiation at stage 15. At stage 21 the gonadal phenotype is irrevocably set and towards the end of molecular changes to achieve differential expression between sexes.

Birds

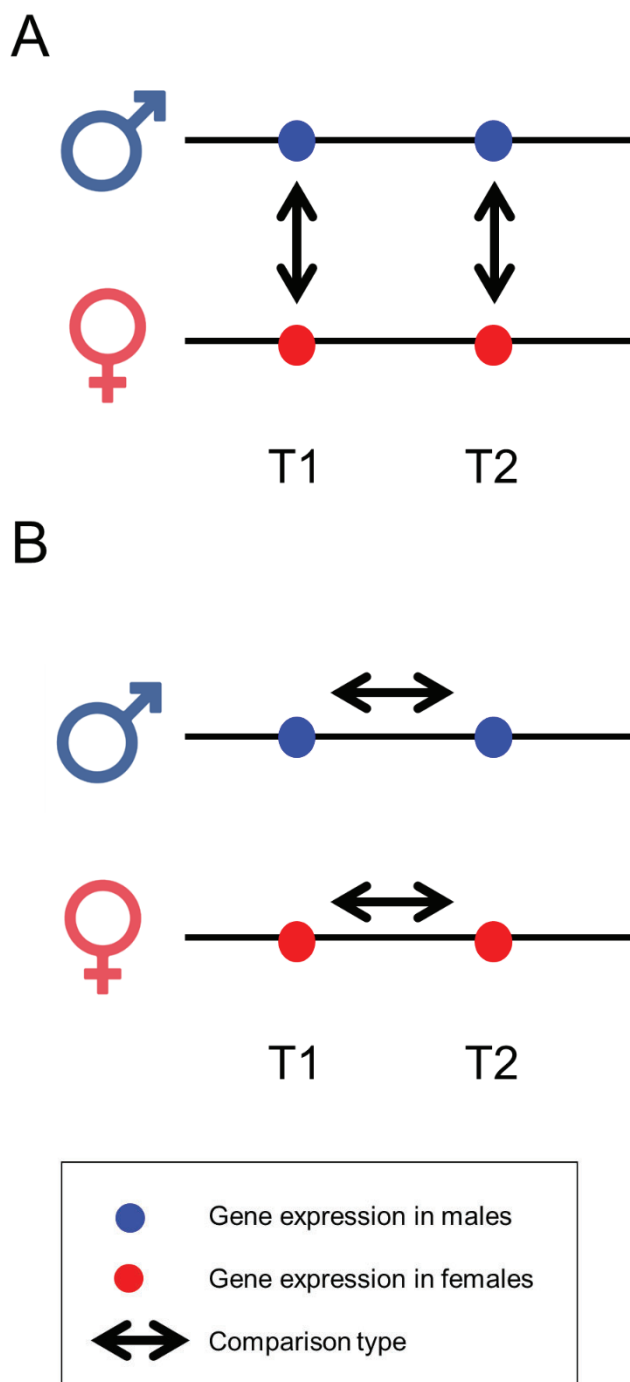
Birds evolved from theropod dinosaurs in the late Jurassic and, similar to some amphibians, some reptiles and numerous fish species, they have female-heterogametic sex chromosomes, *i.e.*, a ZZ/ZW system (Fridolfsson *et al.*, 1998). Yet, unlike in *Xenopus*, a female sex-determining gene has not been discovered. Instead, testis development is initiated by the activity of the double doses of *DMRT1* in the primordial germ cells, followed by coordinated activation of *HEMGN*, *AMH*, and *SOX9*. In contrast, ovarian development is dominated by estrogen production regulated through genes such as *CYP19A1*, *FOXL2*, R-Spondin 1 (*RSPOI*), and *WNT4* (Chue and Smith, 2011; Nagahama *et al.*, 2021). The samples used were studied by (Ayers *et al.*, 2015) who compared embryonic stages E4.5 (stage 26) and E6.0 (stage 29; (Hamburger and Hamilton, 1951) stages). They classified gonads at E4.5 at the beginning of differentiation or “prior to differentiation” and differentiating gonads at E6.0.

Mammals

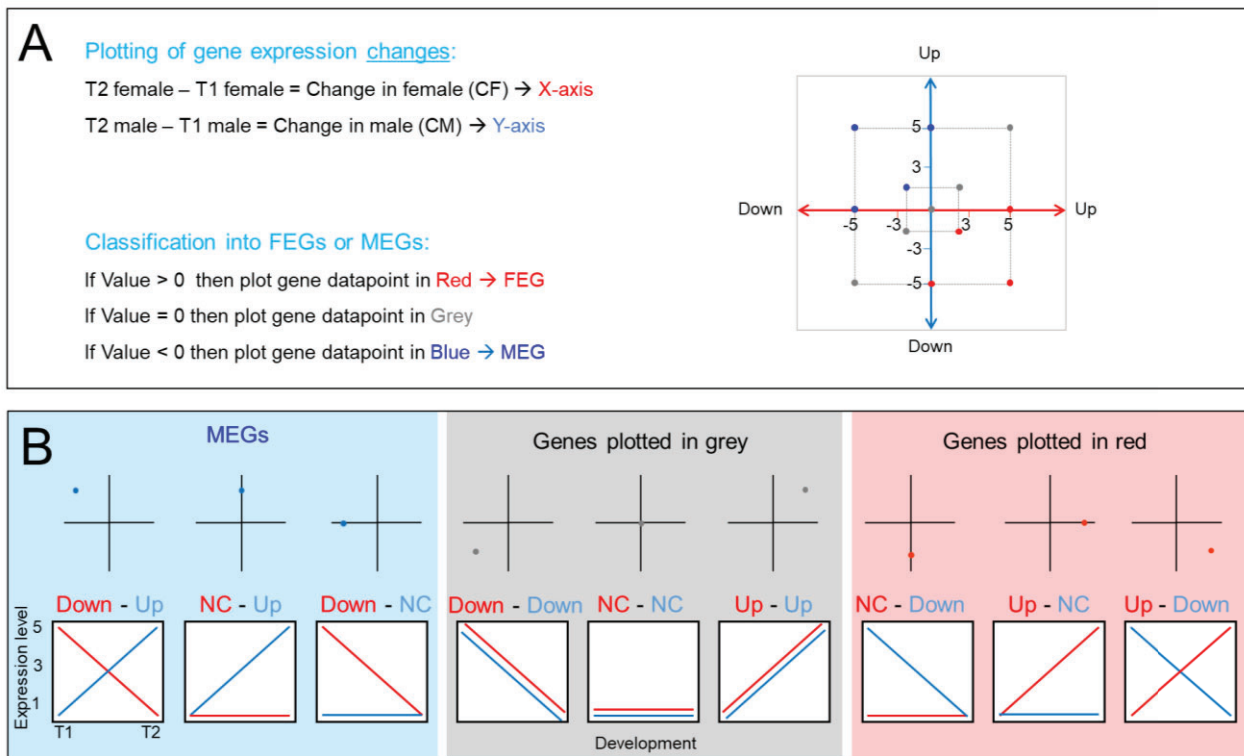
Placental mammals have GSD with heteromorphic sexual chromosomes (XX/XY) system. The mouse, *Mus musculus*, is the model organism for which sexual development is most well-known within vertebrates. Up to the embryonic stage (E.)11.2 the genital ridges are bipotential. The *Sry* determining gene is found in the Y chromosome and initiates the development of testes between E.10.0 and E.11.2. *Sry*, activates *Sox9* with a positive feedback loop existing between *Pdg2/Sox9* and Fibroblast Growth Factor 9 (*Fgf9*)/*Sox9*. In the XX gonads, genes such as *Rspo1*, *Wnt4*, and *Foxl2* are expressed during early ovarian development (Q. Wang *et al.*, 2019; Yildirim *et al.*, 2020). The data used for the present study were obtained from Munger *et al.*, (2013) who sampled individually staged gonads by the number

of somites of the embryos. The selected stages were from individuals at E11.0 and E12.0 of the 129S1 strain.

Supplementary figures



Supplementary figure 1. Diagram of the two comparisons discussed in the present study. **A.** The gonadal transcriptomes of the two sexes are compared at one or more developmental times (T1, T2) or stages. **B.** The gonadal transcriptome at two different developmental times or stages, *e.g.*, at the beginning (T1) and towards the end (T2) of the sex differentiation process are compared within each sex.



Supplementary figure 2. Diagram of the approach used to study transcriptomic dynamics using the “horizontal comparison”, T2 vs. T1, as described in Suppl. Fig. 1B. The two gene expression data points or stages (T1 and T2 in males and females) make up a total of four data points that are converted into a single data point in the X-Y coordinates of a scatterplot, in units of fold change (FC). Abbreviations: MEGs, male-enriched genes; FEGs, female-enriched genes. **A.** The diagram shows where the FC is placed for each sex. Also, it indicates the directions of the regulatory mechanisms (up- and downregulation for each sex, *i.e.*, each axis). **B.** The nine major theoretical possible combinations of regulatory mechanisms for males (blue line) and females (red line) between two developmental time points or stages (T1 and T2) using theoretical gene expression values (0 to 5). These are: 1. Down in female and up in male, 2. No change in female and up in male, 3. Down in female and no change in male, 4. Down in both sexes, 5. No change in both sexes, 6. Up in both sexes, 7. No change in females and down in males, 8. Up in females and no change in males, 9. Up in females and down in males. The cross above represents the X and Y axes shown in A and the dot the corresponding data point. The blue background depicts instances where a given gene is classified as a MEG, and this can be achieved basically in three ways between T1 and T2: 1) upregulation in males and concomitant downregulation in females (left panel), 2) upregulation in males and no change in females (middle panel), and 3) no change in males and downregulation in females (right panel). The same principle but with opposite direction changes apply for the FEG (red background).

Supplementary tables

Supplementary tables are found in an additional Excel file that will be made available online upon publication of the manuscript.

Supplementary table 1. Fishes Differentially expressed genes (DEGs) between 110 and 250 days post fertilization (dpf) in males and females of the European sea bass. The table includes information on gene ID and gene description. It also includes the following information for each sex: $\log_2$ fold change between stages (logFC), average expression (AveExpr), t statistic value (t), *P* - value (P.value), and adjusted *P* - value (adj.P.val). Information for each sex is indicated with .x for data from the female samples and .y for data from the male samples

Supplementary table 2. Fishes DEGs between 17 and 30 dpf in males and females of the platyfish. The table includes information on gene ID and gene description. It also includes the following information for each sex: $\log_2$ fold change between stages (logFC), average expression (AveExpr), t statistic value (t), *P* - value (P.value), and adjusted *P* - value (adj.P.val). Information for each sex is indicated with .x for data from the female samples and .y for data from the male samples

Supplementary table 3. Amphibians DEGs between stages NF50 and NF53 in males and females of the African clawed frog. The table includes information on gene ID and gene description. It also includes the following information for each sex: log₂ fold change between stages (logFC), average expression (AveExpr), t statistic value (t), *P* - value (P.value), and adjusted *P* – value (adj.P.val). Information for each sex is indicated with .x for data from the female samples and .y for data from the male samples

Supplementary table 4. Reptiles DEGs between 15 and 21 days of development in males and females of the red-eared slider turtle. The table includes information on gene ID and gene description. It also includes the following information for each sex: log₂ fold change between stages (logFC), average expression (AveExpr), t statistic value (t), *P* - value (P.value), and adjusted *P* – value (adj.P.val). Information for each sex is indicated with .x for data from the female samples and .y for data from the male samples

Supplementary table 5. Birds DEGs between stages E.4.5 and E.6.0 in males and females of the chicken. The table includes information on gene ID and gene description. It also includes the following information for each sex: log₂ fold change between stages (logFC), average expression (AveExpr), t statistic value (t), *P* - value (P.value), and adjusted *P* – value (adj.P.val). Information for each sex is indicated with .x for data from the female samples and .y for data from the male samples

Supplementary table 6. Mammals DEGs between stages E.11.0 and E.12.0 in males and females of the mouse. The table includes information on gene ID and gene description. It also includes the following information for each sex: log₂ fold change between stages (logFC), average expression (AveExpr), t statistic value (t), *P* - value (P.value), and adjusted *P* – value (adj.P.val). Information for each sex is indicated with .x for data from the female samples and .y for data from the male samples

Acknowledgements

We thank Dr. Susanne Kneitz for her help in obtaining the data and advice on bioinformatic analysis of the RNA sequencing data. We acknowledge Dr. Tapas Chakraborty for helpful discussion.

Authors contributions

NSB and FP conceived and designed the study. AR produced the platy fish data. NSB, IJS and AR compiled the data. NSB and IJS analyzed the data. NSB and FP wrote the paper. NSB, FP and MS edited the paper. All authors approved the final version of the paper.

Funding

NS was supported by a Spanish Ministry of Science (SMS) predoctoral scholarship (BES-2017–079744). Research was supported SMS grant no. PID2019-108888RB-I00 to FP. Also with funding from the Spanish government through the ‘Severo Ochoa Centre of Excellence’ accreditation (CEX2019-000928-S).

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories. The datasets analyzed are available at the GEO database with the accession numbers GSE11584, GSE105103, GSE41948 and the Sequence Read Archive (SRA) database codes SRP079664, SRP014719 and PRJNA944639.

Competing interests

The authors declare no competing interests

Block 2:

***Epigenetic biomarkers and the effect of
sex and temperature***

Chapter 3

*Development of epigenetic biomarkers with
diagnostic and prognostic value to assess the
lasting effects of early temperature changes in
farmed fish*

Aquaculture (2023)

563:738918

doi.org/10.1016/j.aquaculture.2022.738918

Development of epigenetic biomarkers with diagnostic and prognostic value to assess the lasting effects of early temperature changes in farmed fish

Alejandro Valdivieso¹, Núria Sánchez-Baizán¹, Nikos Mitrizakis², Nikos Papandroulakis², Francesc Piferrer¹

1. Institut de Ciències del Mar (ICM), Spanish National Research Council (CSIC), Barcelona, Spain

2. Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Center of Marine Research (HCMR), Heraklion, Greece

Abstract

Epigenetic changes such as DNA methylation modifications at specific loci in the genome in response to environmental cues can appear long before the appearance of transcriptomic or other phenotypic changes with potential consequences for performance. Thus, epigenetic DNA methylation-based biomarkers hold great promise in aquaculture because they can have prognostic value. However, to the best of our knowledge, a clear demonstration of such biomarkers has not been provided yet. In fish farming, temperature during early stages of development is the main abiotic factor determining larval quality, with effects potentially persisting until adulthood and influencing performance. The objective of this study was to test whether epigenetic biomarkers could be developed as reliable indicators of the quality of the thermal environment during European sea bass (*Dicentrarchus labrax*) early development and their relation to long-lasting phenotypic consequences. We exposed European sea bass embryos and newly hatched larvae to different temperature regimes and sampled fish thereafter at three time points: at larval, juvenile and adult stages. Here, we provide an analytical strategy combining Reduced Representation Bisulfite sequencing (RRBS) and RNA-sequencing (RNA-seq) to develop epigenetic and transcriptomic biomarkers with both prognostic and diagnostic value. Specifically, we report a series of DNA-methylation biomarkers linked to short-term, mid-term, long-term, and very-long term effects of temperature in both somatic and reproductive tissues. Among the different biomarkers identified, and a promising example, the methylation status of keratin associated protein 10-4 (*krtap10-4*) fulfills the criteria of an epigenetic biomarker with both prognostic and diagnostic value. This study constitutes a first step toward providing the European sea bass farming sector with a set of epigenetic biomarkers that can be used to screen batches of larvae to determine the quality of their thermal environment and to anticipate their performance when juveniles and adults. Our study provides the way for similar developments for other aquaculture farmed species.

Keywords: DNA methylation-based epigenetic biomarkers; gene expression; thermal stress; key performance indicators; European sea bass aquaculture

1. Introduction

Larval rearing is one of the most crucial and delicate procedures during modern, intensive fish farming production and, thus, many factors, both biotic and abiotic, can influence its success (Conceicao and Tandler, 2018). Temperature affects many biological functions at different levels of organization (Schulte *et al.*, 2011), being considered also the main abiotic factor determining fish larval quality and survival (Pepin, 1991). The early stages of development are particularly the most sensitive to temperature (Burggren and Mueller, 2015), while these effects can persist until adulthood influencing growth performance and product quality (Jonsson and Jonsson, 2014).

The European sea bass, *Dicentrarchus labrax*, is, after the Atlantic salmon, *Salmo salar*, and rainbow trout, *Onchorhynchus mykiss*, the third most important species for finfish aquaculture in Europe in terms of volume, and the most important in the Mediterranean region, with over 208,000 tons produced in 2019 (FEAP, 2020, <https://feap.info/index.php/data/>). The European sea bass is a eurythermal marine teleost (Pickett and Pawson, 1994). Its thermal tolerance in juveniles acclimated to 20°C ranges from 4.7°C to 34.8°C (Dülger *et al.*, 2012). In the Mediterranean Sea, spawning typically takes place during winter. In hatcheries eggs are incubated and hatch at around 13–14°C, and larval rearing takes place usually between 15–21°C (Morretti, 1999). Temperature has a profound effect on the European sea bass ontogenetic plasticity (Koumoundouros *et al.*, 2001). It is a gonochoristic species with sexual size dimorphism in favor of females and a polygenic system of sex determination (Vandeputte *et al.*, 2007), where genetics and temperature contribute equally to sex ratios (Piferrer *et al.*, 2005). Thus, proper temperature control during their embryonic and first larval stages of development is essential to minimize skeletal deformities (Georgakopoulou *et al.*, 2007), prevent precocious male maturation (Felip *et al.*, 2006), maximize growth, optimize sex ratios (Navarro-Martín *et al.*, 2009b), and to increase overall product quality (López-Albors *et al.*, 2008).

Epigenetics concerns the study of phenomena and mechanisms that cause chromosome-bound, heritable changes to gene expression that are not dependent on changes to the DNA sequence (Deans and Maggert, 2015). Epigenetic mechanisms, namely, DNA methylation, histone modifications and non-coding RNAs, regulate gene expression by integrating genomic and environmental information, thus contributing to phenotypic plasticity (Duncan *et al.*, 2014). Epigenetics holds great promise in aquaculture (Gavery and Roberts, 2017; Granada *et al.*, 2018; Piferrer *et al.*, 2015). Epigenetic alterations in response to environmental stimuli, for example, changes in DNA methylation levels in specific loci in the genome that correlate to a certain phenotype, can be used as epigenetic biomarkers. Of note, epigenetic biomarkers not only can have diagnostic but also prognostic value, since sometimes epigenetic changes appear long before gene expression or other phenotypic changes become apparent (Costa-Pinheiro *et al.*, 2015). In the European sea bass, DNA methylation-based biomarkers have already been developed to predict sex (Anastasiadi *et al.*, 2018c), assess the early stages of domestication (Anastasiadi and Piferrer, 2019), determine biological age (Anastasiadi and Piferrer, 2020), assess the consequences of heat waves (Anastasiadi *et al.*, 2021a), and to evaluate the stress response associated with rearing density (Krick *et al.*, 2021). However, and to the best of our knowledge, there are no epigenetic biomarkers in the European sea bass or, for that matter, in any other farmed fish species, that can be used to link the temperature regime during early stages of development with subsequent aquaculture performance. Developing epigenetic biomarkers that link adult phenotypic traits to early environmental conditions is a major goal of epigenetics in aquaculture.

The objective of this study was to assess the effects of temperature changes during embryo and early larval development on the establishment of epigenetic biomarkers that could correlate with phenotypic traits later in development in both somatic and reproductive tissues. Specifically, we tested whether these epigenetic biomarkers could be developed as reliable indicators of the quality of the thermal environment during the most critical stages where temperature may have long-lasting phenotypic consequences. This constitutes a novel approach and a first step towards the ultimate goal of providing the aquaculture sector with a set of epigenetic biomarkers with diagnostic and prognostic value.

2. Materials and methods

2.1. Experimental facilities

Fish were reared at the Hellenic Center of Marine Research (HCMR) in Crete, Greece. The HCMR aquaculture facilities are certified from the national veterinary authority (code GR94FISH0001) and are licensed for operations of breeding and experimentation with fish issued by Region of Crete, General Directorate of Agricultural & Veterinary no. 3989/01.03.2017 (approval codes EL91-BIObr-03 and EL91-BIOexp-04). The experimental protocol was approved by the Veterinarian Authority of the Region of Crete with the 255332/29-11-2017 document. Animal experiments were carried out in accordance with the EU Directive 2010/63/EU.

2.2. Egg incubation and temperature treatments

All eggs used in this study came from the same mass spawn of captive breeders. Approximately 240,000 eggs were distributed into six 500-liter tanks at about 60,000 eggs per tank. Eggs were incubated at 14°C. For the rearing recirculation systems, biological filters were used. Water temperature was adjusted using chillers and heaters, which were automatically controlled with electronic sensors. Hatching and larval rearing were performed by using the “pseudogreen water” methodology (Papandroulakis *et al.*, 2001). With a hatching rate close to 95% the initial number of larvae was approximately 54,000 per tank.

Three groups of fish were subjected in duplicate to 14°C, 17°C, and 20°C (two of the six tanks mentioned above randomly assigned to each group) from the stage of epiboly onset to the end of the yolk-sac larval stage and the beginning of exogenous feeding (**Figure S1A**). In all cases, the experimental temperature was set 1 day post fertilization (dpf) and lasted until 8 dpf in the case of 17°C and 20°C, while for the 14°C the duration was until 12 dpf (different durations compensated for the effect different temperatures on growth). In this study, therefore, and following the ARRIVE guidelines (Percie du Sert *et al.*, 2020), there were three experimental groups each carried out in replicate (being the tanks the biological replicates), and the group exposed to 17°C was the control group. Subsequently, all groups were kept under identical rearing conditions including temperature (17°C) up to the end of the trial. Acclimation of the eggs from the spawning temperature (14°C) to the test temperatures, and then to the common temperature for larval rearing (17°C) was performed at the rate of 0.5°C and 0.2°C h⁻¹, respectively. Fish were weighed periodically at the times indicated below.

2.3. Larval and juvenile rearing

Larvae were successively fed on enriched (DHA Protein Selco, INVE) rotifers until they reached 5.5 mm total length (TL), enriched with rotifers and *Artemia* Instar 1 nauplii (AF, INVE) when they were between 6 and 7.5 mm TL and enriched with *Artemia nauplii* (Easy DHA Selco, INVE) afterwards. Rotifer concentration in the tanks was adjusted three times daily to 6–8 individuals·ml⁻¹. The concentration of *A. nauplii* was adjusted three times daily to 0.5–2 individuals·ml⁻¹. During the rotifer-feeding period, live microalgae *Chlorella sp.* were added three times daily at a concentration of $3 \pm 1 \times 10^5$ cells ml⁻¹. Weaning to inert diets (O-range, INVE) started when fish were approximately 9.5 mm TL. In **Table S1** the feeding regime per experimental condition is shown.

Air-blowing skimmers were installed since first feeding until 29 dpf to keep the water surface free from lipids and enhance swim bladder inflation. Water oxygen concentration was 7.68 ± 0.9 , pH 8.37 ± 0.12 , salinity 35–36‰ and photoperiod 12L:12D. After larval rearing (65 dpf) the survival was estimated and fish were transferred for pre-growing in 10 m³ tanks, supplied with seawater from a deep well and operating in an open water system. The survival rate (mean $\pm$ SD) at the end of larval rearing was $19.6 \pm 5.1\%$ for the control group exposed to 17°C (16.0% and 23.2% for the two replicates); $18.5 \pm 2.4\%$ for the group exposed to 14°C (16.8% and 20.2% for the two replicates) and $17.2 \pm 3.9\%$ for the group exposed to 20°C (14.5% and 20.0% for the two replicates). Water oxygen saturation was 7.0 ± 0.9 , pH 7.6 ± 0.3 , salinity 35–36‰, and photoperiod natural. During pre-growing (65–130 dpf) the larvae were grouped per temperature treatment resulting in one group per experimental temperature. Following the standard husbandry procedures, the juveniles were sorted by size, removing the very small and the very large individuals in order to create a more homogenous size distribution among the three experimental groups and were also selected for swim bladder with a salinity test.

2.4 Coping style test

From each temperature treatment, three subgroups were randomly created when fish had approximately 130 dpf. These groups were transferred to cages for on-growing after being subjected to a conditioned place preference test (CPP) to characterize fish coping style. Briefly, the methodology consisted in: (i) restraining the individuals in a 1 m³ containment, placed within the holding tank, for 20 hours, (ii) allowing the individuals to pass through three open holes (10 cm diameter), constructed close to the water surface while feed was distributed outside the containment by the use of an automatic feeder, and (iii) recording the number of individuals passed through the holes, during a time course of two hours. At the end of the test, six groups were made according to the temperature treatment and the coping style expressed, mentioned as “*Bold*” for the proactive individuals, i.e. the ones exiting the containment and “*Shy*” for the reactive individuals i.e. the ones remaining in the containment.

2.5. Rearing in sea water

The six original groups (from 2,000 to 3,000 individuals each) were then transported for further on-growing in net-open cages. At this time fish were approximately 140 dpf and had a mean weight of 1.72 ± 0.40 g. On-growing lasted for nine months and was performed in the pilot scale cage facility of the Institute of Aquaculture (HCMR) located at the North-West of Crete in Souda Bay (35°28'48.3"N 24°06'41.8"E). The mean water depth at the site is about 60 m (ranging from 30–90 m) with a continuous water circulation. Water temperature varied from 15°C to 28°C, and

salinity was constant all year round at about 39 ‰. Water temperature and dissolved oxygen concentration were monitored daily using digital probes (HQ 40d; Hach SA, Loveland, CO, USA).

Rearing was performed in polyethylene cages (rectangular and/or circular) and standard rearing methodologies were applied (Pavlidis and Mylonas, 2011). The cage size at the beginning of the rearing was approximately 35 m³ (3 m high x 3 m wide x 4 m depth) and increased to approximately 50 m³ (3 x 3 x 6 m) when fish reached a mean weight of about 50 g. Rearing conditions were the standard followed for the European seabass with daily feeding using automatic feeders with standard commercial diets based on feeding tables available/adapted in the farm.

2.6. Samplings

The tissues selected for this study were sampled at three time points. After each sampling, tissues were kept in RNA-later solution (ThermoFisher Scientific) and stored at -80°C until further analysis. Whole larvae at the first feeding stage (8 or 12 dpf depending on the temperature group) were sampled and four pools of ten larvae per temperature group were used for RNA-sequencing. Another four different pools of ten larvae each were made for DNA methylation analysis (see details below). At the past metamorphosis stage (90 dpf) juvenile fish were taken, and two different pools with four fish each were used for RNA-seq and DNA methylation analysis, respectively. Adult fish were sampled at 482 dpf, and four individual fish per each temperature treatment were selected for RNA-seq and DNA methylation analysis. The testis was used for two reasons: first, because we had examined the very long-term effects of temperature on this organ in a previous study (Anastasiadi *et al.*, 2021a), and thus could compare results; and second because epigenetic information can be transmitted through the sires in the European sea bass (Anastasiadi *et al.*, 2018c). To avoid possible size-dependent differences that could confound the results, we selected larvae samples at the first feeding stage of the same size. Similarly, juvenile samples were weighted to ensure that all samples contributed equally to the pool. In addition, during the adult fish sampling at 482 dpf, the total weight was determined in 20 fish (ten males and ten females) randomly taken in each temperature treatment. Phenotypic sex was identified in the same fish by visual inspection, and gonads were weighted. Gonadosomatic index (GSI) was calculated as gonad weight (g) / total fish weight (g) × 100.

2.7. RNA and DNA extraction

Tissues for transcriptomic analysis were removed from the RNeasy[®] solution, samples were immersed into TRIzol[®] Reagent (ThermoFisher Scientific), and homogenized manually. RNA was extracted from larvae, juveniles and adults according to the manufacturer's instructions. RNA quality was evaluated with Agilent RNA 6000 Nano Kit (Agilent), and samples with RNA Integrity Number (RIN) > 8 were used for RNA-seq library preparation. Genomic DNA was extracted using the Blood and Cell Culture DNA Mini kit (13323, Qiagen) according to the manufacturer's instructions. The quality of the DNA was checked with 1000NanoDrop (Thermo Fisher Scientific), and concentration was measured by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

2.8. Sea bass reference genome

The European sea bass reference genome (GCA_000689215.1) (Tine *et al.*, 2014) was used for this study. For methylation analysis, the reference genome was bisulfite converted using Bismarck software (v. 0.22.3) applying the `bismark_genome_preparation` function (Krueger and Andrews, 2011). The start and end positions for each gene (gene body region) in the genome were taken from Biomart (Durinck *et al.*, 2009, 2005).

2.9. RNA-sequencing (RNA-seq) and differential gene expression analysis

The RNA libraries of larvae, juveniles, and adults were prepared using the mRNA-Seq preparation kit (Illumina) according to the manufacturer's protocol at the Centre for Genomic Regulation (CRG) service within the Barcelona Biomedical Research Park (PRBB). Briefly, 0.5 µg of RNA were used for poly-A-based mRNA enrichment selection using oligo-dT magnetic beads followed by fragmentation of divalent cations at elevated temperature resulting into fragments of 80–250 base pairs (bp), with the major peak at 130 bp. First-strand cDNA synthesis by random hexamers and reverse transcriptase was followed by the second-strand cDNA synthesis. Double-stranded cDNA was end-repaired, 3'-adenylated, and the 3'-'T' nucleotide with the Illumina adapter was used to ligate the indexed adapters. The ligation product was amplified using 12 PCR cycles. Each library was sequenced using 50 bp paired-end mode on an Illumina HiSeq 2000 instrument following the manufacturer's protocol. The quality of raw reads was checked using FastQC software (v. 0.11.9) (Wingett and Andrews, 2018). Adapters and sequences with a Phred quality score lower than 30 were trimmed with the *Bbduk* software (v. 38.90) (Bushnell, 2014). The HISAT2 software (v. 2.1.0) (Kim *et al.*, 2019) was used first to index the reference genome using `hisat2-index build` command and then to align the trimmed reads with default parameters. Finally, reads were counted with *featureCounts* software (Y. Liao *et al.*, 2014) (v. 2.0.1) at the gene level. Detailed information of the samples for RNA-seq analysis in each tissue, treatment and control quality statistics are listed in **Table S2**.

Differential gene expression (DEGs) analysis was conducted using the *edgeR* package (v. 3.28.1) (McCarthy *et al.*, 2012; Robinson *et al.*, 2010). The lowly expressed genes were filtered using the “`filterByExpr`” function (Y. Chen *et al.*, 2016) among all samples. Retained counts were normalized using the upper-quartile “`calcNormFactors`” function (Anders and Huber, 2010). Estimates for negative binomial (NB) dispersion parameters were generated using “`estimateDisp`” function (Chen *et al.*, 2014; Phipson *et al.*, 2016). A quasi-likelihood binomial generalized log-linear model was fitted using “`glmQLFit`” (Lun and Smyth, 2016; Lund *et al.*, 2012) and the “`glmQLTest`” function was used to compare the different contrasts within the matrix design (Lund *et al.*, 2012; Chen *et al.*, 2016; Lun *et al.*, 2016; Phipson *et al.*, 2016; Lun and Smyth 2017). Finally, a false discovery rate correction (FDR or Q-value, < 0.05) was applied in all contrasts (Benjamini and Hochberg, 1995).

2.10. Reduced Representation Bisulfite Sequencing (RRBS)

The RRBS libraries of larvae, juveniles, and adults were prepared using the Premium RRBS kit (Diagenode, C02030032) according to the manufacturer's instructions. Qubit DNA High Sensitivity assays (ThermoFisher Scientific) were used to quantify the final concentrations of the libraries and their quality was checked using the DNA High Sensitivity Assay (Agilent). In total, 36 libraries (*i.e.*, four biological replicates x three stages x three temperatures) were grouped in four pools of nine samples. The four pools of libraries were sequenced on an Illumina HiSeq 2500 instrument using single-end mode and a read length of 50 bp.

Quality control of reads were performed using the FastQC software (v. 0.11.9) (Wingett and Andrews, 2018). Trimmomatic software (v.0.36) (Bolger *et al.*, 2014) was used for trimming the adaptors with parameters: sliding window 4, Phred score > 20, and adaptive quality trimming with largest length of 20 bp. Trimmed reads were aligned to the reference genome using the BSMAP software (v. 2.90) (Xi and Li, 2009) with RRBS mode and filtering for at least 10X coverage. In order to extract methylation data only from CG context (CpG) we used the methratio.py function from BSMAP. The bisulfite conversion efficiency of the libraries was checked following the instructions from the Premium RRBS kit.

The *methyKit* package (v. 1.14.2) (Akalın *et al.*, 2012) was used to calculate differentially methylated cytosines (DMCs). To do so, coverage was normalized across samples for each tissue and only CpGs covered in all samples from the same tissue were maintained for further analysis. To detect DMCs a logistic regression model was applied with the F-test, which identifies methylation differences with the following chosen criteria: methylation differences larger than 15% and Q-value < 0.01 adjusted from the SLIM method (Wang *et al.*, 2011). A detailed information of the samples for the methylation analysis in each tissue and treatment and control quality statistics are listed in **Table S3**. Finally, the *GenomicRanges* package (v. 1.38) (Lawrence *et al.*, 2013) was used to identify those genes that contained at least one DMC within their gene body region ± 1000 bp.

2.11. Gene Ontology terms and Kyoto Encyclopedia of Genes and Genomes pathways analysis

The Gene Ontology terms (GO-terms) and the Kyoto Encyclopedia of Genes and Genomes pathways (KEEGs) analyses were carried using the *TopGO* package (v. 2.46.0) (Alexa and Rahnenfuhrer, 2010) adapted for the European sea bass GO-terms database. Only the Biological Process (BP) category was considered relevant for this study. The several GO-terms carried out were simplified using the R code script from Revigo software (Supek *et al.*, 2011) with settings: medium (0.7) and SimRel semantic similarity method (Schlicker *et al.*, 2006).

2.12. Data analysis

Sex ratio analysis and behavior tests were analyzed within each temperature treatment using Chi-squared test. A two-way ANOVA with Tukey post hoc test was applied to weight and GSI data. For methylomics and transcriptomics, the middle temperature, 17°C, was considered the reference or control temperature. Therefore, two different comparisons at each one of the three sampling points were carried along in this study: the effects of cold temperature (*i.e.*, 14°C vs. 17°C) and the effects of elevated temperature (*i.e.*, 20°C vs. 17°C). In all analysis significant differences were considered when $P < 0.05$. Statistical analysis was performed with R software (v. 4.1.2) (R Core Team 2020) using in Rstudio (v. 2021.09.0) (RStudio Team 2020). Graphs were created using the *ggplot2* package (v. 3.3.5) (Wickham, 2021).

Samples for transcriptomic and DNA methylation analysis were selected based on the integrity of RNA (> 8 RIN level), the bisulfite conversion efficiency (> 99.0%), and the aggrupation of the samples in a Principal Component Analyses (PCA) to detect the outliers for the same group. Those samples which did not meet these criteria were discarded. In total, 9 and 6 out of a total of 36 and 36 samples were discarded for RRBS and RNA-seq analysis, respectively. The number of samples used for each comparison by each tissue is presented in **Table S4**.

3. Results

3.1. Growth, gonadal development, sex ratio and coping style

At 95 dpf, the weight (mean $\pm$ SEM) was 1.08 ± 0.26 g for the 17°C-reared fish; 1.07 ± 0.28 g for the 14°C-reared fish; and 1.07 ± 0.26 g for the 20°C-reared fish, without statistically significant differences ($P > 0.05$). At 482 dpf, females tended to have a higher mean weight than males, but no significant sexual dimorphism was observed in any of the three temperatures tested. Although fish exposed to 14°C were bigger than those exposed at 20°C, (Two-way ANOVA: $F = 4.37$, $P = 0.017$) there were no statistical differences between fish exposed at 14°C vs. those exposed at 17°C and the same applied between fish exposed to 20°C vs. those exposed to 17°C (**Figure S1B**). Analysis of the GSI showed that females exposed to 20°C had significantly higher mean values ($P < 0.05$) than females exposed at 14°C, but there were no differences in the GSI of males (**Figure S1C**). Sex ratios progressively increased in favor of males with temperature, without significant differences (**Figure S1D**). Regarding coping style, the number of bold fish increased in the group exposed at 20°C (**Figure S1E**).

3.2. Overview of DNA methylation and gene expression changes as a function of temperature and tissue

We carried out transcriptomic and DNA methylation analyses to identify DEGs and DMCs in larvae, juveniles and adults in two different comparisons: 14°C vs. 17°C, and 20°C vs. 17°C temperatures. Regarding transcriptomic results for larvae, juveniles, and adults, we obtained 8,490, 58 and 28 DEGs between 14°C vs. 17°C (**Table S5**, **S6** and **S7**, respectively), and 8,128, 42 and 53 DEGs between 20°C vs. 17°C (**Table S8**, **S9**, and **S10**, respectively). On the other hand, regarding DNA methylation, the number of DMCs were 27,286 and 22,166 in larvae (**Table S11** and **S12**), 9,884 and 15,886 in juveniles (**Table S13** and **S14**) and 7,254 and 49,610 in adults (**Table S15** and **S16**) from the 14°C vs. 17°C and 20°C vs. 17°C comparisons, respectively. Finally, for each tissue and contrast, we identified 7,231 genes which had at least one DMC (**Table S17**).

Next, taking all gene lists together we identified 163 genes in the 14°C vs. 17°C comparison and 264 genes in the 20°C vs. 17°C comparison with DMCs that were common to larvae, juveniles, and adults (**Figure S2A** and **S2B**). Furthermore, by comparing these 2 lists we identified 70 genes in common regardless of the direction of temperature change (**Table S18**). These are then genes, *e.g.*, amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase (*agl*), the methylation of which is affected in larvae and maintained in both juveniles and adults and represent indicators of long-term epigenetic (but not transcriptomic) effects of temperature on sea bass tissues. This list also includes jumonji domain containing 7 (*jmjd7*) and potassium voltage-gated channel subfamily H member 5 (*kcnh5b*), two genes that are related to environmental signal transduction, and zinc finger FYVE-type containing 9 (*zfvyve9a*), which participates in the TGF- β signaling pathway.

Finally, we were interested to know if there were genes that were considered both DEG and also contained at least one DMC in their body region. Results showed that 1,350 (**Table S19**) and 723 (**Table S20**) genes in larvae, 6 (**Table S21**) and 7 genes (**Table S22**) in juveniles, and the gene potassium voltage-gated channel subfamily H member 3 (*kcnh3*) and two genes (ENSDLAG00005004413 and ENSDLA00005021715) in adults were shared in the 14°C vs. 17°C and 20°C vs. 17°C comparisons, respectively (**Figure S2A** and **S2B**). These are then genes, *e.g.*, potassium voltage-gated channel subfamily H member 2 and 3 (*kcnh2a* and *kcnh3*),

proprotein convertase subtilisin/kexin type 7 (*pcsk7*), protein kinase D4 (*prkd4*), in which temperature affected both their DNA methylation and expression and can be considered biomarkers of the persistent epigenetic and transcriptomic effects of temperature with potential functional consequences.

3.3. Genes whose expression is consistently up- or downregulated by temperature

In larvae, we wanted to know how many genes were consistently upregulated or downregulated by the temperature effects. There were 4,839 DEGs that were common in the list of the 8,490 DEGs between 14°C vs. 17°C and in the list of 8,128 DEGs between 17°C vs. 20°C comparisons. From these 4,839 genes, we filtered those that their transcripts levels were consistently upregulated (*i.e.*, 153 genes) and downregulated (*i.e.*, 58 genes) with the effects of temperature (**Table S23**). The raw transcript values from these 153 and 58 genes were normalized and plotted (**Figure 1A and 1B**). These 211 genes (**Table S23**) included signal transducing adaptor family member 2 (*stap2b*), mannose receptor C-type 1 (*mrc1a*), and DAB adaptor protein 2 (*dab2*). Then, we searched for the biological role of these 153 and 58 list of genes separately and results showed 44 and 33 associated GO-terms, respectively (**Figure 1C and 1D**). The most relevant GO-terms related with genes upregulated by temperature were: response to oxidative stress (GO:0006979), negative regulation of glucagon secretion (GO:0070093) and response to peptide hormone

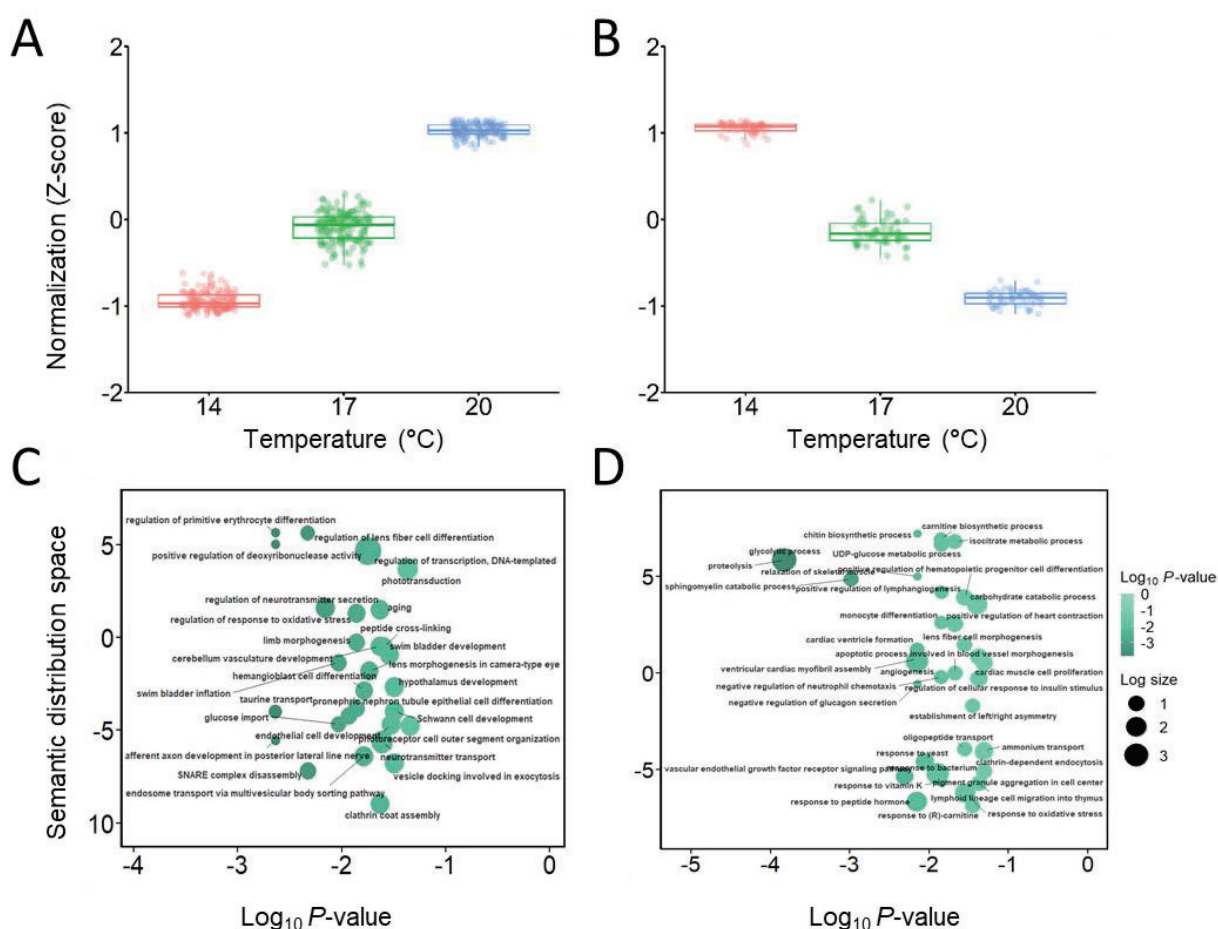


Figure 1. (A-B) Box plots of the normalized (Z-score) gene expression of common differentially expressed genes (DEGs) between 14°C vs. 17°C ($n = 153$) and 20°C vs. 17°C ($n = 58$) that were consistently (A) upregulated (Kruskal-Wallis: chi-squared = 407.11; P-value < 0.01) or (B) downregulated (Kruskal-Wallis: chi-squared = 153.78; P-value < 0.01) along with the effects of temperature. (C-D) Semantic similarity

scatter plots summarizing the enriched Gene Ontology (GO) terms of the Biological Process category (BP) from the (C) 153 DEGs into 44 GO-terms and from the (D) 58 common DEGs into 33 GOs for the genes that consistently increase or decrease expression with temperature, respectively. The scatterplots show GO-terms as circles. Significance was shown with the green gradient color palette ($\log_{10} P$ -value), and the sizes of the plotted circles were scaled by the frequency of GO-terms they represent. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

(GO:0043434). On the other hand, the most relevant GO-terms related with genes downregulated by temperature were: glucose import (GO:0046323), aging (GO:0007568) and photo-transduction (GO:0007602).

3.4. DEGs in larvae that contain at least one DMC regardless of temperature change

Next, we searched in larvae those genes that in addition of being considered DEGs also contained at least one DMC and appeared in both temperature comparisons (*i.e.*, common genes between the 1,350 genes at 14°C vs. 17°C and the 723 genes at 20°C vs. 17°C). We found 210 genes fulfilling this criterion (**Table S24**). To understand their relationship, these genes were filtered and their raw transcript counts data were normalized per individual sample and represented in a heatmap (**Figure 2A**). The heatmap showed that samples clearly grouped according to temperature treatment with a close relationship between the 14°C and 17°C samples compared to samples from 20°C treatment. Thus, of the 210 genes identified, 95 were clearly upregulated at 20°C while they were mostly downregulated at 14°C and 17°C treatments. In contrast, 115 genes had just the opposite pattern, *i.e.*, downregulated at 20°C and upregulated at 14°C and 17°C (**Figure 2A**). These 95 and 115 genes include: major intrinsic protein of lens fiber (*mipb*), ATPase sarcoplasmic/endoplasmic reticulum Ca²⁺ transporting 1 (*atp2a1*), glutathione S-transferase omega 2 (*gstb2*), and calcium binding protein 2 (*cabp2a*).

We also searched for the biological role of these 210 genes to better understand the effects of temperature. Results showed that these genes were participating in 30 GO-terms, of which the most relevant were: DNA methylation on cytosine within a CG sequence (GO:0010424), positive regulation of Wnt signaling pathway involved in dorsal/ventral axis specification (GO:2000055) and cellular response to cold (GO:0070417) (**Figure 2B**).

3.5. Relationship between DNA methylation-gene expression in larvae

The 95 and 115 genes identified above in larvae are genes that not only their expression is affected by temperature but also their methylation levels are affected. However, whether those two changes are linked is not known. Thus, next we were interested to find out those genes whose expression was affected by the temperature treatment as well as their DNA methylation levels. To do so, we started with the list of 1,350 and 723 genes that were DEGs and contained at least one DMC in the 14°C vs. 17°C and the 20°C vs. 17°C comparisons, respectively (**Figure S2A** and **S2B**). For each selected gene, we calculated the average number of transcripts with logarithm transformation and also their mean methylation value by averaging the methylation values of all their DMCs among the samples belonging to the same temperature treatment. Next, for each temperature comparison we identified those genes whose expression was either significantly positively or negatively correlated with DNA methylation levels. Results showed that there were 577 (**Table S25**) and 344 (**Table S26**) genes with negative and positive correlation, respectively, in the 14°C vs. 17°C comparison (**Figure 3A** and **3C**), while there were 102 (**Table S27**) and 250 genes (**Table S28**) with negative and positive correlation in the 20°C vs. 17°C comparison (**Figure 3B** and **3D**). There were no common genes between the 577 and 102 gene lists, whereas

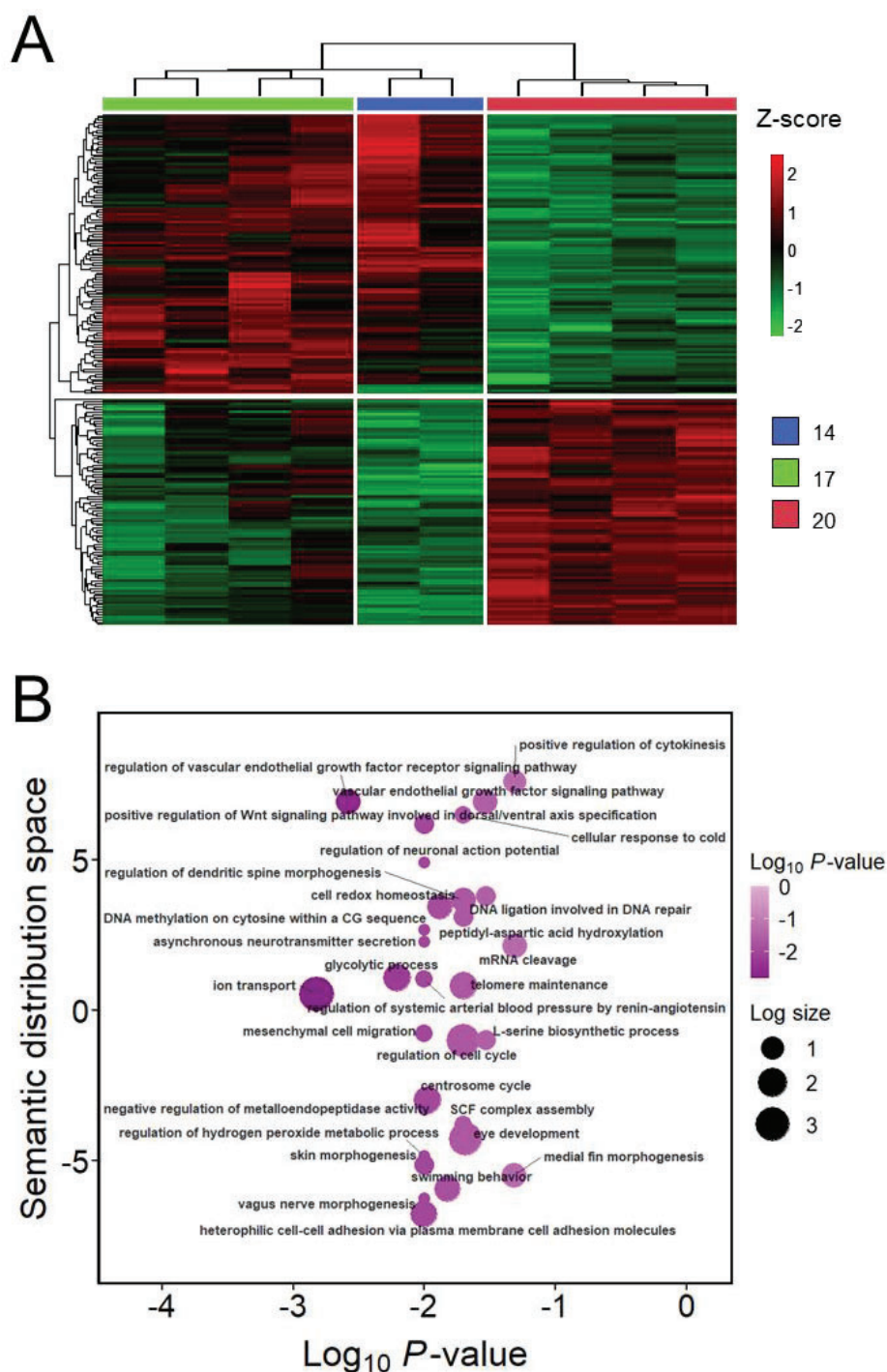


Figure 2. (A) Heatmap of expression levels in larvae tissue from 210 common differentially expressed genes (DEGs) which at least contain one differentially methylated cytosine (DMC) between 14°C vs. 17°C ($n = 1350$) and 20°C vs. 17°C ($n = 723$). The color gradient represents the level of expression scaled by gene (red: high expression and green: low expression). The dendrograms indicate the similarity between genes and between samples. The bands on the top indicate the temperature treatment of each sample (blue: 14°C, green: 17°C, and red: 20°C). (B) Se-mantic similarity scatter plot summarized the enriched GO-terms of the BP category from the 210 common DEGs in both temperature treatments with at least one DMC. The scatterplot showed GO-terms as circles and the significance was shown with the purple gradient color palette ($\log_{10} P$ -value). The sizes of the plotted circles were scaled by the frequency of GO-terms they represent. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

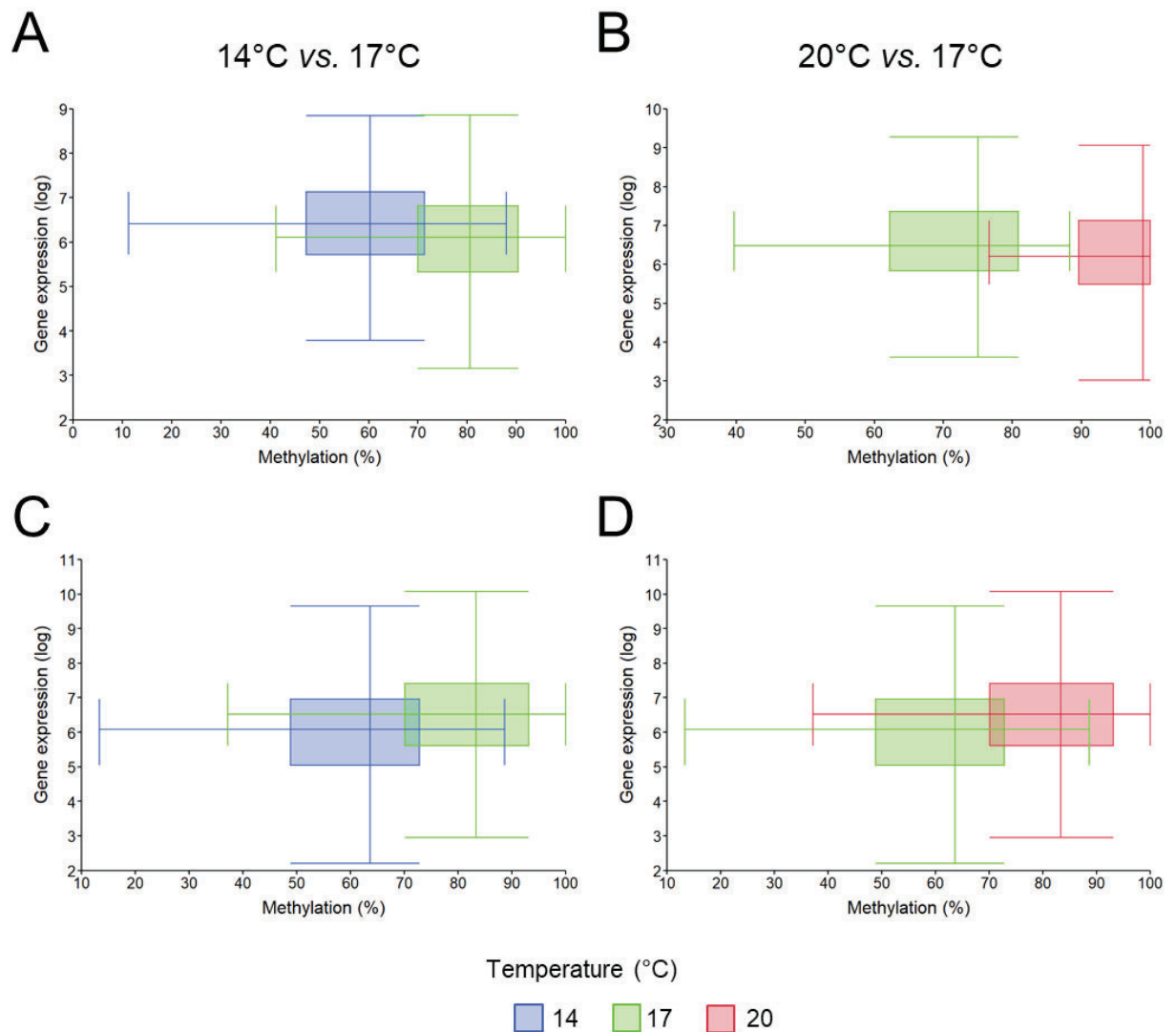


Figure 3. Boxplots of the common differentially expressed genes (DEGs) between 14°C vs. 17°C and 20°C vs. 17°C which at least have one differentially methylated cytosine (DMC) in the gene body region (gene ± 1000 base pairs, bp), and that on top showed DNA methylation negatively correlated with gene expression in (A) 14°C vs. 17°C, and in the (B) 20°C vs. 17°C comparison. The genes that met the same criteria with a positive correlation between DNA methylation and gene expression were plotted in (C) for the 14°C vs. 17°C comparison, and in (D) for the genes from the 20°C vs. 17°C comparison.

only ENSDLAG00005009958 was common in the 344 and 250 gene list that were more methylated and upregulated by temperature.

3.6. Persistent temperature effects on the transcriptome of larvae, juveniles, and adults

Once epigenomic and transcriptomic changes resulting from temperature effects had been well characterized in larvae, we were interested to know how persistent those changes were in other tissues. Specifically, we wanted to know which of the DEGs found in larvae were still affected in juveniles and in reproductive (testis) tissues long after the temperature treatments had ceased during larvae development. For this, we compared the DEGs lists from larvae, juveniles, and adults for each temperature comparison. Results showed that not a single gene was shared among three tissues in both comparisons. However, in the 14°C vs. 17°C comparison, we identified 29 and 6 genes between larvae-juveniles and larvae-adults, respectively (**Figure S2A** and **S2B**). From these 29 DEGs that were common in larvae and juveniles, 16 genes, including chromobox

4 (*cbx4*), showed the same direction of change, either consistently upregulated or consistently downregulated (**Figure S3A**). The same applied to 3 out of the 6 DEGs found in common between larvae and adults (**Figure S3B**). In the 20°C vs. 17°C comparison, 21 DEGs in larvae-juveniles and 14 DEGs in larvae-adults were also identified. From these two gene lists, only 5 and 6 genes showed the same gene expression tendency, respectively (**Figure S3C and D**).

We searched the GO-terms to better understand their biological function due to low temperature (**Figure 4**). From the 29 and 6 common DEGs genes in larvae-juveniles and larvae-adults from 14°C vs. 17°C comparison, we found 10 and 3 GO-terms, respectively (**Figure 4A and 4B**). The most relevant GO-terms from larvae-juveniles were: response to hypoxia (GO:0001666), cell redox homeostasis (GO:0045454) and positive regulation of CREB transcription factor activity (GO:0032793), while from larvae-adults were: integrin-mediated signaling pathway (GO:0007229), cell-matrix adhesion (GO:0007160) and small GTPase mediated signal transduction (GO:0007264). From common genes in larvae-juveniles and larvae-adults in the 20°C vs. 17°C comparison, we identified 21 and 14 genes, respectively (**Figure S2C and S2D**). We also searched the GO-terms to better understand their biological function due to the effects of elevated temperature (**Figure 4**). We found 12 and 3 GO-terms from larvae-juveniles and larvae-adults, respectively (**Figure 4C and 4D**). The most relevant GO-terms from larvae-juveniles were: cellular response to oxidative stress (GO:0034599), juveniles cell cellular homeostasis (GO:0046716), and cardiac juveniles tissue morphogenesis (GO:0055008), while from larvae-adults was: oocyte maturation (GO:0001556).

Additionally, we investigated in which pathways these set of GO-terms were participating. For juveniles-larvae we found metabolic and MAPK signaling pathways in 14°C vs. 17°C and 20°C vs. 17°C, respectively (**Figure 4E**). In larvae-adults, for both comparisons we found 6 KEGG pathways of which the most relevant were: focal adhesion in 14°C vs. 17°C, and oocyte meiosis, P53 signaling and progesterone-mediated oocyte maturation pathways in 20°C vs. 17°C. Together, these results suggested that the biological functions and pathways affected were temperature-dependent and tissue specific.

3.7. Persistent temperature effects on the methylome in larvae, juveniles, and adults

We wanted to know the biological role of the genes found in larvae, juveniles, and adults that had DMCs within their body region due to temperature. For this analysis we selected the 163 and 264 list of genes found previously (**Figure S2A and S2B**).

From the 264 genes in the 14°C vs. 17°C comparison, 25 GO-terms were identified (**Figure 5A**) and the most relevant were: regulation of histone deacetylation (GO:0031063), cell adhesion (GO:0007155), smooth muscle tissue development (GO:0048745) and negative regulation of Wnt signaling pathway involved in dorsal/ventral axis specification (GO:2000054). On the other hand, from the 163 genes found in 20°C vs. 17°C, a total of 45 GO-terms were identified (**Figure 5B**) and the most relevant were: neural crest cell migration (GO:0001755), negative regulation of transforming growth factor beta receptor signaling pathway (GO:0030512) and histone H3-K9 methylation (GO:0051567). Six and one KEGG pathways were present in 14°C vs. 17°C and 20°C vs. 17°C, respectively, for instance included TGF-beta signaling pathway and ECM-receptor interaction pathways (**Figure 5C**). These results pointed at relevant pathways irreversibly affected by temperature changes and, in particular, to the contribution of histone modifications in maintaining certain environmentally-induced epigenetic changes.

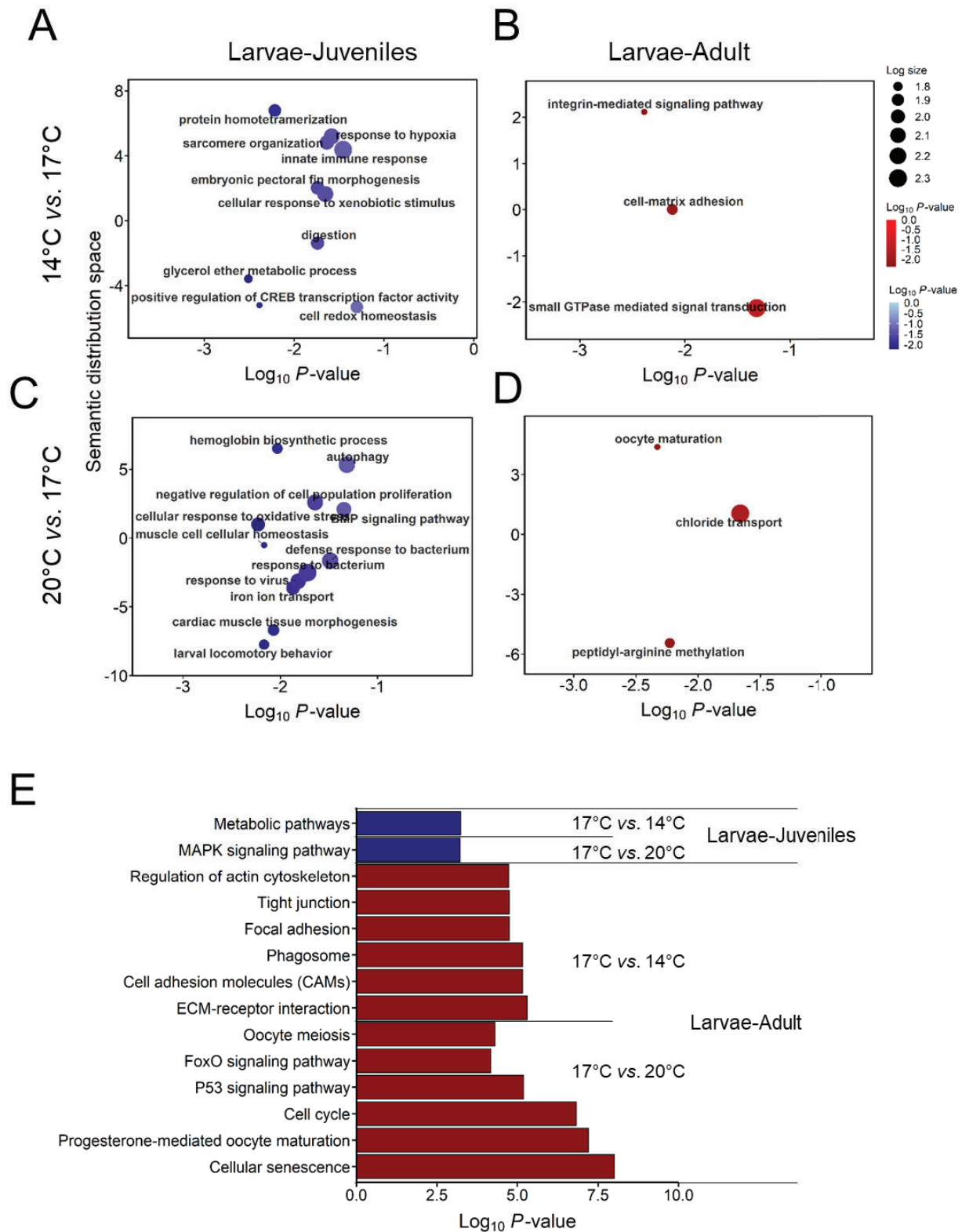


Figure 4. Scatterplots of the enriched Gene Ontology (GO-terms) of the Biological Process (BP) category from the (A) Common differentially expressed genes (DEGs) in larvae and juveniles and (B) Common DEGs in larvae and gonad between 14°C vs. 17°C, (C) common DEGs in larvae and juveniles between 14°C vs. 17°C, and (D) Common DEGs in larvae and gonads between 20°C vs. 17°C. (E) The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment of the DEGs persistently affected by temperature in larvae and juveniles or in larvae and gonads in each temperature comparison.

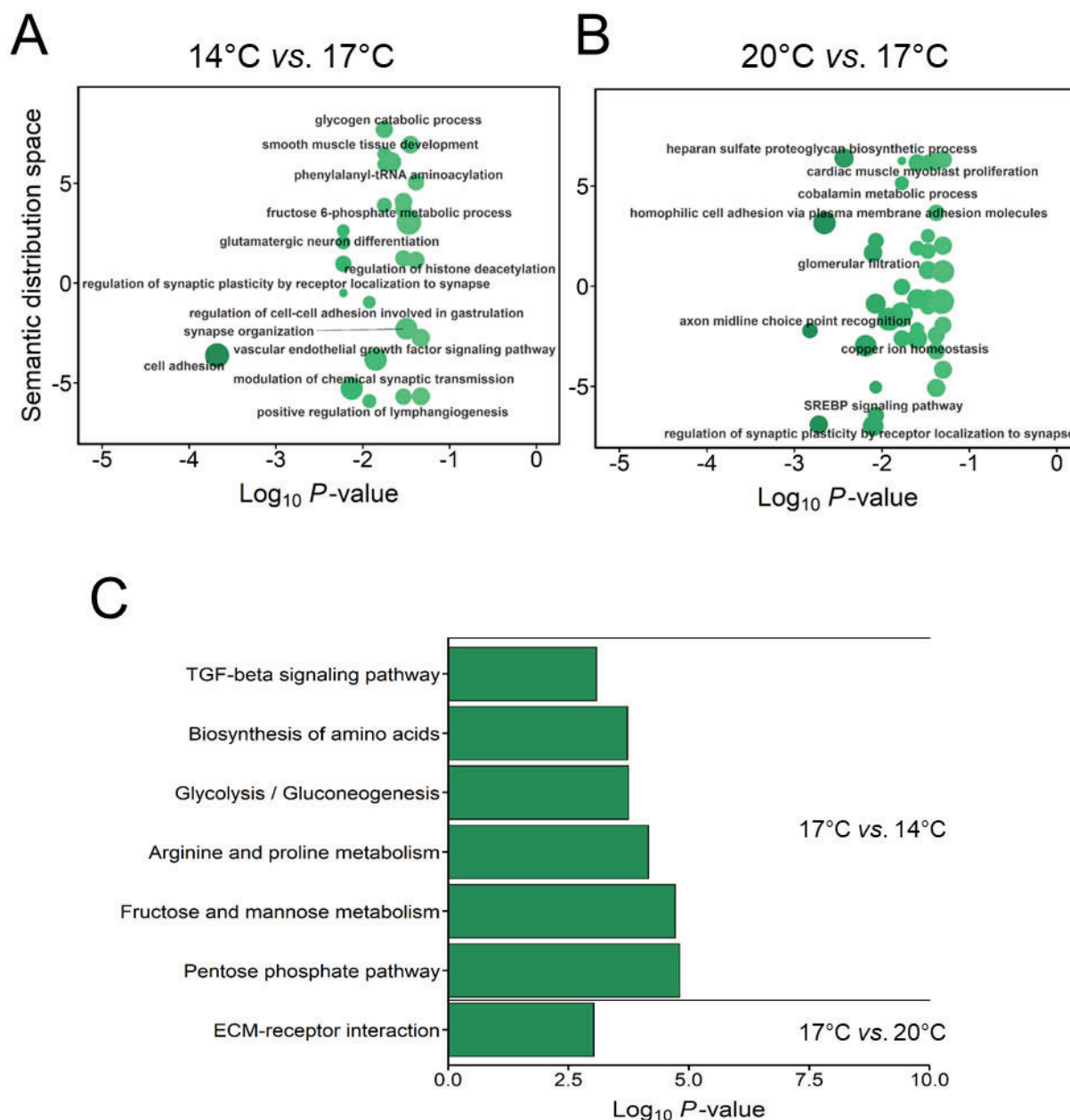


Figure 5. Scatterplots of the enriched Gene Ontology (GO-terms) of the Biological Process (BP) category from the common genes which at least have one differentially methylated cytosine (DMC) in the gene body (gene ± 1000 base pairs, bp) in larvae, gonads, and juveniles in (A) 20°C vs. 17°C comparison and in (B) 14°C vs. 17°C comparison. (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the genes which at least have one DMC in the gene body (gene ± 1000 bp) in larvae, juveniles and gonads in each temperature comparison.

3.8. Persistent temperature effects on transcriptome and methylome in larvae and juveniles

Given the fact that we did analyses of the persistent DEGs between tissues and the persistent genes with affected DMCs between tissues, we were interested in those GO-terms and KEGG pathways. We only selected the 1,350 and 723 genes from larvae and the 6 and 7 genes from juveniles and we discarded the adults results due to very low number of genes found either in 14°C vs. 17°C or 20°C vs. 17°C comparisons (Figure S2A and S2B).

In the 14°C vs. 17°C comparison, and from larvae and juveniles, a total of 36 and 3 GO-terms were identified from the 1,350 and 6 genes, respectively (**Figure 6A** and **6B**). The most relevant GO-terms for larvae were: developmental growth (GO:0048589), somatic muscle development (GO:0007525), calcium ion transmembrane transport (GO:0070588) and regulation of multicellular organism growth (GO:004001), and for juveniles were: digestion (GO:0007586), sarcomere organization (GO:0045214) and embryonic hemopoiesis (GO:0035162). In the 20°C vs. 17°C comparison, and from larvae and juveniles, a total of 52 and 3 GO-terms were identified from the 723 and 7 genes, respectively (**Figure 6C** and **6D**). The GO-terms for larvae included positive regulation of canonical Wnt signaling pathway (GO:0090263), methylation on cytosine within a CG sequence (GO:0010424), histone H3-K9 demethylation (GO:0033169) and neural crest cell differentiation (GO:0014033), and for juvenile tissue were: hemoglobin biosynthetic process (GO:0042541), iron ion transport (GO:0006826) and response to bacterium (GO:0009617).

We only found KEGG pathways in larvae but did not in juveniles due to the low number of genes represented in the two comparisons. Thus, in larvae, 9 and 11 KEGGs pathways were identified in the 14°C vs. 17°C and 20°C vs. 17°C comparisons, respectively (**Figure 6E**). KEGG pathways included Wnt signaling and MAPK signaling pathways in the 14°C vs. 17°C comparison and intestinal immune network for IgA production and biosynthesis of amino acids in the 20°C vs. 17°C comparison. Thus, in particular, Wnt signaling and histone H3K9 demethylation seemed to emerge as key targets of temperature effects.

3.9. Relationship with metastable epialleles and very-long term effects

Next, we wanted to know what of the genes identified in the present study coincide with the genes identified by Anastasiadi *et al.*, (2021) that were differentially methylated still three years after exposing the European sea bass (adulthood) to a simulated heat wave. We did this to further determine the usefulness of the genes identified in the present study as putative biomarkers of the effects of temperature during larvae development that can have very long-term effects and thus potential lifelong consequences, *i.e.*, well beyond the duration of this study.

First, we started with a quite stringent criteria and thus looked which ones of the genes we identified being DEGs and containing at least one DMC coincided with the nine genes containing metastable epialleles, *i.e.*, loci the methylation of which changes in response to temperature and regardless of tissue being considered, in that case brain, liver, muscle and testis (Anastasiadi *et al.*, 2021a), of which the latter two coincide with the tissues investigated in the present study. We found two of these genes, *spata511* and *agl*, among the list of the 1,350 DEGs that also had at least one DMC in larvae but not in juveniles or adults in the 14°C vs. 17°C comparison. None of the 9 metastable epialleles found by Anastasiadi *et al.*, (2021) were found in larvae, juveniles or adults of the 17°C vs. 20°C comparison.

Given the low number of genes identified with the above stringent strategy, we relaxed the criteria and selected the 527 genes that showed DMR in muscle from (Anastasiadi *et al.*, 2021a) and we compared to our 2,199 (14°C vs. 17°C; **Figure S2A**) and 2,990 (20°C vs. 17°C; **Figure S2B**) list of genes with at least one DMC in juveniles. We identified 64 genes in the 14°C vs. 17°C (**Table S29**) and 95 genes in the 20°C vs. 17°C (**Table S30**). Next, we wanted to know the numbers of genes present in juveniles regardless of temperature change comparing these 64 and 95 lists of genes. Results showed a total of 43 genes in common between these two lists (**Table 1**).

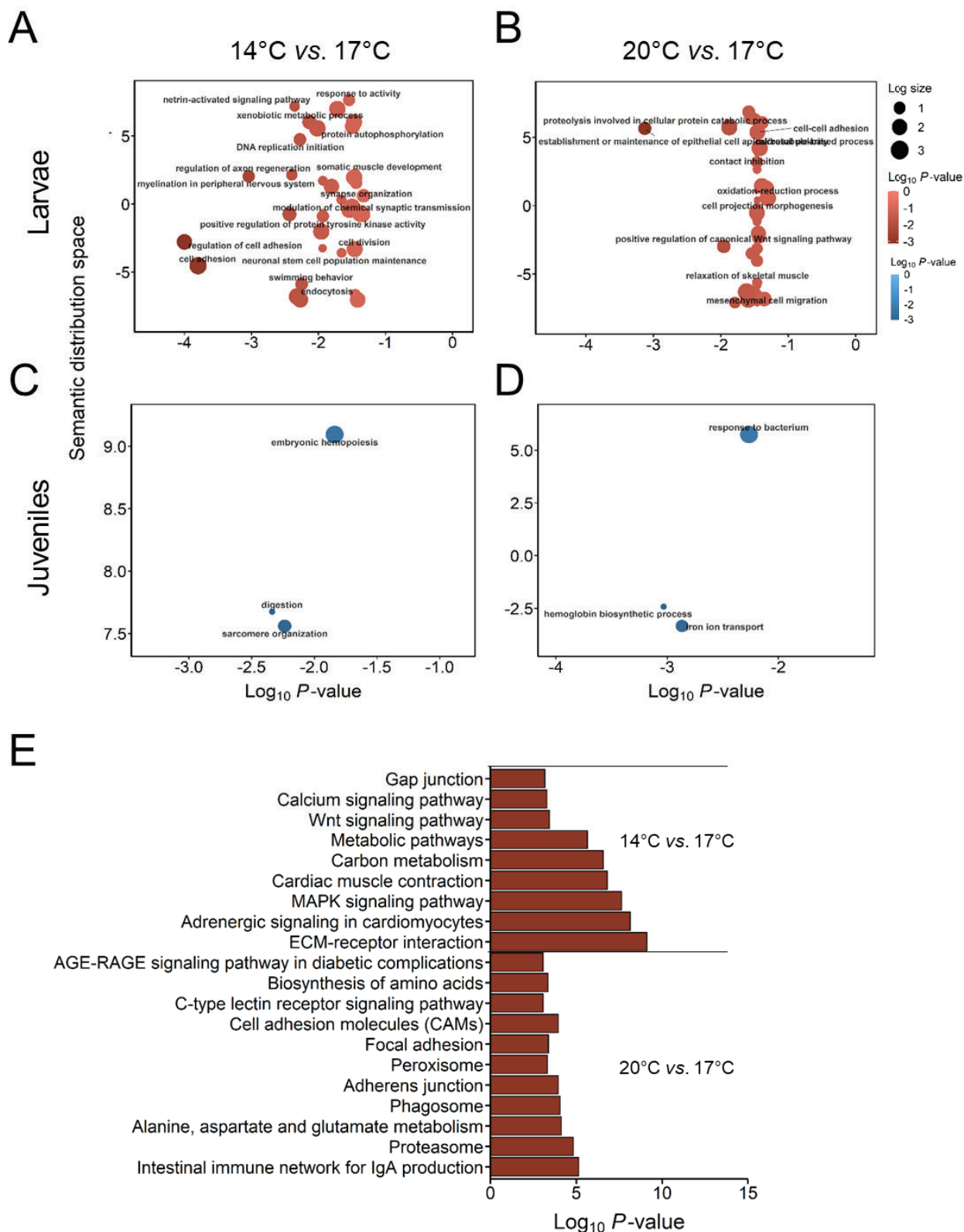


Figure 6. Scatterplots of the enriched GO-terms of the BP category from the DEGs between (A) 14°C vs. 17°C comparison or (B) 20°C vs. 17°C comparison, which at least have one differentially methylated cytosine (DMC) in the gene body region (gene ± 1000 base pairs, bp) in larvae tissue. GO-terms scatterplots of the DEGs between (C) 14°C vs. 17°C or (D) 20°C vs. 17°C comparison, which at least have one DMC in the gene body (gene ± 1000 bp) in juveniles. (E) KEGG enrichment analysis of the DEGs between the two temperature comparisons and which at least contain one DMC in the larvae tissue.

Table 1. List of the 44 genes common between the 64 and 95 genes with at least one DMC in 14°C vs. 17°C and in 20°C vs. 17°C comparisons, respectively, that also contained differentially methylated regions (DMR) in juveniles tissue in the European sea bass after exposure to elevated temperature (Anastasiadi *et al.*, 2021a)

Gene ID	Gene name	Gene description
ENSDLAG00005009668	<i>abcc9</i>	ATP-binding cassette, sub-family C (CFTR/MRP), member 9
ENSDLAG00005019673	<i>agl</i>	amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase
ENSDLAG00005002947	<i>astn1</i>	astrotactin 1
ENSDLAG00005011333	<i>atf7ip</i>	activating transcription factor 7 interacting protein
ENSDLAG00005007690	<i>bcas3</i>	BCAS3 microtubule associated cell migration factor
ENSDLAG00005017692	<i>cacna2d4</i>	calcium voltage-gated channel auxiliary subunit alpha2delta 4
ENSDLAG00005007417	<i>cadm1a</i>	cell adhesion molecule 1a
ENSDLAG00005015894	<i>csmd2</i>	CUB and Sushi multiple domains 2
ENSDLAG00005008332	<i>cxxc4</i>	CXXC finger 4
ENSDLAG00005016884	<i>dhfrs13a.1</i>	dehydrogenase/reductase (SDR family) member 13a, tandem duplicate 1
ENSDLAG00005019105	<i>dusp16</i>	dual specificity phosphatase 16
ENSDLAG00005023582	<i>fgfr3</i>	fibroblast growth factor receptor 3
ENSDLAG00005013628	<i>fstl3</i>	follistatin-like 3 (secreted glycoprotein)
ENSDLAG00005021658	<i>gpc3</i>	glypican 3
ENSDLAG00005002969	<i>hivep1</i>	HIVEP zinc finger 1
ENSDLAG00005002265	<i>hoxc4a</i>	homeobox C4a
ENSDLAG00005004570	<i>hoxd4a</i>	homeobox D4a
ENSDLAG00005017996	<i>inhbaa</i>	inhibin subunit beta Aa
ENSDLAG00005009376	<i>kcnip2</i>	potassium voltage-gated channel interacting protein 2
ENSDLAG00005019431	<i>ksr2</i>	kinase suppressor of ras 2
ENSDLAG00005025443	<i>mcm5</i>	minichromosome maintenance complex component 5
ENSDLAG00005003802	<i>mios</i>	missing oocyte, meiosis regulator, homolog (Drosophila)
ENSDLAG00005011341	<i>mtss1</i>	MTSS I-BAR domain containing 1
ENSDLAG00005001244	<i>nckap5l</i>	NCK-associated protein 5-like
ENSDLAG00005014329	<i>ndufs4</i>	NADH:ubiquinone oxidoreductase subunit S4
ENSDLAG00005006976	<i>nlgn1</i>	neuroligin 1
ENSDLAG00005005359	<i>nup88</i>	nucleoporin 88
ENSDLAG00005023392	<i>ptk7</i>	protein tyrosine kinase 7 (inactive)
ENSDLAG00005003997	<i>rad54l2</i>	RAD54 like 2
ENSDLAG00005010100	<i>raly</i>	RALY heterogeneous nuclear ribonucleoprotein
ENSDLAG00005009049	<i>rbfox1</i>	RNA binding fox-1 homolog 1
ENSDLAG00005022557	<i>rnf25</i>	ring finger protein 25
ENSDLAG00005000657	<i>runx3</i>	RUNX family transcription factor 3
ENSDLAG00005018542	<i>samd4a</i>	sterile alpha motif domain containing 4A
ENSDLAG00005000254	<i>scap</i>	SREBF chaperone
ENSDLAG00005010637	<i>shf</i>	Src homology 2 domain containing F
ENSDLAG00005010165	<i>spata5</i>	spermatogenesis associated 5
ENSDLAG00005018273	<i>strn</i>	striatin, calmodulin binding protein
ENSDLAG00005019987	<i>syde2</i>	synapse defective Rho GTPase homolog 2
ENSDLAG00005009862	<i>tmcc1</i>	transmembrane and coiled-coil domain family 1
ENSDLAG00005009409	<i>ttc7a</i>	tetratricopeptide repeat domain 7A
ENSDLAG00005026035	<i>wnt9a</i>	wingless-type MMTV integration site family, member 9A
ENSDLAG00005013567	<i>xrn2</i>	5'-3' exoribonuclease 2

As we did in the previous section, we did the same analysis but in this case comparing with the list of genes with DMR in the list of 825 (14°C vs. 17°C) and 921 (20°C vs. 17°C) genes with at least one DMC in testis (Anastasiadi *et al.*, 2021a). We identified 16 (**Table S31**) and 20 genes (**Table S32**) in 14°C vs. 17°C and 20°C vs. 17°C lists, respectively. Next, we wanted to know the numbers of genes present in adults regardless of temperature change comparing these 16 and 20 lists of genes. We found seven genes in common between these two lists. These genes were: *agl*, heat shock protein family B (*small*) member 11 (*hspb11*), neuroligin 1 (*nlgn1*), serine/threonine kinase receptor associated protein (*strap*), ADP ribosylation factor guanine nucleotide exchange factor 2 (*arfgef2*), and regulatory factor X5 (*rxf5*).

3.10. Epigenetic biomarkers with both diagnostic and prognostic value

Finally, one of the interesting characteristics of epigenetic biomarkers is that they can have both diagnostic and prognostic value, *i.e.*, epigenetic changes can anticipate transcriptomic changes. In order to explore whether we could find genes fulfilling these characteristics, we first searched genes with at least one DMC in larvae (sampled at 5–8 dpf) and juveniles (sampled at 90 dpf) and differentially expressed in juveniles but not in larvae. In the 14°C vs. 17°C comparison we found one such gene, *prkd4*, while in the 20°C vs. 17°C we found two genes, PDZ Domain Containing 7 (*pdzd7*) and an unannotated gene, ENSDLAG00005003985. As for the adults, we found one gene, *kcnh3* in the 14°C vs. 17°C comparison and two unannotated genes, ENSDLAG00005021715 and ENSDLAG00005004413 in the 20°C vs. 17°C comparison.

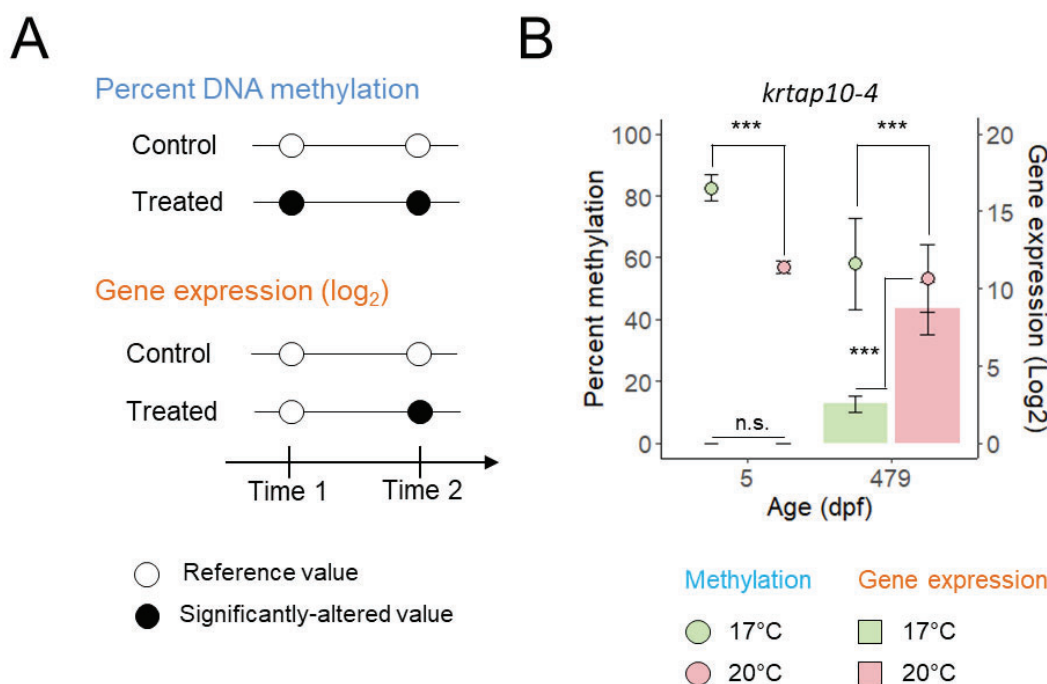


Figure 7. Epigenetic marker with both prognostic and diagnostic value. **(A)** Conceptual diagram of an epigenetic marker. The integration of gene expression data and DNA methylation data in two time points under two conditions can reveal genes, regions or loci with prognostic and diagnostic value. **(B)** The dot plot with the primary y-axis (left) shows the DNA methylation data of the same DMC within ENSDLAG00005004413 gene body region ± 1000 base pairs, bp in larvae at 5 and adults at 482 days post fertilization (dpf). The bar plot with the secondary y-axis (right) shows the gene expression data (log₂) for this gene in larvae at 5 and adults at 482 dpf. The color code indicates the temperature treatment: 17°C in green, and 20°C in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

However, when we looked at the data at the single nucleotide level we found that we had data for the same exact DMC for only one unannotated gene, ENSDLAG00005004413. Regarding the DMC within this gene, the same CpG locus that was hypomethylated in larvae reared at 20°C was also hypomethylated in the adults of fish reared at 20°C at 482 dpf. Therefore, this gene can be considered an epigenetic marker with both prognostic and diagnostic value (**Figure 7**). To investigate further the functionality of the biomarker, we searched for similar protein sequences across all genomes available in NCBI database (“Database resources of the National Center for Biotechnology Information,” 2016) using blastp (Altschul, 1997) with the protein sequence of ENSDLAG00005004413. As a result, the keratin associated protein 10-4 (*krtap10-4*) from the brown trout, *Salmo trutta*, gave the highest sequence similarity (Percentage identity = 43%,) with high statistical significance ($P = 3E^{-25}$). This gene is associated with pathways of keratinization and developmental biology. Although the percentage of identity was low, the alignment was significant, allowing us to suggest the hypothesis that the identified biomarker is likely a coding gene that could be related to keratin production.

4. Discussion

In this study, we provide a set of epigenetic and transcriptomic biomarkers associated with the quality of the thermal regime experienced by the European sea bass during early development, a period when temperature changes can have long-lasting phenotypic consequences for the performance of this species in aquaculture settings.

Before discussing the results proper, however, three considerations on the reliability of the data obtained seem appropriate. First, the temperature treatment was carried out during the thermosensitive period (Piferrer *et al.*, 2005; Vandeputte and Piferrer, 2019) but, as regards to effects on sex ratio, there were no differences among the three tested temperatures, although higher temperature resulted in a slightly higher number of males. The lack of significant differences is likely due to the short duration of the temperature treatment, less than ten days, when treatments of thirty days or longer are necessary to elicit permanent sex ratio responses (Navarro-Martín *et al.*, 2009b), and are regarded as positive in the context of the present study because we wanted to assess the effects of temperature during the most critical phase of development, and not to elicit significant sex ratio changes. Hence, the deliberate short temperature treatment period used. Second, at 482 dpf, fish reared at 14°C until 12 dpf were bigger than those reared at 20°C until 8 dpf, a phenomenon that probably indicates better performance of the cold-exposed fish, a result that should be further studied without excluding the possibility of compensatory growth, as reported previously in the European sea bass (Díaz and Piferrer, 2015). However, the lack of differences in growth between fish reared at 14°C, 17°C and 20°C at 90 dpf, nor between adult fish reared at 14°C vs. 17°C or between 20°C vs. 17°C (*i.e.*, the two comparisons made in this study), coupled with the lack of differences in GSI values among the males of the different temperature regimes, provides confidence that the observed differences in DNA methylation or in gene expression were due only to temperature and not to differences in somatic or gonadal growth. Third, and regarding the analyses performed, transcriptomic data was of good quality since only samples with RIN > 8, average alignment rate > 88% and > 30 million assigned features were used. The same applies to the epigenomic data, based on an average > 70% alignment rate and > 99% bisulfite conversion quality indicators. Taken together, this indicates that epigenetic and transcriptomic results do indeed reflect temperature effects.

It should also be noted that in this study we concentrated on DNA methylation-based biomarkers for their simplicity of measurement (Levenson, 2010) and because, once known the nature of the genes involved, no longer a whole-genome or genome-wide determination method is required. Instead, the measure of their DNA methylation can be achieved by a simple PCR-based method (Hernández *et al.*, 2013), as done, for example, in the development of some epigenetic clocks, another type of DNA methylation-based biomarker (Aliferi *et al.*, 2018; Naue *et al.*, 2017). Other type of epigenetic markers do exist, like those based on histone modifications such as H3K4me3, H3K9me3, H3K79me, and H3K79me3, which have been used, for example, to assess proper oocyte development in cyprinid fishes (Zhou *et al.*, 2019); and those based on the quantification of miRNAs, *e.g.*, the differentially expressed miR-122, novel-miR6, miR-193a-3p, and miR-27a-5p, in the head kidney, the expression levels of which were linked to mechanisms of inflammation and injury during cadmium exposure in the common carp, *Cyprinus carpio* (Liu *et al.*, 2020). Concentrating in these other type of biomarkers was beyond the scope of study but efforts should be aimed at their identification in future studies. Finally, it is worth mentioning that we not only used epigenetic changes but also investigated gene expression changes, meaning that we reported epigenetic changes that also resulted in phenotypic (transcriptomic) change, thus reinforcing the usefulness of the epigenetic biomarkers presented below. However, we are aware that beyond the proof-of-concept that this study represents, additional phenotypic endpoints would be desirable.

Regarding the coping style, we found that in the group of fish treated at 20°C there were more fish that were *Bold* than *Shy*. Therefore, the development of biomarkers that allow selecting fish reared at 14°C, are, at the same time increasing the probability of selecting for *Shy* behavior, a trait preferred in an easily stressed species like the European sea bass. The study of the methylome and the transcriptome of the two different behavior types could hold interesting results. These could be further explored in future studies since recent evidence showed differences between *Bold* and *Shy* copying styles (Sadoul *et al.*, 2022). However, a study using the olive flounder (*Paralichthys olivaceus*) only found 144 DEGs out of the entire transcriptome (Zou *et al.*, 2021). For these reasons, we prioritized the development of biomarkers for thermal stress and not for the coping style.

Among all the genes identified that could potentially be used as biomarkers of the thermal quality of the early development, there are some that repeatedly showed up (**Table 2**) and thus we will focus on these genes. We classified the effects as short-term if they were present only in larvae, right at the end of the temperature treatment, mid-term effects if they were present both in larvae and juveniles at 90 dpf, long-term if they were present both in larvae and adults at ~500 dpf, and very-long term if they were affected in larvae, 90 or ~500 dpf. In addition, coincide with the genes the methylation of which was found to be affected in adults three years after the end of the exposure to elevated temperature as is shown in a previous study with the European sea bass (Anastasiadi *et al.*, 2021a).

Among the 70 genes that were differentially methylated regardless of the direction of the temperature change and that were affected in larvae, juveniles and adults, it is worth mentioning *jmjd7*, *kcnh5b* and *zfyve9a*. The former encodes a highly conserved enzyme with a JmjC domain that recognizes and cleaves monomethylated and dimethylated arginine residues of histones H2, H3 and H4. Other members of this family include (a) *jmjd3* (also known as lysine demethylase 6B, *kdm6b*), which has been implicated in environmental signal transduction in connection to temperature-dependent sex determination in reptiles, including the Australian bearded dragon, *Pogona vitticeps*, (Deveson *et al.*, 2017), the red-eared slider turtle, *Trachemys scripta* (Ge *et al.*,

Table 2 Summary of the biomarkers identified in this study related to the effects of early temperature according to their informative properties. **Abbreviations:** T = transcriptomic; E = epigenetic; B = both; L = larvae; J = juvenile; A = adults

Properties	Type	Tissue	Lower temperature	Elevated temperature
Short-term effects (L)	E	L	<i>agl, kcnh2a, depdc5, prkd2, gemin7</i> ; (Table S 17) *	
	T	L	<i>cabp2a, eloa1, serpina7, mrc1a, gpx1b, sncca</i> ; (Table S23) *	
	B	L	<i>hivep1, kcnq5a, pcdh1b, hmgera, dusp16</i> ; (Table S24) *	
Mid-term effects (L & J)	E	L & J	<i>agl, dligap1a, kcnh2a, scap, rsul</i> ; (Table S17) *	
	T	L & J	<i>ENSDLAG00005005752, ENSDLAG00005023917</i> (Table S5, S6 and S8, S9) *	
	B	L & J	<i>capn9, ENSDLAG000005003493, ENSDLAG00005005752</i> ; (Table S19, S21)	<i>tfa</i> (Table S20, S22)
Long-term effects (L & A or L & J & A)	E	L & A	<i>agl, pvalb7, brinp2, astn1, nup98</i> ; (Table S 17) *	
	E	L & J & A	<i>agl, pvalb7, brinp2, astn1, xdh</i> ; (Table S17) *	
	T	L & A	<i>ENSDLAG00005000425</i> (Table S5, S7 and S8, S10) *	
Very long-term effects (L or J or A & Anastasiadi <i>et al.</i> , 2021)	E	J	<i>agl, spata5, fgfr3, cacna2d4</i> ; (see more examples in Table 1) *	
	E	A	<i>agl, hspb11, nlgn1, strap, arfgef2, rfx5</i> * (Table S31, S32) *	
	B	J	<i>spata5l1, agl</i> , <i>prkd4</i>	N/A <i>pdzd7, ENSDLAG00005003985</i>
Putative prognostic and diagnostic	B	L & J	<i>prkd4</i>	<i>ENSDLAG00005021715, krtap10-4</i>
	B	L & A	<i>kcnh3</i>	<i>ENSDLAG00005021715, krtap10-4</i>
Strong prognostic and diagnostic	B	L & A	N/A	<i>krtap10-4</i>

* Genes affected by both lower (14°C) and higher (20°C) temperatures with respect to the control temperature (17°C).

2018) and involved in sex-related processes in the orange-spotted grouper fish, *Epinephelus coioides* (Zhang *et al.*, 2021); and (b) jumonji and AT-rich interaction domain containing 2, *jarid2a*, upregulated in response to elevated temperature in the European sea bass (Díaz and Piferrer, 2015) but downregulated when fish were also exposed to estrogen (Díaz and Piferrer, 2017). On the other hand, *kcnh5b* encodes a member of voltage-gated potassium channel, also involved in environmental signal transduction in different situations (Yang and Zheng, 2014), but whether they are involved in capturing temperature effects on sex ratios is still not known. Lastly, *zfyve9a* encodes a double zinc finger motif-containing protein that interacts directly with SMAD2, and recruits SMAD2 to the TGFB receptor (Tsukazaki *et al.*, 1998). The *zfyve9a* gene regulates the proliferation of hepatic cells during zebrafish embryogenesis (Liu *et al.*, 2013).

The 211 genes identified in which their expression either consistently increased or decreased from 14°C to 17°C and also from 17°C to 21°C included *stap2b*, *mrc1a* and *dab2*. These genes constitute, therefore, robust short-term only transcriptomic biomarkers of the effects of temperature during European sea bass early development. Their expression levels could be used to assess the on-the-spot quality of the egg incubation thermal regime and larvae until first feeding. Among the most relevant GO-terms related with the above mentioned genes upregulated by elevated temperature was the response to oxidative stress. It is well established that elevated temperature accelerates developmental rate as does mitochondrial respiration, leading to reactive oxygen species (ROS) accumulation and oxidative stress (Sun *et al.*, 2015). Changes in ROS status and oxidative stress have been implicated as key elements in environmental signal transduction in poikilothermic vertebrates (Castelli *et al.*, 2020). Identifying the mechanisms that regulate these molecular oxidative stress responses can help to develop new approaches to monitor the appearance of undesirable conditions, especially in an aquaculture context (Lazado *et al.*, 2020).

There were 210 genes that in addition of being DEG in larvae they also contained DMCs. Of these, 95 genes were upregulated at 20°C and downregulated at 14°C and 17°C treatments while the remaining 115 genes had the opposite pattern. Two considerations are worth being made here. First, the grouping in the heatmap of the 14°C and 17°C treatments, on one hand vs. the 20°C treatment, on the other, suggests that in terms of DNA methylation and gene expression the temperature of 20°C is perceived as altered when compared to the “natural” temperature range of 14–17°C, appropriate for larval development in the European sea bass (Vazquez and Munoz-Cueto, 2014). In fact, this observation matches and supports the observations that temperatures > 17°C alters sex ratios in this species while temperatures < 17°C do not (Navarro-Martín *et al.*, 2009b). Second, these 210 genes, among which we find several genes related to calcium signaling such as *atp2a1*, *gsto2* and *cabp2a*, involved also in environmental signal transduction (Castelli *et al.*, 2020), constituted robust short-term both epigenomic and transcriptomic biomarkers of the effects of temperature during the European sea bass early development. As above, they can be used to assess the quality of the egg incubation thermal regime and larvae until first feeding. Furthermore, the most relevant GO-terms linked to these 210 genes were DNA methylation, regulation of Wnt signaling pathway and cellular response to cold. All of these GO-terms results are consistent with the fact that DNA methylation, Wnt signaling and cellular response to cold has been previously implicated in other studies examining the effects of temperature on animals (Al-Fageeh and Smales, 2006; Risha *et al.*, 2021; Yao *et al.*, 2022).

Regarding the 577 and 344 genes with negative and positive correlation, respectively, in the 14°C vs. 17°C comparison and the 102 and 250 genes with negative and positive correlation,

respectively, in the 20°C vs. 17°C comparison it is interesting to note that gene expression changed on average less than 10-fold (less than one log value). In contrast, DNA methylation changes were around 20–25%. These differences are of sufficient magnitude suitable to use them as biomarkers of effects of temperature during early development. There were no common genes between the 577 and 102 gene lists, meaning there were no genes the expression of which is epigenetically and consistently downregulated in response to temperature. As for the 344 and 250 genes more methylated and upregulated by temperature, only one gene, ENSDLAG00005009958, was common, implying that its expression is epigenetically regulated in a consistent manner in response to temperature changes and thus constitute a powerful biomarkers of the short-term effects of early temperature changes in larvae.

We only found two genes, *spata511* and *agl*, that coincided with the metastable epialleles identified previously in the European sea bass (Anastasiadi *et al.*, 2021a). This low number is not surprising given the stringent criteria we used and the requirements needed to be considered a metastable epiallele. Using instead genes with DMCs in juveniles and/or adults that coincided with genes differentially methylated in the same tissues in the above mentioned study returned other gene lists (43 shown in **Table 1**) that constituted robust epigenetic biomarkers for the very long-term effects of temperature alterations during early development on European sea bass juveniles and adults. Some of them are independent of the direction of temperature change and hence can be used as a reliable method to determine whether animals had the optimal temperature during larval rearing. *Ag1* and *nlg1* genes are present in both lists. *Ag1* encodes the glycogen debrancher, an enzyme involved in glycogen degradation and *agl* mutations are linked with glycogen storage disease (Sentner *et al.*, 2012) found differentially expressed in fast- vs. slow-growing rainbow trout *Onchorynchus mykiss* (Danzmann *et al.*, 2016) but no effects due to temperature have been reported.

One of the most important insight of this study is the report, to our knowledge for the first time, of epigenetic biomarkers with prognostic and diagnostic value. Of the five genes that in principle met the criteria for having these properties DMCs were not consistently on the same CpGs position and thus can be considered as potential epigenetic marker with both prognostic and diagnostic value. The gene ENSDLAG00005004413, for which we had gene expression data and DNA methylation data on exactly the same CpG position in two time points in larvae and adults, was not differentially expressed either at 14°C nor at 20°C when compared to 17°C at 5 dpf but became upregulated in the adults in fish reared at 20°C at 482 dpf. Regarding the DMC within this gene, the same CpG that was hypomethylated in larvae reared at 20°C was also hypomethylated in the adults of fish reared at 20°C at 482 dpf. Thus, hypomethylation of this gene in larvae exposed to elevated temperature precedes higher expression in adults ~500 dpf. Therefore, this gene can be considered an epigenetic biomarker with both prognostic and diagnostic value and therefore their full identification is warranted. However, further analysis on the same gene and all the other biomarkers listed in the present study should be performed in a larger number of individuals from different families and from an independent experiment to confirm its usefulness as biomarker. Given the protein sequence alignment of ENSDLAG00005004413 with *krpa10-4* gene in the brown trout, we suggest that the identified biomarker could be related to keratin production. This could be of relevance for the fitness of the sea bass since they produce keratin in the skin mucus (Cordero *et al.*, 2015). The skin mucosa is a crucial first barrier of protection from both biotic and abiotic factors because it protects the skin epidermis, where fish in general have few or no keratinized cells (Brinchmann, 2016). Thus, since methylation is maintained, measuring the methylation of this gene in fish samples from a given

batch at any time would indicate whether that the fish of that batch had been exposed to an abnormally high temperature during the critical stages of early development.

In this study, we focused on the most studied environmental factor, *i.e.*, temperature, which has a huge impact on stress in many vertebrate and invertebrate marine farming species. However another environmental factor, with detrimental effects on European sea bass development is rearing density (Sammouth *et al.*, 2009), albeit without effects on sex ratio (Saillant *et al.*, 2003b) as temperature does. In an experiment with 6 month-old European sea bass exposed to short intervals of acute density stress, 57 DMCs in 51 distinct genes measured in blood samples were identified (Krick *et al.*, 2021). Interestingly, among these 51 genes, two genes were also present in our list of genes with DMCs in three tissues: roundabout homolog 2-like (*robo3*) in 20°C *vs.* 17°C and zinc finger matrin-type 4b (*smat4*) in 14°C *vs.* 17°C. These two genes seem to be involved in behavior linked with stress conditions (Rey *et al.*, 2013).

In conclusion, we report, to the best of our knowledge, the first identification of epigenetic biomarkers with both diagnostic and prognostic value to anticipate the long-lasting effects of thermal alterations during early development in fish. One can envisage the determination of DNA methylation levels (*e.g.*, by a PCR-based approach) of CpGs within some of the genes identified in this study as a rapid and economic means to monitor the quality of the early environment, in our case the thermal environment, to screen samples of the European sea bass larvae as a diagnostic and prognostic tool for present and future performance. Our study shows the way for similar developments for other farmed species.

Development of epigenetic biomarkers with diagnostic and prognostic value to assess the lasting effects of early temperature changes in farmed fish

Alejandro Valdivieso, Núria Sánchez-Baizán, Nikos Mitrizakis, Nikos Papandroulakis,
Francesc Piferrer

Supplementary Information

Supplementary Figures

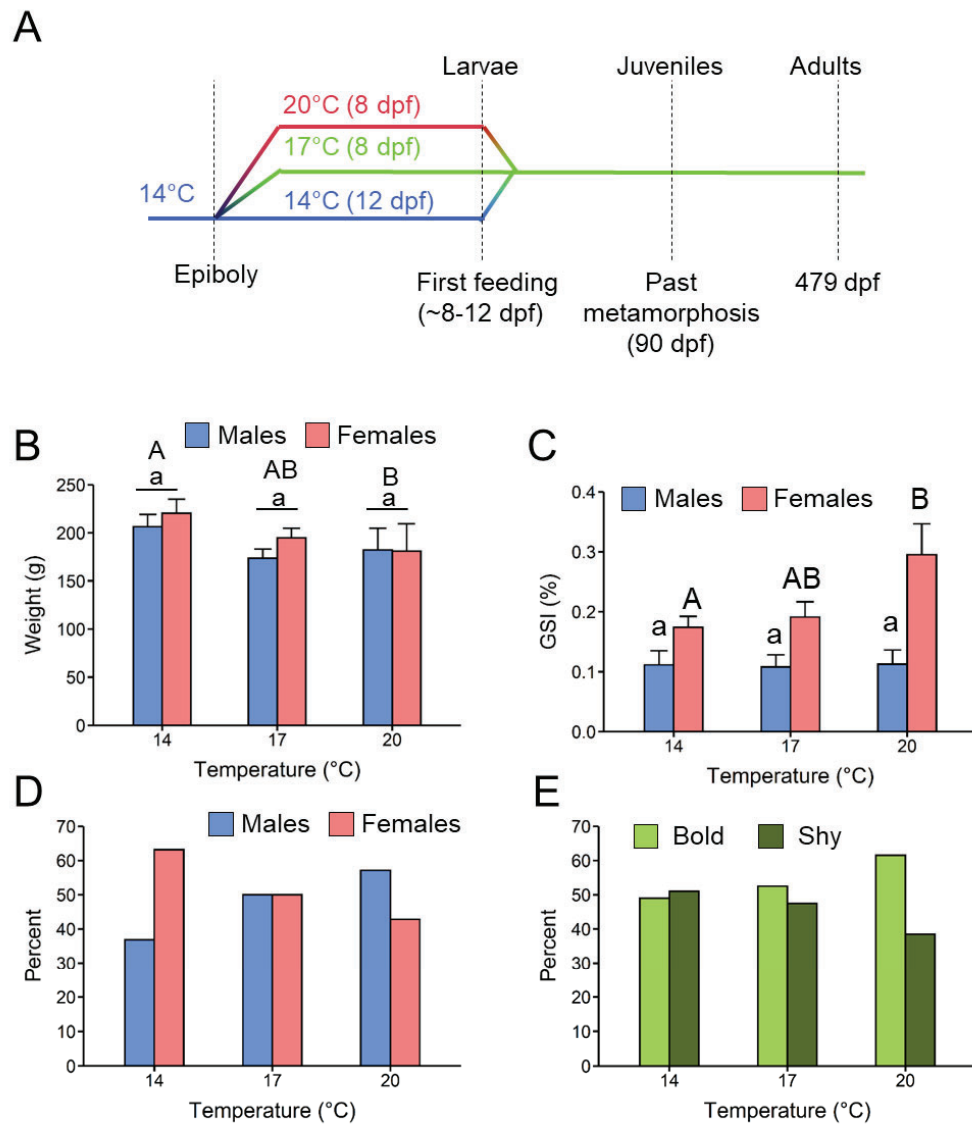


Figure S1. (A) Diagram of the experimental design: the vertical dashed lines indicate the sampling time-points in days post fertilization (dpf) and the tissues taken, while the color of the horizontal lines indicate the temperature treatment (blue: 14°C, green: 17°C, and red: 20°C) and its duration in days (d). Bar plots of: (B) Body weight and (C) Gonadosomatic index (GSI) according to sex, and (D) Number of males and females and (E) Copying style at each temperature treatment.

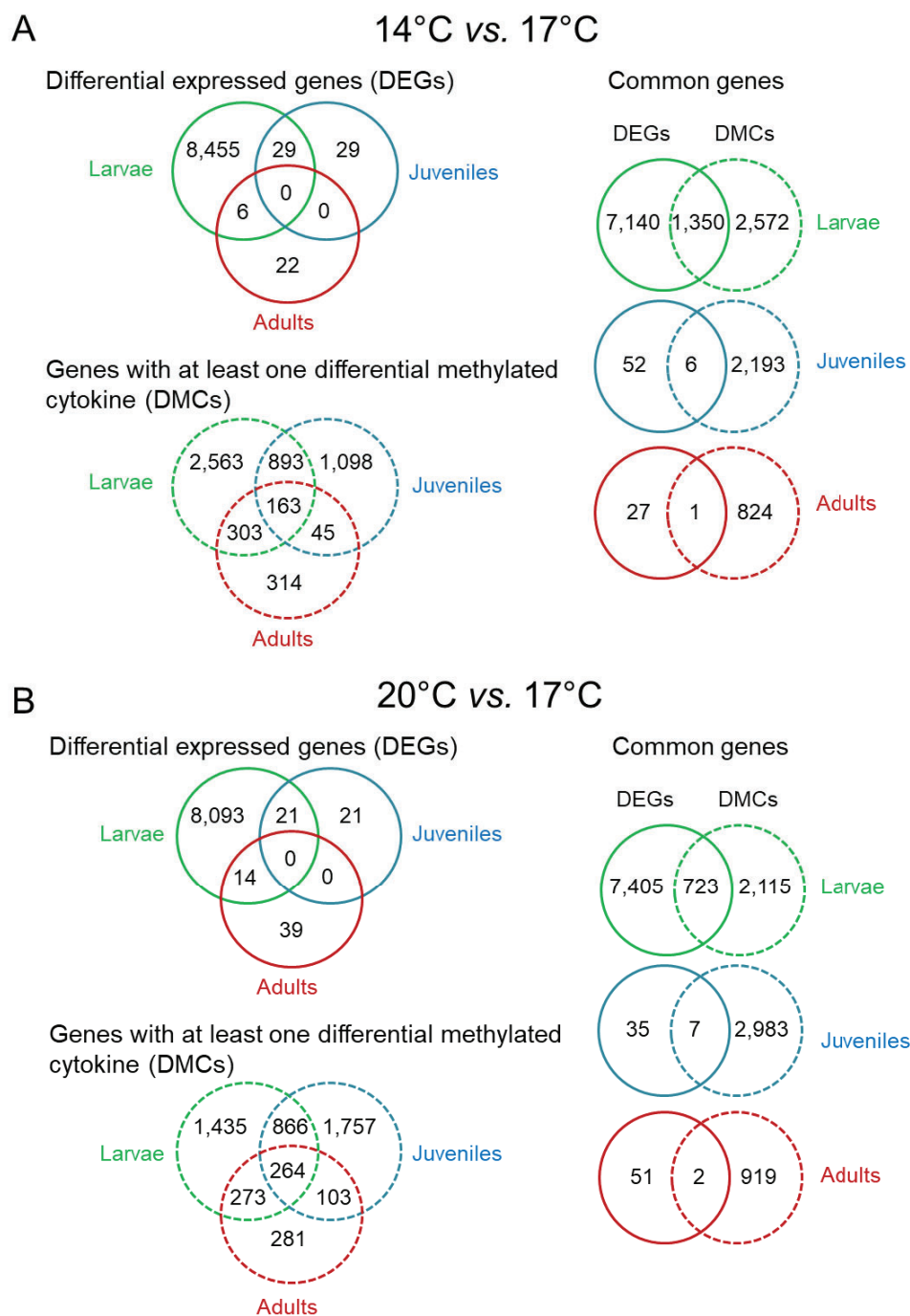


Figure S2. Summary of the number of differentially expressed genes (DEGs, solid circles), genes containing differentially methylated cytosines (DMCs, dashed circles) observed in each tissue or common to more than one tissue according to temperature comparison. Venn diagrams are color-coded in function of tissue (larvae, green; juveniles, blue; adults, red). The number of DEGs with at least one DMC are also indicated.

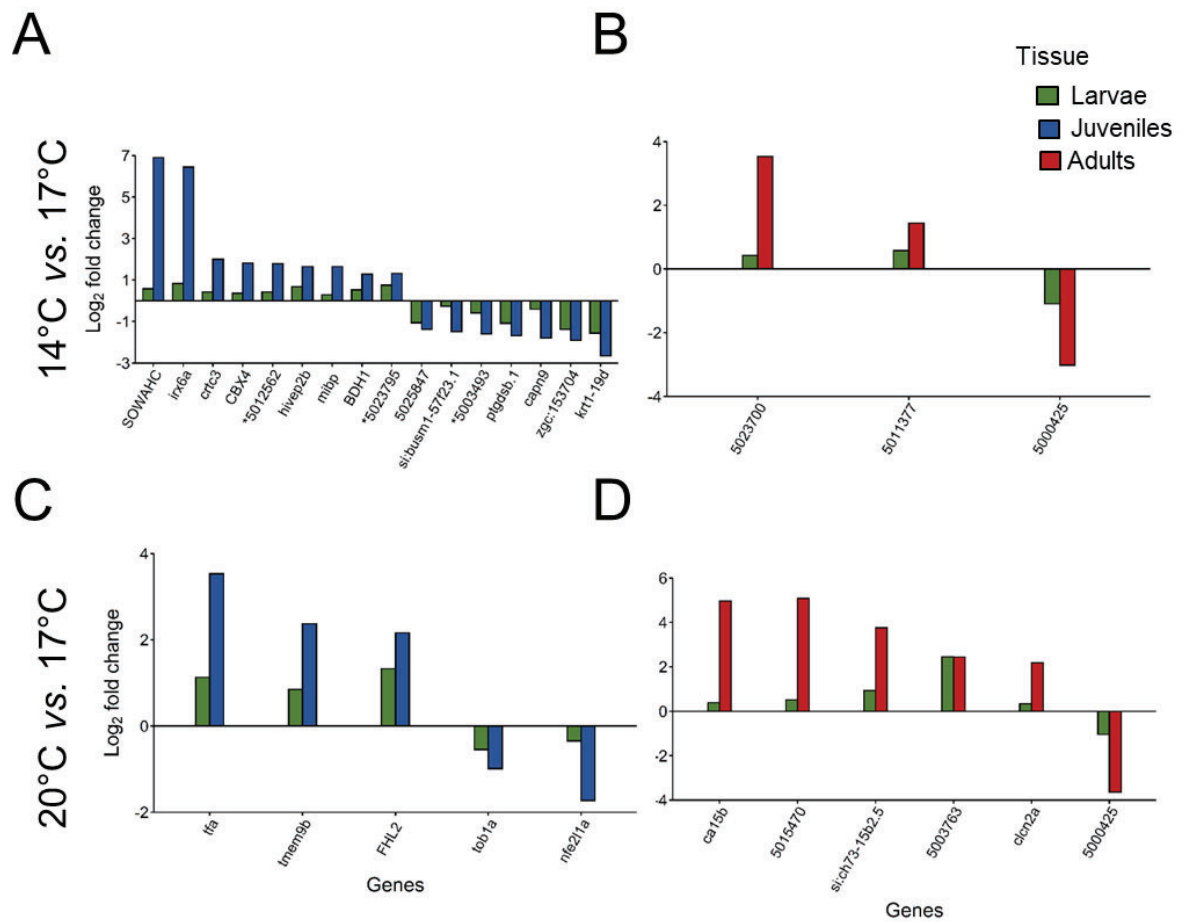


Figure S3. Differentially expressed genes (DEGs) with consistent changes independent of tissue in response to the increase (**A, B**) or decrease (**C, D**) of temperature in larvae and juveniles (**A, C**), or in larvae and adults (**B, D**). Data as bar plots of gene expression (log₂ fold change). Gene IDs are indicated on the x-axis.

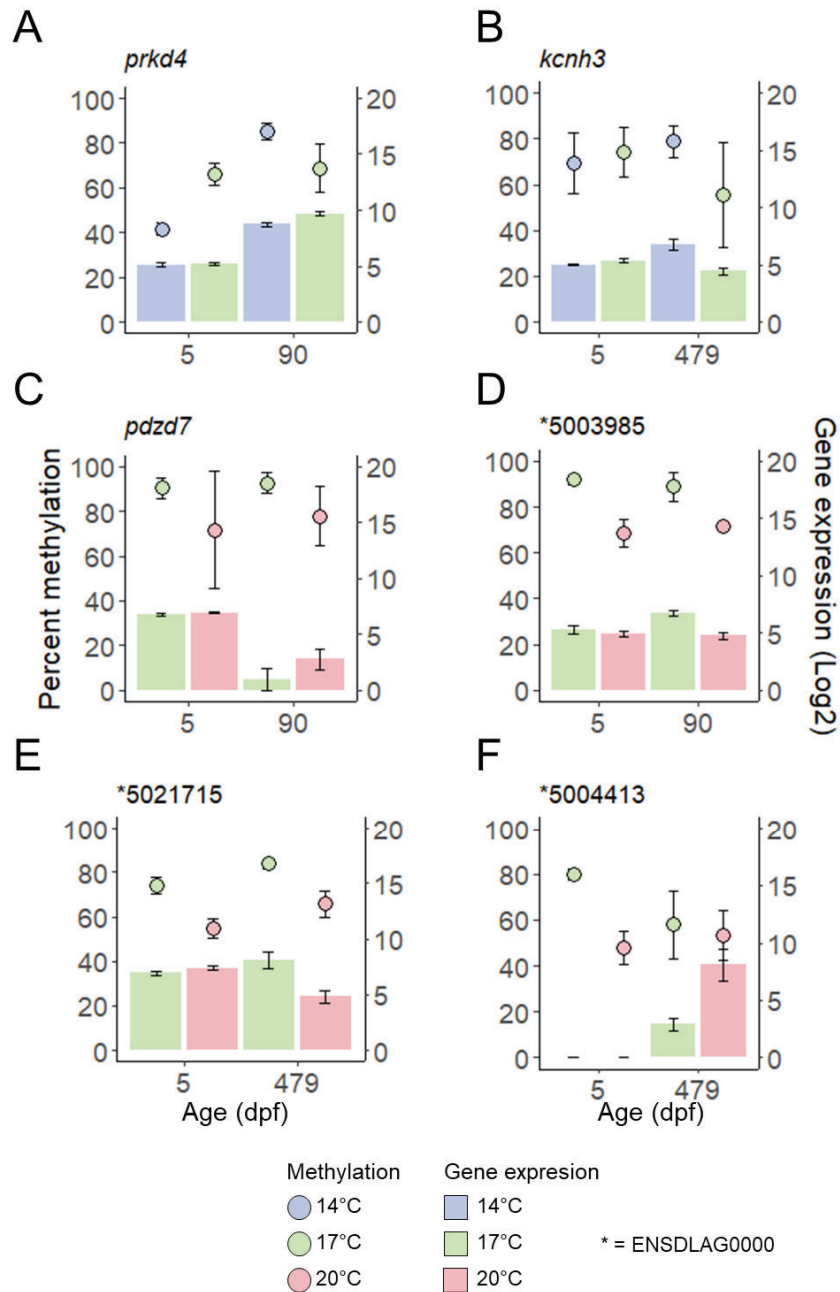


Figure S4. Genes the methylation of which constitute epigenetic markers with both prognostic and diagnostic value: (A) *prkd4*, (B) *kcnh3*, (C) *pdzd7*, (D) ENSDLAG00005003985, (E) ENSDLAG00005021715 and (F) ENSDLAG00005004413. The dot plot with the primary y-axis (left) shows the DNA methylation data of the same DMC within the gene body $\pm$ 1000 base pairs, bp whereas the bar plot with the secondary y-axis (right) shows the gene expression data (log₂) for the same gene.

Supplementary data

Supplementary data to this article can be found online at
<https://doi.org/10.1016/j.aquaculture.2022.738918>.

Supplementary tables

Supplementary tables **Table S1** through **Table S32** are contained in an Excel file as part of the online supplementary information. Their content is described below.

Nutritional feeding regime

Table S1 Feeding regime per experimental temperature after hatching (days post fertilization, dpf)

Quality control statistics of RNA-seq and RRBS analyses

Table S2 RNA-sequencing (RNA-seq) statistics of the 100 paired-end reads for differentially expressed genes (DEGs) analysis

Table S3 Reduced representation bisulfite sequencing (RRBS) statistics for differentially expressed cytosines (DMC) analysis

Table S4 Number of samples used in each tissue and type of analysis for each comparison

RNA-seq results

Table S5. The 8,490 up-and-downregulated DEGs in larvae between 14°C vs. 17°C

Table S6. The 58 up-and-downregulated DEGs in juveniles between 14°C vs. 17°C

Table S7. The 28 up-and-downregulated DEGs in adults between 14°C vs. 17°C

Table S8. The 8,128 up-and-downregulated DEGs in larvae between 20°C vs. 17°C

Table S9. The 42 up-and-downregulated DEGs in juveniles between 20°C vs. 17°C

Table S10. The 53 up-and-downregulated DEGs in adults between 20°C vs. 17°C

RRBS results

Table S11 The 27,286 DMCs in larvae between 14°C vs. 17°C

Table S12 The 22,116 DMCs in larvae between 20°C vs. 17°C

Table S13 The 9,884 DMCs in juveniles between 14°C vs. 17°C

Table S14 The 15,886 DMCs in juveniles between 20°C vs. 17°C

Table S15 The 7,254 DMCs in adults between 14°C vs. 17°C

Table S16 The 49,610 DMCs in adults between 20°C vs. 17°C

Table S17 The 7,231 genes that contain at least one differentially methylated cytosines (DMCs) within their gene body region ($\pm 1,000$ base pairs from start and end gene position)

Table S18 Common genes that contain at least one differentially methylated cytosines (DMCs) within their gene body region between larvae, juveniles and adults in each temperature contrast

Combined RNA-seq and RRBS results

Table S19 The 1,350 genes in larvae identified as DEGs and contained at least one DMCs within their body region between 14°C vs. 17°C

Table S20 The 723 genes in larvae identified as DEGs and contained at least one DMCs within their body region between 20°C vs. 17°C

Table S21 The 6 genes in juveniles identified as DEGs and contained at least one DMCs within their body region between 14°C vs. 17°C

Table S22 The 7 genes in juveniles identified as DEGs and contained at least one DMCs within their body region between 20°C vs. 17°C

- Table S23** The 211 genes in larvae consistently upregulated (153 genes) or downregulated (58 genes) by temperature
- Table S24** The 210 genes in common between the 1,350 and 723 list of genes that were previously identified as DEGs with at least one DMC in larvae between 14°C vs. 17°C and 20°C vs. 17°C, respectively
- Table S25** The 577 genes within the 1,350 DEGs containing at least one DMC in the 14°C vs. 17°C comparison that exhibited a negative relationship between DNA methylation and expression levels
- Table S26** The 344 genes within the 1,350 DEGs containing at least one DMC in the 14°C vs. 17°C comparison that exhibited a positive relationship between DNA methylation and expression levels
- Table S27** The 102 genes within the 723 DEGs containing at least one DMC in the 20°C vs. 17°C comparison that exhibited a negative relationship between DNA methylation and expression levels
- Table S28** The 250 genes within the 723 DEGs containing at least one DMC in the 20°C vs. 17°C comparison that exhibited a positive relationship between DNA methylation and expression levels
- Table S29** The 64 genes that were in common between the list of 527 genes that showed a DMR in juveniles in the study of Anastasiadi *et al.*, (2021) and the 2,199 genes with at least one DMC in juveniles between 14°C vs. 17°C in the present study
- Table S30** The 95 genes that were in common between the list of 527 genes that showed a DMR in juveniles in the study of Anastasiadi *et al.*, (2021) and the 2,990 genes with at least one DMC in juveniles between 20°C vs. 17°C in the present study
- Table S31** The 16 genes that were in common between the list of 589 genes that showed a DMR in testes in the study of Anastasiadi *et al.*, (2021) and the 825 DEGs with at least one DMC in the adults (testis) between 14°C vs. 17°C in the present study
- Table S32** The 20 genes that were in common between the list of 589 genes that showed a DMR in the testes in the study of Anastasiadi *et al.*, (2021) and the 921 genes with at least one DMC in adults (testes) between 20°C vs. 17°C in the present study

Acknowledgements

We thank Dr. Katherina Moutou for providing the adult samples.

Credit authorship contribution statement

Alejandro Valdivieso: Methodology, Investigation, Formal analysis, Writing – original draft. Núria Sánchez Baizán: Methodology, Investigation, Formal analysis, Writing – original draft. Nikos Mitrizakis: Methodology, Investigation, Formal analysis. Nikos Papandroulakis: Conceptualization, Resources, Funding acquisition, Supervision, Methodology, Investigation, Formal analysis, Writing original draft. Francesc Piferrer: Conceptualization, Resources, Funding acquisition, Supervision, Writing – original draft.

Funding

This study was supported by the European Union H2020 project Performfish, reference number 727610 and by Spanish Ministry of Science and Innovation (PID2019-108888RB-100) grant to FP. With funding from the Spanish government through the ‘Severo Ochoa Centre of Excellence’ accreditation

(CEX2019-000928-S). AV was supported by a Performfish contract and NS was supported by a Spanish Ministry of Science and Innovation predoctoral scholarship (BES-2017-079744).

Data availability

BS-Seq and RNA-Seq reads are available at the NCBI-SRA under the BioProject accession number PRJNA888946.

Declaration of competing interests

The authors declare they have no competing interests.

Chapter 4

A cost-effective method for epigenetic biomarker development and their integration with genetic selection in finfish aquaculture

To be submitted

AVÍS IMPORTANT

El capítol següent ha estat retirat seguint instruccions de l'autora de la tesi, per existir la possibilitat de generar patents

AVISO IMPORTANTE

El siguiente capítulo ha sido retirado siguiendo instrucciones de la autora de la tesis, por existir la posibilidad de generar patentes

IMPORTANT NOTICE

The following chapter has been removed following instructions of the author of the thesis, for reasons related to the reproduction rights of the original document

Chapter 5

*Multigenerational temperature additive effects: the
epigenetic regulation of masculinization and its
inheritance in the European sea bass*

To be submitted

Multigenerational temperature additive effects: the epigenetic regulation of masculinization and its inheritance in the European sea bass

Núria Sánchez-Baizán¹, François Allal², Johana Denkena^{3,4}, Maria Colomé-Tatché^{3,4}, Marc Vandeputte^{2,5} and Francesc Piferrer¹

1. Institut de Ciències del Mar (ICM), Spanish National Research Council (CSIC), Barcelona, Spain
2. MARBEC, Univ. Montpellier, Ifremer CNRS-IRD, Palavas-les-Flots, France
3. Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
4. Biomedical Center (BMC), Physiological Chemistry, Faculty of Medicine, LMU Munich, Planegg-Martinsried, Germany.
5. GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

Abstract

The European sea bass exhibits mixed genetic and environmental sex determination, with elevated temperatures during early development influencing sex ratios and causing masculinization. This species serves as a valuable model for studying how epigenetics integrates environmental information to modulate gene expression programs and phenotypes. In previous studies it was observed that the levels of DNA methylation of two sex-related genes, *cyp19a1a* and *dmrt1*, were significantly affected by sex and temperature. However, the broader impact of these factors on methylation levels of other genes and regions involved in sex development and reversal remain unknown. Furthermore, environmentally-induced epigenetic marks may potentially be transmitted to subsequent generations, raising questions about their inheritance patterns. To address these knowledge gaps, we conducted a comprehensive analysis of the genome-wide methylome using Reduced Representation Bisulfite sequencing in 130 individuals from eight families and two generations reared at low (16°C, LT) and high temperature (21°C, HT) during early development (12-60 days post fertilization). Thus fish were exposed zero (control), one (sires, ancestral; offspring, developmental) or two times (developmental and ancestral) to HT. We identified a range of genes and regions associated with masculinization and potential markers of genetic females resistant to masculinizing effects temperature. Thus, our results reveal multigenerational additive effects of temperature and provide insights into the epigenetic mechanisms explaining why females become masculinized while others do not and its potential heritability. Furthermore, we investigated the transmission of epigenetic marks from sires exposed to LT and HT during early development to their offspring. Our analysis demonstrated a large proportion (~85%) of CpGs with stable levels of methylation between sperm of the sires and gonads of the offspring. However, the proportion of temperature-induced marks transmitted to the subsequent generation was relatively low (~5%) suggesting the presence of stochastic source of variation. Overall, this study provides comprehensive insights into the interplay between sex, temperature, and epigenetic regulation in European sea bass. Similar mechanisms likely operate in other species- The findings have practical implications for both aquaculture and conservation biology and enhance our understanding of the mechanisms species have to cope with global warming.

Keywords: DNA methylation, sex reversal, neomales, epigenetic biomarkers, sex ratio, epigenetic inheritance.

1. Introduction

Epigenetics is the study of heritable changes in gene expression that occur without alterations to the DNA sequence, a field that contributes to unraveling the complexity of biological variability (Dupont *et al.*, 2009; Russo *et al.*, 1996). Thus, understanding the sources of variability in epigenetic patterns is crucial for understanding the diverse phenotypic outcomes observed and how a changing environment can affect them. Several factors contribute to the variability of epigenetic marks, including environmental influences, genetic factors, gene-environment interactions (G x E), and stochastic processes (Angers *et al.*, 2020; Massicotte *et al.*, 2011; Richards, 2008). Each of these sources plays a unique role in shaping the epigenome. Furthermore, the proportion of epigenetic marks that can be transmitted from one generation to the next and the mechanisms by which these occur remain unknown in many species.

The level of transmission of epigenetic marks can, in part, be related to epigenetic reprogramming (Ortega-Recalde and Hore, 2019). Epigenetic reprogramming is the process by which epigenetic marks derived from the parents are changed during early embryogenesis and gametogenesis, contributing to the development of the new individual (Jiang *et al.*, 2013). In mammals, it is well characterized by two waves of DNA demethylation: one occurring during early stages of embryo development, and the other during germ cell specification (Messerschmidt *et al.*, 2014; Santos *et al.*, 2002). Recently, in the mouse model, it was confirmed for the first time that DNA methylation of promoter-associated CpG islands can be transmitted from parents to offspring. Specifically, mouse that underwent DNA methylation editing through microinjection of methylated embryonic stem cells showed that acquired methylation was transmitted to offspring through the parental germ line, supporting the concept of epigenetic inheritance in mammals. Additionally, they found that heritable marks could be demethylated in PGCs and yet still be transmitted to the next generation through epigenetic memory mechanisms that are yet to be discovered (Takahashi *et al.*, 2023).

To date, limited research has been conducted on the epigenetic reprogramming of fish, with only a handful of species studied thus far. In zebrafish, there is a gradual transformation of the maternal methylome to resemble that of sperm, which facilitates zygotic genome activation. Unlike in other organisms, fish possess epigenetic marks necessary for germ cell specification that are inherited through the germplasm, rendering demethylation unnecessary in zebrafish (*Danio rerio*) (Jiang *et al.*, 2013; Potok *et al.*, 2013). In medaka (*Oryzias latipes*), however, a strategy similar to that occurring in mammals was found. Methylation reprogramming takes place in both early embryos and PGCs, raising the question of whether this species inherits epigenetic marks from the germplasm or not (Wang and Bhandari, 2019). In the mangrove killifish (*Kryptolebias marmoratus*), a self-fertilizing hermaphrodite fish, epigenetic reprogramming occurs at a later stage, and the entire reprogramming event is more extensive and lasts longer compared to other vertebrates (Fellous *et al.*, 2018). Overall, these findings suggest that epigenetic reprogramming in fish may be species-specific, raising questions about the mechanisms of epigenetic inheritance and to what extent epigenetic marks are transmitted to subsequent generations.

Additionally, regarding the transmission of epigenetic marks to subsequent generations, the available literature on the proportion of epimarks being transmitted in vertebrate species other than humans and mice is relatively limited. The majority of studies primarily utilize bacteria (Grandjean *et al.*, 1998), protists (Grimes and Aufderheide, 1991), fungi (Martienssen and Colot, 2001), plants (Henderson and Jacobsen, 2007; Miryeganeh and Saze, 2020) or invertebrates (Dorn *et al.*, 1993); benefiting from their ability to generate a large number of samples and generations (Jablonka and Raz, 2009). Alternatively,

simulated data has often been employed to model epigenetic inheritance (Ramakrishnan *et al.*, 2022; Slatkin, 2009; Varona *et al.*, 2015). Consequently, investigating epigenetic inheritance in animals with longer life cycles remains challenging. One of the few examples available is the study by Zheng *et al.* (2021), who quantified the proportion of epigenetic alterations in the sperm of the mouse caused by paternal exposure to long-term psychological stress through inter- (11.4%) and transgenerational (0.48%) inheritance of the environmentally-induced marks. Another example providing quantification of epigenetic marks being transmitted across generations was by Hu *et al.*, (2021). They quantified the proportion of marks stable across generations in the threespine stickleback (*Gasterosteus aculeatus*). They found that genetic variance contributed to explain an average of 24-35% of the methylation variance. Further, they identified the genetic architecture underlying two key CpG sites that were differentially methylated between the marine and freshwater ecotypes, suggesting an important role of DNA methylation marks being involved in the adaptation to different environments. Further, in a systematic review on epigenetic inheritance and reproductive modes of plants and animals compiling information on 500 studies, it was concluded that multigenerational inheritance is common in both plants and animals. However, numerous knowledge gaps were also highlighted: sources, sensitive windows, persistence, fitness consequences, and life history implications (Anastasiadi *et al.*, 2021b). Despite being in its early stages, the field of epigenetic inheritance is highly active. Its potential contribution to adaptation and evolution has sparked a growing interest in studying it across a wider range of species, higher number of generations, and diverse tissue types.

The species model of the present study was the European sea bass (*Dicentrarchus labrax*), a fish with mixed genetic and environmental sex determination (Vandeputte *et al.*, 2007). Elevated temperature during early development is well-known to affect sex ratios, causing masculinization (Blázquez *et al.*, 1998, 2009a). In aquaculture, this becomes a problem as this species exhibits female-biased sexual size dimorphism (SSD) (Faggion *et al.*, 2021; Vandeputte *et al.*, 2007) and rearing them at low temperature (LT) is much more expensive. Thus, it is an interesting model to study how epigenetics integrates information from the environment to result into different gene expression programs, phenotypes and ultimately being transmitted to subsequent generations. The effect of sex and temperature in this species is known to significantly affect two sex-related genes: Cytochrome P450 Family 19 Subfamily A Member 1, *cyp19a1a* (Navarro-Martín *et al.*, 2011) and Doublesex And Mab-3 Related Transcription Factor 1, *dmrt1* (Anastasiadi *et al.*, 2018b). Yet, from previous transcriptomic studies performed during sex development, we know that numerous genes are involved in the acquisition and maintenance of the sexual phenotype (Blázquez *et al.*, 2009; Blázquez and Piferrer, 2005; Díaz *et al.*, 2013; Díaz and Piferrer, 2015; Ribas *et al.*, 2019). However, it remains elusive how the methylome might be affected by sex and temperature at the genome-wide level and at other key sex-related genes in this species. In 2014, Shao *et al.*, (2014), revealed that the gonadal DNA methylation levels of genes involved in sex determination pathways were the major targets of substantial methylation modification during sexual reversal in the half-smooth tongue sole (*Cynoglossus semilaevis*).

In addition to understanding the regions involved in sex reversal, our aim was to identify epigenetic biomarkers (EBs) for distinguishing between genetic females that were resistant to temperature-induced masculinization and genetic females that underwent sex reversal (neomales, NM). EBs such as DNA methylation changes at specific loci (CpG), are susceptible to environmental changes and hold promise to complement genetic selection of broodstock (Alfalahi *et al.*, 2022; Crisp *et al.*, 2022; Kim *et al.*, 2022; Lempe *et al.*, 2022). Such tool could help select fish with desired phenotypic traits resulting from the interplay between genetics and the environment. In the sea bass, other EBs have been successfully developed such as those to predict sex (Anastasiadi *et al.*, 2018b), to assess degree of domestication (Anastasiadi and Piferrer, 2019), to determine age (Anastasiadi and Piferrer, 2020), to evaluate the effect

of heat waves (Anastasiadi *et al.*, 2021a), to assess the stress response associated with rearing density (Krick *et al.*, 2021) and to recapitulate past environments (Valdivieso *et al.*, 2023b).

Based on the above, this study therefore had three objectives: First, to identify genes and regions, other than *cyp19a1a* and *dmrt1*, affected by sex and temperature. We investigated whether there was an additive effect of epigenetic marks induced by temperature by comparing the methylome from fish exposed zero, one (sire or offspring) or two times (sires and offspring) to elevated temperature. Additionally, we aimed to identify genes and regions associated with masculinization. Second, to identify EBs to classify fish between neomales and resistant females reared at HT using an unsupervised machine learning algorithm to reduce the dimensionality of the methylome and identify the factors that contribute the most to the observed variance. Third, to identify and quantify the number of temperature-induced and inherited DNA methylation marks acquired in the sperm of the parental generation (F0) being transmitted to the gonads (testis and ovaries) of the unexposed filial generation (F1).

To accomplish these objectives, we crossed one female with eight males and reared each one of the resulting eight families (F0 generation) at different temperature treatments during their early development at the thermos-sensitive period: two at LT, three at high temperature (HT), and three neomales at HT, identified based on the genetic sex tendency (GST) (*i.e.*, the probability based on genetic factors to become a male or a female under the threshold model for sex determination (Hujoel *et al.*, 2020), which was previously estimated by Geffroy *et al.*, (2021). The offspring (F1), in turn, were also exposed to LT and HT during the same thermos-sensitive period, thus creating a total of 32 groups of F1 gonads for analysis: 8parents (of three sire types) x 2 temperatures x 2 sexes =32. The gonads were sampled when F1 fish were over one-year-old. Then, based on the genotypic (sex tendency) and phenotypic data of the offspring, we selected fish representative of each group by sex, temperature, and sire type. In total, we produced 130 libraries by Reduced Representation Bisulfite sequencing (RRBS) to analyze the DNA methylome of grandparents' gametes (F-1, n=2), sires and dam gametes (F0, n=9; gametes of the eight males and one female) and offspring juvenile gonads (F1, n=119). Overall, the objective of the analysis of the genome-wide methylome of 130 fish from eight different families and two generations was to understand the multigenerational temperature effects, the epigenetic regulation of masculinization and its inheritance in the European sea bass, contributing both to the development of tools for better aquaculture practices as well as in the understanding of the adaptation mechanisms to the environment, of particular interest in the current scenario of global warming.

2. Materials and Methods

Fish and rearing conditions

The fish used in this study were European sea bass from the West Mediterranean Sea strain. We used samples from three generations: the grandparents (F-1, which was never exposed), the parental generation (F0, which was the first generation exposed to temperature), and the filial generation (F1). The parental generation was created by artificial fertilization in March 2017 from the crossing of one male and one female (*i.e.*, the grandparents, F-1) with a known tendency to produce an approximately balanced sex ratio in the offspring when reared at LT, based on preliminary crosses (the male used was named sire d in Anastasiadi *et al.*, (2018). The hatched larvae (F0) were split into two groups at 13 days-post-fertilization (dpf): one group was reared at 16.5°C (LT), the control group, and the other group at 21°C (HT). Temperature treatments were applied between 13 and 65 dpf. This thermal regime has been shown to cause environmentally induced masculinization in previous (Anastasiadi *et al.*, 2018b; Blázquez *et al.*, 1998; Navarro-Martín *et al.*, 2009a, 2009b). Both experimental groups were replicated

in two tanks. Afterwards, temperature was raised to 21°C and from day 135 onwards, temperature was maintained at 20-22°C (more details on the rearing conditions of F0 can be found in Anastasiadi *et al.*, (2018). Once the individuals of F0 were two years old, sperm from two males reared at LT, three males and three neomales reared at HT was extracted. One aliquot of sperm was flash frozen and stored at -80°C for further analysis of the methylome and an aliquot was used for artificial fertilization of the next generation, F1. The eight males were crossed with the same female (not the same as in F-1). Their eggs were incubated at 14°C until 2 dpf and were then evenly distributed in six tanks of 500 L each. The temperature was gradually increased from 14°C to 16°C within 1 day. Larvae were exposed to the same temperature treatment as the parental generation: HT and LT groups in triplicates. The temperature treatment was applied to larvae of F1 from 8 to 62 dpf (**Supplementary figure 1**). Fish were reared at Ifremer Plateforme Expérimentale d'Aquaculture (Palavas-les-Flots, France) accredited to use and breed laboratory animals (no. C341926), and the project was approved by the Animal Care Committee no. 36 under project authorization no. APAFIS 19676. After 1 year, 1,256 F1 fish were euthanized using benzocaine (150 mg/L) and randomly sampled. From 476 fish, we measured standard length, weighed, and sexed them by in situ gonad examination. Gonads of the F1 generation were dissected, snap frozen and stored at -80°C until genomic DNA extraction. A total of 1,118 fish were genotyped, including the parents and the 476 offspring individuals for which the gonad tissue was kept. Genotyping was performed using the Thermo Fisher DlabChip European sea bass array of 57,000 SNP markers (Geffroy *et al.*, 2021; Griot *et al.*, 2021).

Sample selection for methylome analysis

To ensure that genome-wide methylome analysis was performed on samples representative of each group (two sexes, two temperatures and three sire types), we selected three individuals from each group based on multiple phenotypic traits: body weight, body length, and GST. First, we selected individuals with body weight and standard body length around the mean of the group. Additionally, we selected fish representative of the distribution of the GST of the group. The GST was estimated based on SNPs data for the same batch of fish in a previous study (Geffroy *et al.*, 2021). The GST can take a range of values around zero, usually from -1.5 to +1.5, where negative values are associated with genetic males and positive values are associated with genetic females (**Supplementary figure 2A**). Thus, we used estimated values of GST to select samples representative of each group. To do so, we calculated the mean and standard deviation of GST from each group. Then, we selected three individuals: 1) the closest sample to the mean, 2) the closest sample to the mean minus the standard deviation, and 3) the closest sample to the mean plus the standard deviation (**Supplementary Figure 2B**). By selecting these samples, we aimed to reduce the potential selection of outliers and to increase the likelihood that our findings accurately reflected the characteristics of each experimental group.

RRBS Library Preparation

Genomic DNA was isolated using the Qiagen Blood and Cell Culture DNA Mini kit (catalog no. 13323) according to the manufacturer's instructions. Quality of DNA was checked with NanoDrop ND-1000 (Thermo Fisher Scientific), and concentration was measured by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). RRBS libraries were prepared using the Premium RRBS kit v.1 (Diagenode, C02030032) according to the manufacturer's instructions. The final concentration of the libraries was measured and their quality checked using the DNA High Sensitivity Assay (Agilent). In total, 130 RRBS libraries were produced (the gametes of the dam and the sire of the F-1, eight samples of sperm from sires of the F0 and the unfertilized eggs of the F0 female and 119 F1 gonad samples: three replicates x two sexes x two temperatures x eight sires, which result in a total of 96 samples (except one female at

HT that was not available, and hence we had 95 samples), plus 24 neomales (three replicates x eight sires x one temperature, HT). The neomales were identified based on the estimated GST value (*i.e.*, males that, nevertheless, had positive GST values). Sequencing of the libraries was performed using Illumina HiSeq 2500 instrument at single-end and read length of 50bp.

Bioinformatic analysis

Quality control of reads were performed using the FastQC software (v. 0.11.9) (Wingett and Andrews, 2018). Trimmomatic software (v.0.36) (Bolger *et al.*, 2014) was used for trimming the adaptors with parameters: sliding window 4, Phred score > 20, and adaptive quality trimming with largest length of 20 bp. Trimmed reads were aligned to the reference genome using the BSMAP software (v. 2.90) (Xi and Li, 2009) with RRBS mode and filtering for at least 10× coverage. In order to extract methylation data only from CG context (CpG) we used the methratio.py function from BSMAP. The bisulfite conversion efficiency of the libraries was calculated following the instructions from the RRBS kit. Stringent criteria were implemented to ensure high-quality data by enforcing a minimum alignment percentage (>85%) and a robust bisulfite conversion rate (>98%). Any samples failing to meet these criteria were excluded from the analysis. Detailed information on the reads, sequencing and control quality statistics for all samples are listed in **Supplementary table 1**. After the bioinformatic analysis all the statistical analysis described in the following sections were performed using R software (v. 4.2.2) (R Core Team, 2022) and Rstudio (v. 2021.9.0.351) (RStudio Team, 2021).

Differentially methylated cytosines (DMCs)

The *methyKit* package (v. 1.14.2) (Akalın *et al.*, 2012) was used to calculate differentially methylated cytosines (DMCs). CpGs with >10 x coverage or with coverage less than 99.9% of the distribution of read counts were kept for further analysis. Then, coverage was normalized across all samples for each comparison. To detect DMCs a logistic regression model with the family information as covariate using the Chisq-test and with the following criteria: methylation differences larger than 15% and *q*-value < 0.01 adjusted from the SLIM method (Wang *et al.*, 2011). The *genomation* package (v.1.30.0) (Akalın *et al.*, 2014) was used to annotate the DMCs to the reference genome annotation file. Finally, the *GenomicRanges* package (v. 1.38) (Lawrence *et al.*, 2013) was used to identify genes that contained at least one DMC within their gene body region $\pm$ 5000 bp.

Differentially methylated regions (DMRs)

DMRs capture the methylation changes within a genomic region, usually providing more biologically relevant information than isolated changes at individual cytosine sites (Gaspar and Hart, 2017). The size of the regions is defined by dividing the genome into non-overlapping tiles or windows of a fixed size. Usually, the size is chosen making a compromise to balance the resolution of methylation patterns and the statistical power to detect significant differences. To choose a region size based on the nature of the data obtained from the sea bass instead of using, for example the most commonly used window length (100 bp), we used the autocorrelation algorithm previously developed to study epimutation rates in *Arabidopsis thaliana* (Denkena *et al.*, 2021). Briefly, this approach initiates by identifying pairs of cytosines that are at the minimal genomic distance from each other. Subsequently, the correlation between methylation levels of the two closest positions is calculated. This iterative process continues by increasing the distance, and hence determining the correlation with increasing number of CpGs until the correlation drops (**Supplementary figure 3**). Thus, we chose to specify the regions every 200 bp without overlap covering at least 4 CpGs. Of note, the median of CpGs included per region was eight. At 200 bp most of the samples showed correlation of CpGs more than 0.75 and after that, a larger drop

in correlation was observed. In this manner, the methylation level obtained was the mean of all reads available for the region. This provides robustness to the analyses. Once the length of the regions was established, we performed two sets of comparisons. The first set included four comparisons of the methylome of ovaries *vs.* testis of F1 individuals: 0, Control (LT); 1, Developmental (HT); 1', ancestral (exposed sires at HT); and 2, developmental and ancestral (F0 and F1 reared at HT) (**Figure 1A**). The second set included five comparisons between the different temperature groups of the F1, for each sex separately. DMRs were defined to when the methylation difference was higher than 10% under a significance threshold of q -value < 0.05 .

Identification of masculinization-associated differentially methylated regions

We performed the principal component analysis (PCA) of the methylation data from all the F1 individuals at group 0. Control (LT), group 1, developmental (HT), group 1', ancestral (exposed sires at HT) and group 2. developmental and ancestral (F0 and F1 reared at HT) using *MethylKit* (v.1.24.0) (Akalin *et al.*, 2012), *ggplot2* (v.3.4.2) (Wickham *et al.*, 2021) and *BiodiversityR* (v.2.15.1) (Kindt and Coe, 2005). The PCA revealed F1 ovaries of females exposed at group 1', ancestral (exposed sires at HT) and group 2, developmental and ancestral (F0 and F1 reared at HT) showing three clusters: some females were clustered at the maximum distance from the males (resistant females, F), some were clustered in the middle (transition 1) and one female overlapped with the males and neomales (transition 2). Thus, to identify the regions associated with masculinization we determined the DMRs between transition 1 females with the females that resisted the most the effect of temperature (cluster at similar position of the females reared at LT). Then we determined the DMRs between the F1 neomales (masculinized genetic females) and the F1 females at the transition 1 cluster. The DMRs ($>10\%$ methylation difference, q -value 0.01) that were found consistently hyper- or hypomethylated from resistant females towards neomales were defined as Masculinization-associated regions (MA-DMRs).

Multi-omics factor analysis (MOFA)

In order to identify biomarkers of resistant females to HT we took advantage of an unsupervised machine learning algorithm to reduce the dimensionality of the DNA methylation data in order to reveal underappreciated axes of variation. For this, we used the Multi-omics factor analysis (MOFA) (Argelaguet *et al.*, 2018). To identify epimarks that contribute to the separation of genetic females that were masculinized from females that resisted the masculinizing effect of temperature, we selected the F1 females and the F1 neomales (from both exposed and unexposed sires, $n=45$). Then we performed several data preprocessing steps and normalization.

To account for tissue-specific differences between females (ovary) and neomales (testis), a filtering step was implemented. Specifically, we excluded all DMCs between genetic males (not neomales) and females with a methylation difference of at least 15% and a q -value below 0.01. By doing so, we aimed to remove CpGs that exhibited tissue-specific inherent differences unrelated to the masculinization process. Next, we constructed a comprehensive matrix comprising all available CpGs for each sample with the methylation data transformed to M-values (*i.e.*, standard transformation of DNA methylation data as $\log_2$ ratio of methylated *vs.* unmethylated reads). This matrix was designed to incorporate multiple layers of information that would be used to train the algorithm and facilitate subsequent analyses. By incorporating diverse sources of information, we aimed to capture a more comprehensive and nuanced representation of the epigenetic landscape associated with masculinization. Then, we filtered the 10,000 CpGs containing the most variable methylation levels across all the samples included for the following layers: promoters, introns, exons and CpG islands. After these pre-processing steps, we had 10,000 features for each of the four different layers of DNA methylation data. Then we trained

the MOFA model using the R/Bioconductor package MOFA2 (version: 1.2.2) with the default model and training parameters (Gaussian distribution, maxiter of 1000 and convergence assessed using the ELBO statistic at “slow” mode, with four factors) (Argelaguet *et al.*, 2020; Velten *et al.*, 2022).

Environmentally induced and inherited epimarks

When studying the levels of methylation between generations, we defined three levels to be considered: stable, environmentally-induced and inherited CpGs (using standard statistical analysis of 1 permutation), plus environmentally-induced, inherited and not random CpGs (using permutation tests to remove stochastic effects). In the first level, to determine the stable CpGs, we used similar criteria than in previous studies to allow for comparison (Hu *et al.*, 2021). Specifically, we considered stable CpGs when not differentially methylated (<15% methylation difference, q -value <0.01). In the second level, in order to identify the transmitted DNA methylation changes induced by the exposure to high temperature during early development in the sires, we determined the DMCs between sires (F0) at LT and HT. Then, we determined the number of DMCs of the F1 gonads of individuals that were reared at LT from unexposed vs. LT from exposed sires, for each sex separately. Then, we checked the overlap of DMCs between LT and HT at the F0 sperm with the DMCs between LT F1 offspring from exposed and unexposed sires. In the third level, we employed the *MethylInheritance* package (v. 1.22.0) (Belleau *et al.*, 2018) to evaluate the number of conserved DMCs between HT and LT in the parental generation (F0) with those of the filial generation (F1) at LT that are associated with an inherited environmental effect rather than being attributed to random chance. This analysis utilized a Monte Carlo algorithm. Specifically, we examined whether the DMCs identified in the F0 sperm samples between LT and HT groups were transmitted to both F1 LT testis and F1 LT ovaries separately. To assess the statistical significance of the observed number of conserved DMCs, we performed 4000 permutations for each sex. The purpose of these permutations was to generate a null distribution by randomly shuffling the sample labels. By comparing the observed number of conserved DMCs to this null distribution, we aimed to reject the null hypothesis ($P < 0.05$), which states that the number of conserved DMCs could have been obtained randomly.

Annotation and enrichment analysis of DMCs and DMRs

For the annotation procedures of the DMCs and DMRs we used the COMBINED ANNOTATION track of the European sea bass genome (dicLab v1.0c, July 2012). For the enrichment analysis we retained only the gene bodies plus regions 5,000 bp upstream of the transcription start site and 5,000 bp downstream of the 3' UTR that contained at least one DMC or DMR. The over-represented Gene Ontology (GO) categories in the biological processes were determined through GO term enrichment analysis using the PANTHER tool (Mi *et al.*, 2017). For this analysis, the background set exclusively included feasible genes, which were the genes captured by RRBS for each comparison, in accordance with the recommendation by Timmons *et al.*, (2015). Human gene database was used as the reference set, and multiple testing correction was applied. Additionally, the PANTHER tool was utilized to map the genes to relevant pathways.

3. Results

Phenotypic traits

Sampling of the gonads of the F1 generation occurred when the fish reached one year of age. Morphologically, all the fish displayed normal gonads, with no discernible differences observed across testis. Among the females, a subset exhibited a higher level of adipose tissue compared to others.

However, detailed histological analysis indicated the absence of intersex characteristics in all fish examined (**Supplementary figure 4**). Regarding body weight, F1 individuals reared at HT exhibited significantly higher weight (99.0 ± 36.2 g) compared to their half-siblings reared at LT (82.13 ± 24.9 g) during early development ($P < 0.05$) (**Supplementary figure 5A**). Females, showed higher weight, particularly those from the HT treatment from the unexposed sires and from neomales ($P < 0.01$) (**Supplementary figure 5B**). No significant differences in mean weight were found between genetic males (95.3 ± 38.5 g) and neomales at HT (93.7 ± 14.2 g). A similar pattern was observed for body length, with the largest females found in the HT treatment (195.69 ± 21.7) and the smaller fish being males reared at LT (184.39 ± 16.6) (**Supplementary figure 5C and D**). No differences between genetic males (192.9 ± 22.9) and neomales (193.9 ± 9.6 cm) were found. Overall, sex, temperature and sire type influenced body weight and body length but without the interaction among these three factors (ANOVA test, $P < 0.01$).

All fish were unambiguously assigned to their respective sires through genotyping. Analysis of the sex ratio among the 476 F1 fish revealed, as expected, a significant increase in males reared at HT (LT = 44.5%; HT = 72.6%; $P < 0.001$) (**Supplementary figure 5E**). When analyzing the sex ratio based on sire type (genetic males reared at LT, genetic males reared at HT, and masculinized genetic females reared at HT), a significant increase in males was observed among the F1 fish reared at HT ($P < 0.001$) (**Supplementary figure 5F**). For instance, F1 fish from LT sires displayed a sex ratio of 53.3% when reared at LT and 77.3% when reared at HT. Interestingly, irrespective of the temperature treatment and the observed increase in males at HT, neomale sires produced less males both at LT (34.6%) and HT (67.7%) compared to LT and HT sire types (53.3% and 45.1% at LT, respectively; and 77.3% and 77.5% at HT, respectively). However, such differences were not significant when comparing sire types within the same temperature treatment.

Global methylation patterns

As expected, the global mean methylation was tissue-specific and with similar levels previously reported for the sea bass. The global mean methylation of the F0 was 93.8% in the sperm and 86.1% in the unfertilized eggs. In the F1 it was 92.2% in the testis and 84.4% in the ovaries. When comparing the four different temperature treatments in F1 (group 0, control; group 1, developmental; group 1', ancestral and group 2, developmental and ancestral) there were no significant differences at the global level of the methylome between half-siblings, neither in males (**Supplementary figure 6A**), nor in females (**Supplementary figure 6B**).

Effect of sex and temperature in the F1

When examining the gonads of the F1 fish, the largest differences were found when comparing ovaries with testis regardless of the temperature treatment. At LT (ovaries vs. testis at 0. control) we found a total of 5,619 DMRs, of which 94% were hypomethylated in ovaries. The hypermethylated DMRs were mostly found at coding gene regions: 15.5% at the promoter, 7.3% at the first exon, 29.9% in the rest of exons, 10.1% in the first intron, and 36.6% in the rest of introns. However, most hypomethylated DMRs (94%) were found at intergenic regions (71.3%) (**Supplementary figure 7A**). In the comparison of ovaries vs. testis at LT (0. control), the pathways with more mapped hypermethylated genes were the gonadotropin releasing hormone (GnRH) receptor pathway, the Wnt signaling pathway, the Heterotrimeric G-protein signaling, and the Integrin signaling and fibroblast growth factor (FGF) signaling pathway. Among the hypomethylated pathways we found the same pathways together with others such as the chemokine and cytokine signaling pathway, the CCKR PDGF cadherin pathway, and the TGF-beta and EGF receptor signaling pathways (**Supplementary figure S8A**). Furthermore, when

comparing ovaries vs. testis at different temperature treatments, the effect of temperature exposure (group 1, Developmental; group 1', ancestral and group 2, developmental and ancestral) resulted in a gradual increase of the number of DMRs (7,049, 7,603 and 8,432, respectively) (**Figure 1B**).

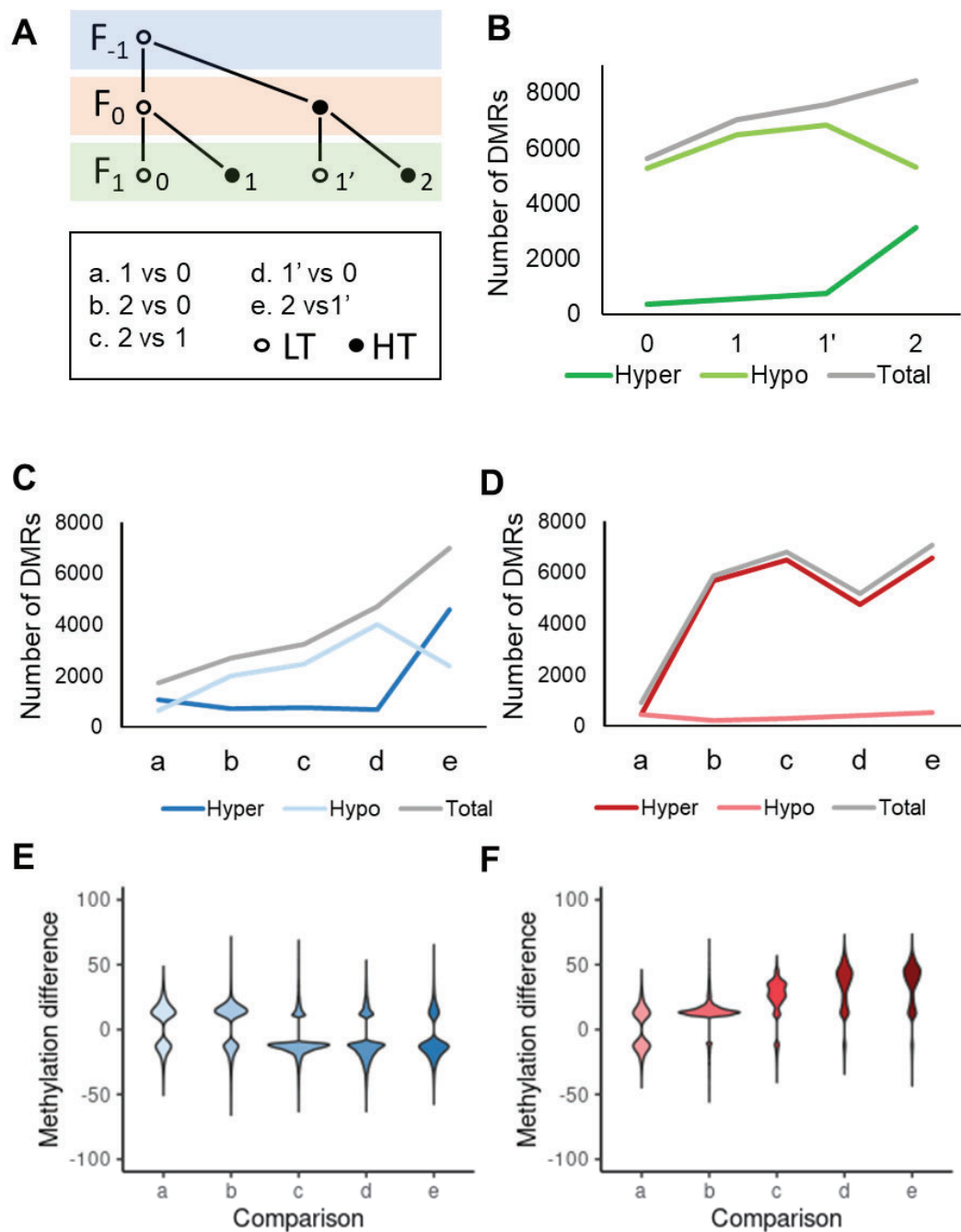


Figure 1. Effect of sex and temperature in the gonads of the European sea bass. LT: low temperature, HT, High temperature. **A.** Experimental design and groups of different temperature treatment: group 0, control; group 1, developmental, reared at HT during early development; group 1', ancestral, F0 at HT and F1 at LT; group 2, combined developmental and ancestral, F0 at HT and F1 at HT. The letters a–e indicate the five comparisons chosen for analysis of the temperature effect for each sex separately. **B.** Number of differentially methylated regions (DMRs) between ovaries and testis for each temperature treatment. **C.** Number of DMRs for the five selected comparisons to study the temperature effect on F1 testis. **D.** Number of DMRs for the five selected comparisons to study the temperature effect on F1 testis.

The analysis of DMRs between groups at different temperature treatments of the same sex, revealed a sex-specific effect. In males, the comparison 'a. 1 vs. 0' (HT vs. LT from unexposed sires) resulted in 1,730 DMRs, of which 61% were hypermethylated (**Figure 1C**). In this comparison, approximately half of the DMRs (both hyper- and hypomethylated) were found within gene regions: ~10% in the promoters, 4.5% in the first exon, ~17% in the rest of exons, 12% in the first intron and 30% in the rest of introns (**Supplementary figure 7B**). Some of the genes with most hypermethylated mapped genes were GnRH receptor pathway, Integrin, and chemokine and cytokine signaling, while the pathways with hypomethylated genes were Wnt signaling, angiogenesis and cholecystokinin receptor (CCKR) signaling pathways (**Supplementary figure 8B**). In females, the comparison 'a. 1 vs. 0' (HT vs. LT from unexposed sires) yielded 909 DMRs, of which 51.2% were hypermethylated (**Figure 1D**).

At the comparison 'b. 2 vs. 0' (HT from exposed sires vs. control), males showed 2,709 DMRs, mostly hypomethylated (73.8%), whereas females had a larger number of DMRs (5,877), predominantly hypermethylated (83.3%). This pattern was consistent in the comparisons 'c. 2 vs. 1' (HT vs. LT from exposed sires) and 'd. 1' vs. 0' (LT from exposed sires vs. LT from unexposed sires) (**Figure 1C and D**). The offspring from exposed sires exhibited the highest number of DMRs when comparing 'e. 2 vs. 1' (HT vs. LT from exposed sires). Males had 7,002 DMRs (65.7% hypermethylated), while females had 7,093 DMRs, of which 96.6% were hypermethylated.

The effect of temperature was not just limited to a higher number of DMRs, but it also had a notable impact on the magnitude of the percentage methylation differences. For instance, in the comparison 'a. 1 vs. 0' (HT vs. LT from unexposed sires), the differences ranged from over 15% to 50% in absolute value (**Figure 1E**). However, in the comparison of testis 'b. 2 vs. 0' (HT from exposed sires vs. control) the range of percent methylation differences increased to 65%. This increasing trend was particularly noticeable in the ovaries (**Figure 1F**). Although a similar range of methylation differences (15% to 65%) was observed, in the comparison 'b. 2 vs. 0' (HT from exposed sires vs. control), most differences were around 15%. However, in comparisons 'd. 1' vs. 0' (LT from exposed sires vs. LT from unexposed sires) and 'e. 2 vs. 1' (HT vs. LT from exposed sires), the majority of DMRs exhibited a higher methylation difference, around 45% (**Figure 1F**).

Masculinization-associated regions

For visualization of the methylomes we produced a series of PCAs based on the DNA methylation level of all the CpGs included in the analysis (*i.e.*, CpGs with more than 10 reads for at least two of the samples per group, including ~700,000 CpGs). The PCA of the sperm samples did not exhibit distinct clustering based on temperature treatment. However, the neomales reared at HT exhibited overlap with genetic males reared at HT and were positioned at negative values along the first component. Conversely, genetic males reared at LT were depicted at positive values along the first component. (**Figure 2A**). The PCA of F1 samples revealed that the first component accounted for most of the variability and effectively distinguished males and females, emphasizing notable differences between the methylomes of testis and ovaries. The first principal component exhibited a separation of 70.2% in offspring at LT (**Figure 2B**) and over 94% in F1 at 1. developmental (**Figure 2C**), group 1', ancestral (**Figure D**), and 2. developmental and ancestral (**Figure 2E**). Interestingly, in the PCA of F1 exposed at 2. developmental and ancestral, some females were positioned midway between the male and female clusters. These particular females will be referred to as females at transition stage 1 (T1). This observation suggests the presence of intermediate methylation patterns in these individuals, likely associated with a mechanism of masculinization. Additionally, there was one female sample that exhibited complete overlap with the male cluster, further supporting the hypothesis that DNA methylation may regulate the process of masculinization. This female will be referred to as female at

transition stage 2 (T2). Furthermore, when examining the global methylation percentage in females grouped based on those observed PCA clusters, a significant gradual increase in methylation was observed (Wilcoxon signed-rank test, $P < 0.05$). The most resistant females (F) exhibited a methylation level around 76.7%, females at T1 showed a global methylation of 81.2%, while females at T2 had a global methylation level of 84.8%, similar to that of neomales (87.9%) and males (87.6%) (**Figure 3A**).

To further understand what regions were most involved in the gradual increase of methylation observed at the global level, we filtered the 4,319 DMRs consistently hypermethylated between resistant females and females at T1 as well as in neomales (*i.e.*, hypermethylated DMRs in T1 vs. F AND hypermethylated in NM vs. T1) (**Figure 3B**).

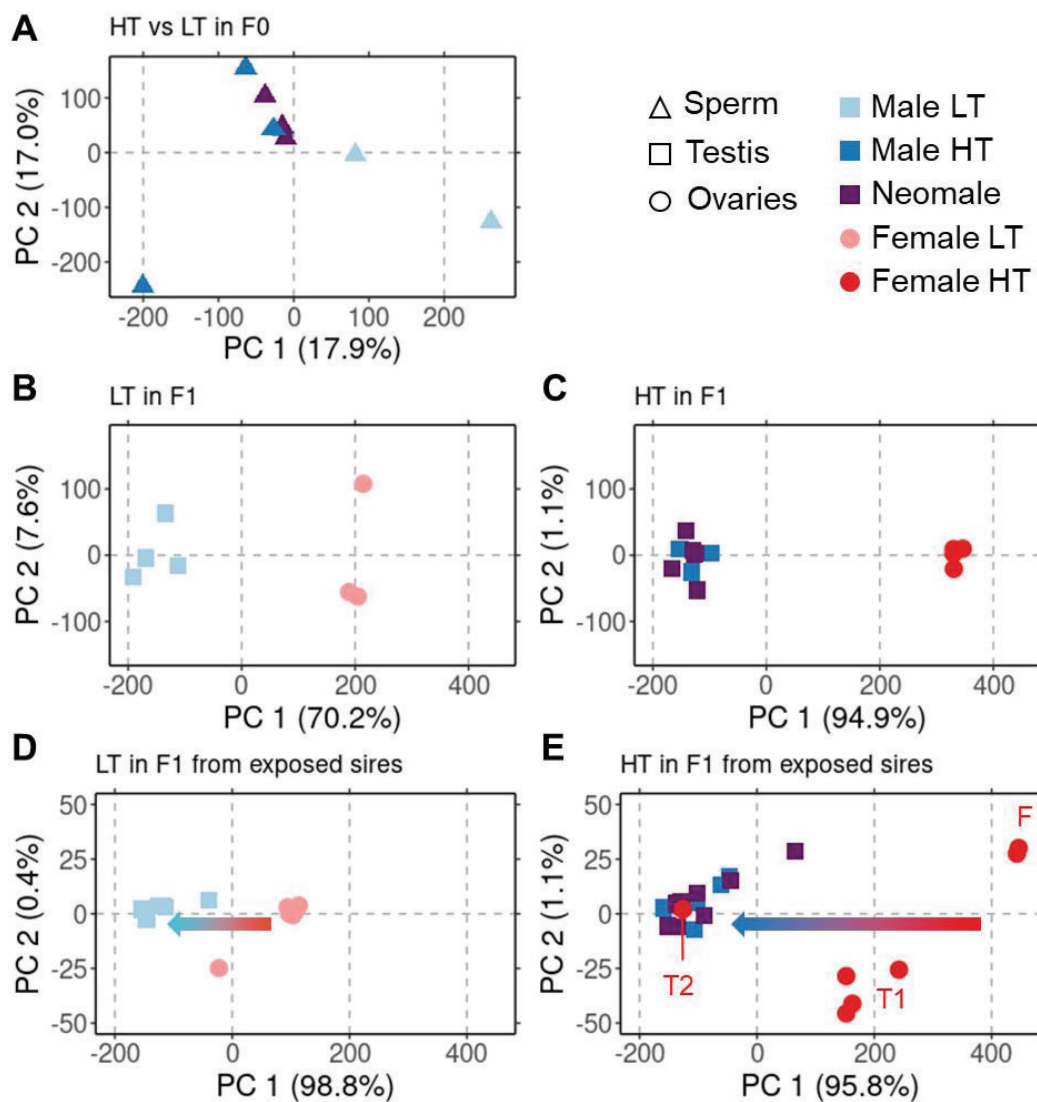


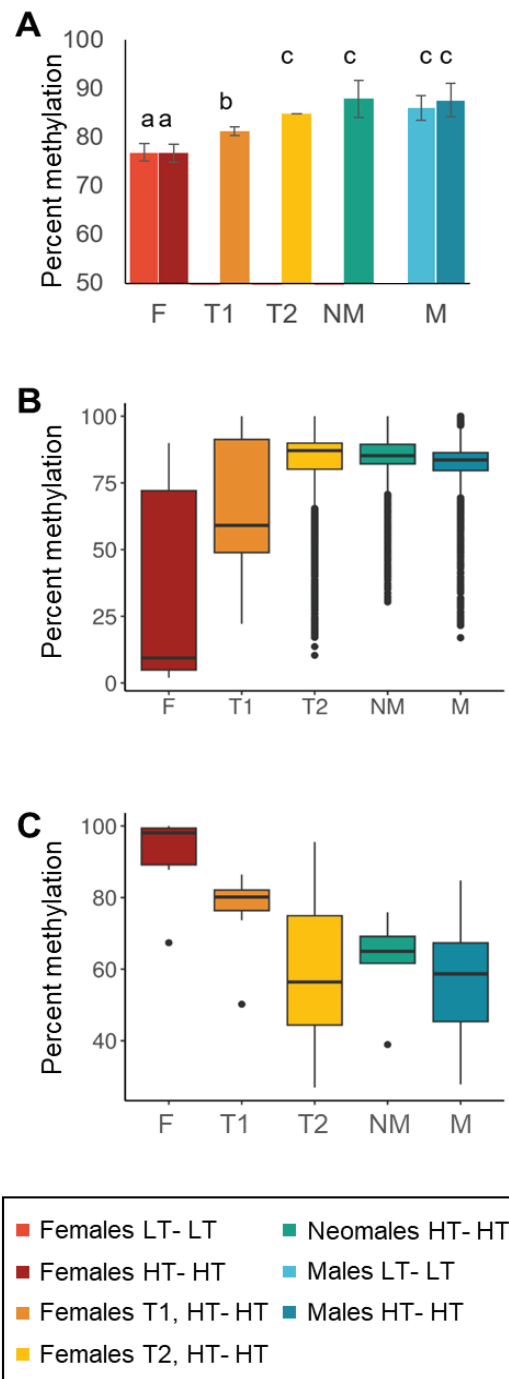
Figure 2. Principal component analysis (PCA) of the DNA methylome from samples analyzed with reduced representation bisulfite sequencing. **A.** PCA of the sperm samples from the eight sires at the parental generation (F0) at low temperature (LT) and high temperature (HT). **B.** PCA of testis and ovaries of the filial generation (F1) at LT. **C.** PCA of testis and ovaries of the filial generation (F1) at HT. **D.** PCA of testis and ovaries of the filial generation (F1) at LT from exposed sires at HT. **E.** PCA of testis and ovaries of the filial generation (F1) at HT from exposed sires at HT. F, females; T1, females at transition stage 1; T2, female at transition stage 2.

Figure 3. Gradual increase of methylation percentage between ovaries of females reared at high temperature (HT) showing a transition towards neomale and male methylome. **A.** Bar plot of the global mean methylation of F1 ovaries (F) at low (LT) and HT, ovaries at transition stage 1 (T1) and transition stage 2 (T2), neomales (NM) and males (M) at LT and HT. Data as mean $\pm$ standard deviation. **B.** Percent methylation of the 4,319 masculinization-associated differentially methylated regions (MA-DMR) that were consistently hypermethylated between resistant females and females at T1 as well as in neomales. **C.** Percent methylation of the 77 MA-DMRs that were consistently hypomethylated between resistant females and females at T1 as well as in neomales. In B and C, the boxes include the values distributed between the lower and upper quartiles, the upper whisker = $\min(\max(x), Q3 + 1.5 * IQR)$, the lower whisker = $\max(\min(x), Q1 - 1.5 * IQR)$, where $IQR = \text{third quartile (Q3)} - \text{first quartile (Q1)}$, the black horizontal line in the box indicates the median and the points outside the boxes represent values higher than the upper whisker. The values represented are from DMRs with a methylation difference of at least 10% and q -value < 0.05 .

Regarding the consistently hypomethylated regions, 77 regions were identified (**Figure 3C**). Interestingly, the hypermethylated MA-DMRs were mostly, 78.7%, found at intergenic regions and 11.3% in promoters, while the hypermethylated MA-DMRS were enriched at exons (37.7%) and introns (16.1%) (**Supplementary figure 7C**).

Furthermore, up to 260 coding genes overlapped with MA-DMRs, of which 240 overlapped with hypermethylated DMRs (**Supplementary table 2**). For instance, one of the genes containing the highest number of DMRs (with up to 55 CpGs covered) was the gene RNA 18S Ribosomal 1 (*18SrRNA*) (**Figure 4A**). This gene clearly showed a gradual increase of methylation in the transition stages from resistant females towards neomales (**Figure 4B**). Another gene overlapping with four MA-DMRs and covering 79 CpGs, was the gene POP4 Homolog, Ribonuclease P/MRP Subunit (*pop4*) (**Figure 4C and D**). Additionally, both genes also overlapped with CpG islands.

The enrichment analysis of the 260 genes overlapping with MA-DMRS revealed physiological processes and mechanisms affected during temperature-induced masculinization. The biological processes category showed significant enrichment of Gene Ontology (GO) terms in genes with hypermethylated DMRs, including synaptic vesicle endocytosis (GO:0048488), response to DNA damage stimulus (GO:0006974), apoptotic signaling pathway (GO:0097190), mannosylation (GO:0097502), and regulation of potassium ion transport (GO:0006813) (**Supplementary figure 9A**).



On the other hand, the 77 genes overlapping with hypomethylated MA-DMRs through the transition from resistant females to neomales, were genes involved in the regulation of trans-synaptic signaling (GO:0099537), modulation of chemical synaptic transmission (GO:0050804), nucleoside diphosphate metabolic process, cholesterol metabolic process (GO:0009132) and maturation of the large subunit of ribosomal RNA (LSU-rRNA; GO:0000470) (**Supplementary figure 9B**). Additionally, the MA-DMRs overlapped with genes found on some of the following pathways: Wnt signaling, Integrin signaling, GnRH receptor pathway, CCKR signaling and p53 signaling pathway (**Supplementary figure 8C**).

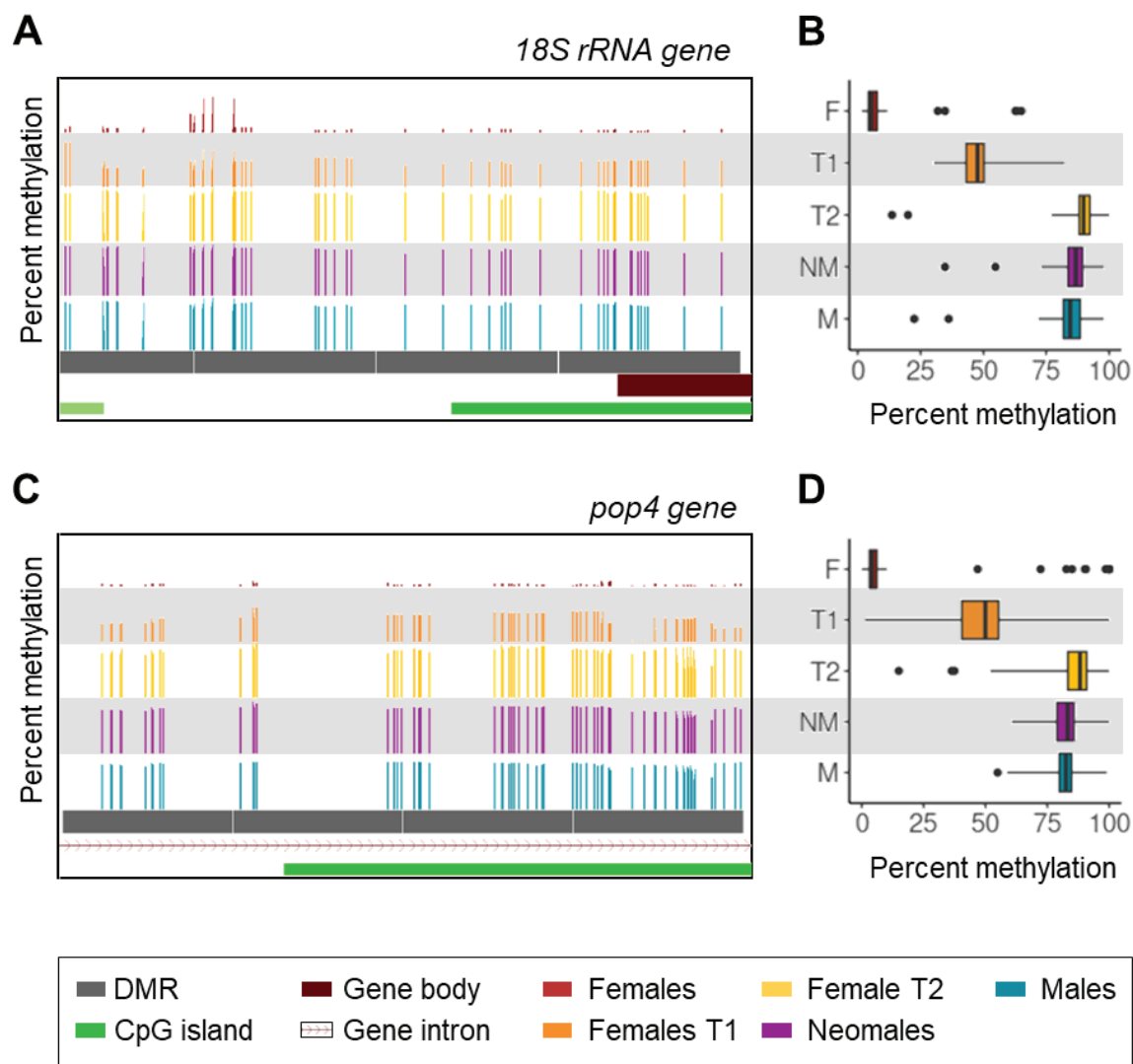


Figure 4. Levels of methylation percentage in masculinization-associated regions (MA-DMRs) found at RNA 18S Ribosomal 1 (*18SrRNA*) and POP4 Homolog, Ribonuclease P/MRP Subunit (*pop4*) genes in the five types of gonad samples observed in fish of the filial generation (F1) from exposed sires at high temperature (HT) and reared at HT. The vertical lines indicate the methylation level of cytosines of: most resistant females, F (red); females at transition stage 1, T1 (orange); female at transition stage 2, T2 (yellow), neomales, NM (purple) and males, M (blue). The dark gray boxes depict differentially methylated regions (DMRs). The annotated gene body (box) or the intron regions (line with arrows) are indicated in dark red below the DMRs and the green box indicates the presence and position of CpG islands. **A.** DNA methylation profile of *18SrRNA* gene. **B.** Boxplot of the methylation levels within the shown region (approximately 800 bp) of the *18SrRNA* gene. **C.** DNA methylation profile of *pop4* gene. **D.** Boxplot of the methylation levels within the shown region (approximately 800 bp) of the *pop4* gene. The list of other genes overlapping with MA-DMRs are presented in **Supplementary table 2**.

Genetic sex tendency

The estimated GST of the F1 fish was obtained from a previous study (Geffroy *et al.*, 2021). In the present study, we carefully selected a smaller set of samples representative of each group based on body weight, body length and GST for the analysis of the methylome at the genome-wide level. The distribution of the GST was studied on the group of the 119 F1 selected gonads. The overall distribution was normal and, as expected, the mean of the genetic males was negative (-0.32) and significantly different from the GST mean of the neomales (0.56, $P < 0.001$) and the females (0.26, $P < 0.05$) (**Figure 5A**). Interestingly, there was no significant correlation between the GST and the global methylation of the genetic females (*i.e.*, both resistant females and neomales) exposed at HT two times (**Figure 5B**).

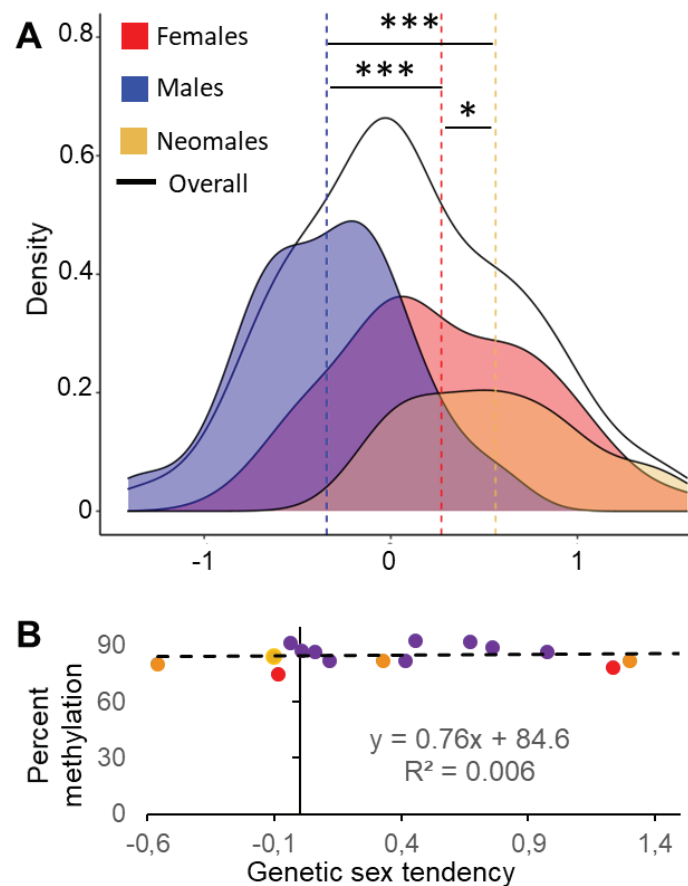


Figure 5. Estimated genetic sex tendency (GST) of the filial generation (F1). **A.** Density plot showing the distribution of estimated GST for each F1 individual from which the methylome were studied. The overall (black) distribution is normal. The mean of the genetic males (blue) is negative (-0.32) and significantly different from the GST mean of the neomales (0.56; yellow) and the females (0.26, red); *** $P < 0.001$; * $P < 0.05$. **B.** Scatterplot of the GST (x-axis) and the global methylation (y-axis) of F1 gonads of fish from group 2 (F0 and F1 exposed at high temperature). The neomales are shown in purple and the females are shown as: most resistant females (red), females at transition stage 1 (T1, orange) and the female at transition stage 2 (T2, yellow). There was no significant correlation between the GST and the global methylation of the genetic females exposed at HT from exposed sires.

Methylation levels of key sex-related genes

To better understand the effect of sex and temperature in the gonads of the sea bass, we checked how many out of 42 key genes previously known to be involved with sex (development and/or maintenance) in sea bass (Ribas *et al.*, 2019, 2017), were captured in the RRBS libraries. Additionally, we checked

whether these genes overlapped or not with DMCs in the comparison 0. control in F1 females vs. males, at the gene body $\pm 5,000$ bp of the transcription starting site (TSS). We found that seven of these genes met the established criteria: follicle stimulating hormone receptor (*fshr*), SRY-box transcription factor 9a (*sox9a*), Wnt family member 4 (*wnt4*), hydroxysteroid 17-beta dehydrogenase 10 (*hsd17 β 10*), estrogen receptor alpha (*era*) and prostaglandin-endoperoxide synthase 2 b (*ptgs2b*). We found that for the genes *fshr*, *sox9a*, *wnt4* and *hsd17 β 10* the levels of methylation were high (>80%) at F0 gametes, except for the levels of *fshr* and *wnt4*, which were hypomethylated in unfertilized eggs (**Figure 6**).

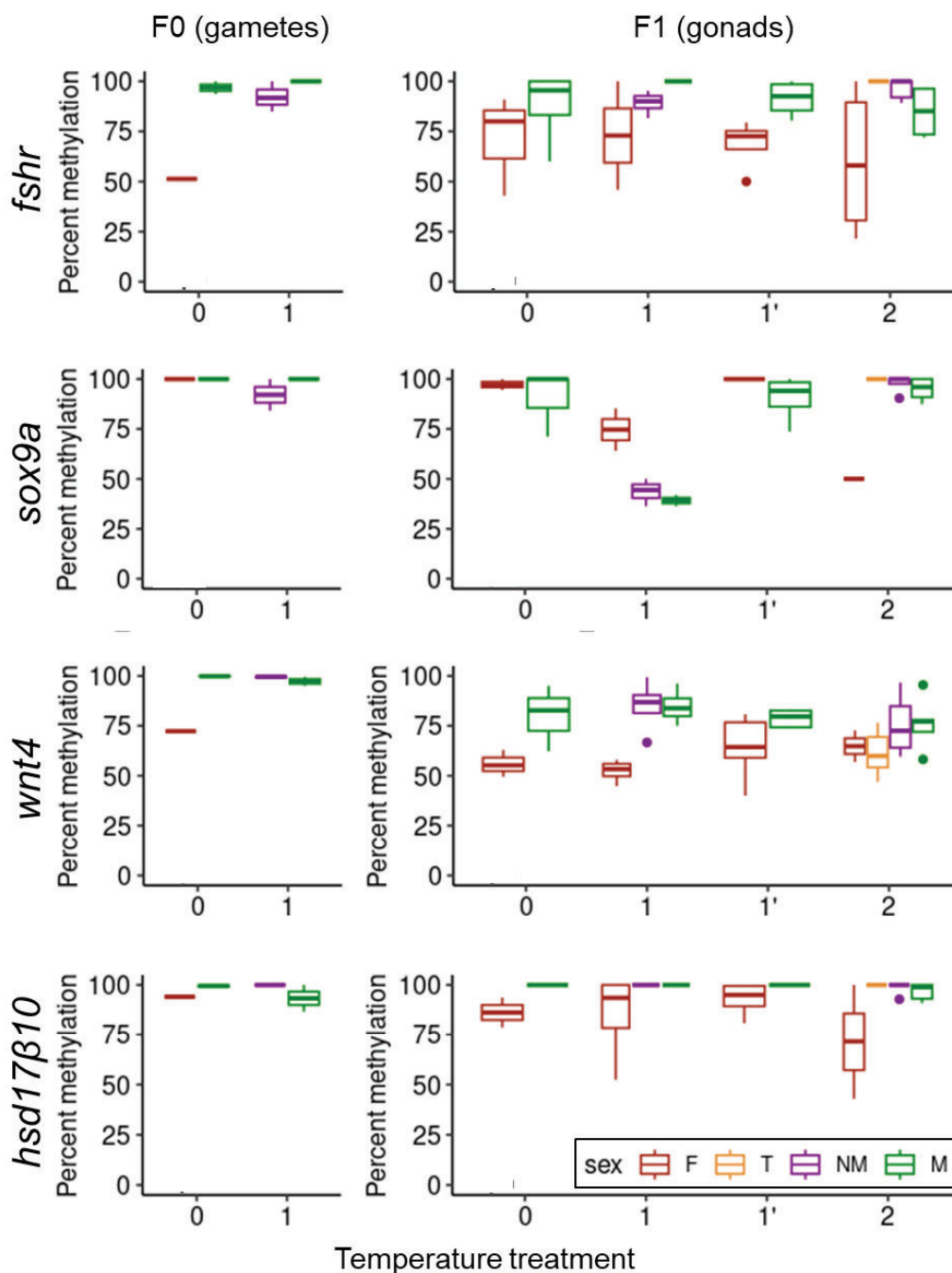


Figure 6. DNA methylation levels of key sex-related genes examined in the gametes of the parents (F0) and the gonads of the offspring (F1). In the left, the DNA methylation in the unfertilized eggs of the dam (red), the sperm of sires (green) and the sperm of the neomale sires (purple) at 0. control and 1. developmental temperature treatment are shown. The right part of the figure illustrates the DNA methylation levels in the F1 gonads at each temperature treatment (0. control and 1. developmental, 1'. ancestral and 2. developmental and ancestral) for ovaries (red), testis (green), females at transition stages (T1 and T2, orange) and neomales (purple).

The levels of methylation in *fshr* were hypomethylated in females and showed a decreasing trend as the number of temperature exposures increased (1, 1', 2). As for *fshr* and *hsd17 β 10*, the fish exposed two times to HT showed hypomethylation in resistant females while in the females at transition stages (T1 and T2) the levels were similar to males and neomales ($P < 0.01$). In *sox9a*, F1 females showed hypomethylation at 0. control and at group 2 developmental and ancestral. At group 1, developmental, the females showed hypermethylation in respect to males and neomales through a decrease of methylation in the males, while no significant differences between males and females were found at group 1', ancestral. For this gene, transitional females showed hypermethylation, similar to levels of males and neomales at comparison 2. developmental and ancestral. In *wnt4* gene, the levels of methylation were hypomethylated in females in respect to the males at control conditions and at 1. Developmental (with the nemoales showing similar levels to males). However, such differences were non-significant at comparisons 1'. ancestral and 2. developmental and ancestral. Additionally, no differences were found between resistant females and transitional females when exposed two times to temperature.

Identification of epigenetic marks to classify resistant vs. non-resistant genetic females

To understand the distinction between genetic females resistant to high temperature and those that experienced sex reversal, we employed Multi-Omics Factor Analysis. This method is usually used to combine different types of data, for instance, gene expression, mutations, methylation, etc. However, here we used this approach to delve deeper into the intricate and high-dimensional DNA methylation data, aiming to uncover latent factors that account for the divergence among genetic females. We divided the DNA methylation data into four layers, using DNA methylation data found at promoters, introns, exons, and CpG islands. We applied data normalization, pre-processing (specifically, removing tissue-specific differences), data transformation to each layer and used a total of 45 F1 gonad samples (24 neomales and 21 females). This resulted in 10,000 features in the integrated dataset (**Figure 7A**). The MOFA model enabled the identification of major axes of variance in the complex dataset. It extracted shared as well as layer-specific combinations of features that describe variation across genetic female types (resistant and neomales). Hence, we utilized MOFA with four factors, aiming to strike a balance between a substantial portion of explained variance and factors that are easy to interpret. The analysis of the percentage of variance explained by each factor in each layer showed that the inferred factors captured variability shared across layers, especially in factor 1 (**Figure 7B**). Thus, since factor 1 explained the largest proportion of variability, we plotted factor 1 for the resistant females and the neomales from different sire types (LT, NM and HT) (**Figure 7C**) and all samples together regardless of the sire type, grouped by female type (**Figure 7D**).

Thus, we found that factor 1 was able to clearly separate resistant females from neomales. We therefore focused on the features contributing the highest factor weights to build factor 1 (**Figure 7E**). For instance, the CpG with the highest weight in factor 1 was found in the linkage group 1B and in the position 12,327,112-12,329,113 overlapping with adp-ribosylation factor-like protein 3-like (*arl3*) and sideroflexin 2 (*sfxn2*) genes (**Figure 7F**).

Temperature-induced and inherited DNA methylation marks

In the present study we assessed the transmission of temperature-induced epimarks to the subsequent generation at three levels: 1) the stable CpGs between F0 and F1; 2) the environmentally-induced marks at the sperm of the F0 transmitted to F1 gonads using common DMCs estimated with standard analysis (1 permutation); and 3) the environmentally-induced marks at the sperm of the F0 transmitted to F1 gonads using a permutation test. In the first level, the stable CpGs (*i.e.*, no DMC between F0 and F1

with a methylation difference $<15\%$ and $q\text{-value} < 0.01$) we found that a large proportion of CpGs showed stable levels of methylation between F0 gametes and the F1 offspring gonads in both comparisons: from the LT sires to the LT offspring (88%) and also from the HT sires to the LT offspring (80%). Also similar values were obtained for both F1 males and F1 females.

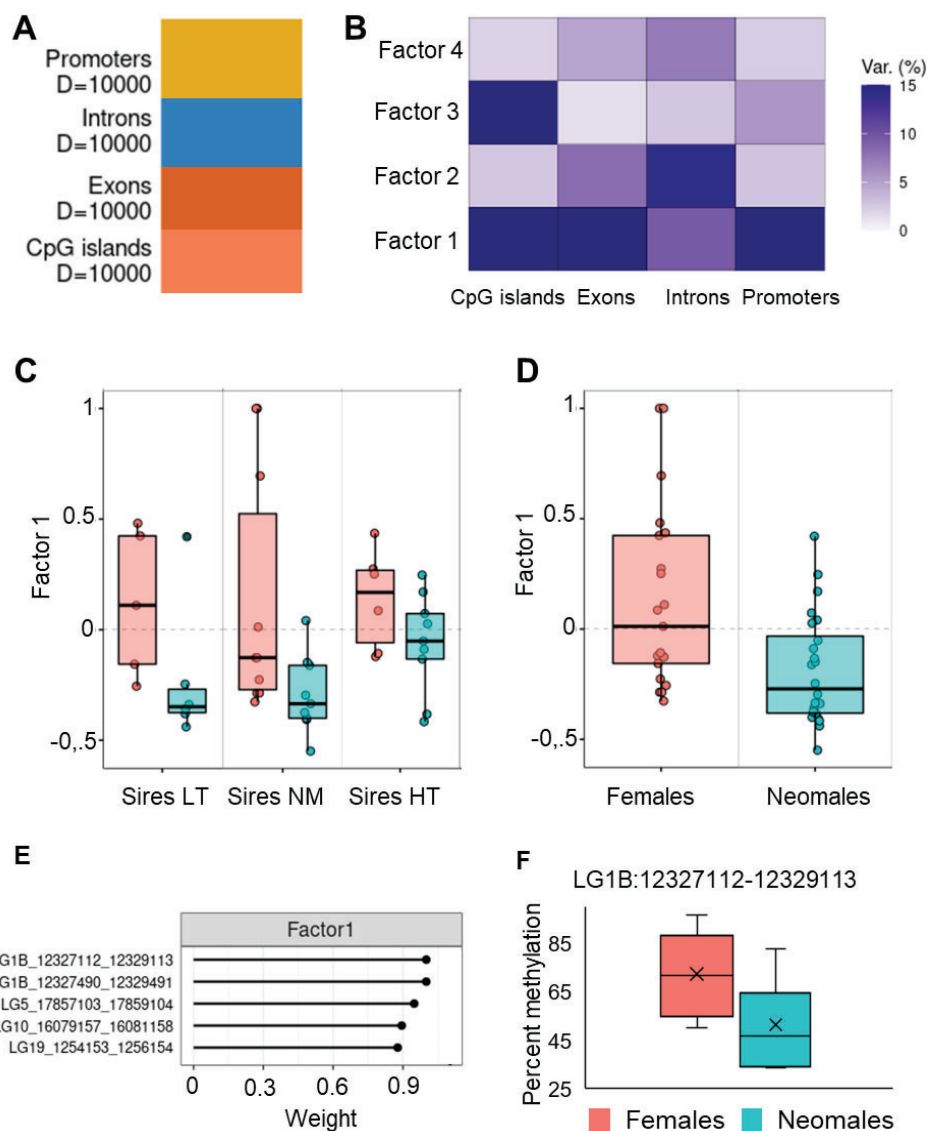


Figure 7. Multi-Omics Factor Analysis applied to F1 ovaries and neomale testis and using different layers of DNA methylation data. **A.** Diagram of the four DNA methylation layers of data used in the study. Data layers are shown in each row (D = number of CpGs) and samples (N) in columns. **B.** Heat map of the proportion of total variance explained (R^2) by individual factors for each layer. **C.** Visualization of resistant females (red) and neomales (blue) by sire type using factor 1. **D.** Visualization of samples by phenotype using factor 1. **E.** Absolute loadings of the top features of factor 1 in the layer of DNA methylation in promoters. **F.** Percent methylation data of the resistant females and neomales at the loci with the highest weight in factor 1.

In the second level, we first determined the number of DMCs induced by the effect of temperature between sperm samples of fish reared at LT and HT. We found 6,120 hypermethylated DMCs and 3,922 hypomethylated DMCs between LT and HT F0 sperm samples. These figures indicate that 1.94% out of the 516,651 CpGs captured by RRBS in F0 samples were significantly affected by the temperature treatment which was applied during their early development and were observed in their gametes when the fish were two years old. The DMCs between LT and HT sperm of the F0 were annotated against the reference genome; 13% were found at the promoters, 5.2% at the first exon and 18.4% in the rest of

exons, 8.9% at the first intron and 26.5% in the rest of introns, while the rest were found in intergenic regions (**Supplementary figure 7D**). Then, we compared the F1 gonads from the control group (0, fish reared at LT from unexposed sires) with F1 gonads from the ancestral group (1', fish reared at LT from exposed sires) to see the CpGs with different methylation level which would be induced by the temperature experienced by the F0 sires. We estimated 26,586 DMCs (of which 96.0% hypomethylated) in the F1 testis and 59,568 DMCs (of which 98.4% hypermethylated) in the F1 ovaries. Nevertheless, only 219 hypermethylated and 202 hypomethylated DMCs were found in common between F0 DMCs of the F0 sperm samples and F1 testis (**Figure 8A and B**) and only 354 hyper- and 149 hypomethylated DMCs were identified in the F1 ovaries and in common with the F0 sperm DMCs. Thus, 4.12% and 5% of the temperature-induced marks in the F0 sperm were transmitted to the F1 testis and ovaries, respectively, when using one permutation test.

In the third level, to determine the number of temperature-induced DNA methylation marks inherited from the F0 sperm to the F1 gonads, we used a permutation test using *MethylInheritance* package. We ran 4000 permutations with the threshold fixed at 0.05, using the sperm samples of the F0 and testis and ovaries of the F1 generation separately. For each permutation, identification of DMCs was performed for each generation separately using the same parameters than previously described. Only 85 hypermethylated DMCs and 53 hypomethylated DMCs were identified as inherited from the F0. However, the level of significance was higher than 0.05 and thus we could not reject the hypothesis that these DMCs being obtained by chance in the F1 testis (**Figure 8C and D**). A similar situation occurred in the F1 ovaries, were 166 hypermethylated DMCs and 7 hypomethylated DMCs were identified as temperature-induced and inherited from the F0, but no significant levels were achieved (**Figure 8E and F**). Statistics were calculated at every 500 permutations during the permutation analysis to check if convergence was reached (**Figure 8G and H**). In the analysis of the hypermethylated DMCs in the males, convergence was achieved, as a flat line was observed after 2000 permutations. However, more permutations will be required for the rest of the analyses because convergence was not acquired.

The distribution on the genome of the DMRs for each generation and sex (a. 1 vs. 0 at F0 sperm, d. 1' vs. 0 at F1 testis and ovaries), as well as the MA-DMRs is shown in **Figure 9**. We found that although few DMRs at F0 sperm induced by temperature overlapped with MA-DMRs (n=9), numerous MA-DMRs overlapped with both testis (n=37) and the ovaries (n=49) from exposed sires compared to their half-siblings from unexposed sires (d.1' vs. 0) (**Supplementary figure 10**). Overall, there was only one DMR overlapping at the four comparisons. It was found at linkage group 8, overlapping with the second exon of the gene Eukaryotic Elongation Factor 2 Lysine Methyltransferase (*eef2kmt*), a gene which enables protein-lysine N-methyltransferase activity and is involved in peptidyl-lysine trimethylation (Davydova *et al.*, 2014).

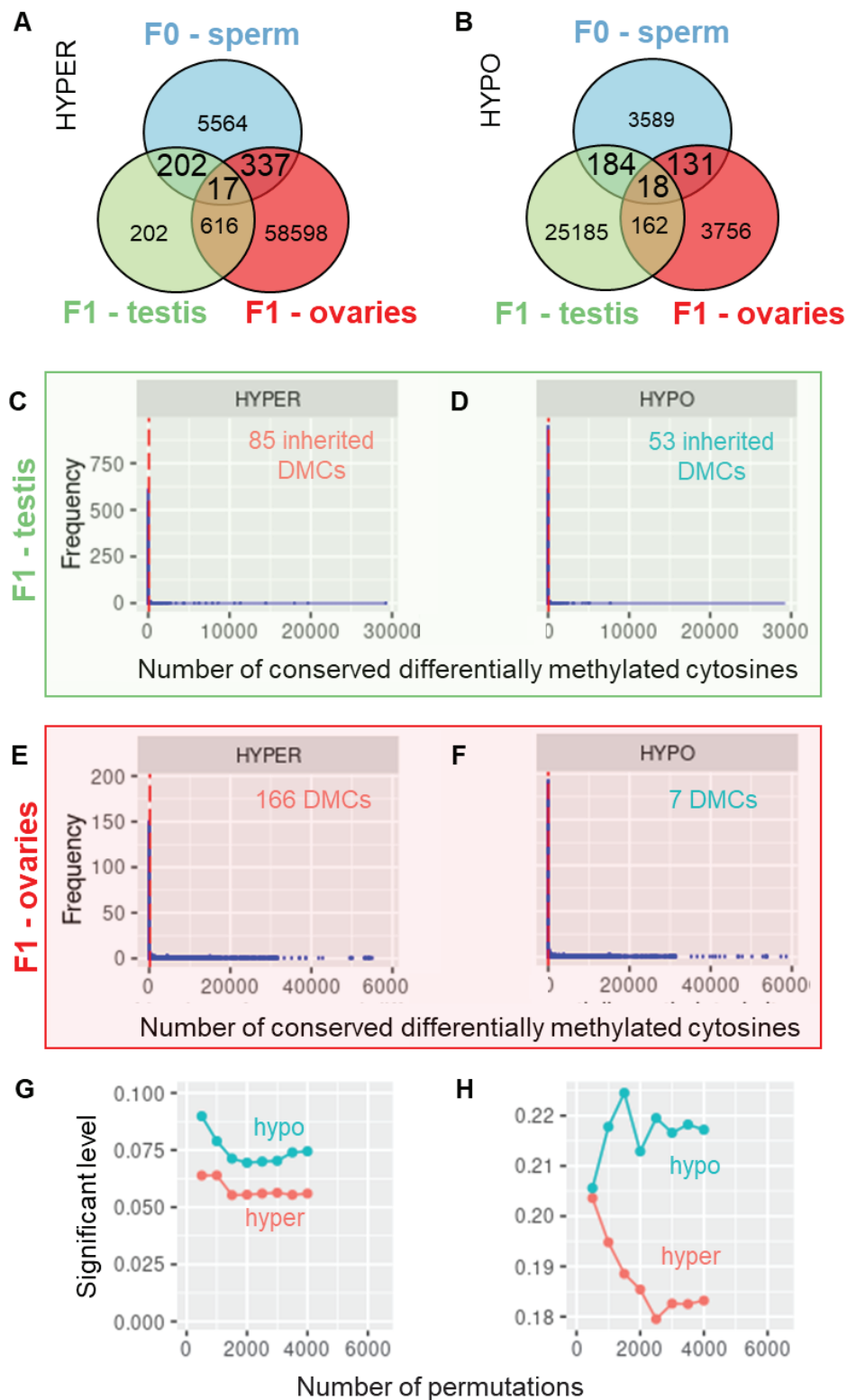


Figure 8. Temperature-induced and inherited differentially methylated cytosines (DMCs) from F0 sperm to F1 gonads of the sea bass. **A.** Overlap of hypermethylated DMCs at F0 and F1 separately (1 permutation analysis). **B.** Overlap of hypomethylated DMCs at F0 and F1 separately (1 permutation analysis). **C-D.** Distribution of the number of conserved DMCs obtained between F0 (sperm) and F1 (testis) generations. Hyper- and hypomethylated conserved DMCs are shown separately. The red dotted lines represent the observed numbers (1 permutation) indicated in the venn diagrams. **E-F.** Distribution of the number of conserved DMCs obtained between F0 (sperm) and F1 (ovaries) generations. **G-H.** Convergence of the permutation analysis with statistics calculated every 500 permutations between F0 and F1 for testis and ovaries, respectively.

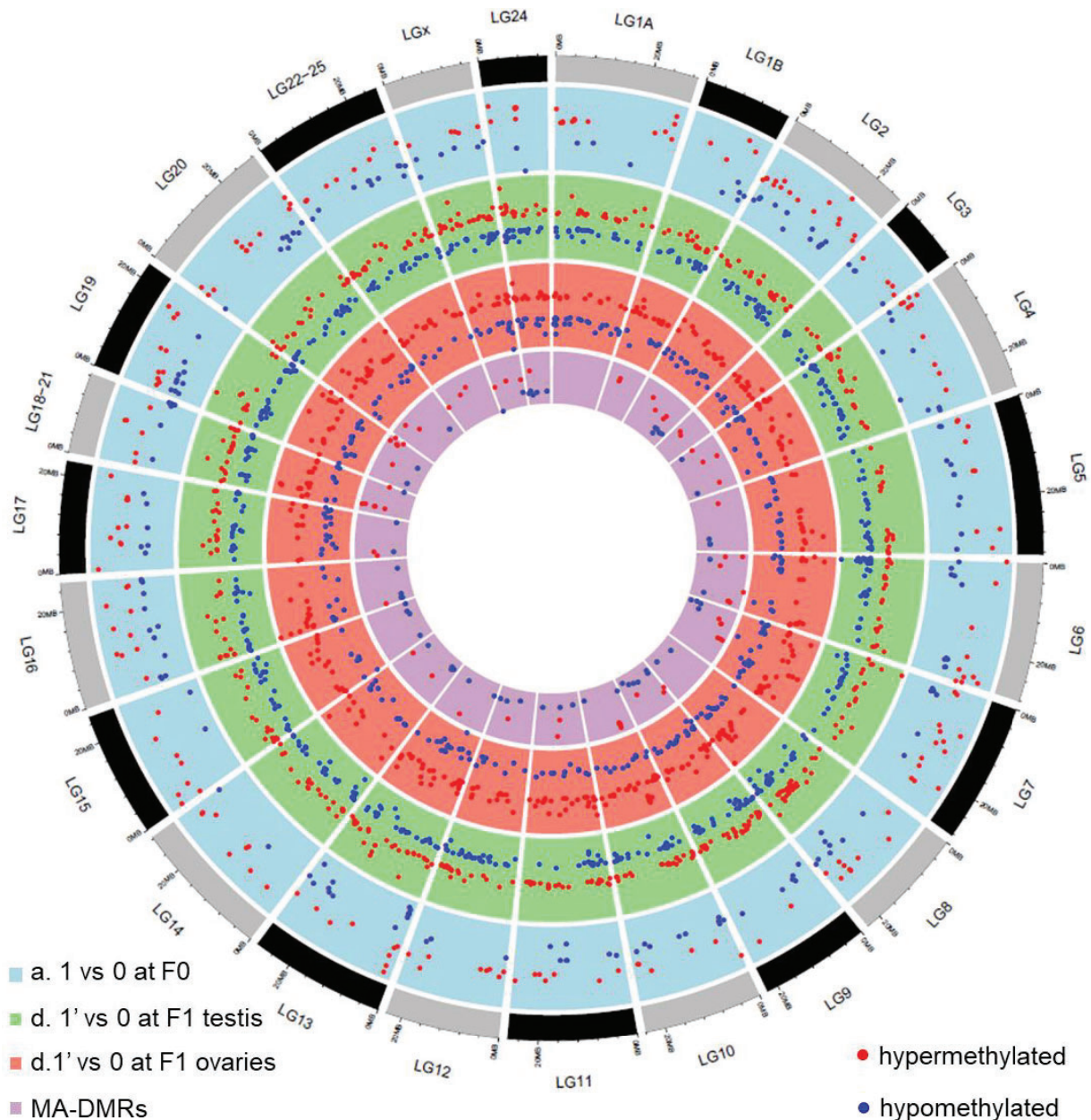


Figure 9. Genome-wide distribution of differentially methylated regions (DMRs; $\geq 10\%$ methylation difference and significance q -value = 0.05) of different comparisons that are, outside to inside: 1. HT vs. LT in F0 sperm (blue), 2. Comparison 'd. 1' vs. 0' in F1 testes (light green), 3. Comparison 'd. 1' vs. 0' in F1 testes ovaries (red) and 4. Masculinization Associated differentially methylated regions (MA-DMRs) in F1 gonads (purple). LG: linkage groups.

4. Discussion

In this study, we used the European sea bass as a model to characterize and compare the epigenetic profiles of fish never exposed to high temperature (controls) with respect to those that were exposed once, either through their parents (ancestral exposure) or during their early stages (development), or exposed twice (ancestral and developmental exposure). As a result of this study, the additive effects of temperature across multiple generations, the epigenetic regulation of masculinization and its inheritance could be elucidated in an unprecedented level of resolution.

The effect of sex and temperature on the sea bass methylome

In the present multigenerational study, sex, temperature and sire type influenced body weight and body length without the interaction among the three factors. The results revealed that the larger individuals were females from neomale sires reared at HT, further highlighting the previously known link between growth and sex in the sea bass (Díaz *et al.*, 2013; Geffroy *et al.*, 2021). The application of HT treatment resulted in significant higher proportion of males in the offspring (F1), as expected from the well-established temperature treatment protocol employed in previous studies (Anastasiadi *et al.*, 2018b; Blázquez *et al.*, 2009, 1998; Navarro-Martín *et al.*, 2009a). Further, F0 neomales (sex reversed genetic females by high temperature) produced more females in the F1 when exposed to either LT or HT. This sex ratio data is in accordance to the Polygenic Sex determination (PSD) system of the sea bass (Vandeputte *et al.*, 2007). Precisely, since the sea bass lacks sexual chromosomes, we identified neomales, for the first time in the sea bass, based on the estimated GST (*i.e.*, a value assigned to each individual representing the probability to become a given sex under a liability threshold model for sex determination). GST was obtained from a model proved to accurately predict the future sex trait based on a combination of SNPs in a previous study (Geffroy *et al.*, 2021). Besides significant mean GST differences, no discernible morphological or phenotypic distinctions were found between the testes of males and neomales, similar to findings observed in other species (Nivelle *et al.*, 2019; Shao *et al.*, 2014).

At the level of methylation profiles between sexes and temperature treatments, the largest differences were found in the comparisons between ovaries and testis (regardless of the temperature treatment), further supporting that DNA methylation is one of the mechanisms playing an important role in the acquisition and maintenance of tissue identity (Recalde *et al.*, 2022; B. Zhang *et al.*, 2013) and, in particular, the sexual phenotype (Piferrer *et al.*, 2019). The second largest change on the DNA methylation profiles were found between different temperature treatments for each sex separately (*i.e.*, fish reared at zero (control), one (in the sire, ancestral; or in offspring, developmental) or two times (sires and offspring; developmental and ancestral) at HT. We found that temperature at multiple generations showed an additive effect and, further, that such temperature effect was sex-specific. An additive effect of developmental and reproductive exposure to high temperature was observed as a decline in reproduction from females and reduced reproductive output from males (Spinks *et al.*, 2021), suggesting that continuous exposure to high temperatures could significantly impact population sustainability. However, this was not associated with increasing number of DNA methylation changes. An environmentally-induced additive effect on DNA methylation was observed before in a different context, *e.g.*, in mouse exposed during multiple generations to dietary zearalenone, an estrogenic mycotoxin, causing a cumulative impairment of fertility and puberty in females (Zhao *et al.*, 2014). In another study, Ryu *et al.*, (2020), evaluated if changes in DNA methylation were associated with the ability for within-generational thermal plasticity of a coral reef fish (*Acanthochromis polyacanthus*) and compared these methylation profiles with those in another population of the same species which showed transgenerational thermal plasticity. In both populations HT induced DMRs in genes related to metabolism, development, cardiovascular and heat-response functions. However, little overlap between the two populations of the same species was found, suggesting that a different epigenetic toolkit was used to cope with HT. Nevertheless, no cumulative effects were reported from the population that exhibits transgenerational thermal plasticity. Thus, here we report, to the best of our knowledge, the first additive effect of temperature on the methylome, a relevant observation that paves the way to find similar findings in a variety of other contexts including both animal production and conservation biology.

Ultimately, the third largest differences at the level of methylome were found in the comparison of offspring gonads from fish that experienced HT twice with their half-siblings that did not experience HT themselves but their sires did (comparison 'd. group 2 vs. group 1'), especially in F1 testis. This result brings to light the concept of epigenetic buffering and the epigenetic trap hypothesis. Epigenetic buffering refers to the role of epigenetic modifications enabling evolutionary rescue (*i.e.*, the phenomenon where genetic adaptation prevents a population from extinction), hypothesizing that these modifications can support the inheritance of new phenotypic variations resulting from environmental changes, allowing for a strategy known as "heritable bet-hedging." This process helps maintain and enhance the evolutionary capacity of a population (O'Dea *et al.*, 2016). However, inherited epigenetic variation could also be maladaptive and thus become an epigenetic trap (O'Dea *et al.*, 2016).

Epigenetic traps have been associated to environmentally-driven changes in population sex ratio, suggesting they could pose a risk under a global warming scenario (Consuegra and Rodríguez López, 2016). Although this is certainly a possibility, other events such as frequency-selection of sex ratios may offset the effects of epigenetic traps (Piferrer, 2016). In any case, similar to what was proposed to occur in species with Environmental Sex Determination (ESD), in which sex determination can be epigenetically regulated and adapt to fast environmental changes, this could also be the epigenetic mechanism by which the sexual phenotype is regulated in the sea bass, where multiple genetic factors and environmental influences (temperature) determine sex. Thus, in the context of the present study, the most different methylome was that of the testis at F1 from exposed sires that were reared at LT, which suggest a miss-match with the expected environment from the ancestors. This could be the first evidence of the epigenetic trap hypothesis in a species with PSD of which DNA methylation changes were transmitted to the next generation (likely causing maladaptive changes), as previously suggested by (Piferrer, 2016).

Epigenetic regulation of sex reversal

Sexual development is known to be regulated by a network of genes of pro-male and pro-female genes exerting opposing forces until a threshold is reached and tilts the balance towards one sexual phenotype or the other (Capel, 2017; Munger *et al.*, 2013). However, it is also known that early temperature influences phenotypic plasticity (*i.e.*, the ability of individual genotypes to produce altered phenotypes when exposed to different conditions) (Pigliucci *et al.*, 2006). This includes the capacity to modify developmental trajectories, activities, and resource allocation (Jonsson and Jonsson, 2019). In 2014, Shao *et al.*, (2014), described one of the primary mechanisms underlying phenotypic plasticity in the sex reversal of the half-smooth tongue sole as epigenetic modification and its inheritance. Also, in the Nile tilapia (*Oreochromis niloticus*), the transcriptome and methylome of neomales were very similar to those of genetic males (Wang *et al.*, 2019a), as well as in the tiger pufferfish (*Takifugu rubripes*) (Zhou *et al.*, 2019). In sea bass, we also observed that methylation patterns of neomales were highly similar to those of genetic males but distinct from females. These results suggest that the temperature-induced sex reversal in sea bass is, at least, partly driven by epigenetic factors. Regarding global mean methylation, no significant differences between F1 gonads of the same sex at different temperature treatments (exposed zero, once developmental, ancestral, twice) were found. However, from the PCAs of the fish from experimental groups 1' (ancestral) and 2 (developmental and ancestral), we identified some females that we call transitional at a mid-way distance between the most resistant females and the males (both genetic males and neomales) in the PCA cartesian space. When considering such classification, a significant gradual increase of the global mean methylation between the most resistant females, females at transition stages and neomales, was observed. Furthermore, neither the females reared at HT two times (developmental and ancestral), including resistant, T1 and T2 females nor the neomales, showed correlation between the level of global mean methylation and the GST. Taken

together, these findings indicate that the GST value is not informative as to which females will undergo sex reversal and provide additional support for the crucial role of epigenetics, at least involving DNA methylation modifications, in the temperature-induced masculinization process of sea bass.

Sex reversal is a phenomenon observed in various species including fish, amphibians, reptiles, birds, and mammals (Piferrer and Anastasiadi, 2021). Interestingly, Li *et al.*, (2021) created Animal Sex Reversal (ASER), a comprehensive database integrating sex reversal-related information from 18 species, aiding understanding of the molecular basis. However, an increasing number of studies have shed light on the epigenetic mechanisms through which environmental factors can influence and override the sexual fate determined by genetic factors. Thus, we propose expanding ASER to incorporate epigenetic data, enabling a deeper exploration of the interplay between genetics and epigenetics in sex reversal across species.

The study of sex reversal, whether artificially- or naturally-induced, provides a valuable system to identify sex-determining genes as well as molecular cascades in sex differentiation. In the sea bass, we found a total of 4,396 MA-DMRs. Notably, 78.7% of these DMRs were located in intergenic regions. These regions hold significant potential as candidate factors and regions involved in the sex determination of sea bass and warrant further investigation (Shao *et al.*, 2014). Among those MA-DMRs, 4,319 were consistently hypermethylated between most resistant females, females at transition stages and neomales. Only 77 MA-DMRs were consistently hypomethylated between the same genetic female groups. The progressive increase in methylation levels of pro-female genes, coupled with the hypomethylation of pro-male genes found during sex reversal in the sea bass (gonochoristic), resembles the gradual genetically-regulated mechanism observed in protogynous species undergoing sex change, where it was shown that sex-specific genes and steroids underlie the process of sex change. For instance, in protogynous sex change, upregulation of pro-male genes such as *dmrt1* occur, which are followed by inhibition of pro-female genes, such as *cyp19a1a*. This is reflected by a decrease in estrogen levels followed a concomitant increased in androgen levels. The opposite pattern occurs in species with protandrous sex change (Todd *et al.*, 2019, 2016).

A total of 260 coding genes overlapped with MA-DMRS. Among them, twelve genes were previously known to be differentially expressed between males and females during early sex differentiation of the sea bass (Ribas *et al.*, 2019). Some of them upregulated in females, such as Lysine Acetyltransferase 8 (*kat8*) and Zinc Finger and BTB Domain Containing 14 (*zbtb14*). And others, upregulated in males: Receptor Activity Modifying Protein 3 (*ramp3*) and ELAV Like RNA Binding Protein 1 (*elavl1*). The enrichment analysis of those 260 genes overlapping with MA-DMRS revealed physiological processes and mechanisms affected during temperature-induced masculinization. Some of them were previously related to gonad development, such as the cholesterol metabolic process, necessary for the biosynthesis of steroids (Gwynne and Strauss, 1982), as well as the link between sex-related genes and apoptosis signaling involved, for example, in the sex change of the protogynous ricefield eel (*Monopterus albus*) (He *et al.*, 2022). Other biological process more recently involved in the sex reversal of other species is the regulation of ion transport, as found to play important roles in thermally induced sex reversal of the Nile tilapia (Wang *et al.*, 2019a). In relation to the GO terms enriched in the sea bass MA-DMRs, several were related to synaptic signaling: modulation of chemical synaptic transmission, synaptic vesicle endocytosis and trans-synaptic signaling. These are biological processes that can certainly be relevant within the neuroactive ligand-receptor pathway, an essential pathway in the sex reversal of the Nile tilapia (Sun *et al.*, 2016; Yao *et al.*, 2022). This pathway includes relevant sex-related genes such as GnRH receptor type 2 (*gnrhr2*), Thyrotropin receptor (*tgh-rII*) and KISS1 Receptor (*kiss1r* also named *gpr54*), which mediate the release of gonadotropin releasing hormone and gonadotropin hormone. In

fact, several studies reported that this pathway plays an important role at the transcriptomic and epigenetic levels during high temperature-induced masculinization in Nile tilapia (Sun *et al.*, 2016; Teng *et al.*, 2020; Yao *et al.*, 2022).

Regarding the masculinization-associated genes which mapped to pathways in the sea bass, we found many of them that were likewise identified at the protein level in the sex reversal of the blackhead seabream (*Acanthopagrus schlegelii*): the integrin signaling pathway, inflammation mediated by the chemokine and cytokine signaling pathways, CCKR signaling map, the heterotrimeric G-protein signaling pathway, Gi alpha- and Gs alpha-mediated pathways as well as proteins involved in the Wnt signaling pathway (Chen *et al.*, 2022). Additionally, two of the genes with highest number MA-DMRs were *18SrRNA* and *pop4* genes. The former was identified as a marker of sex and is used as a marker of ovarian maturation in fish (Rojo-Bartolomé *et al.*, 2017; Shen *et al.*, 2017). The latter, *pop4*, was identified as a hub gene (*i.e.*, highly connected gene) in the network of ovarian differentiation in the sea bass (Sánchez-Baizán *et al.*, 2022). Thus, hypermethylation during masculinization of this gene associated with ovarian development allowed us to hypothesize a relevant role in the gonadal development through negative correlation of their methylation levels and gene expression in the gonads of the sea bass.

Finally, we also characterized the methylation levels of key genes previously known to be involved in early gonadal development of the sea bass and other vertebrates (Nagahama *et al.*, 2021; Ribas *et al.*, 2019). We paid particular attention to the biomarkers of early ovarian and testis differentiation: *hsd17 β 10* (Ribas *et al.*, 2019) and *sox9a* (Sánchez-Baizán *et al.*, 2022), respectively. The interaction with DNA methylation and gene expression is not completely understood, however, an inverse relationship between epigenetic silencing and gene expression has been observed in some instances. Based on that and several observations in the literature, the conserved epigenetic regulation of sex (CERS) model was proposed. According to that model, specific epigenetic and gene expression patterns are more closely associated with the development of one sex or the other rather than with the intrinsic or extrinsic factors leading to the development of this phenotype (Piferrer *et al.*, 2019). Thus, genes whose expression is enriched in males (“pro-male” genes) show high expression in males with concomitant low levels of DNA methylation and the contrary in females. In contrast, genes whose expression is enriched in females (“pro-females” genes) would exhibit high expression in females combined with hypomethylation when compared to males.

The *sox9a* gene showed no DMCs when comparing the gametes of the parents, only the in the sperm of the neomales, which were hypomethylated in comparison to the sperm of males at HT. In the offspring, this gene was hypomethylated in the ovaries of adult F1 fish at control treatment. Taking into account that this gene showed higher expression in testis than in ovaries in adult fish at control conditions (Ribas *et al.*, 2019), one would suspect this gene was not in accordance to the CERS model. However, when looking at the other comparisons we found that at HT, the levels of DNA methylation were decreased in both ovaries and testis but with higher levels in the females (resistant females) than in the males and with neomales showing similar levels to the males. This result suggests that the levels of DNA methylation decreased as an effect of HT and could lead to higher expression of *sox9a* in the genetic females, with a proportion of them undergoing sex reversal. Given that this gene is a biomarker of early testis differentiation and a hub gene of the network involved (Sánchez-Baizán *et al.*, 2022), this could show the mechanism by which masculinization is partly driven and that in fact this gene would follow the CERS model but only when environmental changes were induced (*i.e.*, lower methylation levels induced by HT cause that a proportion of the females expresses higher *sox9a* gene, leading to testis differentiation). Otherwise, under control conditions this situation was not observed; the levels of DNA

methylation were higher in males at LT, suggesting that the previously reported higher expression of this gene could be differently regulated at LT than at HT.

The *hsd17 β 10* gene is a biomarker of ovarian differentiation and thus a pro-female gene (Ribas *et al.*, 2019). The levels of DNA methylation were lower in the unfertilized eggs than in the sperm of the F0 generation. The neomale sires showed higher levels of DNA methylation. In the F1, similar results were found when comparing ovaries and testis at LT (hypomethylation in females). As the number of temperatures increased the level of DNA methylation in this gene increased in the females towards similar levels of those in the males and neomales. Moreover, when looking at the effect of temperature two times (2. developmental and ancestral) we found that resistant females showed hypomethylation, with levels even lower than the control females at LT, while the females at transition stages and neomales showed the same levels of methylation than in the testis of genetic males. Together, these results suggest that this gene would be in accordance with the CERS model and that could play an important role during temperature-induced sex reversal of the sea bass.

Epigenetic biomarkers of HT resistant females

To develop EBs allowing to classify genetic females between neomales and HT resistant females, we used MOFA. This approach allowed the identification of a set of CpG loci that can effectively distinguish between resistant and non-resistant individuals, even when removing tissue specific differences, *i.e.*, testis *vs.* ovary DMCs. This opens the door for the development practical applications for aquaculture. However, further research is needed to refine and optimize this EBs. Specifically, we must investigate whether these markers are also present in tissues such as sperm, fin or blood, the methylation of which can be assessed in a non-invasive or lethal manner and determine their prognostic value by testing them in younger individuals, before gonadal development occurs. Additionally, assessing the heritability of these markers across generations is crucial for future breeding strategies (Anastasiadi and Beemelmans, 2023).

Epigenetic inheritance of temperature-induced epimarks

Epigenetic inheritance is a burgeoning field, but there is still much work to be done in establishing mathematical models and criteria to accurately identify and classify instances of epigenetic inheritance, particularly in animals. Specially, for species of longer life cycle and for whose it is difficult to obtain large sample sizes. Currently, some studies claim epigenetic inheritance based solely on the detection of differentially methylated regions (DMCs) between experimental groups and generations, without fully considering underlying genetic factors, genetic-environment interactions, or randomly acquired changes. For instance, a study on the inland silverside (*Menidia beryllina*) reported that early-life exposure to endocrine disrupting chemicals (EDCs) resulted in epigenetic changes that were inherited at multigenerational and transgenerational levels (Major *et al.*, 2020). However, the analysis was based on identifying DMCs common across generations, and thus, overlooking the complex interplay between genetic, environmental factors and random factors. Additionally, the concept of epigenetic inheritance is often conflated with stable epimarks, which are loci that do not exhibit differential methylation.

In our study, we quantified epigenetic marks inherited across generations at three levels: 1) stable CpGs, 2) environmentally-induced marks at the parental generation found in common at the filial generation and 3) environmentally-induced marks at the parental generation found in common at the filial generation without random effects. Thus, we found that 88 and 80% of CpG sites were not differentially methylated across generations from both LT and HT, suggesting the majority of sites showed stable levels of methylation across generations. This was in accordance with the results found in the threespine

stickleback (96% of stable CpGs), where a similar analysis with the same DMC statistical thresholds were chosen to allow for comparison of results (Hu *et al.*, 2021). However, they compared the sperm tissue in both generations while we compared the sperm of the F0 with the testis of the F1 generation, which could perhaps explain our lower number of shared CpGs.

In terms of environmentally-induced DMCs, our study revealed that 4.1% of temperature-induced marks in F0 sperm were inherited by F1 testis, while 5% were transmitted to F1 ovaries. These findings align with results from previous studies (considering that in our studies different tissues were compared while sperm was always used both in F0 and F1 in the other studies). As described above in the introduction, 11.36% of the DMRs induced by the long-term physiological stress in the sperm of the mouse were transmitted to F1 (Zheng *et al.*, 2021). In the Atlantic salmon, (Rodriguez Barreto *et al.*, 2019), focused on the DNA methylation in sperm of wild and hatchery-reared males and their offspring. In such study, only six DMRs (13.9% of the total DMRs between wild and hatchery parents) were reported to be transmitted to the subsequent generation. Together, these results indicate that a small proportion of environmentally-induced marks can be transmitted across generations, potentially serving as an epigenetic mechanism for inheriting phenotypic responses to the environment and possibly contributing to rapid adaptation.

Furthermore, we could assume that the genetic variation underlying these epigenetic variation was very little because all the groups compared in the F1 were half-siblings (same dam) and from very close related sires (the sires were full siblings). Additionally, we only found a total of 5 DMRs overlapping with SNPs. Therefore, if the proportion of environmentally induced marks was relatively low and the underlying genetic variation was minimal, the stochastic source of variation could be important. For this reason, we wanted to identify those marks that could not be achieved by change. However, at the third level of our study, we did not observe significant inheritance when considering the random source of variation. Nevertheless, it is important to note that further refinement in our analysis (using *methylinheritance* package) is needed before drawing definitive conclusions. One potential improvement could be increasing the number of permutations, especially in the analysis of female offspring. Although we used the recommended number of permutations by the developers (Belleau *et al.*, 2018), we did not achieve convergence of the significance level in the analysis of inheritance with female offspring. Thus, higher number of permutations could improve the results. Additionally, it is worth mentioning that the *methylinheritance* package was designed to study inheritance across three generations, while our study only included two generations. This difference in generation number could have unknown effects on our results. Moreover, the method was originally developed to examine epigenetic inheritance within the same tissue across generations. However, in our study, we used sperm in the parental generation and gonadal tissue in the offspring, potentially introducing bias due to tissue-specific differences. Nonetheless, it is noteworthy that a study by Belleau *et al.*, only identified three inherited sites, indicating that the number of actually environmentally inherited marks in the F1 generation is likely minimal.

Limitations of the study, future perspectives and conclusions

Although the analysis of the 130 methylomes has yielded valuable insights, there remain areas that can be further improved and investigated. To complement our understanding of the effects of sex, temperature, and masculinization phenomena in sea bass, it would be beneficial to study the gonadal transcriptome of the same experimental groups. Moreover, exploring additional epigenetic mechanisms, such as non-coding RNAs and histone modifications, could also enhance our understanding of the processes discussed above. For instance, a study by (Feng *et al.*, 2021), highlighted the involvement of the long non-coding RNA DMRT2-AS as a transcriptional regulator of *dmrt2*, contributing to sex

differentiation in the half smooth tongue sole. Incorporating such investigations will help broaden our knowledge and provide a more comprehensive understanding of the underlying mechanisms at play.

It will also be crucial to examine the methylome and transcriptome of the next filial generation (F2) to elucidate the extent of transgenerational epigenetic inheritance. This comprehensive analysis will contribute to a more holistic understanding of the entire processes involved in the effect of temperature and its inheritance across generations in the sea bass. Furthermore, unravelling the dynamics of DNA methylation reprogramming of the sea bass would help to explain how and to what extent the epigenetic marks can be transmitted from the gametes to the developing embryos. Lastly, it is worth noting that a significant number of MA-DMRs were found in intergenic regions. With this knowledge, it becomes evident that more well-defined layers of DNA methylation data could have been utilized in the MOFA analysis. For instance, in the future it would be useful to include a layer specifically focused on intergenic regions that overlap with transcription factor binding sites, which could be particularly relevant as the activity of transcription factors is known to be crucial during early development (Pérez-Lluch *et al.*, 2015). Incorporating such refined layers of analysis could provide better EBs for the identification of resistant females to the effect of HT.

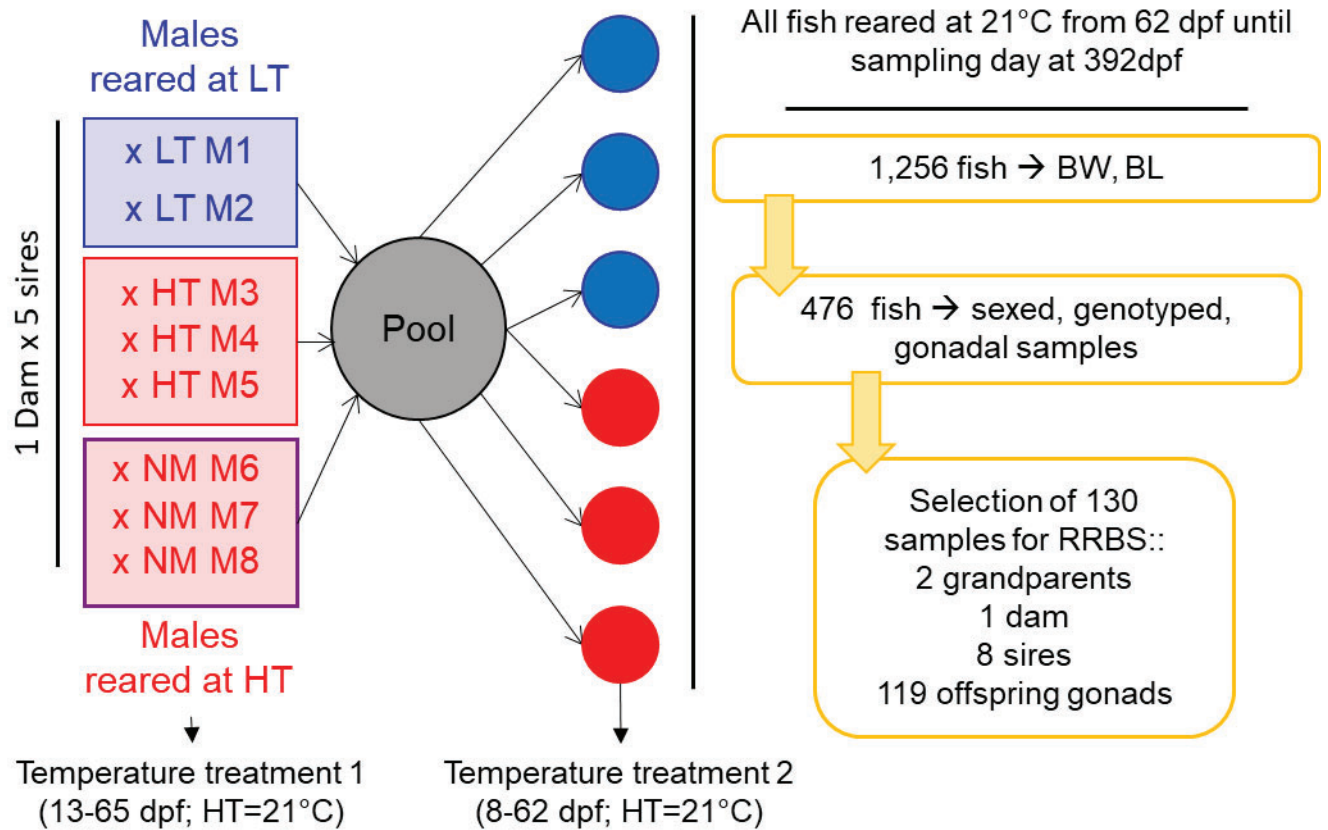
Nevertheless, the analysis of the genome-wide methylome of 130 sea bass individuals from eight families and two generations has provided significant insights into the interplay between sex, temperature, and epigenetic regulation in this species. Our findings elucidated multigenerational cumulative temperature effects, the epigenetic control of masculinization, and its potential inheritance in the European sea bass. These discoveries have important implications for the development of improved aquaculture practices and ultimately enhance our understanding of the species' adaptation mechanisms, particularly relevant in the current context of global warming (Swaegers *et al.*, 2023).

Multigenerational temperature additive effects: the epigenetic regulation of masculinization and its inheritance in the European sea bass

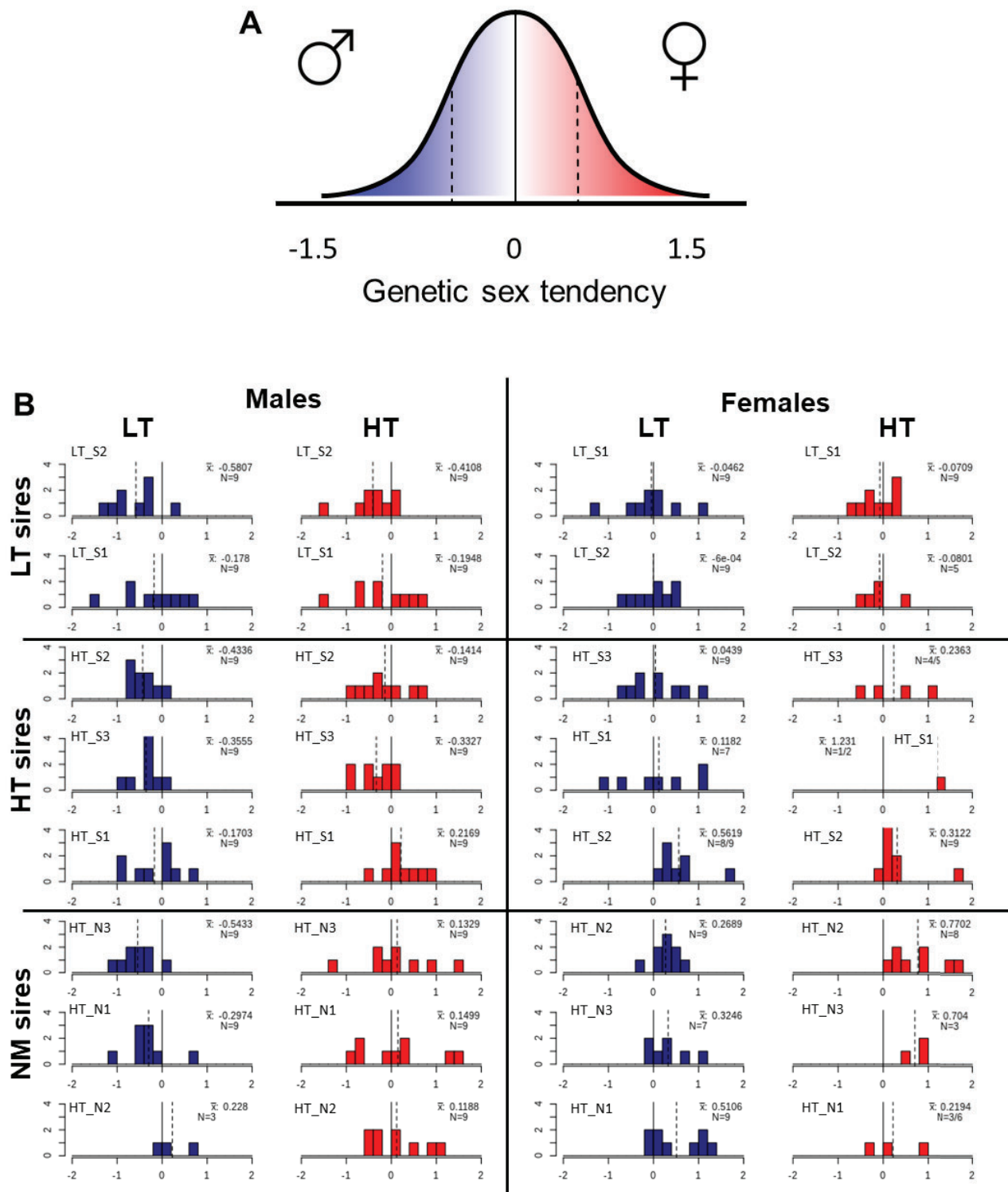
Núria Sánchez-Baizán, François Allal, Johana Denkena, Maria Colomé-Tatché, Marc Vandeputte and Francesc Piferrer

Supplementary information

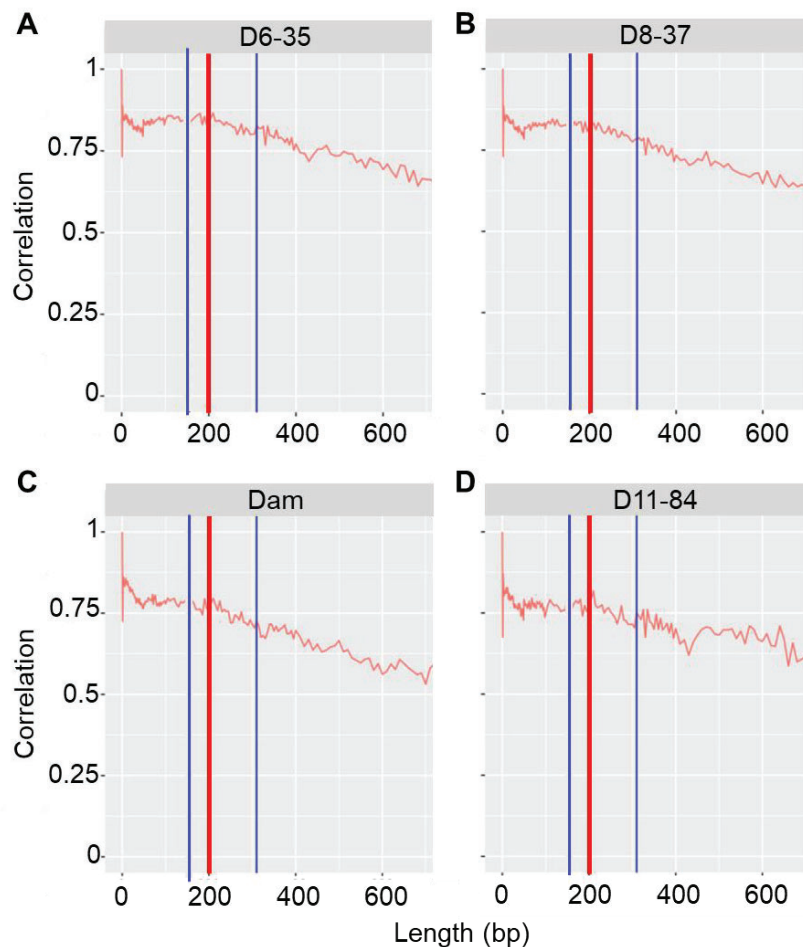
Supplementary figures



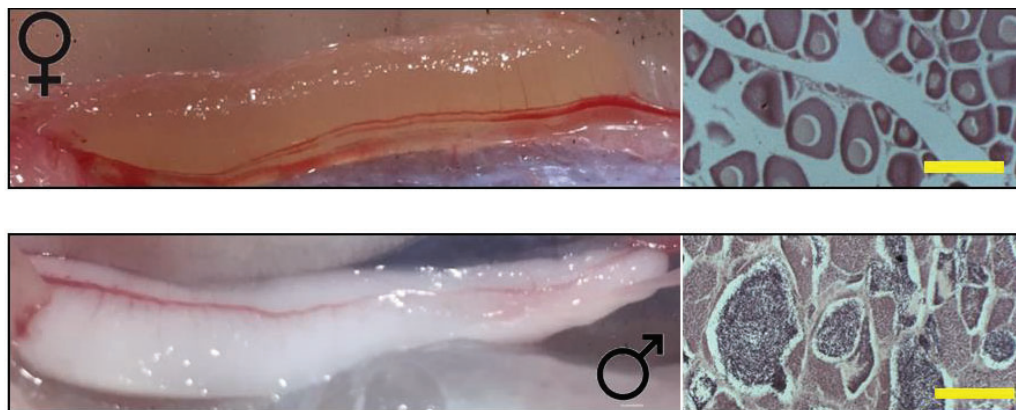
Supplementary figure 1. Diagram of the experimental design, temperature treatments and number of samples processed at the end of the experiment for the different measurements. LT, low temperature; HT, high temperature; NM, neomale; dpf: days post fertilization; BW, body weight; BL, body length; RRBS, Reduced Representation Bisulfite Sequencing.



Supplementary figure 2. Genetic sex tendency. **A.** Diagram of the theoretical distribution of the genetic sex tendency under a threshold, set at zero. **B.** Histograms of the estimated genetic sex tendency for the samples of each group (2 sexes, 2 temperatures, 3 sire types). LT, low temperature (blue); HT, high temperature (red); NM, neomales.

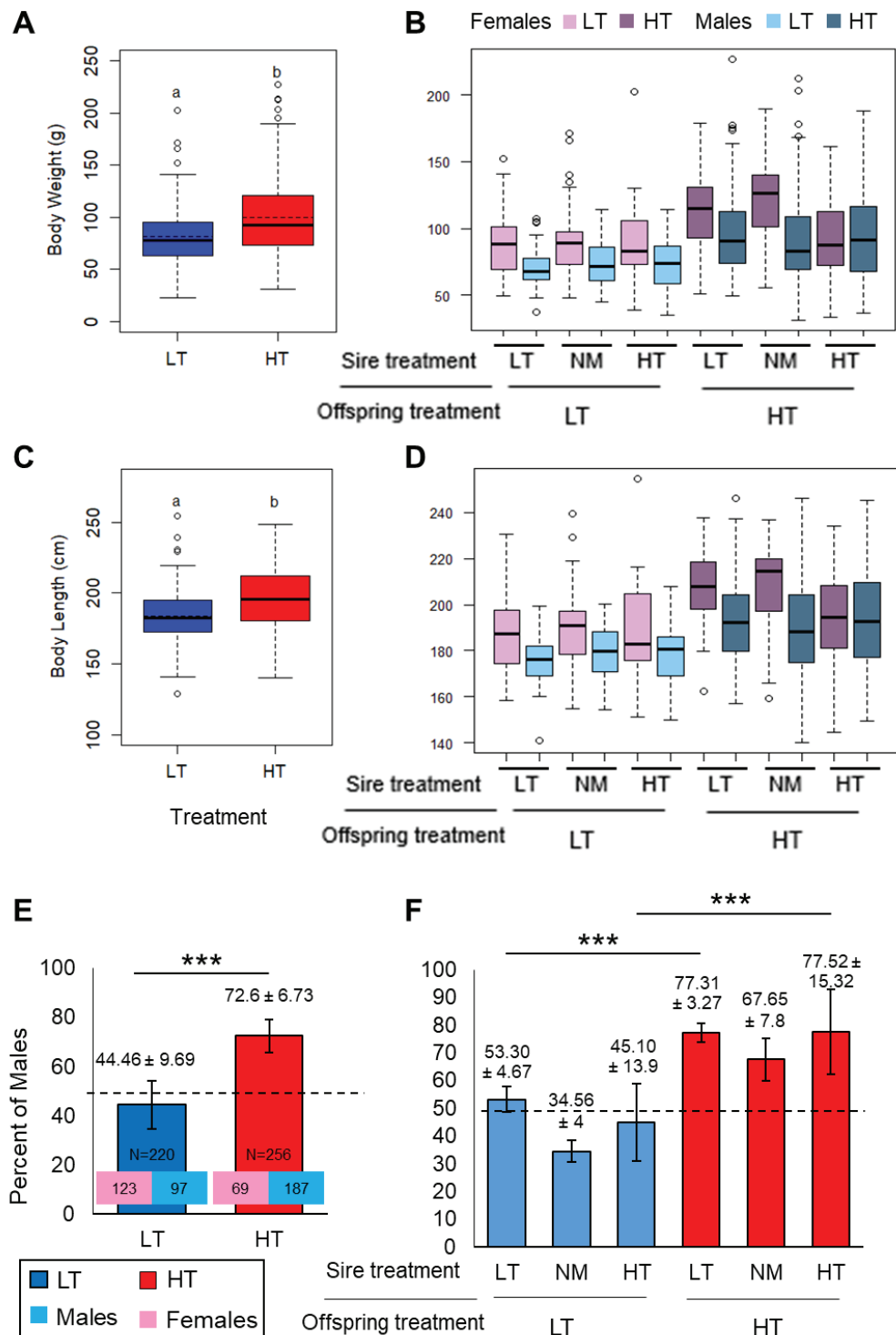


Supplementary figure 3. Autocorrelation profiles. Autocorrelation is the approach used to determine the region length according to correlation of the methylation levels between adjacent CpGs. The length of the regions (in bp) is shown in the x-axis and the Pearson correlation measure between all the CpGs covered within the region is indicated in the y-axis. Several tiles were considered (blue vertical lines). The selected region length was 200 bp (red vertical lines). The figure shows only four of the samples as example but autocorrelation was performed in 20 (15%) randomly selected samples (including fish from F-1, F0 and F1). **A.** Profile of the sample D6-35, the testis of a neomale reared at high temperature. **B.** Profile of the sample D8-37, testis of a male reared at low temperature. **C.** Profile of the unfertilized eggs of the dam (F0) reared at low temperature. **D.** Profile of the sample D11-84, testis of a male reared at low temperature.

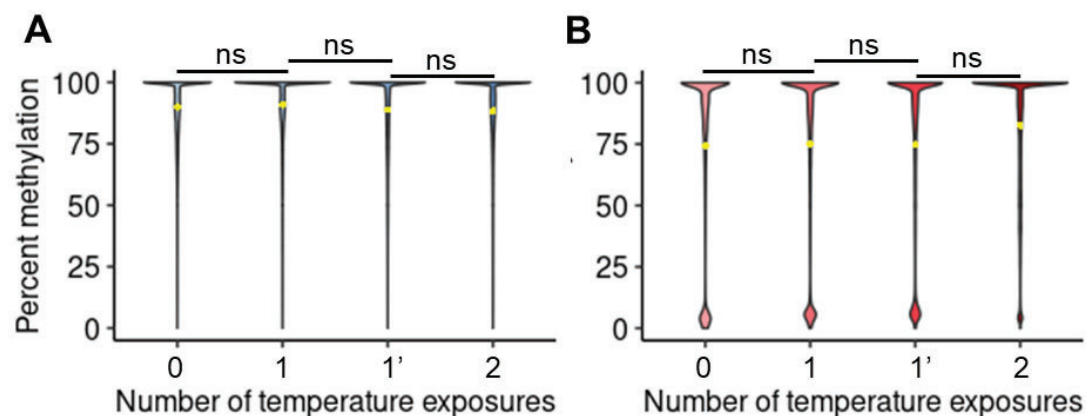


Supplementary figure 4. Gonads of the F1. Left panels, photographs of F1 ovaries (top) and testes (bottom). Right panels, photomicrographs of the corresponding histological sections. Some of the ovaries showed white

adipose tissue but the hypothesis of having intersex individuals was rejected after histological analysis. Yellow bar = 50 μm . No major phenotypic differences were found across testis, neither in the morphology nor regarding body weight or body length.



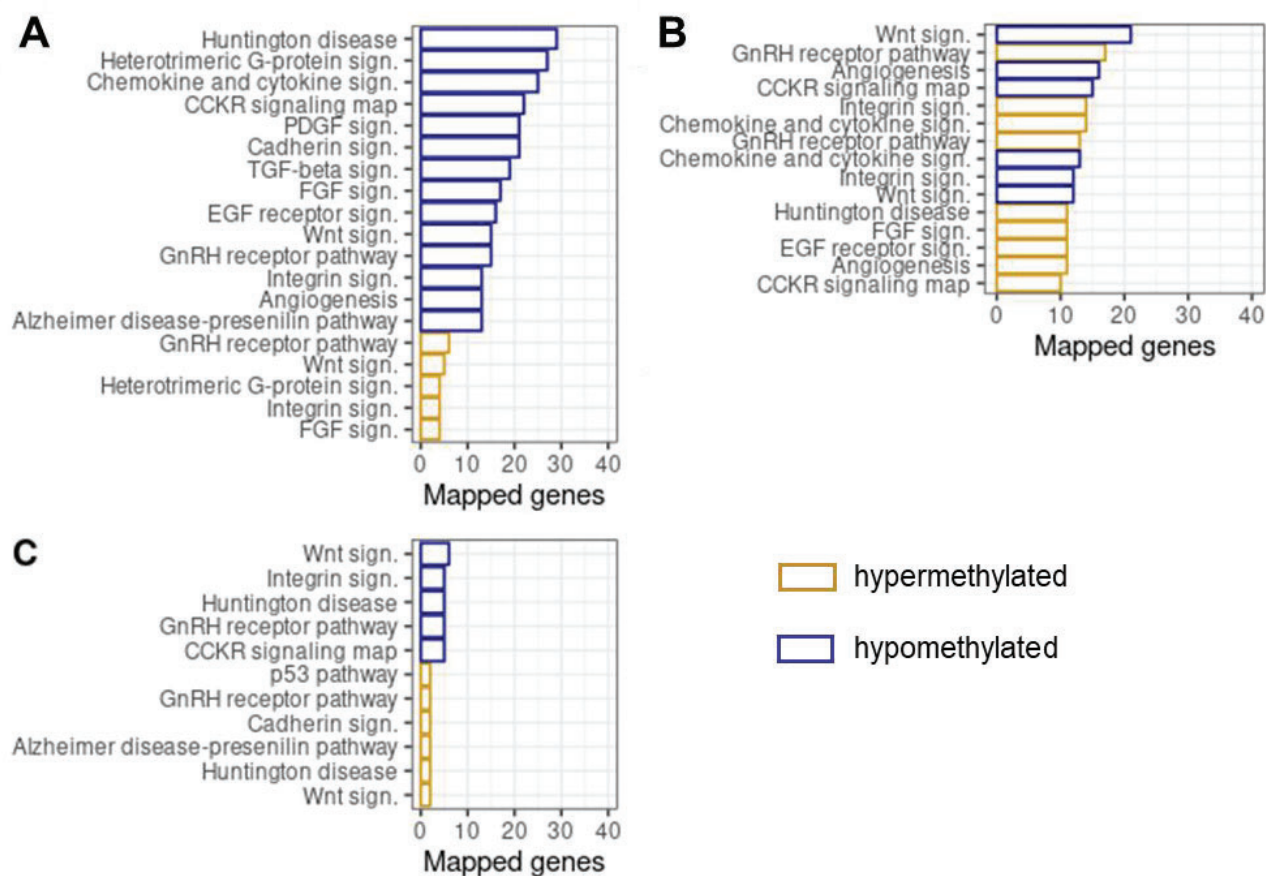
Supplementary figure 5. Biometry and sex ratios of the F1). **A.** Body weight per temperature treatment. **B.** Body weight per phenotypic sex and temperature treatment. **C.** Body length per temperature treatment. **D.** Body length per phenotypic sex and temperature treatment. **E.** Sex ratio (proportion of males) per temperature treatment. **F.** Sex ratio (proportion of males) per temperature treatment and for each sire type: reared at low temperature (LT), high temperature (HT) or neomales (NM) reared at HT. *** $P < 0.001$.



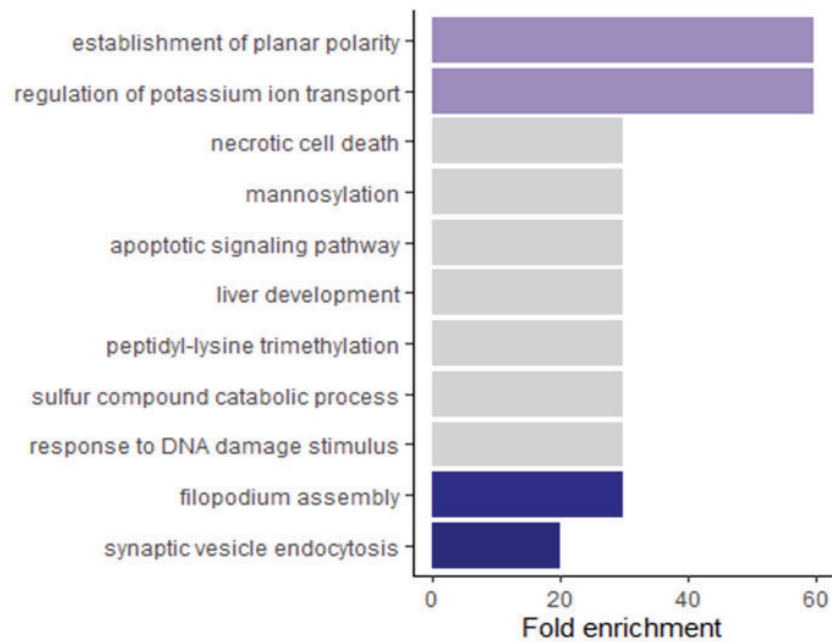
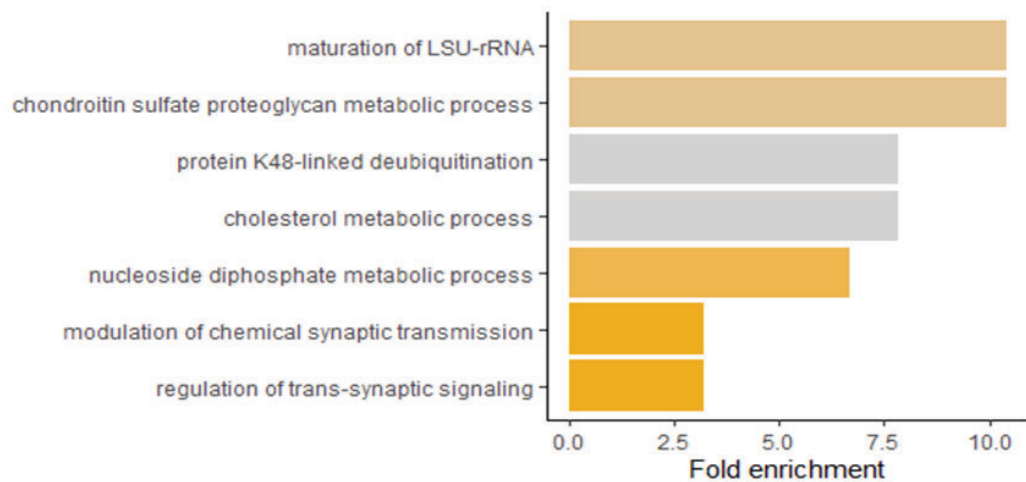
Supplementary figure 6. Violin plot of the percent methylation of all CpGs captured by Reduced Representation Bisulfite Sequencing of testis (**A**) and ovary (**B**) samples of the F_1 . The number of temperature exposure corresponds to the temperature treatment at which fish were exposed. 0, control or low temperature; 1, high temperature during early development; 1', offspring reared at low temperature from sires exposed to high temperature (*i.e.*, ancestral); 2, fish from sires reared at high temperature that were also exposed to high temperature during their early development. No significant differences were found at the global level (ns, $P > 0.05$).



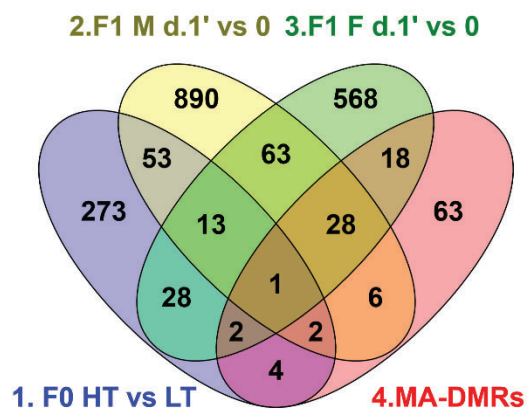
Supplementary figure 7. Proportion of differentially methylated regions (DMRs) found at genomic features (promoters, exons, introns or intergenic regions). The outer and inner rings describe the locations of hypermethylated and hypomethylated DMRs, respectively. **A.** DMRs of males (M) vs. females (F) gonads from fish reared at low temperature (LT). **B.** DMRs between testis from testis from F1 males LT vs. HT. **C.** Regions identified as masculinization-associated DMRs (MA-DMRs). **D.** DMRs in the sperm of F0 LT vs. HT sires.



Supplementary figure 8. Bar plot of the number of mapped genes into pathways for the comparisons: **A.** Differentially methylated regions (DMRs) in F1 ovaries vs. testis at group 0, control. **B.** Comparison 'a. 1 vs. 0' (high vs. low temperature from unexposed sires) **C.** Genes overlapping with regions identified as masculinization-associated DMRs (MA-DMRs).

A**B**

Supplementary figure 9. Functional enrichment analysis of the masculinization-associated differentially methylated regions (MA-DMRs). The results show the gene ontology terms of the biological processes category for **A.** hypermethylated and **B.** hypomethylated regions.



Supplementary figure 10. Venn diagram of the overlap of differentially methylated regions (DMRs). 1. Number of DMRs in the sperm of LT vs. HT parents (F0) (blue), 2. Number of DMRs at comparison 'd. 1' vs. 0' in offspring (F1) testes (yellow), 3. Number of DMRs at comparison 'd. 1' vs. 0' in F1 ovaries (green) and 4. Masculinization-associated differentially methylated regions (MA-DMRs) in F1 gonads (red).

Supplementary tables

Supplementary table 1. Sequencing statistics of the samples retained after quality control for analysis

Treat ment	Sire type	Mean Raw Reads	Mean Trimming (%)	Mean Alignment (%)	Mean CpGs > 10 reads	Mean Fold coverage	Mean Bisulfite conversion
HT	HTN1	24,126,831.2	98.4	91.4	680,526.2	75.6	99.1
	HTN2	22,391,964.0	97.2	91.1	659,972.8	76.9	99.0
	HTN3	23,069,451.4	98.9	89.4	704,214.0	82.6	99.2
	HTS1	23,312,367.3	99.1	91.3	657,048.2	79.8	98.5
	HTS2	33,948,170.3	99.2	91.2	922,292.7	74.5	99.0
	HTS3	21,506,007.5	97.8	91.7	608,741.0	95.3	98.8
	LTS1	22,386,060.4	97.5	91.1	721,095.2	74.3	99.0
	LTS2	23,054,189.0	96.8	92.7	597,925.3	102.1	98.6
	sires	19,706,406.9	99.5	90.5	792,780.9	48.5	98.8
	dam	21,374,081.0	99.1	87.9	851,104.0	48.6	98.8
	grandfa ther	17,484,226.0	99.4	90.0	820,023.0	39.7	99.4
	grandm other	15,472,946.0	99.6	83.3	702,443.0	41.2	98.5
	HTN1	26,384,469.0	95.9	91.5	653,724.5	128.4	98.9
LT	HTN2	27,016,071.7	98.8	92.7	762,174.5	86.1	99.1
	HTN3	28,554,741.3	98.7	91.5	779,045.0	93.1	98.8
	HTS1	25,100,267.5	97.8	92.3	597,622.3	123.1	98.7
	HTS2	24,529,972.0	98.8	93.3	694,238.5	83.8	98.8
	HTS3	30,862,407.2	98.8	91.9	656,313.2	149.8	98.9
	LTS1	16,064,978.8	98.5	91.2	678,731.8	51.1	98.9
	LTS2	19,457,836.7	97.4	92.5	694,775.0	77.1	98.4
HT(N M)	sires	19,170,212.0	99.1	89.9	809,880.0	44.9	98.5
	HTN1	20,254,802.3	98.4	89.1	846,306.7	44.9	99.0
	HTN2	19,390,052.0	99.3	89.1	871,139.0	42.0	98.6
	HTN3	17,244,078.5	96.6	86.2	807,215.0	36.6	98.9
	HTS1	16,225,854.7	97.3	88.5	834,700.3	36.7	98.7
	HTS2	15,411,535.3	98.0	88.8	761,533.7	40.7	98.7
	HTS3	16,537,461.3	97.8	86.9	705,278.7	46.7	98.8
	LTS1	13,805,316.5	93.1	87.5	681,356.5	37.0	98.6
	LTS2	10,723,113.0	98.7	88.4	638,692.5	30.5	99.5
Total		22,459,963.9	98.2	90.7	719,627.2	76.4	98.9

Supplementary table 2. Information of the 260 gene transcripts that overlap with masculinization-associated differentially methylated regions (MA-DMRs)

Transcript ID	Genome browser description	Gene symbol	Type
DLAgn_00002290	novel sal-like protein	-	hypermethylated
DLAgn_00074550	receptor activity-modifying protein 3-like	RAMP3	hypermethylated
DLAgn_00083330	af4 fmr2 family member 1-like isoform x1	AFF1	hypermethylated
DLAgn_00147580	peroxisomal membrane protein 11c	PEX11G	hypermethylated
DLAgn_00155660	sal-like protein 1	SALL1	hypermethylated
DLAgn_00164710	seven in absentia homolog 1	SIAH1	hypermethylated
DLAgn_00164720	chk1 checkpoint-like protein	CHEK1	hypermethylated
DLAgn_00173630	protein spire homolog 2	SPIRE2	hypermethylated
DLAgn_00196890	germ cell-specific gene 1-like	GSG1L	hypermethylated
DLAgn_00210990	glycosyltransferase-like protein large1	LARGE	hypermethylated
DLAgn_00211000	zgc:165536 protein	OTUD4	hypermethylated
DLAgn_00211010	heat shock protein 30	HSP30	hypermethylated
DLAgn_00211080	transmembrane protease serine 9-like	-	hypermethylated
DLAgn_00211150	---NA---	-	hypermethylated
DLAgn_00211160	novel sal-like protein	-	hypermethylated
DLAgn_00211190	olfactory receptor 10j4-like	-	hypermethylated
DLAgn_00211200	zgc:165536 protein	OTUD4	hypermethylated
DLAgn_00211210	anti-streptococcal anti-lysoganglioside immunoglobulin heavy chain variable partial	-	hypermethylated
DLAgn_00211250	l-rhamnose-binding lectin csl2-like	SAL	hypermethylated
DLAgn_00211260	---NA---	-	hypermethylated
DLAgn_00211380	hemoglobin subunit alpha-2	ACTG1	hypermethylated
DLAgn_00211390	zgc:165536 protein	OTUD4	hypermethylated
DLAgn_00211410	zgc:165536 protein	OTUD4	hypermethylated
DLAgn_00211420	uncharacterized protein art2-like	-	hypermethylated
DLAgn_00211490	diverse immunoglobulin domain-containing protein partial	DICP3.3	hypermethylated
DLAgn_00211660	novel sal-like protein	-	hypermethylated
DLAgn_00211740	zinc finger protein 436- partial	-	hypermethylated
DLAgn_00211750	transmembrane protease serine 9-like	PRS27	hypermethylated
DLAgn_00211770	cell division cycle protein 27 homolog isoform 1	CDC27	hypermethylated
DLAgn_00212060	icos ligand-like isoform x1	-	hypermethylated
DLAgn_00213270	alpha-2-macroglobulin-like protein 1-like	A2ML	hypermethylated
DLAgn_00213280	---NA---	-	hypermethylated
DLAgn_00213480	immunoglobulin light chain	-	hypermethylated
DLAgn_00213560	phosphoenolpyruvate carboxykinase	PCK2	hypermethylated
DLAgn_00213600	uncharacterized protein art2-like	-	hypermethylated
DLAgn_00214560	---NA---	-	hypermethylated
DLAgn_00214710	zgc:165536 protein	OTUD4	hypermethylated
DLAgn_00214800	zgc:165536 protein	OTUD4	hypermethylated
DLAgn_00214840	intersectin-2-like	ITSN2	hypermethylated
DLAgn_00214850	---NA---	-	hypermethylated
DLAgn_00215910	reverse transcriptase-like protein	HEL_DR3	hypermethylated
DLAgn_00215920	uncharacterized protein art2-like	-	hypermethylated
DLAgn_00216030	---NA---	-	hypermethylated
DLAgn_00216060	---NA---	-	hypermethylated

DLAgn_00216080	---NA---	-	hypermethylated
DLAgn_00216340	immunoglobulin light chain precursor	-	hypermethylated
DLAgn_00216370	protein nlrc5	NLRC5	hypermethylated
DLAgn_00216740	protocadherin gamma-a8-like	-	hypermethylated
DLAgn_00217080	---NA---	-	hypermethylated
DLAgn_00217130	oocyte zinc finger protein 6-like	-	hypermethylated
DLAgn_00217390	sh3 domain-containing ysc84-like protein 1	SH3YL1	hypermethylated
DLAgn_00217500	synaptojanin-1-like isoform x1	SYNJ1	hypermethylated
DLAgn_00218360	stromal interaction molecule 1-like	STIM1A	hypermethylated
DLAgn_00218500	dachshund homolog 2-like	DACH2	hypermethylated
DLAgn_00218770	myosin heavy chain	MYHM86-1	hypermethylated
DLAgn_00218800	nuclear factor ovary-like	-	hypermethylated
DLAgn_00218860	---NA---	-	hypermethylated
DLAgn_00218940	reverse transcriptase-like protein fragile x mental retardation syndrome-related protein 1-like	YO84 FXR1	hypermethylated
DLAgn_00219100	olfactory receptor 10j4-like	-	hypermethylated
DLAgn_00219220	immunoglobulin light chain precursor	-	hypermethylated
DLAgn_00219420	gametogenetin binding protein 2	GGNBP2	hypermethylated
DLAgn_00219690	secretory phospholipase a2 receptor-like	-	hypermethylated
DLAgn_00219900	sodium hydrogen exchanger 5	SLC9A5	hypermethylated
DLAgn_00219960	uncharacterized protein loc101169743	-	hypermethylated
DLAgn_00220430	semaphorin-7a-like isoform x1	-	hypermethylated
DLAgn_00220490	novel sal-like protein	-	hypermethylated
DLAgn_00220680	uncharacterized protein loc101482357	-	hypermethylated
DLAgn_00220690	voltage-gated potassium channel subunit beta-3-like	KCNAB3	hypermethylated
DLAgn_00220760	ig kappa chain v-ii region rpmi 6410-like	-	hypermethylated
DLAgn_00222160	nuclease harbi1-like	-	hypermethylated
DLAgn_00222170	protein nlrc3-like	-	hypermethylated
DLAgn_00222400	histone h3c	HIST2H3C	hypermethylated
DLAgn_00222530	protocadherin-10-like	-	hypermethylated
DLAgn_00222730	uncharacterized protein partial low quality protein: rap guanine nucleotide exchange factor 6-like	-	hypermethylated
DLAgn_00223430	---NA---	-	hypermethylated
DLAgn_00223700	---NA---	-	hypermethylated
DLAgn_00223920	myst histone acetyltransferase 1	KAT8	hypermethylated
DLAgn_00223930	protein nlrc3-like	-	hypermethylated
DLAgn_00226020	cytochrome c oxidase subunit mitochondrial precursor	-	hypermethylated
DLAgn_00226300	translation initiation factor if-2-like	-	hypermethylated
DLAgn_00226310	olfactory receptor 10j4-like	-	hypermethylated
DLAgn_00226320	trace amine-associated receptor 7e-like	-	hypermethylated
DLAgn_00226680	---NA---	-	hypermethylated
DLAgn_00227080	myosin heavy fast skeletal muscle-like	-	hypermethylated
DLAgn_00227320	cd82 antigen	-	hypermethylated
DLAgn_00227370	chk1 checkpoint-like protein	CHEK1	hypermethylated
DLAgn_00227520	protein fam26e-like scavenger receptor cysteine-rich type 1 protein m130- like	FAM26E	hypermethylated
DLAgn_00227530	novel sal-like protein	-	hypermethylated
DLAgn_00227650	novel sal-like protein	-	hypermethylated
DLAgn_00228130	calcium and integrin-binding family member 2-like	CIB2	hypermethylated

DLAgn_00228380	transcription factor hes-5-like	HER12	hypermethylated
DLAgn_00229300	olfactory receptor 10a6-like	-	hypermethylated
DLAgn_00230030	olfactory receptor 10j4-like	-	hypermethylated
DLAgn_00230230	ig heavy chain v-iii region vh26-like	-	hypermethylated
DLAgn_00230300	rab proteins geranylgeranyltransferase component a 2	CHM	hypermethylated
DLAgn_00230770	immunoglobulin light chain	-	hypermethylated
DLAgn_00230840	protein nlrc3-like	-	hypermethylated
DLAgn_00230850	---NA---	-	hypermethylated
DLAgn_00230860	---NA---	-	hypermethylated
DLAgn_00231150	PREDICTED: hypothetical protein LOC795321 isoform 1 [Danio rerio]	-	hypermethylated
DLAgn_00231770	extracellular calcium-sensing receptor-like	-	hypermethylated
DLAgn_00231900	e3 ubiquitin-protein ligase trim21-like	-	hypermethylated
DLAgn_00231950	cytochrome c oxidase subunit iii	COX3	hypermethylated
DLAgn_00231960	tetratricopeptide repeat protein 28-like	-	hypermethylated
DLAgn_00232030	lrr and pyd domains-containing protein 12-like	-	hypermethylated
DLAgn_00232080	fc receptor-like b- partial	-	hypermethylated
DLAgn_00232400	fc receptor-like protein 5-like	-	hypermethylated
DLAgn_00232810	laminin subunit gamma-1-like	LAMC1	hypermethylated
DLAgn_00232880	ig kappa chain v-ii region rpmi 6410-like	-	hypermethylated
DLAgn_00233000	thymidine phosphorylase-like	TYMP	hypermethylated
DLAgn_00233590	protein nlrc3-like	-	hypermethylated
DLAgn_00233970	---NA---	-	hypermethylated
DLAgn_00234440	protein nlrc3-like	-	hypermethylated
DLAgn_00235170	uncharacterized protein mj0240-like	-	hypermethylated
DLAgn_00236020	gamma-glutamyltransferase 5-like	GGT1	hypermethylated
DLAgn_00236630	extracellular calcium-sensing receptor-like	OLFCD2	hypermethylated
DLAgn_00237600	gtpase imap family member 8- partial	-	hypermethylated
DLAgn_00237620	anion exchange protein 2-like	SLC4A3	hypermethylated
DLAgn_00237630	mam and ldl-receptor class a domain-containing protein c10orf12-like	-	hypermethylated
DLAgn_00237700	ig kappa chain v-iv region sth	-	hypermethylated
DLAgn_00238040	immunoglobulin heavy chain variable region	-	hypermethylated
DLAgn_00238070	transposase	-	hypermethylated
DLAgn_00238980	uncharacterized protein art2-like	-	hypermethylated
DLAgn_00239020	uncharacterized protein loc101167763	-	hypermethylated
DLAgn_00239630	novel immune-type receptor 30	NITR11	hypermethylated
DLAgn_00240750	---NA---	-	hypermethylated
DLAgn_00241340	t-cell receptor alpha chain	-	hypermethylated
DLAgn_00241350	uncharacterized protein art2-like	-	hypermethylated
DLAgn_00241360	uncharacterized protein loc101071402	WDFY4	hypermethylated
DLAgn_00241550	elav-like protein 1-like isoform 3	HUG	hypermethylated
DLAgn_00241620	lrr and pyd domains-containing protein 12-like	-	hypermethylated
DLAgn_00242620	protein nlrc3-like	-	hypermethylated
DLAgn_00243070	bile salt-activated lipase-like	CEL	hypermethylated
DLAgn_00243140	uncharacterized protein loc101483497	-	hypermethylated
DLAgn_00243880	gtp-binding nuclear protein	VANGL2	hypermethylated
DLAgn_00244390	leucine-rich repeat-containing protein 45	LRRC45	hypermethylated
DLAgn_00244550	reverse transcriptase-like protein	-	hypermethylated

DLAgn_00244670	ubiquitin carboxyl-terminal hydrolase cyld-like	CYLD	hypermethylated
DLAgn_00244760	tubulin-specific chaperone e-like	TBCE	hypermethylated
DLAgn_00244770	loc553308 protein	-	hypermethylated
DLAgn_00245030	t cell receptor alpha chain	-	hypermethylated
DLAgn_00245100	ribonuclease p protein subunit p29	POP4	hypermethylated
DLAgn_00245110	bromodomain containing 2	BRD2A	hypermethylated
DLAgn_00245550	hemoglobin subunit beta-2	-	hypermethylated
DLAgn_00245630	tubulin alpha-1c chain	TUBA1C	hypermethylated
DLAgn_00245710	uncharacterized protein loc101477029	-	hypermethylated
DLAgn_00246090	disintegrin and metalloproteinase domain-containing protein 17-like	ADAM17A	hypermethylated
DLAgn_00246410	protein nlr3-like	-	hypermethylated
DLAgn_00246670	general transcription factor ii-i repeat domain-containing protein 2-like	GTF2IRD2B	hypermethylated
DLAgn_00247120	histone-lysine n-methyltransferase suv420h2	SUV420H2	hypermethylated
DLAgn_00247130	zinc bed-type containing 1	ZBED1	hypermethylated
DLAgn_00247340	pol polyprotein	-	hypermethylated
DLAgn_00247680	---NA---	-	hypermethylated
DLAgn_00248140	neural cell adhesion molecule l1-like	-	hypermethylated
DLAgn_00248810	cytochrome p450 2j2-like	CYP2P3	hypermethylated
DLAgn_00249720	uncharacterized protein loc101480956	-	hypermethylated
DLAgn_00249730	sorting nexin-6-like	SNX6	hypermethylated
DLAgn_00249980	retrotransposable element tf2 155 kda protein type 1-like	-	hypermethylated
DLAgn_00250240	probable ubiquitin carboxyl-terminal hydrolase faf-x	USP9X	hypermethylated
DLAgn_00250270	inositol monophosphatase	IMPA1	hypermethylated
DLAgn_00250380	immunoglobulin light chain precursor	-	hypermethylated
DLAgn_00250450	2-acylglycerol o-acyltransferase 2-a-like	DGAT2L1-A	hypermethylated
DLAgn_00250540	protein nlr3-like	-	hypermethylated
DLAgn_00250570	membrane-spanning 4-domains subfamily a member 4a-like	-	hypermethylated
DLAgn_00250590	t cell receptor alpha	-	hypermethylated
DLAgn_00251040	zinc finger mym-type protein 1-like	ZMYM1	hypermethylated
DLAgn_00251690	p2y purinoceptor 4-like	-	hypermethylated
DLAgn_00252120	immunoglobulin light chain	-	hypermethylated
DLAgn_00252350	butyrophilin-like protein 2-like	-	hypermethylated
DLAgn_00252500	zinc finger swim domain-containing protein kiaa0913-like	ZSWIM8	hypermethylated
DLAgn_00252820	transposable element tc1 transposase	-	hypermethylated
DLAgn_00252830	tripartite motif-containing protein 16-like	-	hypermethylated
DLAgn_00253160	ig kappa chain v-ii region rpmi 6410-like	-	hypermethylated
DLAgn_00253220	myosin heavy chain	MMYHL1	hypermethylated
DLAgn_00254460	immunoglobulin light chain	-	hypermethylated
DLAgn_00254510	rna-directed dna polymerase from mobile element jockey-like	PTPRZ1	hypermethylated
DLAgn_00254610	loc562064 protein	CAMK2B1	hypermethylated
DLAgn_00254620	---NA---	-	hypermethylated
DLAgn_00254750	uncharacterized protein loc101156939	-	hypermethylated
DLAgn_00256500	---NA---	-	hypermethylated
DLAgn_00256520	tcr beta chain	-	hypermethylated

DLAgn_00256530	protein nlrc3-like	-	hypermethylated
DLAgn_00257060	protein jtb	JTB	hypermethylated
DLAgn_00257080	lily-type lectin	PUFFERIN	hypermethylated
DLAgn_00257170	PREDICTED: hemicentin-2-like [Maylandia zebra]	-	hypermethylated
DLAgn_00257180	---NA---	-	hypermethylated
DLAgn_00257190	interleukin-18 receptor 1	IL1RL1	hypermethylated
DLAgn_00257200	mhc class i antigen	-	hypermethylated
DLAgn_00257260	nuclease harbi1-like	-	hypermethylated
DLAgn_00257350	---NA---	-	hypermethylated
DLAgn_00257700	protein nlrc3-like	-	hypermethylated
DLAgn_00257710	protein nlrc3-like brain-specific angiogenesis inhibitor 1-associated protein 2-like	K31 BAIAP2	hypermethylated
DLAgn_00258250	b-cell receptor cd22-like	-	hypermethylated
DLAgn_00258280	tetratricopeptide repeat protein 28-like	-	hypermethylated
DLAgn_00258750	lim domain-binding protein 1-like	LDB1	hypermethylated
DLAgn_00259070	mhc class i antigen	-	hypermethylated
DLAgn_00259180	kelch-like protein 18	KLHL18	hypermethylated
DLAgn_00259360	butyrophilin-like protein 2-like	-	hypermethylated
DLAgn_00259370	zgc:165536 protein	OTUD4	hypermethylated
DLAgn_00259400	zgc:165536 protein	OTUD4	hypermethylated
DLAgn_00259710	nucleic acid binding protein	-	hypermethylated
DLAgn_00259800	diverse immunoglobulin domain-containing protein alpha-ketoglutarate-dependent dioxygenase alkb homolog 2-like	DICP3.3 ALKBH2	hypermethylated
DLAgn_00260330	reverse transcriptase-like protein	-	hypermethylated
DLAgn_00260630	protein nlrc3-like	-	hypermethylated
DLAgn_00260640	myosin heavy chain	MMYHL2	hypermethylated
DLAgn_00260930	protocadherin beta-16-like	-	hypermethylated
DLAgn_00261190	ig heavy chain v-iii region hil-like	-	hypermethylated
DLAgn_00261640	protein nlrc3-like	-	hypermethylated
DLAgn_00261880	gtpase imap family member 8- partial	-	hypermethylated
DLAgn_00262260	protein nlrc3-like	-	hypermethylated
DLAgn_00262370	c-type lectin domain family 4 member e-like	C209D	hypermethylated
DLAgn_00262450	immunoglobulin light chain	-	hypermethylated
DLAgn_00262460	hepcidin-like precursor	-	hypermethylated
DLAgn_00262470	gtpase imap family member 4- partial acidic leucine-rich nuclear phosphoprotein 32 family member b	- AN32B	hypermethylated
DLAgn_00262870	uncharacterized protein loc101479447 isoform x2	-	hypermethylated
DLAgn_00263070	uncharacterized protein loc101481070 arf-gap with sh3 ank repeat and ph domain-containing protein 2-like	- ASAP2	hypermethylated
DLAgn_00263100	b-cell receptor cd22-like	-	hypermethylated
DLAgn_00263670	zinc finger protein 161 homolog	ZBTB14	hypermethylated
DLAgn_00263680	---NA---	-	hypermethylated
DLAgn_00263690	---NA---	-	hypermethylated
DLAgn_00263700	---NA---	-	hypermethylated
DLAgn_00263740	interferon-induced very large gtpase 1-like	-	hypermethylated
DLAgn_00264060	echinoderm microtubule-associated 6-like	EML6	hypermethylated

DLAgn_00264080	---NA---	-	hypermethylated
DLAgn_00264640	lrr and pyd domains-containing protein 3-like	-	hypermethylated
DLAgn_00264760	atpase mitochondrial precursor	ATPIF1B	hypermethylated
DLAgn_00264770	60s ribosomal protein l13a	-	hypermethylated
DLAgn_00265000	---NA---	-	hypermethylated
DLAgn_00265710	von willebrand factor a domain-containing protein 5a-like	VWA5A	hypermethylated
DLAgn_00265720	uncharacterized protein loc101170193	-	hypermethylated
DLAgn_00265750	uncharacterized protein loc101071402	-	hypermethylated
DLAgn_00265830	urokinase plasminogen activator surface receptor-like	PLAUR	hypermethylated
DLAgn_00266890	uncharacterized protein loc101472247	-	hypermethylated
DLAgn_00266900	ac1147-like protein	-	hypermethylated
DLAgn_00267170	butyrophilin subfamily 1 member a1-like	-	hypermethylated
DLAgn_00014310	dna polymerase eta	POLH	hypomethylated
DLAgn_00017210	d-like dopamine receptor-like	-	hypomethylated
DLAgn_00038120	heterogeneous nuclear ribonucleoproteins c1 c2-like	HNRNPC	hypomethylated
DLAgn_00053770	frizzled homolog 8c	FZD5	hypomethylated
DLAgn_00056040	leucine-rich repeat-containing protein 72	LRRC72	hypomethylated
DLAgn_00139870	gamma-crystallin m2-like	CRYGS3	hypomethylated
DLAgn_00163100	deubiquitinating protein vcip135	VCPIP1	hypomethylated
DLAgn_00174860	multidrug resistance-associated protein 7-like	ABCC10	hypomethylated
DLAgn_00175010	rho-related gtp-binding	RHOA	hypomethylated
DLAgn_00194660	protein fam86a-like	FAM86A	hypomethylated
DLAgn_00205050	uncharacterized protein loc101486585	-	hypomethylated
DLAgn_00205060	---NA---	-	hypomethylated
DLAgn_00212510	p2y purinoceptor 14-like	P2RY14	hypomethylated
DLAgn_00215150	neural-cadherin-like	-	hypomethylated
DLAgn_00222530	protocadherin-10-like	-	hypomethylated
DLAgn_00237870	pleiotrophic factor-alpha-2 precursor	MDKB	hypomethylated
DLAgn_00238800	protein nlrc3-like	-	hypomethylated
DLAgn_00242560	carcinoembryonic antigen-related cell adhesion molecule 3-like isoform x1	-	hypomethylated
DLAgn_00260800	dickkopf-related protein 2-like	DKK2	hypomethylated
DLAgn_00260810	---NA---	UBE2Q2	hypomethylated

Acknowledgements

We would like to thank Dr Ricard Argelaguet for his invaluable advice on MOFA as well as for fostering a supportive community through available tutorials, Slack group and GitHub.

Author contributions

Núria Sánchez Baizán: Methodology, Investigation, Formal analysis, Writing – original draft. Johana Denkena: Methodology, Formal analysis. François Allal and Marc Vandeputte: Provided the genotyping data, Conceptualization, Resources, Funding acquisition. Francesc Piferrer: Conceptualization, Resources, Funding acquisition, Supervision, Investigation, Writing – original draft.

Funding

The study was supported by the AQUAEXCEL2020 Transnational access grant “Transsexbass” to F.P. and M.V. Research at the F.P. laboratory was supported by Spanish Ministry of Science Grant No. PID2019-108888RB-I00 and with funding from the Spanish government through the ‘Severo Ochoa Centre of Excellence’ accreditation (CEX2019-000928-S). NS was supported by a Spanish Ministry of Science and Innovation predoctoral scholarship (BES-2017-079744).

Availability of data

The data associated with this study will be made available upon publication through the NCBI platform.

Competing interests

The authors declare they have no competing interest

General discussion

General discussion

The aim of this thesis was to investigate the transcriptome, epigenome (methylome), and the impact of intrinsic (sex) and environmental factors (temperature) in the European sea bass (*Dicentrarchus labrax*). The title emphasizes the development of DNA methylation biomarkers, the study epigenetic inheritance and the integration of environmental stimuli. However, to achieve this it was essential first to establish the background and thoroughly examine the genes involved in gonadal development under control conditions. Accordingly, the thesis is divided into two blocks. In the first block, comprising two chapters, the focus was on understanding gene expression in the gonads during early sex differentiation. The second block consists of three chapters dedicated to the development of epigenetic biomarkers. These chapters explore the differences in DNA methylation marks between males and females, as well as the effects of different temperature treatments during early development on DNA methylation and its transmission to the offspring.

In chapter 1, the main objective was to study the genes involved in sexual differentiation in sea bass. Nevertheless, when trying to obtain the most out of the transcriptome we found a way to improve the transcriptomic analysis. Thus, we propose a novel approach for analyzing it, combining weighted gene co-expression network analysis (WGCNA) with differential gene expression analysis (DEGs). By applying WGCNA to the entire dataset before filtering by DEGs, the approach yields improved network architecture, increased identification of key genes and modules associated with the trait of interest, enhanced representation of biological processes, and facilitates biomarker discovery. In addition to providing a comprehensive description of the sea bass transcriptome during early development, we validated our method by applying it to mouse and human transcriptomes. This validation confirmed that our approach is a suitable method for transcriptomic studies, irrespective of the biological systems, species, or research questions being investigated.

To delve deeper on the process of gonadal development, in chapter 2, we carefully selected data from other vertebrate species for comparison with the transcriptome of the sea bass. The aim was to investigate gonadal sex differentiation in vertebrates and characterize the transcriptome dynamics involved. The results revealed significant variability across species, with different proportions of genes being activated and repressed in males and females. Key genes like Doublesex and Mab-3 Related Transcription Factor 1 (*dmrt1*), SRY-Box Transcription Factor 9 (*sox9*), and Anti-Müllerian Hormone (*amh*) were consistently enriched only in males, while genes like Forkhead Box L2 (*foxl2*), Cytochrome P450 Family 19 Subfamily A Member 1 (*cyp19a1a*), and Follistatin (*fst*) were enriched only in females. The study also identified potential novel markers of early sex differentiation that were conserved across species.

In chapters 3 and 4, the aim was to identify epigenetic biomarkers. In chapter 3, DNA methylation marks were identified as indicators of the thermal environment during early development. The methylome and the transcriptome were combined to identify biomarkers with prognostic and diagnostic value. These type of biomarkers with prognostic value have implications for aquaculture and conservation biology, offering the potential to screen larval batches and anticipate the performance of juveniles and adults based on the quality of their thermal environment. In chapter 4, a new cost-effective method for epigenetic biomarker development and their integration with genetic selection in finfish aquaculture was developed. To overcome the challenge of having the high sample size required for the use of machine learning algorithms, a method was developed to provide a “short-cut” where less samples are required at the stage of genome screening. As a result, we identified three specific CpGs that predicted the best fish for aquaculture production based on biomass, sex proportion, and resistance to the masculinizing

effects of high temperature. These CpGs demonstrated high heritability and were successfully validated in larger sample sets and future broodstock.

In chapter 5, the impact of sex and temperature on the methylome of the sea bass was explored. In this study, the results revealed multigenerational effects of temperature and provided insights into the epigenetic mechanisms underlying masculinization. The study also investigated the transmission of epigenetic marks from sires to their offspring and identified potential markers of genetic females resistant to temperature-induced masculinization.

Overall these findings have implications for aquaculture and expand our understanding of adaptive mechanisms in response to environmental changes. In the following sub-sections, we discuss some of the considerations and limitations to be taken into account. Additionally, we present an integrative review of all the results. Finally, we explore some open questions, future perspectives and ideas for future research.

Critical assessment of the methods used

In recent years, the relationship between DNA methylation, gene expression, and phenotypes has led to the development of DNA methylation-based epigenetic markers (Jin and Liu, 2018). In the field of medicine, the methylation status of certain genes has been associated with various diseases (Jones and Baylin, 2007). As a result, specific CpGs' methylation levels have been utilized in the development of diagnostic and prognostic tests for diseases (Costa-Pinheiro *et al.*, 2015). From that, the exploration of epigenetic biomarkers gained significant attention. Some markers have been linked with age, lifestyle, or environmental factors (Marsit, 2015), and have been developed to predict age in humans and other vertebrates, including fish (Anastasiadi and Piferrer, 2020; Bors *et al.*, 2021; Horvath, 2013; Prado *et al.*, 2021). They have also been used to predict sex in fish (Anastasiadi *et al.*, 2018b; Valdivieso *et al.*, 2023a), domestication level (Anastasiadi and Piferrer, 2019), early life stress in chickens (Pétille *et al.*, 2017), and thermal regime in zebrafish (*Danio rerio*) (Valdivieso *et al.*, 2023a, 2023b).

Recently, several studies have highlighted the potential of EBs to enhance breeding strategies of plants and animals in the near future (Alfalahi *et al.*, 2022; Crisp *et al.*, 2022; Kim *et al.*, 2022; Lempe *et al.*, 2022). However, prior to the widespread adoption of these markers, it remains crucial to establish the economic feasibility of large-scale implementation. Furthermore, it is unclear how these markers will persist across generations within breeding programs or in the wild. Hence, as additional epigenetic biomarkers are discovered, there is the need to concurrently investigate their outcomes at large scale, both in animal production and conservation applications.

The limitations of each study were discussed separately within each chapter, but here we will mention some general limitations. For example, the compromise between the amount of data being sequenced (considering the time and costs of sequencing and its analysis) and the number of samples to process (Zhang *et al.*, 2023). In our studies, we decided to use Reduced Representation Bisulfite Sequencing (RRBS) or Multiplex Bisulfite Sequencing (MBS) instead of Whole Genome Bisulfite Sequencing (WGBS). The cost per sample in RRBS is 2.5 and in MBS 5 times cheaper than WGBS. Taking into account that we did a total of 194 RRBS libraries in this thesis, the savings is substantial. Thus, by using these approaches we were able to sequence a large number of samples while generating minimal but highly informative data. The downside was that there were cases where we could not capture important information from target regions. For example, when doing RRBS, some of the key sex-related genes such as *cyp19a1a* or *dmrt1* were sometimes not captured. Nevertheless, RRBS provided an average of 570,000 CpGs per sample (**Figure 1**), which represents 5.5% of the total number of CpGs (10,325,421)

in the sea bass reference genome (dicLab v1.0c, July 2012). Moreover, the aim of this thesis was to conduct a comprehensive genome-wide screening, which serves as an essential initial step in directing subsequent investigations towards specific regions or genes that warrant further exploration. Genome-wide screening studies primarily rely on association rather than establishing direct cause-and-effect relationships. Nevertheless, these studies are useful in guiding subsequent detailed investigations that delve into the elucidation of gene regulatory networks (Graham and Root, 2015). Subsequent functional studies can employ other techniques, including gene editing or, more recently, epigenome editing, to further elucidate the functional roles of these previously filtered candidates (Graham and Root, 2015; Takahashi *et al.*, 2023).

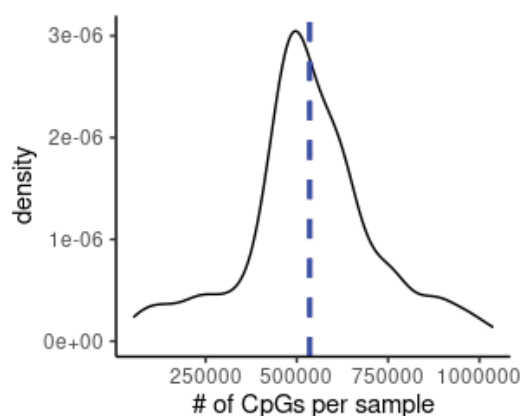


Figure 1. Density plot of the number of CpGs captured for each of the 130 samples analyzed by RRBS in chapter 5.

Another aspect to be considered, is the selection of methods based on bisulfite conversion reaction. Bisulfite conversion (BC) is considered the “gold standard” method to study DNA methylation because it allows obtaining improved sensitivity, coverage and resolution of DNA methylation detection coupled with standard sequencing technologies.

Nevertheless, this technique has some limitations: 1) bisulfite treatment drastically damages DNA (Grunau *et al.*, 2001), which causes high requirements of starting DNA concentration, 2) induces Polymerase Chain Reaction (PCR) bias, which is that preferential amplification of unmethylated alleles occurs. However, this could be corrected with strict filtering parameters and normalization algorithms designed for BC data (Akalın *et al.*, 2012b; Moskalev *et al.*, 2011).

Recently, a new technique called Enzymatic Methyl-seq (EM-seq) was developed by New England Biolabs to overcome some of these limitations (Gouil and Keniry, 2019; Sun and Zhu, 2022; Vaisvila *et al.*, 2021). EM-seq consists in the performance of two sets of enzymatic reactions. The first reaction uses Tet Methylcytosine Dioxygenase 2 (TET2) and T4 Phage β -glucosyltransferase (T4-BGT) enzymes to convert 5-methylcytosine (5mC) and 5-Hydroxymethylcytosine (5hmC) into products that are resistant to deamination by Apolipoprotein B MRNA Editing Enzyme Catalytic Subunit 3A (APOBEC3A). In the second reaction, APOBEC3A converts unmodified cytosine to uracil through deamination. Through the combination of these three enzymes, the same sequence as when using BC treatment is produced but without the use of harsh chemicals (Vaisvila *et al.*, 2021). Thus, with the continuous development of methods and sequencing technologies it is likely that significant improvement on the techniques will occur during the next years (Sun and Zhu, 2022). Particularly, the future perspectives are found towards the emerging long-read sequencing methods which provide promising opportunities to investigate a variety of modifications to existing methods with the inherent benefits of single-molecule sequencing and longer reads (Gouil and Keniry, 2019). For a recent review on the advances in measuring DNA methylation see (Sun and Zhu, 2022).

Another limitation of this thesis is that the majority of studies were focused on the thorough analysis of a single “omic” or layer, with the exception of chapter 3. To complement transcriptomic data with DNA methylation information and vice versa provides many interesting insights that greatly enhance our

understanding of the interplay between DNA methylation and gene expression in shaping phenotypic outcomes (Hawkins *et al.*, 2010; Minton, 2023). Furthermore, by studying multiple layers of information from the same individuals, it enables the utilization of multi-omics analyses. Such approaches integrate data from different layers to pinpoint information that would otherwise be overlooked (Argelaguet *et al.*, 2020, 2018).

The following aspects to be discussed involve some of the decisions taken during the experimental design: the temperature range and the tissues studied. The temperature range used in the experiments of chapter 4 and 5 was chosen because in aquaculture the temperature used to rear larvae is 21°C. Further, this protocol is well-known from previous research on the sea bass to induce masculinization (Anastasiadi *et al.*, 2018b; Blázquez *et al.*, 1998; Navarro-Martín *et al.*, 2011, 2009a). Therefore, while the temperature treatment initially selected was not chosen for its direct biological relevance in the wild, it is increasingly apparent that the ongoing global warming phenomenon, prompts us to reconsider the possibility of observing similar effects in the natural environment, at least to some extent. In the Mediterranean, for example, significant temperature range shifts have been projected, which are expected to rise 1°C by 2050, and 1.9°C by 2100 (Cascarano *et al.*, 2021). Also, regarding heat-waves, in the Mediterranean there is prediction of increase in the length, the number of events and the increase of temperature (Molina *et al.*, 2020), which could coincide with the thermosensitive period of the developing sea bass in the wild. Thus, the relevance of studying the observed effects of high temperature gains importance in light of these changing environmental circumstances.

Regarding the tissues used to develop EBs, the principal limitation is that in some instances they were developed using tissues that require the sacrifice of the individuals to access them (chapter 3, whole larvae, body trunk and gonad; chapter 5, gonads). However, it is important to further explore the development of EBs in non-invasive tissues. In chapter 4, for example, we focused our efforts on validating the EBs using sperm samples due to their non-lethal nature and relative ease of manipulation. To continue the development of EBs considering no-lethal tissues, one potential avenue is to continue seeking for metastable epialleles (MEs) more extensively (Anastasiadi *et al.*, 2021a). A ME is a locus, the methylation of which can vary even in genetically identical individuals, but responds to environmental cues in the same way in all tissues within a given individual (Piferrer, 2023). Additionally, a promising direction is to consider that while each tissue has its specific epigenome, some observations suggest the existence of environmentally induced responses shared by multiple tissues (Bysani *et al.*, 2017; Ewald *et al.*, 2014; Rönn *et al.*, 2015; Stueve *et al.*, 2017). Therefore, the ability to infer epigenetic patterns in internal tissues from liquid tissues like sperm or blood would be highly beneficial, as it would enable sequential and non-lethal sampling from the same individual.

Another aspect that warrants critical assessment is the number of loci utilized for predicting traits in EBs development. It is well-established that combining methylation levels of CpGs from different genes into a single panel enhances the sensitivity and specificity of diagnosis (Lind *et al.*, 2011). In fish, some studies have used the levels for specific CpGs to facilitate the identification of Essential Epigenetic Marks (EEMs) associated with a particular phenotype of interest (Piferrer *et al.*, 2019), which enabled its prediction using machine learning algorithms. **Table 1** provides a summary of the number of loci employed in predicting phenotypic traits in fish using a panel of CpGs, along with the accuracy achieved in other studies and in this thesis. In the table, only some instances of EBs to predict age in fish species are shown, but as reviewed elsewhere, in piscine epigenetic clocks the number of loci ranges from 26 to 199 CpGs, with median number of CpGs of 30 and mean accuracy of 93%. In the rest of studies (aimed at prediction of traits other than age), the range of CpGs used was from 15 to 26, and the accuracy was from 71 to 90%.

Comparatively, in chapter 4, a much lower number of CpGs was employed while the accuracy was amongst the highest. This can be justified by two aspects. On one hand, the techniques employed for the development of EBs for prediction of sex or thermal regime in other species was performed using targeted-based techniques. Hence, the selection of the EEM was limited to a number of starting targeted regions. In Chapter 4, the selection of EEMs was carried out through a genome-wide screening, enabling the identification of the most informative features at the global genomic level. Consequently, no additional information from other loci was required. On the other hand, this explanation does not hold true when comparing it to the estimation of age in fish using EEMs, as these were developed based on prior genome screening. Age estimation is a unique trait as it involves varying levels of methylation at specific loci throughout an individual's lifespan due to the gradual decrease in DNA methyltransferase activity over time (Y. Li *et al.*, 2010). Therefore, it is a trait characterized by continuous demethylation without a clearly defined threshold. Moreover, the algorithms utilized for age prediction, based on regression, need a minimum number of samples and time-points to achieve accuracy and statistical power (Anastasiadi and Piferrer, 2023). In contrast, the support vector machine algorithm, employed in Chapter 4, has demonstrated statistical reliability even with the use of a limited number of CpGs, similar to approaches employed in cancer research (Das *et al.*, 2020; Huang *et al.*, 2018; Zhi *et al.*, 2018).

Table 1. Compilation of epigenetic biomarkers for prediction of phenotypic traits in fish using machine learning algorithms

Species	Prediction	N° of CpGs	Accuracy (%)	Reference
<i>Dicentrarchus labrax</i>	Growth and sex ratio	3	88-100	Chapter 4 (<i>In preparation</i>)
	Resistance to masculinization	Undergoing further analysis and evaluation of prediction		Chapter 5 (<i>In preparation</i>)
	Sex	23	90	(Anastasiadi <i>et al.</i> , 2018b)
	Age	28	82	(Anastasiadi and Piferrer, 2020)
<i>Epinephelus morio</i>	Age	49	98	(Weber <i>et al.</i> , 2022)
<i>Danio rerio</i>	Age	29	92	(Mayne <i>et al.</i> , 2020)
	Sex	15	88	(Valdivieso <i>et al.</i> , 2023a)
	Past thermal regime	15 (testes), 26 (ovaries)	71 and 78, respectively	

Strengths and impact of the results

Despite the limitations considered above, in all chapters of this thesis we identified a series of biomarkers for different applications. In **Table 2**, a summary of all the biomarkers related to the effects of sex and temperature according to their application and characteristics is shown. The integrative view of these results allowed us to compare them with other EBs developed in the sea bass.

In our studies, the number of samples used to develop EBs ranged from 45 to 140, a similar range of samples used in other studies. For example, the EBs developed to assess the early stages of domestication (Anastasiadi and Piferrer, 2019) were developed using 24 samples (analyzed by RRBS) and validated with 16 more samples. Hence, using a total of 40 samples. Similarly, in the EBs developed to predict sex, 94 samples were used to analyze six of the genes and 183 samples were used to analyze methylation levels of *cyp19a1a*. Another of the strengths of the studies performed in this thesis for the development of biomarkers is that they stand out for the inclusion of high number of families, which

Table 2. Summary of all the biomarkers identified in this thesis related to effects of sex and temperature according to application and characteristics

Chapter	N° of biomarkers	Application	Name / description	Type	Tissue	Stage	Age (dpf)	n° of samples
1	1	Early testis differentiation	SRY-box transcription factor 9a (<i>sox9a</i>)	T				
	2,089	Early gonad differentiation (testis and ovaries)	Genes within the network module significantly associated with sex which additionally were DE between sexes at 110 dpf	T	gonad	J	110	22
2	16	Early gonad differentiation (testis and ovaries)	Genes within the top 10% most expressed in differentiating gonads and found in common in 6 vertebrates, including sea bass	T	gonad	J	110-250	22
3	210	Short-term effects of LT and HT	Gene DE between temperature treatments and with at least one DMC between 14°C vs. 17°C and 20°C vs. 17°C	B	whole larvae	L	08-12	12 pools of 10 larvae
3	4	Mid-term effects of LT or HT	Genes DE between temperature treatments and with at least one DMC between 14°C vs. 17° or between 20°C vs. 17°C	B	whole larvae & body trunk	L & J	8-12 & 90	12 pools of 10 larvae + 12 pools of 4 fish
	4	Long-term effects of LT or HT	Genes DE between temperature treatments and with at least one DMC between 14°C vs. 17° or between 20°C vs. 17°C	B	whole larvae & testis	L & A	8-12 & 482	12 pools of 10 larvae + 12 individual testis
4	3	To select fish for growth and resistance to HT	CpG13, CpG31, CpG30	E	sperm and gonad	A	402	140
5	5	CpGs indicators of masculinization	5 CpGs found at promoters			A		45
5	260	Candidates of sex determination and indicators of masculinization	Genes overlapping with MA-DMRs	E	gonad		392	
						A		14

Abbreviations: days post fertilization (dpf), differentially expressed (DE), differentially methylated cytosines (DMC), masculinization-associated differentially methylated regions (MA-DMRs), low temperature (LT), high temperature (HT), transcriptomic (T), epigenetic (E), both (B), larvae (L), juvenile (J), adult (A)

allowed considering more biological variability. In chapter 3 we used 12 pools of 10 larvae, 12 pools of 4 juveniles and testis from 12 adults, which were obtained from mass spawn of captive breeders. In chapter 4, up to 59 families were used to develop the EBs and in chapter 5, a total of 8 different families. However, in previous studies the number of families was not considered (Mayne *et al.*, 2020), or was four at the most (Anastasiadi *et al.*, 2018b; Valdivieso *et al.*, 2023a).

Besides the comparison to other studies, we wanted to know if there were any of the EBs related to temperature effects that could coincide with the list of genes overlapping with Masculinization-associated differentially methylated regions (MA-DMRs). Interestingly, we found that one gene overlapping with MA-DMRs coincided with one out of the nine previously identified MEs, which on top was indicator of very long-term effects of temperature in the European sea bass (Anastasiadi *et al.*, 2021a; Valdivieso *et al.*, 2023b). On one hand, that gene is an uncharacterized protein in the first version of the genome, but in the newest version of the reference genome (V.2, 2021), it appeared as stem-loop binding protein 2 (*slbp2*). This gene is closely linked with oocyte development in several vertebrates, including mammals (Thelie *et al.*, 2012), amphibians (Liu *et al.*, 2015; Wang *et al.*, 1999) and fish (He *et al.*, 2018). During zebrafish oogenesis, the maternal *slbp2* protein plays a crucial role in the storage of replication-dependent histone mRNAs and proteins (He *et al.*, 2018). In a study using the knockout CRISPR/Cas9 approach in zebrafish, it was confirmed that the expression changes of replication-dependent histones (necessary for packaging the newly replicated chromosomal DNA) were influenced by *slbp2*. On the other hand, the identification of an overlap between MA-DMRs and a gene containing a ME presents an exciting opportunity to develop an epigenetic biomarker that can be analyzed in non-invasive tissues. This finding is an example of how the biomarkers identified in this thesis have direct applications but also pave the way for future investigation.

Another strength of this thesis is that, when possible, we always tried to compare the results of each specific chapter with those described in the literature for other species. This approach aligns with the emerging perspective emphasized in high-impact papers within biomedical research which underscore the importance of embracing an integrative view of biodiversity to comprehensively understand the genetic architecture underlying complex phenotypes. For instance, comparisons across hundreds of mammals revealed associations between noncoding sequences and exceptional traits in humans, such as brain size and vocal learning, as demonstrated by (Christmas *et al.*, 2023). Such strategy was particularly considered in chapter 2, where six vertebrate species, from fish to mammals with different sex determining system were used.

Consequently, we saw that in the mouse and the platyfish (coinciding with the two species with XX/XY sex determining system), active downregulation in males was necessary to achieve higher expression in females for about half of the differentially expressed genes in females. However, we did not observe similar transcriptomic dynamics in other species with different sex determining systems (*e.g.*, TSD, PSD, ZZ/ZW). These results support the importance of conducting further investigations across a broader range of species to gain a comprehensive understanding of this phenomenon.

Furthermore, another aspect that was considered when designing the goals of this thesis was to evaluate whether to respond to all questions it was really necessary to sequence and produce new data. Nowadays we are in the era of next-generation sequencing and the costs of sequencing are dropping daily. Yet, it has long been recognized that as the cost of sequencing continuously decreases, new challenges for data storage and its analysis are increasingly important (Muir *et al.*, 2016). Hence, to ensure a sustainable and efficient use of resources in research it is necessary to question the generation of new data. In this manner, it is advisable to begin by formulating the specific biological question that needs to be

addressed. Subsequently, examine the existing data that could be potentially used to achieve this objective. Finally, to evaluate if additional experiments are necessary to validate the obtained results. This was the idea behind chapters 1 and 2.

The final noteworthy aspect of this thesis regards to its contribution to the field of epigenetic inheritance. To date, the amount of information on intergenerational and transgenerational inheritance in animals is not abundant. Additionally, most of the studies were performed in invertebrate species or in the mouse, as reviewed elsewhere (Fitz-James and Cavalli, 2022; Perez and Lehner, 2019). In this section, we will compare the results of this thesis with a few of the existing studies that have specifically investigated DNA methylation inheritance in vertebrates. When attempting to investigate epigenetic inheritance in sea bass, our initial challenge arose from the fact that different ways to assess it exist and that no standard pipeline is available in the literature. To show that, we summarized some of the information in **Table 3**.

From the summary, we concluded that transgenerational epigenetic inheritance should no longer be a matter of debate, as Takahashi *et al.*, (2023) demonstrated that DNA methylation of promoter-associated CpG islands can be transmitted from parents to offspring, in a mouse model with an edited epigenome. However, whether environmentally-induced epimarks are inherited in animals still remains a subject of debate, the proportion of inheritance and the mechanisms remain unknown. Consequently, there is need for further research to shed light on this topic. In particular, it is essential to address two critical aspects: the consistent definition and usage of several key concepts throughout the existing literature, and the development of a standardized method to determine the inheritance of epigenetic marks.

Regarding the first aspect, we highlight three specific terms that lack clarity or are easy to mislead. The first is the concept of stable CpGs. Stable CpGs are loci the methylation of which is not differentially methylated between parents and subsequent generations (Hu *et al.*, 2021). However, it is important to keep in mind that similar levels do not imply inheritance. In fact, those similar levels of DNA methylation could be constitutive or being acquired by chance (Angers *et al.*, 2020; Massicotte *et al.*, 2011; Richards, 2008). The second term is the concept of inherited DMC/DMR. In the literature, the most common form to refer to these loci are those identified as DMC in both the exposed parental (F0) and the unexposed filial generations. However, it is necessary to specify whether the determination of DMCs was conducted using a statistical test with or without permutations. This clarification is necessary to discern whether the observed common DMCs could potentially be attributed to stochasticity (Belleau *et al.*, 2018). Finally, it is important to highlight the distinction between heritability and epigenetic inheritance, as these are different concepts. Heritability refers to the probability of a particular trait, including DNA methylation levels at a specific locus, being transmitted to subsequent generations (Plongthongkum *et al.*, 2014). Heritability estimates provide insights into the genetic basis of a trait and indicate the degree to which parental variation (genetic or epigenetic) contributes to phenotypic differences among individuals (Plongthongkum *et al.*, 2014).

On the other hand, epigenetic inheritance involves the transfer of non-genetic information. According to the number of generations where the effects of a stimulus can be found, there are different types of epigenetic inheritance: intragenerational and intergenerational (Piferrer, 2023). The intragenerational inheritance consists in the transmission of environmentally-induced changes that occur within the directly exposed generation (F0). These changes which may or may not be epigenetic are transmitted from mother to daughter cells during mitosis and can result into long-lasting phenotypic changes. The intergenerational inheritance occurs when the effects induced by a stimulus are found across two or more generations (*i.e.*, at F0 and F1 or more). In turn, intergenerational inheritance can be classified into

Table 3. Summary of studies related to epigenetic inheritance of animal species.

Species	Unit	Type	Tissue	Environmental factor	N° of samples	Method used to determine inheritance	Reference
	DMR	T	Sperm	Pesticide	80	Common DMC between generations and overlap with other omics	Beck <i>et al.</i> , 2021
Mammals	DMR	I & T	Sperm	Long-term psychological stress	36	Wilcoxon signed-rank tests of mean DNA methylation levels of candidate regions & unpaired t-tests of mean DNA methylation levels of DMRs	Zheng <i>et al.</i> , 2021
	single nucleotide resolution	T	Mouse tails, mouse embryonic limbs and stem cells	Acquired epigenetic marks through epigenetic editing	238	Unpaired Student's t-test of DNA methylation levels between wild-type and chimeric mouse	Takahashi <i>et al.</i> , 2023
	DMC	I & T	Sperm	Persistent organic pollutants	36	Common DMC between F0 and F1 & common DMCs with permutation test	Belleau <i>et al.</i> , 2018)
	Global methylation	T	Whole blood	Genistein injection	62	Shapiro–Wilk test of normality and two sample t test	Leroux <i>et al.</i> , 2017
Birds	DMC & DMR	H	Whole blood	NA	96	MPO- regression & meQTLs	Plongthongkum <i>et al.</i> , 2014
	DMC	H	Sperm & gonad	Temperature	95	MPO- regression	Chapter 4
	DMC	I	Sperm	Captivity	24	Common DMC between F0 and F1	Rodriguez Barreto <i>et al.</i> , 2019
	DMC	I & T	Gonads	Cadmium, temperature, density	77	MPO- regression	Pierron <i>et al.</i> , 2021
Fish	DMC	H	Caudal fin	Marine vs. freshwater	94	Stable CpGs & meQTLs	Hu <i>et al.</i> , 2021
	DMC & DMR	I	Sperm & gonad	Temperature	130	Stable CpGs, common DMCs between F0 and F1 & common DMCs with permutation test	Chapter 5

Abbreviations: intergenerational (I), transgenerational (T), heritability (H).

*Here we provide examples of some of the most compelling reports of intergenerational and transgenerational epigenetic inheritance. Other studies related to the topic of interest do exist

multigenerational or transgenerational. Multigenerational inheritance occurs when the effects of a stimulus are observed in all generations that are directly exposed to it. In a pregnant mammal, this includes F0, F1 and F2 generations. In non-pregnant female or male mammals as well as fish, it affects the F0 and F1 generations. These effects can be either epigenetic or non-epigenetic. In the contrary, transgenerational inheritance, which can only be epigenetic, can be considered when the effects induced by the stimulus in the first exposed generation are found in subsequent generations that were never exposed directly. Therefore, epigenetic inheritance can be considered when the effects can be observed up to at least the F3 in pregnant mammals or up to at least the F2 in non-pregnant mammals, males, or in fish (Piferrer, 2023).

As a result of this revision and considering the lessons learned from this thesis we propose that to study environmentally-induced epigenetic inheritance the following steps could be considered: 1) to quantify the proportion of CpGs with stable levels of methylation between F0 and subsequent generations (Hu *et al.*, 2021). 2) To identify if there are DNA methylation levels associated to SNPs using permutation tests such as *regioneR* package (Gel *et al.*, 2016) as done by Anastasiadi and Piferrer, (2019) or to perform methylation quantitative trait loci (meQTL) analysis, as done by Plongthongkum *et al.*, (2014) and Hu *et al.*, (2021). These approaches would aid to identify which part of the methylation variance is explained by genetics (Hu *et al.*, 2021). 3) To investigate the genotype and the transcriptome to reveal its interaction with the environment (Geffroy *et al.*, 2021). 4) To identify DMCs in common between the exposed F0 and unexposed following generations. 5) To identify which of those are not randomly acquired (Belleau *et al.*, 2018). Finally, the ultimate goal would be to build a model in which all information could be integrated in order to predict the phenotype of an individual from known ancestors and given a specific environmental factor. However, to achieve all these steps, a method applied in a consistent manner that can robustly determine the degree of inheritance of epimutations and their association with a given phenotype has yet to be developed.

Future directions and open questions

This section presents a range of ideas for potential research in the field and open questions. In the first block of the thesis we studied the genes involved in sexual differentiation and we were able to work with gonads from fish smaller than 8 cm at 110 dpf of age, enabling us to study sex differentiation of the sea bass. However, the study of sex determination of this species remains elusive. Partly, because it is difficult to sample gonadal tissue from smaller fish and also because as a species with PSD, it might be more difficult to capture multiple factors playing together than a single sex determining gene. One potential idea to investigate the genes and mechanisms involved in the early stages of gonadal development could be to perform isolation of the primordial germ cells (PGCs) using flow cytometry. Similar to what was done in the zebrafish, medaka and goldfish (Goto-Kazeto *et al.*, 2011). They used a construct of green fluorescent protein (gfp) and zebrafish nos 1 3'UTR mRNA to label PGCs in the sea bass. Then, by Fluorescence-activated cell sorting (FACS) we could analyze gene expression of these cells using single-cell RNA sequencing at several time-points to investigate the gene expression dynamics of the differentiating PGCs during sex determination of the sea bass. This approach, however, implies that the polygenic factors act on the fate of PGC and could allow to see changes in gene expression, but since isolating PGCs is necessarily lethal, it would not allow for a proper time course study during sexual development. Further, it would not capture the changes occurring in supporting cell lines, where, as in other species, are expressed the sex determining genes.

To continue research on the comparative transcriptome analysis across vertebrates during gonadal development two specific ideas are provided. First, we could perform the functional enrichment analysis of each species separately. Then we could explore newly developed bioinformatic tools such as

GOcompare, which provide quantitative procedures to compare Gene Ontology terms among species (Sosa *et al.*, 2023). Additionally, the use of WGCNA has provided relevant insights that could not be uncovered by identification of DEGs. Therefore, another idea to further investigate the gonadal transcriptome across species could be to benefit from methods developed for comparative analysis of gene co-expression networks (Ovens *et al.*, 2021a). For instance, using methods based on alignment of genes at the network level to reveal the evolution of their function. Such methods are based on generative probabilistic models of module evolution to represent the expression modules in each species and assign genes to modules in each extant and ancestral species (Koch *et al.*, 2017; Nguyen *et al.*, 2019; Ovens *et al.*, 2021b). Other methods to compare networks are based on differential network connectivity and module conservation, which are useful to identify pivotal genes in evolutionary changes (Ahn *et al.*, 2023; Qiu *et al.*, 2018; Xie *et al.*, 2020). This would provide the similarities and differences in the modules associated with sex among species. Thus, we could identify common modules and modules composed of similar or different genes and to reveal conserved hub genes in vertebrates.

Finally, an additional area worth considering for future research is the mechanism by which temperature actually modifies DNA methylation within the cell. Nowadays, the “epigenetators”, “initiators” and “maintainers” of the operational epigenetic process remain unknown (Berger *et al.*, 2009). However, several hypotheses exist. For example, several studies related DNA methylation changes induced by temperature linked to stress (Dorts *et al.*, 2016; Goikoetxea *et al.*, 2017; Lallias *et al.*, 2021). In reptiles, species with Temperature Sex determining (TSD) system, a recent hypothesis proposed is that conserved cellular mechanisms (calcium and redox) sense temperature and identify genes and regulatory pathways involved in driving sexual development (Whiteley *et al.*, 2021). In brief, temperature regulates expression of the Lysine Demethylase 6B (*kdm6b*) gene, which leads to testis development. Weber *et al.*, (2020), found that the signal transducer and activator of transcription 3 (STAT3) was the link between temperature and the sex-specific expression of *kdm6b*. Following an influx of calcium ions (Ca²⁺) at high temperature (Female producing temperature), the phosphorylation of STAT3 occurs, leading to the inhibition of *kdm6b* transcription and subsequent inhibition of testis development. In that regard, when examining the list of MA-DMRs, we found that up to three genes with hypermethylated DMRs were directly involved in calcium sensing: Calcium and Integrin Binding Family Member 2 (*cib2*) and two extra cellular calcium-sensor receptor-like. Furthermore, one of the significantly enriched terms of the genes overlapping with MA-DMRs was regulation of potassium ion transport, which could be the link to the dynamics of [Ca²⁺] as a result from cooperative Ca²⁺/K⁺ ion-exchange process (Nguyen *et al.*, 1998). However, more evidence is required to show the link between these processes and the influence of temperature in the sea bass.

Another alternative hypothesis could be related to other factors, for example, the concentration of dissolved oxygen in ocean water. It is well-known that there is an inverse correlation between temperature and dissolved oxygen (**Figure 2**) (Chapra *et al.*, 2021). Thus, similar to the alternative hypothesis proposed by Pauly, (2021) to explain why the fish mature when they do. According to this hypothesis, there could be a link between temperature, the available dissolved oxygen and subsequent changes in other factors of the environment such as pH. Alternatively, physical factors could also be explored. Heat, for instance, can modify protein structure by disrupting noncovalent bonds, a process known as denaturation (Kim *et al.*, 2016). Additionally, the 3D chromatin structure can play a crucial role in gene expression regulation. In plants, changes in light and temperature have been observed to impact chromatin organization (Perrella *et al.*, 2020).

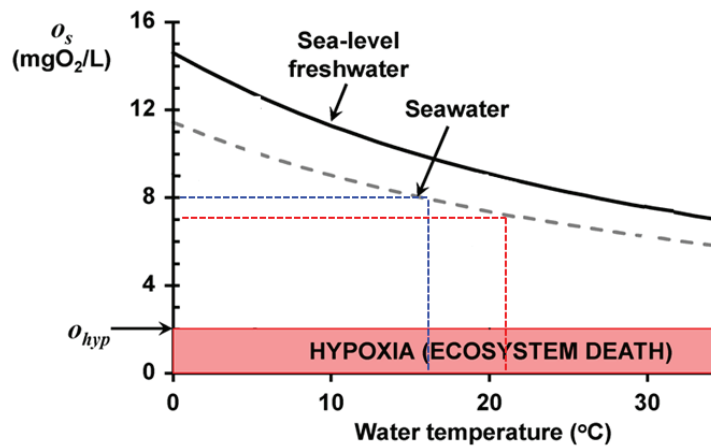


Figure 2. Dissolved oxygen saturation (mgO_2/L) versus temperature ($^\circ\text{C}$) for sea-level freshwater and seawater. The blue dashed line indicates the dissolved oxygen at 16 $^\circ\text{C}$ (low temperature treatment, LT) and the red dashed line indicates the corresponding dissolved oxygen in seawater at 21 $^\circ\text{C}$ (high temperature, HT). Modified from Chapra *et al.*, 2021.

Overall, the mechanisms by which temperature affect DNA methylation levels remain unknown and whether the DNA methylation changes are the cause or the consequence of other factors will warrant further investigation like in many other fields of research (Kim, 2019; Rang and Boonstra, 2014; Scott and De Sario, 2020).

Conclusions

Conclusions

1. In transcriptomic data analysis, the common practice of using only differentially expressed genes in co-expression network analysis instead of the entire dataset easily leads to biased results in the resulting topology of the network and on the identification of hub genes. We empirically demonstrate that constructing a co-expression network using the entire dataset and only subsequently filtering by differentially expressed genes improves model fit, better identifies relevant genes and thus leads to more biologically meaningful results. It also facilitates biomarker discovery. Thus, we propose to use such an approach in transcriptomics regardless of the biological system, species, or question being investigated.
2. We examined, for the first time in a truly comparative manner, the gonadal transcriptome dynamics during sex differentiation of six vertebrate species, ranging from fish to mammals. Instead of comparing females to males at a given developmental stage, we analyzed gene expression changes between two developmental stages within each sex. This approach revealed an extraordinary variability in gene transcriptomic dynamics across species without relationship to the position in the phylogenetic tree. This was true even between the two fish species examined.
3. We found consistent enrichment of key genes in sex differentiation within the same sex across species. However, our approach revealed that a significant number of genes involved in differentiation were not conserved, including the top 10% upregulated genes. This suggests that the conservation of genes involved in sex differentiation may be lower than previously assumed. While the concept of gonadal sex differentiation as two opposing networks with pro-male and pro-female genes remains valid, we propose perceiving sex differentiation as the result of opposing hierarchical networks with hub genes and less connected nodes. This perspective suggests that evolutionary pressures may influence genes based on their level of connectivity within the network.
4. We validate an approach to identify epigenetic biomarkers with both diagnostic and prognostic value as indicators of the long-lasting effects of an environmental factor, in our case the thermal regime during early development in sea bass. Such approach is based on the evaluation of the DNA methylation levels of specific regions that change before the transcriptomic changes become evident. One example is the methylation status of keratin-associated protein 10–4 (*krtap10–4*). The epigenetic biomarkers identified in this study enable screening of larval batches to determine the quality of their thermal environment and to anticipate their performance as juveniles and adults.
5. Machine learning algorithms typically developed in the biomedical field require a large number of samples, limiting their applicability to address practical problems in animal production. We demonstrate the possibility of using machine learning with a small initial sample size for genome screening. This approach involves carefully selecting samples and implementing various filtering, evaluation, and validation steps. By employing selection criteria focused on maximizing biomass and the desired sex, we could successfully integrate epigenetic biomarkers-assisted selection along with classical genetic selection. The phenotypic variation captured by epigenetic variation will aid aquaculture selection programs.

6. We carried out a comprehensive study on the sea bass gonadal methylome, examining both intrinsic (sex) and extrinsic (temperature) effects. We demonstrated additive effects of multiple high temperature exposures, providing evidence both at genome-wide and as well as focusing on sex-related genes. Using Multi Omics Factor Analysis we developed epigenetic biomarkers for the identification of sex-reversed females (neomales). Further, we revealed for the first time the epigenetic signature (DNA methylation profile) of sex reversal in a species with polygenic sex determination. Lastly, we found that a very small proportion of temperature-induced marks were inherited to the next generation, highlighting the potential role of environmentally-induced epimutations in shaping phenotypic variation across generations.
7. Overall, this thesis provides new approaches to solve old questions and to answer new ones. Specifically, it contributes to reveal the extraordinary variability in gene activation and repression programs during vertebrate gonadal sex differentiation and to the understanding of the role of environmental influences, DNA methylation and transcriptome on the gonadal phenotype in the sea bass. We introduce innovative epigenetic biomarker tools for monitoring and enhancing finfish production. Furthermore, our findings reveal the cumulative effects of elevated temperature, raising concerns about its potential impact on wild population sex ratios under global warming. On a broader scale, this thesis contributes to our understanding of epigenetic inheritance, revealing underlying mechanisms and patterns in this exciting field of contemporary biology.

References

References

A

- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*. 252(5013), 1651–1656. <https://doi.org/10.1126/science.2047873>
- Ah-King, M., Nylin, S., 2010. Sex in an evolutionary perspective: Just another reaction norm. *Evolutionary Biology*. 37, 234–246. <https://doi.org/10.1007/s11692-010-9101-8>
- Ahn, S., Grimes, T., Datta, S., 2023. A pseudo-value regression approach for differential network analysis of co-expression data. *BMC Bioinformatics*. 24(1), 8. <https://doi.org/10.1186/s12859-022-05123-w>
- Akalın, A., Franke, V., Vlahovicek, K., Mason, C., Schubeler, D., 2014. genomation: A toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics*. 31(7), 1127–1129. <https://doi.org/10.1093/bioinformatics/btu775>
- Akalın, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., Mason, C.E., 2012. methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology* 13(10), 1–9. <https://doi.org/10.1186/gb-2012-13-10-r87>
- Akazome, Y., Abe, T., Mori, T., 2002. Differentiation of chicken gonad as an endocrine organ: expression of LH receptor, FSH receptor, cytochrome P450c17 and aromatase genes. *Reproduction*. 123(5), 721–728. <https://pubmed.ncbi.nlm.nih.gov/12006100/>
- Alexa, A., Rahnenfuhrer, J., 2010. topGO: Enrichment analysis for gene ontology. R package version 2.16.0.
- Al-Fageeh, M.B., Smales, C.M., 2006. Control and regulation of the cellular responses to cold shock: The responses in yeast and mammalian systems. *Biochemical Journal* 397(2), 247–259. <https://doi.org/10.1042/BJ20060166>
- Alfalahi, A.O., Hussein, Z.T., Khalofah, A., Sadler, M.T., Qasem, J.R., Al-Khayri, J.M., Jain, S.M., Almeheemdi, A.F., 2022. Epigenetic variation as a new plant breeding tool: A review. *Journal of King Saud University - Science*. 102302. <https://doi.org/10.1016/j.jksus.2022.102302>
- Aliferi, A., Ballard, D., Gallidabino, M.D., Thurtle, H., Barron, L., Syndercombe Court, D., 2018. DNA methylation-based age prediction using massively parallel sequencing data and multiple machine learning models. *Forensic Science International: Genetics*. 37, 215–226. <https://doi.org/10.1016/j.fsigen.2018.09.003>
- Allal, F., Nguyen, N.H., 2022. Genomic selection in aquaculture species. In: *Genomic Prediction of Complex Traits: Methods and Protocols*. 469–491 pp. Humana, New York, NY: Springer US. ISBN: 9781071622049. https://doi.org/10.1007/978-1-0716-2205-6_17
- Altschul, S., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*. 25(17)3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Amiri, A., Coppola, G., Scuderi, S., Wu, F., Roychowdhury, T., Liu, F., Pochareddy, S., Shin, Y., Safi, A., Song, L., Zhu, Y., Sousa, A.M.M., Gerstein, M., Crawford, G.E., Sestan, N., Abyzov, A., Vaccarino, F.M., 2018. Transcriptome and epigenome landscape of human cortical development modeled in brain organoids. *Science*. 362(6420), eaat6720. <https://doi.org/10.1126/science.aat6720>
- Anastasiadi, D., Díaz, N., Piferrer, F., 2017. Small ocean temperature increases elicit stage-dependent changes in DNA methylation and gene expression in a fish, the European sea bass. *Scientific Reports*. 7(1), 12401. <https://doi.org/10.1038/s41598-017-10861-6>
- Anastasiadi, D., Esteve-Codina, A., Piferrer, F., 2018a. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics & Chromatin*. 11(1), 37. <https://doi.org/10.1186/s13072-018-0205-1>
- Anastasiadi, D., Piferrer, F., 2023. Bioinformatic analysis for age prediction using epigenetic clocks: Application to fisheries management and conservation biology. *Frontiers in Marine Science*. 10:1096909. <https://doi.org/10.3389/fmars.2023.1096909>
- Anastasiadi, D., Piferrer, F., 2020. A clockwork fish: Age prediction using DNA methylation-based biomarkers in the European seabass. *Molecular Ecology Resources*. 20(2)387–97. <https://doi.org/10.1111/1755-0998.13111>
- Anastasiadi, D., Piferrer, F., 2019. Epimutations in developmental genes underlie the onset of domestication in farmed European sea bass. *Molecular Biology and Evolution*. 36(10), 2252–2264. <https://doi.org/10.1093/molbev/msz153>
- Anastasiadi, D., Shao, C., Chen, S., Piferrer, F., 2021a. Footprints of global change in marine life: Inferring past environment based on DNA methylation and gene expression marks. *Molecular Ecology*. 30(3), 747–760. <https://doi.org/10.1111/mec.15764>
- Anastasiadi, D., Vandeputte, M., Sánchez-Baizán, N., Allal, F., Piferrer, F., 2018b. Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a fish with mixed genetic and environmental sex determination. *Epigenetics*. 13(9), 988–1011. <https://doi.org/10.1080/15592294.2018.1529504>

- Anastasiadi, D., Venney, C.J., Bernatchez, L., Wellenreuther, M., 2021b. Epigenetic inheritance and reproductive mode in plants and animals. *Trends in Ecology & Evolution*. 36(12), 1124–1140. <https://doi.org/10.1016/j.tree.2021.08.006>
- Anastasiadi, D., Beemelmans A., 2023. Development of epigenetic biomarkers in aquatic organisms. In: Piferrer F., Wang, H.P., eds. 2023. *Epigenetics in Aquaculture*. John Wiley and Sons, Hoboken, New Jersey. 450 pp. ISBN: 9781119821915. <https://doi.org/10.1002/9781119821946.ch18>
- Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. *Nature Precedings*. 1-1. <https://doi.org/10.1186/gb-2010-11-10-r106>
- Ando, Y., Hayashizaki, Y., 2006. Restriction landmark genomic scanning. *Nature Protocols*. 1(6), 2774–2783. <https://doi.org/10.1038/nprot.2006.350>
- Angers, B., Perez, M., Menicucci, T., Leung, C., 2020. Sources of epigenetic variation and their applications in natural populations. *Evolutionary Applications*. 13(6), 1262–1278. <https://doi.org/10.1111/eva.12946>
- Arechavala-Lopez, P., Fernandez-Jover, D., Black, K.D., Ladoukakis, E., Bayle-Sempere, J.T., Sanchez-Jerez, P., Dempster, T., 2013. Differentiating the wild or farmed origin of Mediterranean fish: A review of tools for sea bream and sea bass. *Reviews in Aquaculture*. 5(3), 137–157. <https://doi.org/10.1111/raq.12006>
- Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J.C., Stegle, O., 2020. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*. 21(1), 1-7. <https://doi.org/10.1186/s13059-020-02015-1>
- Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J.C., Buettner, F., Huber, W., Stegle, O., 2018. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology* 14(6), e8124. <https://doi.org/10.15252/msb.20178124>
- Arias, A., 1980. Crecimiento, régimen alimentario y reproducción de la dorada (*Sparus aurata* L.) y del robalo (*Dicentrarchus labrax* L.) en los esteros de Cádiz. *INVEST. PESQ.*; ISSN 0020-9953. 44(1), 59-83.
- Armstrong, K.M., Bermingham, E.N., Bassett, S.A., Treloar, B.P., Roy, N.C., Barnett, M.P.G., 2011. Global DNA methylation measurement by HPLC using low amounts of DNA. *Biotechnology Journal*. 6(1), 113–117. <https://doi.org/10.1002/biot.201000267>
- Arslan, E., Schulz, J., Rai, K., 2021. Machine learning in epigenomics: Insights into cancer biology and medicine. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 1876(2), 188588. <https://doi.org/10.1016/j.bbcan.2021.188588>
- Ayers, K.L., Davidson, N.M., Demiyah, D., Roeszler, K.N., Grützner, F., Sinclair, A.H., Oshlack, A., Smith, C.A., 2013. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biology*. 14:1-7. <https://doi.org/10.1186/gb-2013-14-3-r26>
- Ayers, K.L., Lambeth, L.S., Davidson, N.M., Sinclair, A.H., Oshlack, A., Smith, C.A., 2015. Identification of candidate gonadal sex differentiation genes in the chicken embryo using RNA-seq. *BMC Genomics*. 16(1), 1-9. <https://doi.org/10.1186/s12864-015-1886-5>

B

- Bai, K.H., He, S.Y., Shu, L.L., Wang, W.D., Lin, S.Y., Zhang, Q.Y., Li, L., Cheng, L., Dai, Y.J., 2020. Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index. *Cancer Medicine*. 9(12)4290–4298. <https://doi.org/10.1002/cam4.3047>
- Banta, J.A., Richards, C.L., 2018. Quantitative epigenetics and evolution. *Heredity*. 121(3), 210–224. <https://doi.org/10.1038/s41437-018-0114-x>
- Bao, L., Tian, C., Liu, S., Zhang, Y., Elswad, A., Yuan, Z., Khalil, K., Sun, F., Yang, Y., Zhou, T., Li, N., 2019. The Y chromosome sequence of the channel catfish suggests novel sex determination mechanisms in teleost fish. *BMC biology*. 17, 1–16. <https://doi.org/10.1186/s12915-019-0627-7>
- Barahona-Fernandes, M.H., 1982. Body deformation in hatchery reared European sea bass *Dicentrarchus labrax* (L). Types, prevalence and effect on fish survival. *Journal of Fish Biology*. 21(3), 239–249. <https://doi.org/10.1111/j.1095-8649.1982.tb02830.x>
- Barnabe, G., 1974. Mass rearing of the bass *Dicentrarchus labrax* L., In: *The Early Life History of Fish: The Proceedings of an International Symposium Held at the Dunstaffnage Marine Research Laboratory of the Scottish Marine Biological Association at Oban, Scotland, from May 17–23, 1973*. Springer, 749–753 pp.
- Baroiller, J.F., D’Cotta, H., Saillant, E., 2009. Environmental effects on fish sex determination and differentiation. *Sexual Development*. 3(2-3), 118–135. <https://doi.org/10.1159/000223077>
- Batut, P.J., Bing, X.Y., Sisco, Z., Raimundo, J., Levo, M., Levine, M.S., 2022. Genome organization controls transcriptional dynamics during development. *Science*. 375(6580), 566–570. <https://doi.org/10.1126/science.abi7178>
- Beck, D., Ben Maamar, M., Skinner, M.K., 2021. Integration of sperm ncRNA-directed DNA methylation and DNA methylation-directed histone retention in epigenetic transgenerational inheritance. *Epigenetics & Chromatin*. 14(1), 1-4. <https://doi.org/10.1186/s13072-020-00378-0>

- Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., Weigel, D., 2011. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*. 480(7376), 245–249. <https://doi.org/10.1038/nature10555>
- Behringer, R.R., Finegold, M.J., Cate, R.L., 1994. Müllerian-inhibiting substance function during mammalian sexual development. *Cell*. 79(3), 415–425. [https://doi.org/10.1016/0092-8674\(94\)90251-8](https://doi.org/10.1016/0092-8674(94)90251-8)
- Belleau, P., Deschênes, A., Scott-Boyer, M.-P., Lambrot, R., Dalvai, M., Kimmins, S., Bailey, J., Droit, A., 2018. Inferring and modeling inheritance of differentially methylated changes across multiple generations. *Nucleic Acids Research*. 46(14), e85. <https://doi.org/10.1093/nar/gky362>
- Bender, J., Fink, G.R., 1995. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell*. 83(5), 725–734.
- Benhaïm, D., Péan, S., Lucas, G., Blanc, N., Chatain, B., Bégout, M.-L., 2012. Early life behavioural differences in wild caught and domesticated sea bass (*Dicentrarchus labrax*). *Applied Animal Behaviour Science* 141(1–2), 79–90. <https://doi.org/10.1016/j.applanim.2012.07.002>
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berger, S.L., Kouzarides, T., Shiekhhattar, R., Shilatifard, A., 2009. An operational definition of epigenetics. *Genes & development*. 23(7)781–783. <https://genesdev.cshlp.org/content/23/7/781>
- Bertho, S., Herpin, A., Scharthl, M., Guiguen, Y., 2021. Lessons from an unusual vertebrate sex-determining gene. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 376(1832), 20200092. <https://doi.org/10.1098/rstb.2020.0092>
- Bertucci, E.M., Mason, M.W., Rhodes, O.E., Parrott, B.B., 2021. Exposure to ionizing radiation disrupts normal epigenetic aging in Japanese medaka. *Aging (Albany NY)*. 13(19), 22752–22771. <https://doi.org/10.18632/aging.203624>
- Beukeboom, L.W., Perrin, N., 2014. The evolution of sex chromosomes, In: Beukeboom, L.W., Perrin, N. (Eds.), *The Evolution of Sex Determination*. Oxford University Press. 89–114 pp. <https://doi.org/10.1093/acprof:oso/9780199657148.003.0005>
- Bezault, E., Clota, F., Derivaz, M., Chevassus, B., Baroiller, J.F., 2007. Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture*. 272(1), S3–S16. <https://doi.org/10.1016/j.aquaculture.2007.07.227>
- Bibikova, M., Fan, J.-B., 2009. GoldenGate assay for DNA methylation profiling. *Methods in Molecular Biology* (Clifton, N.J.) 507, 149–163. https://doi.org/10.1007/978-1-59745-522-0_12
- Black, M.B., Parks, B.B., Pluta, L., Chu, T.-M., Allen, B.C., Wolfinger, R.D., Thomas, R.S., 2014. Comparison of microarrays and RNA-seq for gene expression analyses of dose-response experiments. *Toxicological Sciences*. 137(2), 385–403. <https://doi.org/10.1093/toxsci/kft249>
- Blázquez, M., Carrillo, M., Zanuy, S., Piferrer, F., 1999. Sex ratios in offspring of sex-reversed sea bass and the relationship between growth and phenotypic sex differentiation. *Journal of Fish Biology*. 55(5), 916–930. <https://doi.org/10.1111/j.1095-8649.1999.tb00730.x>
- Blázquez, M., González, A., Papadaki, M., Mylonas, C., Piferrer, F., 2008. Sex-related changes in estrogen receptors and aromatase gene expression and enzymatic activity during early development and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *General and Comparative Endocrinology*. 158(1), 95–101. <https://doi.org/10.1016/j.ygcen.2008.06.001>
- Blázquez, M., Navarro-Martín, L., Piferrer, F., 2009. Expression profiles of sex differentiation-related genes during ontogenesis in the European sea bass acclimated to two different temperatures. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. 312(7), 686–700. <https://doi.org/10.1002/jez.b.21286>
- Blázquez, M., Piferrer, F., 2005. Sea bass (*Dicentrarchus labrax*) androgen receptor: cDNA cloning, tissue-specific expression, and mRNA levels during early development and sex differentiation. *Molecular and Cell Endocrinology*. 237(1–2), 37–48. <https://doi.org/10.1016/j.mce.2005.04.001>
- Blázquez, M., Zanuy, S., Carillo, M., Piferrer, F., 1998. Effects of rearing temperature on sex differentiation in the European sea bass (*Dicentrarchus labrax* L.). *Journal of Experimental Zoology*. 281(3), 207–216. [https://doi.org/10.1002/\(SICI\)1097-010X\(19980615\)281:3<207::AID-JEZ6>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-010X(19980615)281:3<207::AID-JEZ6>3.0.CO;2-R)
- Bogdanović, O., Long, S.W., van Heeringen, S.J., Brinkman, A.B., Gómez-Skarmeta, J.L., Stunnenberg, H.G., Jones, P.L., Veenstra, G.J.C., 2011. Temporal uncoupling of the DNA methylome and transcriptional repression during embryogenesis. *Genome Research*. 21(8), 1313–1327. <https://doi.org/10.1101/gr.114843.110>
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

- Booth, M.J., Ost, T.W.B., Beraldi, D., Bell, N.M., Branco, M.R., Reik, W., Balasubramanian, S., 2013. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nature Protocols*. 8(10), 1841–1851. <https://doi.org/10.1038/nprot.2013.115>
- Bors, E.K., Baker, C.S., Wade, P.R., O'Neill, K.B., Shelden, K.E.W., Thompson, M.J., Fei, Z., Jarman, S., Horvath, S., 2021. An epigenetic clock to estimate the age of living beluga whales. *Evolutionary Applications*. 14(5), 1263–1273. <https://doi.org/10.1111/eva.13195>
- Bowler, S., Papoutsoglou, G., Karanikas, A., Tsamardinos, I., Corley, M.J., Ndhlovu, L.C., 2022. A machine learning approach utilizing DNA methylation as an accurate classifier of COVID-19 disease severity. *Scientific Reports*. 12(1), 17480. <https://doi.org/10.1038/s41598-022-22201-4>
- Brinchmann, M.F., 2016. Immune relevant molecules identified in the skin mucus of fish using -omics technologies. *Molecular BioSystems*. 12(7), 2056–2063. <https://doi.org/10.1039/C5MB00890E>
- Brinkman, A.B., Simmer, F., Ma, K., Kaan, A., Zhu, J., Stunnenberg, H.G., 2010. Whole-genome DNA methylation profiling using MethylCap-seq. *Methods*. 52(3), 232–236. <https://doi.org/10.1016/j.ymeth.2010.06.012>
- Brunner, A.L., Johnson, D.S., Kim, S.W., Valouev, A., Reddy, T.E., Neff, N.F., Anton, E., Medina, C., Nguyen, L., Chiao, E., Oyolu, C.B., Schroth, G.P., Absher, D.M., Baker, J.C., Myers, R.M., 2009. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Research*. 19(6), 1044–1056. <https://doi.org/10.1101/gr.088773.108>
- Bull, J.J., 1983. *Evolution of sex determining mechanisms*. The Benjamin/Cummings Publishing Company, Inc.
- Burggren, W.W., Mueller, C.A., 2015. Developmental critical windows and sensitive periods as three-dimensional constructs in time and space. *Physiological and Biochemical Zoology*. 88(2), 91–102. <https://doi.org/10.1086/679906>
- Bushnell, B., 2014. BBMap: A fast, accurate, splice-aware aligner (No. LBNL-7065E). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).
- Bysani, M., Perfilyev, A., de Mello, V.D., Rönn, T., Nilsson, E., Pihlajamäki, J., Ling, C., 2017. Epigenetic alterations in blood mirror age-associated DNA methylation and gene expression changes in human liver. *Epigenomics*. 9(2), 105–122. <https://doi.org/10.2217/epi-2016-0087>

C

- Campbell, R.R., Wood, M.A., 2019. How the epigenome integrates information and reshapes the synapse. *Nature Reviews Neuroscience* 20(3), 133–147. <https://doi.org/10.1038/s41583-019-0121-9>
- Capel, B., 2017. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nature Reviews Genetics*. 18(11), 675–689. <https://doi.org/10.1038/nrg.2017.60>
- Carlberg, C., Molnár, F., 2014. *Mechanisms of gene regulation*. Dordrecht: Springer. pp. 57–73.
- Carrillo, M., Zanuy, S., Prat, F., Cerdà, J., Ramos, J., Mañanós, E.L., Bromage, N., 1995. Sea bass (*Dicentrarchus labrax*). In: N.R. Bromage, R.J. Roberts (Eds.), *Broodstock Management and Egg and Larval Quality*, Blackwell, Oxford. 138–168 pp. ISBN: 0412400901.
- Cascarano, M.C., Stavrakidis-Zachou, O., Mladineo, I., Thompson, K.D., Papandroulakis, N., Katharios, P., 2021. Mediterranean aquaculture in a changing climate: temperature effects on pathogens and diseases of three farmed fish species. *Pathogens* 10(9), 1205. <https://doi.org/10.3390/pathogens10091205>
- Castelli, M.A., Whiteley, S.L., Georges, A., Holleley, C.E., 2020. Cellular calcium and redox regulation: the mediator of vertebrate environmental sex determination? *Biological Reviews*. 95(3), 680–695. <https://doi.org/10.1111/brv.12582>
- Castro-Muñoz, L.J., Ulloa, E.V., Sahlgren, C., Lizano, M., De La Cruz-Hernández, E., Contreras-Paredes, A., 2023. Modulating epigenetic modifications for cancer therapy. *Oncology Reports*. 49(3), 1–23. <https://doi.org/10.3892/or.2023.8496>
- Cavalli, G., Heard, E., 2019. Advances in epigenetics link genetics to the environment and disease. *Nature*. 571(7766), 489–499. <https://doi.org/10.1038/s41586-019-1411-0>
- Cerdá-Reverter, J.M., Anglade, I., Martínez-Rodríguez, G., Mazurais, D., Muñoz-Cueto, J.A., Carrillo, M., Kah, O., Zanuy, S., 2000. Characterization of neuropeptide Y expression in the brain of a perciform fish, the sea bass (*Dicentrarchus labrax*). *Journal of Chemical Neuroanatomy*. 19(4), 197–210. [https://doi.org/10.1016/s0891-0618\(00\)00063-6](https://doi.org/10.1016/s0891-0618(00)00063-6)
- Cesaroni, M., Powell, J., Sapienza, C., 2014. Validation of methylation biomarkers that distinguish normal colon mucosa of cancer patients from normal colon mucosa of patients without cancer. *Cancer Prevention Research*. 7(7), 717–726. <https://doi.org/10.1158/1940-6207.CAPR-13-0407>
- Chang, X., Kobayashi, T., Kajiura, H., Nakamura, M., Nagahama, Y., 1997. Isolation and characterization of the cDNA encoding the tilapia (*Oreochromis niloticus*) cytochrome P450 aromatase (P450arom): changes in P450arom mRNA, protein and enzyme activity in ovarian follicles during oogenesis. *Journal of Molecular Endocrinology*. 18(1), 57–66. <https://doi.org/10.1677/jme.0.0180057>

- Chapra, S.C., Camacho, L.A., McBride, G.B., 2021. Impact of global warming on dissolved oxygen and BOD assimilative capacity of the world's rivers: modeling analysis. *Water*. 13(17), 2408. <https://doi.org/10.3390/w13172408>
- Charnov, E.L., Bull, J., 1977. When is sex environmentally determined? *Nature*. 266(5605), 828–830. <https://doi.org/10.1038/266828a0>
- Chen, S., Yang, D., Lei, C., Li, Y., Sun, X., Chen, M., Wu, X., Zheng, Y., 2019. Identification of crucial genes in abdominal aortic aneurysm by WGCNA. *PeerJ*. 7, e7873. <https://doi.org/10.7717/peerj.7873>
- Chen, S., Yang, Y., Gao, B., Jia, C., Zhu, F., Meng, Q., Zhang, Zhiwei, Zhang, Zhiyong, Xu, S., 2022. Comparative proteomics of the *Acanthopagrus schlegelii* gonad in different sex reversal. *Genes*. 13(2), 253. <https://doi.org/10.3390/genes13020253>
- Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., Li, E., 2003. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by *Dnmt3a* and *Dnmt3b*. *Molecular and cellular biology*. 23(16), 5594–5605. <https://doi.org/10.1128/MCB.23.16.5594-5605.2003>
- Chen, Y., Lun, A.T.L., Smyth, G.K., 2014. Differential expression analysis of complex RNA-seq experiments using edgeR. In: Datta, S., Nettleton, D. (eds) *Statistical Analysis of Next Generation Sequencing Data*. *Frontiers in Probability and the Statistical Sciences*. Springer, 51–74 pp. https://doi.org/10.1007/978-3-319-07212-8_3
- Chen, Y., Lun, A.T.L., Smyth, G.K., 2016. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*. 5, 1438. <https://doi.org/10.12688/f1000research.8987.2>
- Chen, Y.-C., Guo, Y.-F., He, H., Lin, X., Wang, X.-F., Zhou, R., Li, W.-T., Pan, D.-Y., Shen, J., Deng, H.-W., 2016. Integrative analysis of genomics and transcriptome data to identify potential functional genes of BMDs in females. *Journal of Bone Mineral Research*. 31(5), 1041–1049. <https://doi.org/10.1002/jbmr.2781>
- Christmas, M.J., Kaplow, I.M., Genereux, D.P., Dong, M.X., Hughes, G.M., Li, X., Sullivan, P.F., Hindle, A.G., Andrews, G., Armstrong, J.C., Bianchi, M., Breit, A.M., Diekhans, M., Fanter, C., Foley, N.M., Goodman, D.B., Goodman, L., Keough, K.C., Kirilenko, B., Kowalczyk, A., Lawless, C., Lind, A.L., Meadows, J.R.S., Moreira, L.R., Redlich, R.W., Ryan, L., Swofford, R., Valenzuela, A., Wagner, F., Wallerman, O., Brown, A.R., Damas, J., Fan, K., Gatesy, J., Grimshaw, J., Johnson, J., Kozyrev, S.V., Lawler, A.J., Marinescu, V.D., Morrill, K.M., Osmanski, A., Paulat, N.S., Phan, B.N., Reilly, S.K., Schäffer, D.E., Steiner, C., Supple, M.A., Wilder, A.P., Wirthlin, M.E., Xue, J.R., ZOOONOMIA CONSORTIUM, Birren, B.W., Gazal, S., Hubley, R.M., Koepfli, K.-P., Marques-Bonet, T., Meyer, W.K., Nweeia, M., Sabeti, P.C., Shapiro, B., Smit, A.F.A., Springer, M.S., Teeling, E.C., Weng, Z., Hiller, M., Levesque, D.L., Lewin, H.A., Murphy, W.J., Navarro, A., Paten, B., Pollard, K.S., Ray, D.A., Ruf, I., Ryder, O.A., Pfenning, A.R., Lindblad-Toh, K., Karlsson, E.K., 2023. Evolutionary constraint and innovation across hundreds of placental mammals. *Science*. 380(6643), eabn3943. <https://doi.org/10.1126/science.abn3943>
- Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C., 2012. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *NeuroImage*. 60(1), 59–70. <https://doi.org/10.1016/j.neuroimage.2011.11.066>
- Chue, J., Smith, C.A., 2011. Sex determination and sexual differentiation in the avian model. *FEBS Journal*. 278(7), 1027–1034. <https://doi.org/10.1111/j.1742-4658.2011.08032.x>
- Chwialkowska, K., Korotko, U., Kosinska, J., Szarejko, I., Kwasniewski, M., 2017. Methylation sensitive amplification polymorphism sequencing (MSAP-Seq)- A method for high-throughput analysis of differentially methylated CCGG sites in plants with large genomes. *Frontiers in Plant Science*. 8, 2056. <https://doi.org/10.3389/fpls.2017.02056>
- Claireaux, G., Couturier, C., Groison, A.-L., 2006. Effect of temperature on maximum swimming speed and cost of transport in juvenile European sea bass (*Dicentrarchus labrax*). *Journal of Experimental Biology*. 209(17), 3420–3428. <https://doi.org/10.1242/jeb.02346>
- Clement, T.M., Anway, M.D., Uzumcu, M., Skinner, M.K., 2007. Regulation of the gonadal transcriptome during sex determination and testis morphogenesis: comparative candidate genes. *Reproduction*. 134(3), 455–472. <https://doi.org/10.1530/REP-06-0341>
- Colot, V., Rossignol, J.L., 1999. Eukaryotic DNA methylation as an evolutionary device. *BioEssays*. 21(5), 402–411. [https://doi.org/10.1002/\(SICI\)1521-1878\(199905\)21:5<402::AID-BIES7>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1521-1878(199905)21:5<402::AID-BIES7>3.0.CO;2-B)
- Colvin, J.S., Green, R.P., Schmahl, J., Capel, B., Ornitz, D.M., 2001. Male-to-female sex reversal in mice lacking fibroblast growth factor 9. *Cell*. 104(6), 875–889. [https://doi.org/10.1016/S0092-8674\(01\)00284-7](https://doi.org/10.1016/S0092-8674(01)00284-7)
- Conceicao, L., Tandler, A., 2018. *Success factors for fish larval production*. John Wiley & Sons. ISBN: 9781119072140.
- Conover, D.O., Van Voorhees, D.A., 1990. Evolution of a balanced sex ratio by frequency-dependent selection in a fish. *Science*. 250(4987), 1556–1558. <https://doi.org/10.1126/science.250.4987.1556>
- Consuegra, S., Rodríguez López, C.M., 2016. Epigenetic-induced alterations in sex-ratios in response to climate change: An epigenetic trap? *BioEssays*. 38(10), 950–958. <https://doi.org/10.1002/bies.201600058>

- Cordero, H., Brinchmann, M.F., Cuesta, A., Meseguer, J., Esteban, M.A., 2015. Skin mucus proteome map of European sea bass (*Dicentrarchus labrax*). *Proteomics*. 15(23-24), 4007–4020. <https://doi.org/10.1002/pmic.201500120>
- Costa-Pinheiro, P., Montezuma, D., Henrique, R., Jerónimo, C., 2015. Diagnostic and prognostic epigenetic biomarkers in cancer. *Epigenomics*. 7(6), 1003–1015. <https://doi.org/10.2217/epi.15.56>
- Cox, B., Tsamou, M., Vrijens, K., Neven, K.Y., Winkelmans, E., de Kok, T.M., Plusquin, M., Nawrot, T.S., 2019. A co-expression analysis of the placental transcriptome in association with maternal pre-pregnancy BMI and newborn birth weight. *Frontiers in Genetics*. 10:354. <https://doi.org/10.3389/fgene.2019.00354>
- Crespo, B., Gómez, A., Mazón, M.J., Carrillo, M., Zanuy, S., 2013. Isolation and characterization of *Ffl* and *Gsdf* family genes in European sea bass and identification of early gonadal markers of precocious puberty in males. *General and Comparative Endocrinology*. 191, 155–167. <https://doi.org/10.1016/j.ygcen.2013.06.010>
- Crisp, P.A., Bhatnagar-Mathur, P., Hundleby, P., Godwin, I.D., Waterhouse, P.M., Hickey, L.T., 2022. Beyond the gene: Epigenetic and cis-regulatory targets offer new breeding potential for the future. *Current Opinion in Biotechnology*. 73, 88–94. <https://doi.org/10.1016/j.copbio.2021.07.008>
- Cucchi, P., Sucré, E., Santos, R., Leclère, J., Charmantier, G., Castille, R., 2012. Embryonic development of the sea bass *Dicentrarchus labrax*. *Helgoland Marine Research*. 66(2), 199–209. <https://doi.org/10.1007/s10152-011-0262-3>
- Cui, W., Ma, A., 2020. Transcriptome analysis provides insights into the effects of myo-inositol on the turbot *Scophthalmus maximus*. *Fish & Shellfish Immunology*. 106, 691–704. <https://doi.org/10.1016/j.fsi.2020.07.019>
- Cui, Z., Liu, Y., Wang, W., Wang, Q., Zhang, N., Lin, F., Wang, N., Shao, C., Dong, Z., Li, Y., 2017. Genome editing reveals *dmrt1* as an essential male sex-determining gene in Chinese tongue sole (*Cynoglossus semilaevis*). *Scientific reports*. 7(1) 42213. <https://doi.org/10.1038/srep42213>
- Czerwinski, M., Natarajan, A., Barske, L., Looger, L.L., Capel, B., 2016. A timecourse analysis of systemic and gonadal effects of temperature on sexual development of the red-eared slider turtle *Trachemys scripta elegans*. *Developmental Biology*. 420(1), 166–177. <https://doi.org/10.1016/j.ydbio.2016.09.018>

D

- Dai, R., Xia, Y., Liu, C., Chen, C., 2019. csuWGCNA: A combination of signed and unsigned WGCNA to capture negative correlations. *bioRxiv*. 288225. <https://doi.org/10.1101/288225>
- Danzmann, R.G., Kocmarek, A.L., Norman, J.D., Rexroad, C.E., Palti, Y., 2016. Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics*. 17, 1-18. <https://doi.org/10.1186/s12864-016-2363-5>
- Darst, R.P., Pardo, C.E., Ai, L., Brown, K.D., Kladde, M.P., 2010. Bisulfite sequencing of DNA. *Current Protocols in Molecular Biology*. 91(1), 7-9. <https://doi.org/10.1002/0471142727.mb0709s91>
- Das, P., Roychowdhury, A., Das, S., Roychowdhury, S., Tripathy, S., 2020. sigFeature: novel significant feature selection method for classification of gene expression data using support vector machine and t statistic. *Frontiers in Genetics*. 11,247. <https://doi.org/10.3389/fgene.2020.00247>
- Database resources of the National Center for Biotechnology Information, 2016. *Nucleic Acids Research*. 44, D7–D19. <https://doi.org/10.1093/nar/gkv1290>
- De Bustos, C., Ramos, E., Young, J.M., Tran, R.K., Menzel, U., Langford, C.F., Eichler, E.E., Hsu, L., Henikoff, S., Dumanski, J.P., Trask, B.J., 2009. Tissue-specific variation in DNA methylation levels along human chromosome 1. *Epigenetics & chromatin*. 2(1), 1–21. <https://doi.org/10.1186/1756-8935-2-7>
- de la Herrán, R., Hermida, M., Rubiolo, J.A., Gómez-Garrido, J., Cruz, F., Robles, F., Navajas-Pérez, R., Blanco, A., Villamayor, P.R., Torres, D., Sánchez-Quinteiro, P., Ramirez, D., Rodríguez, M.E., Arias-Pérez, A., Cross, I., Duncan, N., Martínez-Peña, T., Ríaza, A., Millán, A., De Rosa, M.C., Pirolli, D., Gut, M., Bouza, C., Robledo, D., Rebordinos, L., Alioto, T., Ruiz-Rejón, C., Martínez, P., 2023. A chromosome-level genome assembly enables the identification of the follicle stimulating hormone receptor as the master sex-determining gene in the flatfish *Solea senegalensis*. *Molecular Ecology Resources*. 23(4), 886–904. <https://doi.org/10.1111/1755-0998.13750>
- de Oliveira, N.F.P., Coelho, M. de C., Viana Filho, J.M.C., 2020. Chapter Five - ELISA analysis of global methylation levels, in: Tollefsbol, T. (Ed.), *Epigenetics Methods, Translational Epigenetics*. Academic Press. pp. 83–92. <https://doi.org/10.1016/B978-0-12-819414-0.00005-7>
- De Paoli-Iseppi, R., Deagle, B.E., Polanowski, A.M., McMahon, C.R., Dickinson, J.L., Hindell, M.A., Jarman, S.N., 2019. Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Molecular ecology resources*. 19(2), 411–425. <https://doi.org/10.1111/1755-0998.12981>
- Deans, C., Maggert, K.A., 2015. What do you mean, “epigenetic”? *Genetics*. 199(4), 887–896. <https://doi.org/10.1534/genetics.114.173492>

- Debie, E., Shafi, K., 2019. Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses. *Pattern Analysis and Applications*. 22, 519–536. <https://doi.org/10.1007/s10044-017-0649-0>
- DeFalco, T., Capel, B., 2009. Gonad morphogenesis in vertebrates: divergent means to a convergent end. *Annual Review of Cell Developmental Biology*. 25, 457–482. <https://doi.org/10.1146/annurev.cellbio.042308.13350>
- Deist, M.S., Gallardo, R.A., Bunn, D.A., Dekkers, J.C.M., Zhou, H., Lamont, S.J., 2017. Resistant and susceptible chicken lines show distinctive responses to Newcastle disease virus infection in the lung transcriptome. *BMC Genomics*. 18(1), 1–5. <https://doi.org/10.1186/s12864-017-4380-4>
- Deng, S.P., Zhu, L., Huang, D.S., 2015. Mining the bladder cancer-associated genes by an integrated strategy for the construction and analysis of differential co-expression networks. *BMC Genomics*. 16(3), 1–10. <https://doi.org/10.1186/1471-2164-16-S3-S4>
- Deng, T., Liang, A., Liang, S., Ma, X., Lu, X., Duan, A., Pang, C., Hua, G., Liu, S., Campanile, G., Salzano, A., Gasparrini, B., Neglia, G., Liang, X., Yang, L., 2019. Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. *Frontiers in Genetics*. 10, 36. <https://doi.org/10.3389/fgene.2019.00036>
- Denkena, J., Johannes, F., Colomé-Tatché, M., 2021. Region-level epimutation rates in *Arabidopsis thaliana*. *Heredity*. 127(2), 190–202. <https://doi.org/10.1038/s41437-021-00441-w>
- Dergunova, L.V., Filippenkov, I.B., Stavchansky, V.V., Denisova, A.E., Yuzhakov, V.V., Mozerov, S.A., Gubsky, L.V., Limborska, S.A., 2018. Genome-wide transcriptome analysis using RNA-Seq reveals a large number of differentially expressed genes in a transient MCAO rat model. *BMC Genomics*. 19(1), 655. <https://doi.org/10.1186/s12864-018-5039-5>
- Deveson, I.W., Holleley, C.E., Blackburn, J., Marshall Graves, J.A., Mattick, J.S., Waters, P.D., Georges, A., 2017. Differential intron retention in Jumonji chromatin modifier genes is implicated in reptile temperature-dependent sex determination. *Science Advances*. 3(6), e1700731. <https://doi.org/10.1126/sciadv.1700731>
- Devlin, R.H., Nagahama, Y., 2002. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture*. 208(3–4), 191–364. [https://doi.org/10.1016/S0044-8486\(02\)00057-1](https://doi.org/10.1016/S0044-8486(02)00057-1)
- Di, Y., Chen, D., Yu, W., Yan, L., 2019. Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis. *Heredity*. 156(1)1–11. <https://doi.org/10.1186/s41065-019-0083-y>
- Díaz, N., Piferrer, F., 2017. Estrogen exposure overrides the masculinizing effect of elevated temperature by a downregulation of the key genes implicated in sexual differentiation in a fish with mixed genetic and environmental sex determination. *BMC Genomics*. 18(1), 1–3. <https://doi.org/10.1186/s12864-017-4345-7>
- Díaz, N., Piferrer, F., 2015. Lasting effects of early exposure to temperature on the gonadal transcriptome at the time of sex differentiation in the European sea bass, a fish with mixed genetic and environmental sex determination. *BMC genomics*. 16, 1–16. <https://doi.org/10.1186/s12864-015-1862-0>
- Díaz, N., Ribas, L., Piferrer, F., 2013. The relationship between growth and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *Aquaculture*. 408–409, 191–202. <https://doi.org/10.1016/j.aquaculture.2013.05.032>
- Ding, M., Li, F., Wang, B., Chi, G., Liu, H., 2019. A comprehensive analysis of WGCNA and serum metabolomics manifests the lung cancer-associated disordered glucose metabolism. *Journal of Cellular Biochemistry*. 120(6), 10855–10863. <https://doi.org/10.1002/jcb.28377>
- Ding, W., Chen, J., Feng, G., Chen, G., Wu, J., Guo, Y., Ni, X., Shi, T., 2020. DNMIVD: DNA methylation interactive visualization database. *Nucleic Acids Research*. 48(D1), D856–62. <https://doi.org/10.1093/nar/gkz830>
- Docherty, S.J., Davis, O.S., Haworth, C.M., Plomin, R., Mill, J., 2009. Bisulfite-based epotyping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. *Epigenetics & Chromatin*. 2(1), 1–8. <https://doi.org/10.1186/1756-8935-2-3>
- Doncheva, N.T., Assenov, Y., Domingues, F.S., Albrecht, M., 2012. Topological analysis and interactive visualization of biological networks and protein structures. *Nature Protocols*. 7(4), 670–685. <https://doi.org/10.1038/nprot.2012.004>
- Dorn, R., Krauss, V., Reuter, G., Saumweber, H., 1993. The enhancer of position-effect variegation of *Drosophila*, E(var)3-93D, codes for a chromatin protein containing a conserved domain common to several transcriptional regulators. *Proceedings of the National Academy of Sciences*. 90(23), 11376–11380. <https://doi.org/10.1073/pnas.90.23.11376>
- Dorts, J., Falisse, E., Schoofs, E., Flamion, E., Kestemont, P., Silvestre, F., 2016. DNA methyltransferases and stress-related genes expression in zebrafish larvae after exposure to heat and copper during reprogramming of DNA methylation. *Scientific Reports*. 6(1), 34254. <https://doi.org/10.1038/srep34254>
- Du, Z., Zhang, K., Xie, W., 2022. Epigenetic reprogramming in early animal development. *Cold Spring Harbor Perspectives in Biology*. 14(6), a039677. <https://doi.org/10.1101/cshperspect.a039677>
- Duan, H., Ge, W., Zhang, A., Xi, Y., Chen, Z., Luo, D., Cheng, Y., Fan, K.S., Horvath, S., Sofroniew, M.V., Cheng, L., Yang, Z., Sun, Y.E., Li, X., 2015. Transcriptome analyses reveal molecular mechanisms underlying

- functional recovery after spinal cord injury. *Proceedings of the National Academy of Sciences*. 112(43), 13360–13365. <https://doi.org/10.1073/pnas.1510176112>
- Duan, K.B., Rajapakse, J.C., Wang, H., Azuaje, F., 2005. Multiple SVM-RFE for gene selection in cancer classification with expression data. *IEEE transactions on nanobioscience*. 4(3), 228–234. <https://doi.org/10.1109/TNB.2005.853657>.
- Dülger, N., Kumlu, M., Türkmen, S., Ölçülü, A., Tufan Eroldoğan, O., Asuman Yılmaz, H., Öçal, N., 2012. Thermal tolerance of European sea bass (*Dicentrarchus labrax*) juveniles acclimated to three temperature levels. *Journal of Thermal Biology*. 37(1), 79–82. <https://doi.org/10.1016/j.jtherbio.2011.11.003>
- Duncan, E.J., Gluckman, P.D., Dearden, P.K., 2014. Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype? *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. 322(4), 208–220. <https://doi.org/10.1002/jez.b.22571>
- Dupont, C., Armant, D.R., Brenner, C.A., 2009. Epigenetics: definition, mechanisms and clinical perspective. *Seminars in reproductive medicine*. 27(5), 351–357. <https://doi.org/10.1055/s-0029-1237423>
- Dupont, S., Capel, B., 2021. The chromatin state during gonadal sex determination. *Sex Development*. 15(5-6), 308–316. <https://doi.org/10.1159/000520007>
- Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., Huber, W., 2005. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*. 21(16), 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525>
- Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*. 4(8), 1184–1191. <https://doi.org/10.1038/nprot.2009.97>

E

- Eads, C.A., Laird, P.W., 2002. Combined bisulfite restriction analysis (COBRA). *DNA methylation protocols*. 71–85. <https://doi.org/10.1385/1-59259-182-5:071>
- Edvardsen, R.B., Wallerman, O., Furmanek, T., Kleppe, L., Jern, P., Wallberg, A., Kjærner-Semb, E., Mæhle, S., Olausson, S.K., Sundström, E., Harboe, T., Mangor-Jensen, R., Møgster, M., Perrichon, P., Norberg, B., Rubin, C.-J., 2022. Heterochiasmy and the establishment of *gsdf* as a novel sex determining gene in Atlantic halibut. *PLoS Genetics*. 18(2), e1010011. <https://doi.org/10.1371/journal.pgen.1010011>
- Emam, B., Shahsavani, A., Khodagholi, F., Zarandi, S.M., Hopke, P.K., Hadei, M., Behbahani, H., Yarahmadi, M., 2020. Effects of PM_{2.5} and gases exposure during prenatal and early-life on autism-like phenotypes in male rat offspring. *Particle and Fibre Toxicology*. 17(1), 1–16. <https://doi.org/10.1186/s12989-020-0336-y>
- Ewald, E.R., Wand, G.S., Seifuddin, F., Yang, X., Tamashiro, K.L., Potash, J.B., Zandi, P., Lee, R.S., 2014. Alterations in DNA methylation of *Fkbp5* as a determinant of blood-brain correlation of glucocorticoid exposure. *Psychoneuroendocrinology*. 44, 112–122. <https://doi.org/10.1016/j.psyneuen.2014.03.003>
- Ezaz, T., Stiglec, R., Veyrunes, F., Marshall Graves, J.A., 2006. Relationships between vertebrate ZW and XY sex chromosome systems. *Current Biology*. 16(17), R736–743. <https://doi.org/10.1016/j.cub.2006.08.021>

F

- Faggion, S., Vandeputte, M., Vergnet, A., Clota, F., Blanc, M.-O., Sanchez, P., Ruelle, F., Allal, F., 2021. Sex dimorphism in European sea bass (*Dicentrarchus labrax* L.): New insights into sex-related growth patterns during very early life stages. *PloS One*. 16(4), e0239791. <https://doi.org/10.1371/journal.pone.0239791>
- Felip, A., Zanuy, S., Carrillo, M., 2006. Comparative analysis of growth performance and sperm motility between precocious and non-precocious males in the European sea bass (*Dicentrarchus labrax*, L.). *Aquaculture*. 256(1-4), 570–578. <https://doi.org/10.1016/j.aquaculture.2006.02.014>
- Fellous, A., Labed-Veydert, T., Locrel, M., Voisin, A.-S., Earley, R.L., Silvestre, F., 2018. DNA methylation in adults and during development of the self-fertilizing mangrove rivulus, *Kryptolebias marmoratus*. *Ecology and Evolution*. 8(12), 6016–6033. <https://doi.org/10.1002/ece3.4141>
- Feltrin, A.S., Tahira, A.C., Simões, S.N., Brentani, H., Martins, D.C., 2019. Assessment of complementarity of WGCNA and NERI results for identification of modules associated to schizophrenia spectrum disorders. *PLoS One* 14(1), e0210431. <https://doi.org/10.1371/journal.pone.0210431>
- Feng, B., Li, S., Wang, Q., Tang, L., Huang, F., Zhang, Z., Mahboobe, S., Shao, C., 2021. lncRNA DMRT2-AS acts as a transcriptional regulator of *dmrt2* involving in sex differentiation in the Chinese tongue sole (*Cynoglossus semilaevis*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*. 253, 110542. <https://doi.org/10.1016/j.cbpb.2020.110542>
- Feng, L., Lou, J., 2019. DNA methylation analysis, in: Zhang, Q. (Ed.), *Nanotoxicity: Methods and Protocols*, *Methods in Molecular Biology*. Springer, New York, NY, pp. 181–227. https://doi.org/10.1007/978-1-4939-8916-4_12

- Feron, R., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., Ech , C., Vali re, S., Donnadi u, C., Haffray, P., Bestin, A., Morvezen, R., Acloque, H., Euclide, P.T., Wen, M., Jouano, E., Scharlt, M., Postlethwait, J.H., Schraidt, C., Christie, M.R., Larson, W.A., Herpin, A., Guiguen, Y., 2020. Characterization of a Y-specific duplication/insertion of the anti-Mullerian hormone type II receptor gene based on a chromosome-scale genome assembly of yellow perch, *Perca flavescens*. *Molecular Ecology Resources*. 20, 531–543. <https://doi.org/10.1111/1755-0998.13133>
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., Reik, W., 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*. 473(7347), 398–402. <https://doi.org/10.1038/nature10008>
- Finn, R.N., Kristoffersen, B.A., 2007. Vertebrate vitellogenin gene duplication in relation to the “3R Hypothesis”: correlation to the pelagic egg and the oceanic radiation of teleosts. *PloS One*. 2(1), e169. <https://doi.org/10.1371/journal.pone.0000169>
- Fisher, R.A., 1930. The genetical theory of natural selection. Clarendon.
- Fitz-James, M.H., Cavalli, G., 2022. Molecular mechanisms of transgenerational epigenetic inheritance. *Nature Reviews Genetics*. 23(6), 325–341. <https://doi.org/10.1038/s41576-021-00438-5>
- Freyhof, J., Kottelat, M., 2007. Handbook of European freshwater fishes. ISBN: 9782839902984.
- Fridolfsson, A.K., Cheng, H., Copeland, N.G., Jenkins, N.A., Liu, H.C., Raudsepp, T., Woodage, T., Chowdhary, B., Halverson, J., Ellegren, H., 1998. Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proceedings of the National Academy of Sciences*. 95(14), 8147–8152. <https://doi.org/10.1073/pnas.95.14.8147>
- Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*. 33(1), 1. <https://doi.org/10.18637/jss.v033.i01>
- Frimodt, C., 1995. Multilingual illustrated guide to the world’s commercial coldwater fish. Fishing News Books Ltd.
- Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*. 89(5), 1827–1831. <https://doi.org/10.1073/pnas.89.5.1827>

G

- Garcia de Leon, F.J., Canonne, M., Quillet, E., Bonhomme, F., Chatain, B., 1998. The application of microsatellite markers to breeding programmes in the sea bass, *Dicentrarchus labrax*. *Aquaculture*. 159(3-4), 303–316. [https://doi.org/10.1016/S0044-8486\(97\)00188-9](https://doi.org/10.1016/S0044-8486(97)00188-9)
- Garz n-Porras, A.M., Chory, E., Gryder, B.E., 2023. Dynamic Opposition of Histone Modifications. *ACS Chemical Biology*. 18(4), 1027–1036. <https://doi.org/10.1021/acscchembio.1c01000>
- Gaspar, J.M., Hart, R.P., 2017. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinformatics*. 18, 1-8. <https://doi.org/10.1186/s12859-017-1909-0>
- Gavery, M.R., Roberts, S.B., 2017. Epigenetic considerations in aquaculture. *PeerJ*. 5, e4147. <https://doi.org/10.7717/peerj.4147>
- Ge, C., Ye, J., Weber, C., Sun, W., Zhang, H., Zhou, Y., Cai, C., Qian, G., Capel, B., 2018. The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science*. 360(6389), 645–648. <https://doi.org/10.1126/science.aap8328>
- Geffroy, B., Besson, M., S nchez-Baiz n, N., Clota, F., Goikoetxea, A., Sadoul, B., Ruelle, F., Blanc, M.-O., Parrinello, H., Hermet, S., Blondeau-Bidet, E., Pratlong, M., Piferrer, F., Vandeputte, M., Allal, F., 2021. Unraveling the genotype by environment interaction in a thermosensitive fish with a polygenic sex determination system. *Proceedings of the National Academy of Sciences*. 118(50), e2112660118. <https://doi.org/10.1073/pnas.2112660118>
- Gel, B., D  ez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., Malinverni, R., 2016. regioneR: An R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32(2), 289–291. <https://doi.org/10.1093/bioinformatics/btv562>
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*. 5(10), R80. <https://doi.org/10.1186/gb-2004-5-10-r80>
- Georgakopoulou, E., Sfakianakis, D.G., Kouttoui, S., Divanach, P., Kentouri, M., Koumoundouros, G., 2007. The influence of temperature during early life on phenotypic expression at later ontogenetic stages in sea bass. *Journal of Fish Biology*. 70(1), 278–291. <https://doi.org/10.1111/j.1095-8649.2007.01305.x>

- Gessler, T.B., Wu, Z., Valenzuela, N., 2023. Transcriptomic thermal plasticity underlying gonadal development in a turtle with ZZ/ZW sex chromosomes despite canalized genotypic sex determination. *Ecology and Evolution*. 13(2), e9854. <https://doi.org/10.1002/ece3.9854>
- Gjedrem, T., 2010. The first family-based breeding program in aquaculture. *Reviews in Aquaculture*. 2(1), 2–15. <https://doi.org/10.1111/j.1753-5131.2010.01011.x>
- Gjedrem, T., Aulstad, D., 1974. Selection experiments with salmon: I. Differences in resistance to vibrio disease of salmon parr (*Salmo salar*). *Aquaculture* 3(1), 51–59. [https://doi.org/10.1016/0044-8486\(74\)90098-2](https://doi.org/10.1016/0044-8486(74)90098-2)
- Gjedrem, T., Baranski, M., 2009. Domestication and the Application of Genetic Improvement in Aquaculture, In: Gjedrem, T., Baranski, M. (Eds.), *Selective Breeding in Aquaculture: An Introduction, Reviews: Methods and Technologies in Fish Biology and Fisheries*. Springer Netherlands, Dordrecht, 5–11 pp. https://doi.org/10.1007/978-90-481-2773-3_2
- Goikoetxea, A., Todd, E.V., Gemmell, N.J., 2017. Stress and sex: does cortisol mediate sex change in fish? *Reproduction*. 154(6), R149–R160. <https://doi.org/10.1530/REP-17-0408>
- Gong, Z., Zhu, J.-K., 2011. Active DNA demethylation by oxidation and repair. *Cell Research*. 21(12), 1649–1651. <https://doi.org/10.1038/cr.2011.140>
- Gonzalzo, M.L., Liang, G., 2007. Methylation-sensitive single-nucleotide primer extension (Ms-SNuPE) for quantitative measurement of DNA methylation. *Nature Protocols* 2(8), 1931–1936. <https://doi.org/10.1038/nprot.2007.271>
- Gornati, R., Terova, G., Vigetti, D., Prati, M., Saroglia, M., Bernardini, G., 2004. Effects of population density on seabass (*Dicentrarchus labrax*, L.) gene expression. *Aquaculture*. 230(1-4), 229–239. [https://doi.org/10.1016/S0044-8486\(03\)00426-5](https://doi.org/10.1016/S0044-8486(03)00426-5)
- Goto-Kazeto, R., Saito, T., Takagi, M., Arai, K., Yamaha, E., 2011. Isolation of teleost primordial germ cells using flow cytometry. *International Journal of Developmental Biology*. 54(10), 1487–1492. <https://doi.org/10.1387/ijdb.092914rg>
- Gouil, Q., Keniry, A., 2019. Latest techniques to study DNA methylation. *Essays in Biochemistry*. 63(6), 639–648. <https://doi.org/10.1042/EBC20190027>
- Grafodatskaya, D., Rens, W., Wallis, M.C., Trifonov, V., O'Brien, P.C., Clarke, O., Graves, J.A., Ferguson-Smith, M.A., 2007. Search for the sex-determining switch in monotremes: mapping *WT1*, *SFI*, *LHX1*, *LHX2*, *FGF9*, *WNT4*, *RSP01* and *GATA4* in platypus. *Chromosome Research*. 15, 777–785. <https://doi.org/10.1007/s10577-007-1161-y>
- Graham, D.B., Root, D.E., 2015. Resources for the design of CRISPR gene editing experiments. *Genome Biology*. 16(1), 260. <https://doi.org/10.1186/s13059-015-0823-x>
- Graham, P., Penn, J.K.M., Schedl, P., 2003. Masters change, slaves remain. *BioEssays*. 25(1), 1–4. <https://doi.org/10.1002/bies.10207>
- Granada, L., Lemos, M.F.L., Cabral, H.N., Bossier, P., Novais, S.C., 2018. Epigenetics in aquaculture – the last frontier. *Reviews in Aquaculture*. 10(4), 994–1013. <https://doi.org/10.1111/raq.12219>
- Grandjean, V., Hauck, Y., Beloin, C., Le, H.F., Hirschbein, L., 1998. Chromosomal inactivation of *Bacillus subtilis* exfusants: a prokaryotic model of epigenetic regulation. *Biological Chemistry*. 379(4-5), 553–557.
- Grehl, C., Kuhlmann, M., Becker, C., Glaser, B., Grosse, I., 2018. How to design a whole-genome bisulfite sequencing experiment? *Epigenomes*. 2(4), 21. <https://doi.org/10.3390/epigenomes2040021>
- Griffin, P.T., Kane, A.E., Trapp, A., Li, J., McNamara, M.S., Meer, M.V., MacArthur, M.R., Mitchell, S.J., Mueller, A.L., Carmody, C., Vera, D.L., 2021. Ultra-cheap and scalable epigenetic age predictions with TIME-Seq. *BioRxiv* 25, 465725. <https://doi.org/10.1101/2021.10.25.465725>
- Grimes, G.W., Aufderheide, K.J., 1991. Cellular aspects of pattern formation: The problem of assembly. *Monographs in Developmental Biology*. 22:1-94. Karger Publishers.
- Griot, R., Allal, F., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A., Haffray, P., François, Y., Morin, T., Poncet, C., Vergnet, A., Cariou, S., Brunier, J., Bruant, J.S., Peyrou, B., Gagnaire, P.A., Vandeputte, M., 2021. Genome-wide association studies for resistance to viral nervous necrosis in three populations of European sea bass (*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip. *Aquaculture*. 530, 735930. <https://doi.org/10.1016/j.aquaculture.2020.735930>
- Grunau, C., Clark, S.J., Rosenthal, A., 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Research*. 29(13), e65. <https://doi.org/10.1093/nar/29.13.e65>
- Guevara, M.Á., de María, N., Sáez-Laguna, E., Vélez, M.D., Cervera, M.T., Cabezas, J.A., 2017. Analysis of DNA Cytosine Methylation Patterns Using Methylation-Sensitive Amplification Polymorphism (MSAP). *Plant Epigenetics: Methods and Protocols*. 99–112. https://doi.org/10.1007/978-1-4899-7708-3_9
- Guiguen, Y., Fostier, A., Piferrer, F., Chang, C.-F., 2010. Ovarian aromatase and estrogens: A pivotal role for gonadal sex differentiation and sex change in fish. *General and Comparative Endocrinology, Fish Reproduction* 165(3), 352–366. <https://doi.org/10.1016/j.ygcen.2009.03.002>
- Guillotin, D., Taylor, A.R., Platé, M., Mercer, P.F., Edwards, L.M., Haggart, R., Miele, G., McNulty, R.J., Maher, T.M., Hynds, R.E., Jamal-Hanjani, M., Marshall, R.P., Fisher, A.J., Blanchard, A.D., Chambers, R.C., 2021.

- Transcriptome analysis of IPF fibroblastic foci identifies key pathways involved in fibrogenesis. *Thorax*. 76(1), 73–82. <https://doi.org/10.1136/thoraxjnl-2020-214902>
- Guo, S., Yan, F., Xu, J., Bao, Y., Zhu, J., Wang, X., Wu, J., Li, Y., Pu, W., Liu, Y., Jiang, Z., Ma, Y., Chen, X., Xiong, M., Jin, L., Wang, J., 2015. Identification and validation of the methylation biomarkers of non-small cell lung cancer (NSCLC). *Clinical Epigenetics*. 7, 1–10. <https://doi.org/10.1186/s13148-014-0035-3>
- Guo, T., Li, X., 2023. Machine learning for predicting phenotype from genotype and environment. *Current Opinion in Biotechnology*. 79, 102853. <https://doi.org/10.1016/j.copbio.2022.102853>
- Guo, Y., Yin, J., Dai, Y., Guan, Y., Chen, P., Chen, Y., Huang, C., Lu, Y.-J., Zhang, L., Song, D., 2021. A novel CpG methylation risk indicator for predicting prognosis in bladder cancer. *Frontiers in Cell and Developmental Biology*. 9, 642650. <https://doi.org/10.3389/fcell.2021.642650>
- Gupta, S., Ellis, S.E., Ashar, F.N., Moes, A., Bader, J.S., Zhan, J., West, A.B., Arking, D.E., 2014. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nature Communications*. 5(1), 5748. <https://doi.org/10.1038/ncomms6748>
- Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. *Machine Learning*. 46, 389–422. <https://doi.org/10.1023/A:1012487302797>
- Gwynne, J.T., Strauss, J.F., III, 1982. The role of lipoproteins in steroidogenesis and cholesterol metabolism in steroidogenic glands. *Endocrine Reviews* 3(3), 299–329. <https://doi.org/10.1210/edrv-3-3-299>
- ## H
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., Surani, M.A., 2002. Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of Development*. 117(1-2), 15–23. [https://doi.org/10.1016/S0925-4773\(02\)00181-8](https://doi.org/10.1016/S0925-4773(02)00181-8)
- Hamburger, V., Hamilton, H.L., 1951. A series of normal stages in the development of the chick embryo. *Journal of Morphology*. 88(1), 49–92. <https://doi.org/10.1002/jmor.1050880104>
- Hansen, C.L., Pelegri, F., 2021. Primordial germ cell specification in vertebrate embryos: phylogenetic distribution and conserved molecular features of preformation and induction. *Frontiers in Cell and Developmental Biology*. 9, 730332. <https://doi.org/10.3389/fcell.2021.730332>
- Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., Strüssmann, C.A., 2012. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proceedings of the National Academy of Sciences*. 109(8), 2955–2959. <https://doi.org/10.1073/pnas.1018392109>
- Hawe, J.S., Wilson, R., Schmid, K.T., Zhou, L., Lakshmanan, L.N., Lehne, B.C., Kühnel, B., Scott, W.R., Wielscher, M., Yew, Y.W., Baumbach, C., Lee, D.P., Marouli, E., Bernard, M., Pfeiffer, L., Matias-García, P.R., Autio, M.I., Bourgeois, S., Herder, C., Karhunen, V., Meitinger, T., Prokisch, H., Rathmann, W., Roden, M., Seibert, S., Shin, J., Strauch, K., Zhang, W., Tan, W.L.W., Hauck, S.M., Merl-Pham, J., Grallert, H., Barbosa, E.G.V., Illig, T., Peters, A., Paus, T., Pausova, Z., Deloukas, P., Foo, R.S.Y., Jarvelin, M.-R., Kooner, J.S., Loh, M., Heinig, M., Gieger, C., Waldenberger, M., Chambers, J.C., 2022. Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. *Nature Genetics*. 54(1), 18–29. <https://doi.org/10.1038/s41588-021-00969-x>
- Hawkins, R.D., Hon, G.C., Ren, B., 2010. Next-generation genomics: An integrative approach. *Nature Reviews Genetics*. 11(7), 476–486. <https://doi.org/10.1038/nrg2795>
- He, W.-X., Wu, M., Liu, Z., Li, Z., Wang, Y., Zhou, J., Yu, P., Zhang, X.-J., Zhou, L., Gui, J.-F., 2018. Oocyte-specific maternal *Slbp2* is required for replication-dependent histone storage and early nuclear cleavage in zebrafish oogenesis and embryogenesis. *RNA* 24(12), 1738–1748. <https://doi.org/10.1261/rna.067090.118>
- He, X., Chang, S., Zhang, J., Zhao, Q., Xiang, H., Kusonmano, K., Yang, L., Sun, Z.S., Yang, H., Wang, J., 2008. MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Research*. 36(suppl. 1), D836–D841. <https://doi.org/10.1093/nar/gkm730>
- He, Zhi, Deng, F., Yang, D., He, Zhide, Hu, J., Ma, Z., Zhang, Q., He, J., Ye, L., Chen, H., He, L., Luo, J., Xiong, S., Luo, W., Yang, S., Gu, X., Yan, T., 2022. Crosstalk between sex-related genes and apoptosis signaling reveals molecular insights into sex change in a protogynous hermaphroditic teleost fish, ricefield eel *Monopterus albus*. *Aquaculture*. 552, 737918. <https://doi.org/10.1016/j.aquaculture.2022.737918>
- Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., Lumey, L.H., 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences*. 105(4), 17046–17049. <https://doi.org/10.1073/pnas.0806560105>
- Henderson, I.R., Jacobsen, S.E., 2007. Epigenetic inheritance in plants. *Nature*. 447(7143), 418–424. <https://doi.org/10.1038/nature05917>
- Hernández, H.G., Tse, M.Y., Pang, S.C., Arboleda, H., Forero, D.A., 2013. Optimizing methodologies for PCR-based DNA methylation analysis. *BioTechniques*. 55, 181–197. <https://doi.org/10.2144/000114087>

- Herpin, A., Scharl, M., 2015. Plasticity of gene-regulatory networks controlling sex determination: Of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO reports*. 16(10), 1260–1274. <https://doi.org/10.15252/embr.201540667>
- Herpin, A., Scharl, M., Depincé, A., Guiguen, Y., Bobe, J., Hua-Van, A., Hayman, E.S., Octavera, A., Yoshizaki, G., Nichols, K.M., Goetz, G.W., Luckenbach, J.A., 2021. Allelic diversification after transposable element exaptation promoted *gsdf* as the master sex determining gene of sablefish. *Genome Research*. 31(8), 1366–1380. <https://doi.org/10.1101/gr.274266.120>
- Heule, C., Salzburger, W., Böhne, A., 2014. Genetics of sexual development: an evolutionary playground for fish. *Genetics*. 196(3), 579–591. <https://doi.org/10.1534/genetics.114.161158>
- Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J.V., Setien, F., Carmona, F.J., Puca, A.A., Sayols, S., Pujana, M.A., Serra-Musach, J., Iglesias-Platas, I., Formiga, F., Fernandez, A.F., Fraga, M.F., Heath, S.C., Valencia, A., Gut, I.G., Wang, J., Esteller, M., 2012. Distinct DNA methylomes of newborns and centenarians. *Proceedings of the National Academy of Sciences*. 109(26), 10522–10527. <https://doi.org/10.1073/pnas.1120658109>
- Hirst, C.E., Major, A.T., Smith, C.A., 2018. Sex determination and gonadal sex differentiation in the chicken model. *International Journal of Developmental Biology*. 62(1-3), 153–166. <https://doi.org/10.1387/ijdb.170319cs>
- Horvath, S., 2013. DNA methylation age of human tissues and cell types. *Genome biology*. 14(10), 1–20. <https://doi.org/10.1186/gb-2013-14-10-r115>
- Horvath, S., Dong, J., 2008. Geometric interpretation of gene coexpression network analysis. *PLoS Computational Biology*. 4(8), e1000117. <https://doi.org/10.1371/journal.pcbi.1000117>
- Hosseini, M., Lotfi-Shahreza, M., Nikpour, P., 2023. Integrative analysis of DNA methylation and gene expression through machine learning identifies stomach cancer diagnostic and prognostic biomarkers. *Journal of Cellular and Molecular Medicine*. 27(5), 714–726. <https://doi.org/10.1111/jcmm.17693>
- Houston, R.D., Bean, T.P., Macqueen, D.J., Gundappa, M.K., Jin, Y.H., Jenkins, T.L., Selly, S.L.C., Martin, S.A.M., Stevens, J.R., Santos, E.M., Davie, A., Robledo, D., 2020. Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*. 21(7), 389–409. <https://doi.org/10.1038/s41576-020-0227-y>
- Hu, J., Wuitchik, S.J.S., Barry, T.N., Jamniczky, H.A., Rogers, S.M., Barrett, R.D.H., 2021. Heritability of DNA methylation in threespine stickleback (*Gasterosteus aculeatus*). *Genetics*. 217(1), iyab001. <https://doi.org/10.1093/genetics/iyab001>
- Huang, H.Y., Li, J., Tang, Y., Huang, Y.X., Chen, Y.G., Xie, Y.Y., Zhou, Z.Y., Chen, X.Y., Ding, S.Y., Luo, M.F., Jin, C.N., Zhao, L.S., Xu, J.T., Zhou, Y., Lin, Y.C.D., Hong, H.C., Zuo, H.L., Hu, S.Y., Xu, P.Y., Li, X., Huang, H.D., 2021. MethHC 2.0: information repository of DNA methylation and gene expression in human cancer. *Nucleic Acids Research*. 49(D1), D1268–D1275. <https://doi.org/10.1093/nar/gkaa1104>
- Huang, S., Cai, N., Pacheco, P.P., Narrandes, S., Wang, Y., Xu, W., 2018. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer genomics & proteomics*. 15(1), 41–51. <https://cgp.iiarjournals.org/content/15/1/41.short>
- Huang, Z., Ma, A., Yang, S., Liu, X., Zhao, T., Zhang, J., Wang, X., Sun, Z., Liu, Z., Xu, R., 2020. Transcriptome analysis and weighted gene co-expression network reveals potential genes responses to heat stress in turbot *Scophthalmus maximus*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. 33, 100632. <https://doi.org/10.1016/j.cbd.2019.100632>
- Hughes, I.A., Houk, C., Ahmed, S.F., Lee, P.A., 2006. Consensus statement on management of intersex disorders. *Journal of Pediatric Urology*. 2(3), 148–162. <https://doi.org/10.1016/j.jpuro.2006.03.004>

I

- Iancu, O.D., Colville, A., Oberbeck, D., Darakjian, P., McWeeney, S.K., Hitzemann, R., 2015. Cosplicing network analysis of mammalian brain RNA-Seq data utilizing WGCNA and Mantel correlations. *Frontiers in Genetics*. 6, 174. <https://doi.org/10.3389/fgene.2015.00174>
- International Protein Nomenclature Guidelines [WWW Document], 2020. URL https://www.ncbi.nlm.nih.gov/genome/doc/internatprot_nomenguide/ (accessed 2.23.23).
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., Zhang, Y., 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*. 466(7310), 1129–1133. <https://doi.org/10.1038/nature09303>
- Ito, T., 2007. Role of histone modification in chromatin dynamics. *The Journal of Biochemistry*. 141(5), 609–614. <https://doi.org/10.1093/jb/mvm091>

J

- Jablonka, E., Lamb, M.J., 2007. Précis of evolution in four dimensions. *Behavioral and brain sciences*. 30(4), 353–365. <https://doi.org/10.1017/S0140525X07002221>
- Jablonka, E., Lamb, M.J., 2006. The evolution of information in the major transitions. *Journal of Theoretical Biology*. 239(2), 236–246. <https://doi.org/10.1016/j.jtbi.2005.08.038>
- Jablonka, E., Lamb, M.J., 2002. The changing concept of epigenetics. *Annals of the New York Academy of Sciences*. 981(1), 82–96. <https://doi.org/10.1111/j.1749-6632.2002.tb04913.x>
- Jablonka, E., Raz, G., 2009. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly Review of Biology*. 84(2), 131–176. <https://doi.org/10.1086/598822>
- Jiang, C., Li, Z., Wu, Z., Liang, Y., Jin, L., Cao, Y., Wan, S., Chen, Z., 2020. Integrated bioinformatics analysis of hub genes and pathways associated with a compression model of spinal cord injury in rats. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*. 26, e927107-1-e927107-13. <https://doi.org/10.12659/MSM.927107>
- Jiang, D.-N., Huang, Y.-Q., Zhang, J.M., Mustapha, U.F., Peng, Y.X., Huang, H., Li, G.L., 2022. Establishment of the Y-linked *Dmrt1Y* as the candidate sex determination gene in spotbanded scat (*Selenotoca multifasciata*). *Aquaculture Reports*. 23, 101085. <https://doi.org/10.1016/j.aqrep.2022.101085>
- Jiang, L., Zhang, Jing, Wang, J.J., Wang, L., Zhang, L., Li, G., Yang, X., Ma, X., Sun, X., Cai, J., Zhang, Jun, Huang, X., Yu, M., Wang, X., Liu, F., Wu, C.I., He, C., Zhang, B., Ci, W., Liu, J., 2013. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*. 153(4), 773–784. <https://doi.org/10.1016/j.cell.2013.04.041>
- Jin, Z., Liu, Y., 2018. DNA methylation in human diseases. *Genes & Diseases*. 5(1), 1–8. <https://doi.org/10.1016/j.gendis.2018.01.002>
- Johnson, A.D., Alberio, R., 2015. Primordial germ cells: the first cell lineage or the last cells standing? *Development*. 142(16), 2730–2739. <https://doi.org/10.1242/dev.113993>
- Jones, P.A., Baylin, S.B., 2007. The epigenomics of cancer. *Cell*. 128(4), 683–692. <https://doi.org/10.1016/j.cell.2007.01.029>
- Jonsson, B., Jonsson, N., 2019. Phenotypic plasticity and epigenetics of fish: Embryo temperature affects later-developing life-history traits. *Aquatic Biology*. 28, 21–32. <https://doi.org/10.3354/ab00707>
- Jonsson, B., Jonsson, N., 2014. Early environment influences later performance in fishes: Effects of early experiences. *Journal of Fish Biology*. 85(2), 151–188. <https://doi.org/10.1111/jfb.12432>

K

- Kakati, T., Bhattacharyya, D.K., Barah, P., Kalita, J.K., 2019. Comparison of methods for differential co-expression analysis for disease biomarker prediction. *Computers in Biology and Medicine*. 113, 103380. <https://doi.org/10.1016/j.compbiomed.2019.103380>
- Kakati, T., Kashyap, H., Bhattacharyya, D.K., 2016. THD-Module Extractor: An application for CEN module extraction and interesting gene identification for Alzheimer's disease. *Scientific Reports*. 6(1), 38046. <https://doi.org/10.1038/srep38046>
- Kallman, K.D., 1984. A new look at sex determination in *Poeciliid* fishes, in: Turner, B.J. (Ed.), *Evolutionary Genetics of Fishes*, Monographs in Evolutionary Biology. Springer US, Boston, MA, pp. 95–171. https://doi.org/10.1007/978-1-4684-4652-4_3
- Kamińska, K., Nalejska, E., Kubiak, M., Wojtyśiak, J., Żoła, Ł., Kowalewski, J., Lewandowska, M.A., 2019. Prognostic and predictive epigenetic biomarkers in oncology. *Molecular Diagnosis & Therapy*. 23, 83–95. <https://doi.org/10.1007/s40291-018-0371-7>
- Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., Kikuchi, K., 2012. A trans-species missense SNP in *amhr2* is associated with sex determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genetics*. 8(7), e1002798. <https://doi.org/10.1371/journal.pgen.1002798>
- Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., Henikoff, S., 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications*. 10(1), 1–10. <https://doi.org/10.1038/s41467-019-09982-5>
- Keating, S.E., Fenelon, J.C., Pyne, M., Pinto, B.J., Guzmán-Méndez, I.A., Johnston, S.D., Renfree, M.B., Gamble, T., 2022. Genetic sex test for the short-beaked echidna (*Tachyglossus aculeatus*). *Conservation Genetics Resources*. 14(3), 271–278. <https://doi.org/10.1007/s12686-022-01258-3>
- Kerachian, M.A., Javadmanesh, A., Azghandi, M., Mojtabanezhad Shariatpanahi, A., Yassi, M., Shams Davodly, E., Talebi, A., Khadangi, F., Soltani, G., Hayatbakhsh, A., Ghaffarzadegan K., 2020. Crosstalk between DNA

- methylation and gene expression in colorectal cancer, a potential plasma biomarker for tracing this tumor. *Scientific reports*. 10(1), 1–13. <https://doi.org/10.1038/s41598-020-59690-0>
- Khazaei, T., Barlow, J.T., Schoepp, N.G., Ismagilov, R.F., 2018. RNA markers enable phenotypic test of antibiotic susceptibility in *Neisseria gonorrhoeae* after 10 minutes of ciprofloxacin exposure. *Scientific Reports*. 8(1), 11606. <https://doi.org/10.1038/s41598-018-29707-w>
- Kim, A., Yoon, D., Lim, Y., Roh, H.J., Kim, S., Park, C.I., Kim, H.S., Cha, H.J., Choi, Y.H., Kim, D.-H., 2020. Co-Expression Network Analysis of Spleen Transcriptome in Rock Bream (*Oplegnathus fasciatus*) Naturally Infected with Rock Bream Iridovirus (RBIV). *International Journal of Molecular Sciences*. 21(5), E1707. <https://doi.org/10.3390/ijms21051707>
- Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*. 12(4), 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*. 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kim, E.Y., Kim, K.D., Cho, J., 2022. Harnessing epigenetic variability for crop improvement: current status and future prospects. *Genes and Genomics*. 44, 259–266. <https://doi.org/10.1007/s13258-021-01189-7>
- Kim, H., Kim, S., Jung, Y., Han, J., Yun, J.H., Chang, I., Lee, W., 2016. Probing the folding-unfolding transition of a thermophilic protein, MTH1880. *PLoS One* 11(1), e0145853. <https://doi.org/10.1371/journal.pone.0145853>
- Kim, M., 2019. DNA methylation: a cause and consequence of type 2 diabetes. *Genomics and Informatics*. 17(4), e38. <https://doi.org/10.5808/GI.2019.17.4.e38>
- Kim, T.H., Dekker, J., 2018. ChIP-chip. *Cold Spring Harbor Protocols*. 5, pdb-rot082628. <https://doi.org/10.1101/pdb.prot082636>
- Kindt, R., Coe, R., 2005. Tree diversity analysis. A manual and software for common statistical methods for ecological and biodiversity studies. World Agroforestry Centre (ICRAF), Nairobi (Kenya).
- Kirubakaran, T.G., Andersen, Ø., De Rosa, M.C., Andersstuen, T., Hallan, K., Kent, M.P., Lien, S., 2019. Characterization of a male specific region containing a candidate sex determining gene in Atlantic cod. *Scientific Reports*. 9(1), 116. <https://doi.org/10.1038/s41598-018-36748-8>
- Klingenberg, C.P., 2019. Phenotypic plasticity, developmental instability, and robustness: the concepts and how they are connected. *Frontiers in Ecology and Evolution*. 7, 7-56. <https://doi.org/10.3389/fevo.2019.00056>
- Klughammer, J., Romanovskaia, D., Nemc, A., Posautz, A., Seid, C.A., Schuster, L.C., Keinath, M.C., Lugo Ramos, J.S., Kosack, L., Evankow, A., Printz, D., Kirchberger, S., Ergüner, B., Datlinger, P., Fortelny, N., Schmidl, C., Farlik, M., Skjærven, K., Berghaler, A., Liedvogel, M., Thaller, D., Burger, P.A., Hermann, M., Distel, M., Distel, D.L., Küber-Heiss, A., Bock, C., 2023. Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. *Nature Communications*. 14(1), 232. <https://doi.org/10.1038/s41467-022-34828-y>
- Kobayashi, T., Matsuda, M., Kajiura-Kobayashi, H., Suzuki, A., Saito, N., Nakamoto, M., Shibata, N., Nagahama, Y., 2004. Two DM domain genes, *DMY* and *DMRT1*, involved in testicular differentiation and development in the medaka, *Oryzias latipes*. *Developmental dynamics: an official publication of the American Association of Anatomists*. 231(3), 518–526. <https://doi.org/10.1002/dvdy.20158>
- Kobayashi, Y., Alam, M.A., Horiguchi, R., Shimizu, A., Nakamura, M., 2010. Sexually Dimorphic Expression of Gonadotropin Subunits in the Pituitary of Protogynous Honeycomb Grouper (*Epinephelus merra*): Evidence That Follicle-Stimulating Hormone (FSH) Induces Gonadal Sex Change1. *Biology of Reproduction*. 82(6), 1030–1036. <https://doi.org/10.1095/biolreprod.109.080986>
- Koch, C., Konieczka, J., Delorey, T., Lyons, A., Socha, A., Davis, K., Knaack, S.A., Thompson, D., O'Shea, E.K., Regev, A., Roy, S., 2017. Inference and evolutionary analysis of genome-scale regulatory networks in large phylogenies. *Cell Systems*. 4(5), 543-558.e8. <https://doi.org/10.1016/j.cels.2017.04.010>
- Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., Lovell-Badge, R., 1991. Male development of chromosomally female mice transgenic for *Sry*. *Nature*. 351(6322), 117–121. <https://doi.org/10.1038/351117a0>
- Kosswig, C., 1964. Polygenic sex determination. *Experientia*. 20(4), 190–199. <https://doi.org/10.1007/BF02135395>
- Kottelat, M., Freyhof, J., 2007. Handbook of European freshwater fishes. Publications Kottelat. ISBN: 9782839902984.
- Koumoundouros, G., Divanach, P., Anezaki, L., Kentouri, M., 2001. Temperature-induced ontogenetic plasticity in sea bass (*Dicentrarchus labrax*). *Marine Biology*. 139, 817–830. <https://doi.org/10.1007/s002270100635>
- Koumoundouros, G., Sfakianakis, D., Divanach, P., Kentouri, M., 2002. Effect of temperature on swimming performance of sea bass juveniles. *Journal of Fish Biology*. 60(4), 923–932. <https://doi.org/10.1111/j.1095-8649.2002.tb02418.x>
- Koyama, T., Nakamoto, M., Morishima, K., Yamashita, R., Yamashita, T., Sasaki, K., Kuruma, Y., Mizuno, N., Suzuki, M., Okada, Y., Ieda, R., Uchino, T., Tasumi, S., Hosoya, S., Uno, S., Koyama, J., Toyoda, A., Kikuchi,

- K., Sakamoto, T., 2019. A SNP in a steroidogenic enzyme is associated with phenotypic sex in *Seriola* Fishes. *Current Biology*. 29(11), 1901–1909.e8. <https://doi.org/10.1016/j.cub.2019.04.069>
- Kratochvíl, L., Gamble, T., Rovatsos, M., 2021. Sex chromosome evolution among amniotes: Is the origin of sex chromosomes non-random? *Philosophical Transactions of the Royal Society B: Biological Sciences*. 376(1833), 20200108. <https://doi.org/10.1098/rstb.2020.0108>
- Krick, M.V., Desmarais, E., Samaras, A., Guéret, E., Dimitroglou, A., Pavlidis, M., Tsigenopoulos, C., Guinand, B., 2021. Family-effects in the epigenomic response of red blood cells to a challenge test in the European sea bass (*Dicentrarchus labrax*, L.). *BMC Genomics*. 22(1), 111. <https://doi.org/10.1186/s12864-021-07420-9>
- Kroon, H., van Groenendaal, J., (eds.). 1997. The ecology and evolution of clonal plants. Backhuys Publishers, Leiden. The New Phytologist, 143(2), 239–241. ISBN: 9073348730. <https://doi.org/10.1046/j.1469-8137.1999.00462.x>
- Krueger, F., Andrews, S.R., 2011a. Bismark: a flexible aligner and methylation caller for Bisulfite-seq applications. *Bioinformatics* 27, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>
- Kuhn, M., 2022. caret: Classification and regression training. R package version 6.0-93.
- Kuhn, M., 2015. caret: Classification and regression training. *Astrophysics Source Code. Library ascl-1505*, 003.
- Kuhn, M., 2008. Building predictive models in R using the caret package. *Journal of Statistical Software*. 28, 1–26. <https://doi.org/10.18637/jss.v028.i05>

L

- Lallias, D., Bernard, M., Ciobotaru, C., Dechamp, N., Labbé, L., Goardon, L., Le Calvez, J.-M., Bideau, M., Fricot, A., Prézélin, A., Charles, M., Moroldo, M., Cousin, X., Bouchez, O., Roulet, A., Quillet, E., Dupont-Nivet, M., 2021. Sources of variation of DNA methylation in rainbow trout: combined effects of temperature and genetic background. *Epigenetics*. 16(9), 1031–1052. <https://doi.org/10.1080/15592294.2020.1834924>
- Langfelder, P., Horvath, S., 2017. WGCNA package: Frequently Asked Questions [WWW Document]. URL <https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html> (accessed 7.2.23).
- Langfelder, P., Horvath, S., 2012. Fast R functions for robust correlations and hierarchical clustering. *Journal of Statistical Software*. 46(11). <https://doi.org/10.18637/jss.v046.i11>
- Langfelder, P., Horvath, S., 2008. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. 9(1), 559. <https://doi.org/10.1186/1471-2105-9-559>
- Lata, C., Sahu, P.P., Prasad, M., 2010. Comparative transcriptome analysis of differentially expressed genes in foxtail millet (*Setaria italica* L.) during dehydration stress. *Biochemical and Biophysical Research Communications*. 393(4), 720–727. <https://doi.org/10.1016/j.bbrc.2010.02.068>
- Law, C.W., Alhamdoosh, M., Su, S., Smyth, G.K., Ritchie, M.E., 2016. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research*. 5. <https://doi.org/10.12688/f1000research.9005.3>
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M., Carey, V., 2013. Software for computing and annotating genomic ranges. *PLoS Computational Biology*. 9(8). <https://doi.org/10.1371/journal.pcbi.1003118>
- Lazado, C.C., Voldvik, V., Breiland, M.W., Osório, J., Hansen, M.H.S., Krasnov, A., 2020. Oxidative chemical stressors alter the physiological state of the nasal olfactory mucosa of Atlantic salmon. *Antioxidants*. 9(11), 1144. <https://doi.org/10.3390/antiox9111144>
- Lecluze, E., Rolland, A.D., Filis, P., Evrard, B., Leverrier-Penna, S., Maamar, M.B., Coiffec, I., Lavoué, V., Fowler, P.A., Mazaud-Guittot, S., Jégou, B., Chalmel, F., 2020. Dynamics of the transcriptional landscape during human fetal testis and ovary development. *Human Reproduction*. 35(5), 1099–1119. <https://doi.org/10.1093/humrep/deaa041>
- Leisch, F., Hornik, K., Ripley, B.D., Narasimhan B. Hastie, T., Tibshirani, R. 2022. mda: Mixture and flexible discriminant analysis. R package version 0.5-3. <https://CRAN.R-project.org/package=mda>
- Lempe, J., Flachowsky, H., Peil, A., 2022. Exploring epigenetic variation for breeding climate resilient apple crops. *Physiologia Plantarum*. 174(5), e13782. <https://doi.org/10.1111/ppl.13782>
- Leroux, S., Gourichon, D., Leterrier, C., Labrune, Y., Coustham, V., Rivière, S., Zerjal, T., Coville, J.L., Morisson, M., Minvielle, F., Pitel, F., 2017. Embryonic environment and transgenerational effects in quail. *Genetics Selection Evolution*. 49(1), 14. <https://doi.org/10.1186/s12711-017-0292-7>
- Levenson, V.V., 2010. DNA methylation as a universal biomarker. *Expert Review of Molecular Diagnostics* 10(4), 481–488. <https://doi.org/10.1586/erm.10.17>
- Li, D., Brown, J.B., Orsini, L., Pan, Z., Hu, G., He, S., 2022. MODA: MOdule Differential Analysis for weighted gene co-expression network. Preprint arXiv:1605.04739. <https://doi.org/10.18129/B9.bioc.MODA>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

- Li, L.-C., Dahiya, R., 2002. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18(11), 1427–1431. <https://doi.org/10.1093/bioinformatics/18.11.1427>
- Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., Jiang, D., Zhou, L., Sun, L., Tao, W., Nagahama, Y., 2015. A tandem duplicate of anti-Müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLoS Genetics* 11(11), e1005678. <https://doi.org/10.1371/journal.pgen.1005678>
- Li, N., Ye, M., Li, Y., Yan, Z., Butcher, L.M., Sun, J., Han, X., Chen, Q., Zhang, X., Wang, J., 2010. Whole genome DNA methylation analysis based on high throughput sequencing technology. *Methods*. 52(3), 203–212. <https://doi.org/10.1016/j.ymeth.2010.04.009>
- Li, Y., Chen, Z., Liu, H., Li, Q., Lin, X., Ji, S., Li, R., Li, S., Fan, W., Zhao, H., Zhu, Z., Hu, W., Zhou, Y., Luo, D., 2021. ASER: Animal Sex Reversal Database. *Genomics, Proteomics & Bioinformatics*. 19(6), 873–881. <https://doi.org/10.1016/j.gpb.2021.10.001>
- Li, Y., Liu, Y., Strickland, F.M., Richardson, B., 2010. Age-dependent decreases in DNA methyltransferase levels and low transmethylation micronutrient levels synergize to promote overexpression of genes implicated in autoimmunity and acute coronary syndromes. *Experimental Gerontology*. 45(4), 312–322. <https://doi.org/10.1016/j.exger.2009.12.008>
- Liao, Yang, Smyth, G.K., Shi, W., 2014. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 30(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 30(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Lillicrop, K.A., Phillips, E.S., Jackson, A.A., Hanson, M.A., Burdge, G.C., 2005. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *The Journal of Nutrition*. 135(6), 1382–1386. <https://doi.org/10.1093/jn/135.6.1382>
- Lim, H.N., van Oudenaarden, A., 2007. A multistep epigenetic switch enables the stable inheritance of DNA methylation states. *Nature Genetics*. 39(2), 269–275. <https://doi.org/10.1038/ng1956>
- Lin, X., Li, J., Zhao, Q., Feng, J.R., Gao, Q., Nie, J.Y., 2018. WGCNA reveals key roles of IL8 and MMP-9 in progression of involvement area in colon of patients with ulcerative colitis. *Current Medical Science*. 38, 252–258. <https://doi.org/10.1007/s11596-018-1873-6>
- Liu, B., Zhan, Y., Chen, X., Hu, X., Wu, B., Pan, S., 2021. Weighted gene co-expression network analysis can sort cancer-associated fibroblast-specific markers promoting bladder cancer progression. *Journal of Cellular Physiology*. 236, 1321–1331. <https://doi.org/10.1002/jcp.29939>
- Liu, F., Xu, H., Ni, W., Wang, Y., Hong, X., Li, W., Yu, L., Chen, C., Wei, C., Liu, X., Zhu, X., 2022. Temporal variation in DNA methylation during gonadal development in a reptile with temperature-dependent sex determination. *Biology of Reproduction*. 107(5), 1217–1227. <https://doi.org/10.1093/biolre/ioac142>
- Liu, N., Li, Z., Pei, D., Shu, X., 2013. Zfyve9a regulates the proliferation of hepatic cells during zebrafish embryogenesis. *International Journal of Developmental Biology*. 57(9-10), 773–778. <https://doi.org/10.1387/ijdb.130065xs>
- Liu, Q., Yang, J., Gong, Y., Cai, J., Zheng, Y., Zhang, Y., Yu, D., Zhang, Z., 2020. MicroRNA profiling identifies biomarkers in head kidneys of common carp exposed to cadmium. *Chemosphere*. 247, 125901. <https://doi.org/10.1016/j.chemosphere.2020.125901>
- Liu, S., Ye, T., Li, Z., Li, J., Jamil, A.M., Zhou, Y., Hua, G., Liang, A., Deng, T., Yang, L., 2019. Identifying hub genes for heat tolerance in water buffalo (*Bubalus bubalis*) using transcriptome data. *Frontiers in Genetics*. 10, 209. <https://doi.org/10.3389>
- Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., Song, C.X., 2019. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature Biotechnology*. 37(4), 424–429. <https://doi.org/10.1038/s41587-019-0041-2>
- Liu, Z., Zhang, X.J., Wang, W., Zhang, J., Li, Z., Gui, J.F., 2015. Molecular characterization and expression of an oocyte-specific histone stem-loop binding protein in *Carassius gibelio*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*. 190, 46–53. <https://doi.org/10.1016/j.cbpb.2015.08.011>
- Locke, W.J., Guanzone, D., Ma, C., Liew, Y.J., Duesing, K.R., Fung, K.Y.C., Ross, J.P., 2019. DNA methylation cancer biomarkers: translation to the clinic. *Frontiers in Genetics*. 10, 1150. <https://doi.org/10.3389/fgene.2019.01150>
- López-Albors, O., Abdel, I., Periago, Ma.J., Ayala, Ma.D., Alcázar, A.G., Graciá, C.M., Nathanailides, C., Vázquez, J.Ma., 2008. Temperature influence on the white muscle growth dynamics of the sea bass *Dicentrarchus labrax*, L. Flesh quality implications at commercial size. *Aquaculture*. 277(1-2), 39–51. <https://doi.org/10.1016/j.aquaculture.2008.02.015>
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>

- Lovell-Badge, R., Robertson, E., 1990. XY female mice resulting from a heritable mutation in the primary testis-determining gene, Tdy. *Development*. 109(3), 635–646. <https://doi.org/10.1242/dev.109.3.635>
- Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T., 2017. Transcriptomics technologies. *PLoS Computational Biology*. 13(5), e1005457. <https://doi.org/10.1371/journal.pcbi.1005457>
- Loyfer, N., Magenheim, J., Peretz, A., Cann, G., Bredno, J., Klochendler, A., Fox-Fisher, I., Shabi-Porat, S., Hecht, M., Pelet, T., Moss, J., Drawshy, Z., Amini, H., Moradi, P., Nagaraju, S., Bauman, D., Shveiky, D., Porat, S., Dior, U., Rivkin, G., Or, O., Hirshoren, N., Carmon, E., Pikarsky, A., Khalaileh, A., Zamir, G., Grinbaum, R., Abu Gazala, M., Mizrahi, I., Shussman, N., Korach, A., Wald, O., Izhar, U., Erez, E., Yutkin, V., Samet, Y., Rotnemer Golinkin, D., Spalding, K.L., Druid, H., Arner, P., Shapiro, A.M.J., Grompe, M., Aravanis, A., Venn, O., Jamshidi, A., Shemer, R., Dor, Y., Glaser, B., Kaplan, T., 2023. A DNA methylation atlas of normal human cell types. *Nature*. 613(7943), 355–364. <https://doi.org/10.1038/s41586-022-05580-6>
- Lubzens, E., Young, G., Bobe, J., Cerdà, J., 2010. Oogenesis in teleosts: How fish eggs are formed. *General and Comparative Endocrinology*. 165(3), 367–389. <https://doi.org/10.1016/j.ygcen.2009.05.022>
- Lun, A.T., Smyth, G.K., 2016. csaw: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic acids research*. 44(5), e45–e45. <https://doi.org/10.1093/nar/gkv1191>
- Lund, S.P., Nettleton, D., McCarthy, D.J., Smyth, G.K., 2012. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Statistical applications in genetics and molecular biology*. 11(5). <https://doi.org/10.1515/1544-6115.1826>
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., Maleszka, R., 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biology* 8(11), e1000506. <https://doi.org/10.1371/journal.pbio.1000506>
- Lynch, M., 2007. The evolution of genetic networks by non-adaptive processes. *Nature Reviews Genetics*. 8(10), 803–813. <https://doi.org/10.1038/nrg2192>

M

- Mahmoud, A.M., Ali, M.M., 2019. Methyl donor micronutrients that modify DNA methylation and cancer outcome. *Nutrients*. 11(3), 608. <https://doi.org/10.3390/nul1030608>
- Major, K.M., DeCourten, B.M., Li, J., Britton, M., Settles, M.L., Mehinto, A.C., Connon, R.E., Brander, S.M., 2020. Early life exposure to environmentally relevant levels of endocrine disruptors drive multigenerational and transgenerational epigenetic changes in a fish model. *Frontiers in Marine Science*. 471. <https://doi.org/10.3389/fmars.2020.00471>
- Mank, J., E., Promislow, D., E.L., Avise, J., C., 2006. Evolution of alternative sex-determining mechanisms in teleost fishes. *Biological Journal of the Linnean Society*. 87(1), 83–93. <https://doi.org/10.1111/j.1095-8312.2006.00558.x>
- Marshall, A., Lukk, M., Kutter, C., Davies, S., Alexander, G., Odom, D.T., 2013. Global gene expression profiling reveals SPINK1 as a potential hepatocellular carcinoma marker. *PLoS One* 8(3), e59459. <https://doi.org/10.1371/journal.pone.0059459>
- Marsit, C.J., 2015. Influence of environmental exposure on human epigenetic regulation. *Journal of Experimental Biology*. 218(1), 71–79. <https://doi.org/10.1242/jeb.106971>
- Martienssen, R.A., Colot, V., 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science*. 293(5532), 1070–1074. <https://doi.org/10.1126/science.293.5532.1070>
- Martínez, P., Robledo, D., Taboada, X., Blanco, A., Moser, M., Maroso, F., Hermida, M., Gómez-Tato, A., Álvarez-Blázquez, B., Cabaleiro, S., Piferrer, F., Bouza, C., Lien, S., Viñas, A.M., 2021. A genome-wide association study, supported by a new chromosome-level genome assembly, suggests *sox2* as a main driver of the undifferentiated ZZ/ZW sex determination of turbot (*Scophthalmus maximus*). *Genomics*. 113(4), 1705–1718. <https://doi.org/10.1016/j.ygeno.2021.04.007>
- Martínez, P., Viñas, A.M., Sánchez, L., Díaz, N., Ribas, L., Piferrer, F., 2014. Genetic architecture of sex determination in fish: applications to sex ratio control in aquaculture. *Frontiers in Genetics*. 5, 340. <https://doi.org/10.3389/fgene.2014.00340>
- Martinez-Bengochea, A.L., Kneitz, S., Herpin, A., Nóbrega, R.H., Adolphi, M.C., Scharl, M., 2022. Sexual development dysgenesis in interspecific hybrids of Medaka fish. *Scientific Reports*. 12(1), 5408. <https://doi.org/10.1038/s41598-022-09314-6>
- Mason, M.J., Fan, G., Plath, K., Zhou, Q., Horvath, S., 2009. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics*. 10(1), 327. <https://doi.org/10.1186/1471-2164-10-327>
- Massicotte, R., Whitelaw, E., Angers, B., 2011. DNA methylation: A source of random variation in natural populations. *Epigenetics*. 6(4), 421–427. <https://doi.org/10.4161/epi.6.4.14532>
- Mateos, J., Mañanos, E., Carrillo, M., Zanuy, S., 2002. Regulation of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) gene expression by gonadotropin-releasing hormone (GnRH) and sexual steroids in

- the Mediterranean Sea bass. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*. 132(1), 75–86. [https://doi.org/10.1016/S1096-4959\(01\)00535-8](https://doi.org/10.1016/S1096-4959(01)00535-8)
- Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., Hori, H., 2002. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*. 417(6888), 559–563. <https://doi.org/10.1038/nature751>
- Matsumoto, Y., Buemio, A., Chu, R., Vafaee, M., Crews, D., 2013. Epigenetic control of gonadal aromatase (*cyp19a1*) in temperature-dependent sex determination of red-eared slider turtles. *PLoS One* 8(6), e63599. <https://doi.org/10.1371/journal.pone.0063599>
- Matsushita, S., Suzuki, K., Murashima, A., Kajioaka, D., Acebedo, A.R., Miyagawa, S., Haraguchi, R., Ogino, Y., Yamada, G., 2018. Regulation of masculinization: androgen signalling for external genitalia development. *Nature Reviews Urology*. 15(6), 358–368. <https://doi.org/10.1038/s41585-018-0008-y>
- Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V.M., Rowitch, D.H., Xing, X., Fiore, C., Schillebeekx, M., Jones, S.J.M., Haussler, D., Marra, M.A., Hirst, M., Wang, T., Costello, J.F., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*. 466(7303), 253–257. <https://doi.org/10.1038/nature09165>
- Mayer, W., Niveleau, A., Walter, J., Fundele, R., Haaf, T., 2000. Demethylation of the zygotic paternal genome. *Nature*. 403(6769), 501–502. <https://doi.org/10.1038/35000656>
- Mayne, B., Espinoza, T., Roberts, D., Butler, G.L., Brooks, S., Korbie, D., Jarman, S., 2021. Nonlethal age estimation of three threatened fish species using DNA methylation: Australian lungfish, Murray cod and Mary River cod. *Molecular Ecology Resources* 21(7), 2324–2332. <https://doi.org/10.1111/1755-0998.13440>
- Mayne, B., Korbie, D., Kenchington, L., Ezzy, B., Berry, O., Jarman, S., 2020. A DNA methylation age predictor for zebrafish. *Aging (Albany NY)*. 12(24), 24817–24835. <https://doi.org/10.18632/aging.202400>
- Mayne, B., Mustin, W., Baboolal, V., Casella, F., Ballorain, K., Barret, M., Vanderkluft, M.A., Tucker, A.D., Korbie, D., Jarman, S., Berry, O., 2022. Age prediction of green turtles with an epigenetic clock. *Molecular Ecology Resources*. 22(6), 2275–2284. <https://doi.org/10.1111/1755-0998.13621>
- McCarthy, F.M., Jones, T.E.M., Kwitek, A.E., Smith, C.L., Vize, P.D., Westerfield, M., Bruford, E.A., 2023. The case for standardizing gene nomenclature in vertebrates. *Nature*. 614(7948), E31–E32. <https://doi.org/10.1038/s41586-022-05633-w>
- Meehan, R.R., Stancheva, I., 2001. DNA methylation and control of gene expression in vertebrate development. *Essays in biochemistry*. 37, 59–70.
- Meisel, R.P., 2021. The maintenance of polygenic sex determination depends on the dominance of fitness effects which are predictive of the role of sexual antagonism. *G3*. 11(7), jkab149. <https://doi.org/10.1093/g3journal/jkab149>
- Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., Jaenisch, R., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*. 33(18), 5868–5877. <https://doi.org/10.1093/nar/gki901>
- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., Heintz, N., 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*. 151(7), 1417–1430. <https://doi.org/10.1016/j.cell.2012.11.022>
- Men, X., Ma, J., Wu, T., Pu, J., Wen, S., Shen, J., Wang, X., Wang, Y., Chen, C., Dai, P., 2017. Transcriptome profiling identified differentially expressed genes and pathways associated with tamoxifen resistance in human breast cancer. *Oncotarget*. 9(3), 4074–4089. <https://doi.org/10.18632/oncotarget.23694>
- Menke, D.B., Koubova, J., Page, D.C., 2003. Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave. *Developmental Biology*. 262(2), 303–312. [https://doi.org/10.1016/S0012-1606\(03\)00391-9](https://doi.org/10.1016/S0012-1606(03)00391-9)
- Messerschmidt, D.M., Knowles, B.B., Solter, D., 2014. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes and Development*. 28(8), 812–828. <https://doi.org/10.1101/gad.234294.113>
- Meuwissen, T.H., Hayes, B.J., Goddard, M., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157(4):1819–29. <https://doi.org/10.1093/genetics/157.4.1819>
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2021. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-13. <https://CRAN.R-project.org/package=e1071>
- Meyer, H., Milà, C., Ludwig, M., Linnenbrink, J., 2020. CAST: 'caret' Applications for Spatial-Temporal Models. R package version 0.8.1. <https://CRAN.R-project.org/package=CAST>
- Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., Thomas, P.D., 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*. 45(D1), D183–D189. <https://doi.org/10.1093/nar/gkw1138>

- Migeot, V., Hermand, D., 2018. Chromatin immunoprecipitation-polymerase chain reaction (ChIP-PCR) detects methylation, acetylation, and ubiquitylation in *S. pombe*. In: Methods in Molecular Biology. Methods and Protocols. 1721, 25–34 pp. https://doi.org/10.1007/978-1-4939-7546-4_3
- Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R., Horvath, S., 2011. Strategies for aggregating gene expression data: The collapseRows R function. BMC Bioinformatics. 12(1), 1–3. <https://doi.org/10.1186/1471-2105-12-322>
- Miller, J.A., Horvath, S., Geschwind, D.H., 2010. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proceedings of the National Academy of Sciences. 107(28), 12698–12703. <https://doi.org/10.1073/pnas.0914257107>
- Millot, S., Péan, S., Chatain, B., Bégout, M.L., 2011. Self-feeding behavior changes induced by a first and a second generation of domestication or selection for growth in the European sea bass, *Dicentrarchus labrax*. Aquatic Living Resources. 24(1), 53–61. <https://doi.org/10.1051/alr/2011103>
- Minton, K., 2023. Layering epigenomic and transcriptomic space. Nature Reviews Genetics. 24(5), 273–273. <https://doi.org/10.1038/s41576-023-00596-8>
- Miryeganeh, M., Saze, H., 2020. Epigenetic inheritance and plant evolution. Population Ecology. 62(1), 17–27. <https://doi.org/10.1002/1438-390X.12018>
- Mittwoch, U., 2006. Sex is a threshold dichotomy mimicking a single gene effect. Trends in Genetics. 22(2), 96–100. <https://doi.org/10.1016/j.tig.2005.12.003>
- Miura, T., Yamauchi, K., Takahashi, H., Nagahama, Y., 1991. Human chorionic gonadotropin induces all stages of spermatogenesis in vitro in the male Japanese eel (*Anguilla japonica*). Developmental biology. 146(1), 258–262. [https://doi.org/10.1016/0012-1606\(91\)90468-I](https://doi.org/10.1016/0012-1606(91)90468-I)
- Mizoguchi, B., Valenzuela, N., 2020. Alternative splicing and thermosensitive expression of *Dmrt1* during urogenital development in the painted turtle, *Chrysemys picta*. PeerJ. 8, e8639. <https://doi.org/10.7717/peerj.8639>
- Modhukur, V., Sharma, S., Mondal, M., Lawarde, A., Kask, K., Sharma, R., Salumets, A., 2021. Machine learning approaches to classify primary and metastatic cancers using tissue of origin-based DNA methylation profiles. Cancers. 13(15), 3768. <https://doi.org/10.3390/cancers13153768>
- Mohn, F., Weber, M., Schübeler, D., Roloff, T.C., 2009. Methylated DNA immunoprecipitation (MeDIP). DNA methylation: methods and protocols. 507, 55–64. https://doi.org/10.1007/978-1-59745-522-0_5
- Mohsen, K., Johansson, S., Ekström, T.J., 2006. Using LUMA: a luminometric-based assay for global DNA-methylation. Epigenetics. 1(1), 46–49. <https://doi.org/10.4161/epi.1.1.2587>
- Molina, M.O., Sánchez, E., Gutiérrez, C., 2020. Future heat waves over the Mediterranean from a Euro-CORDEX regional climate model ensemble. Scientific Reports. 10(1), 8801. <https://doi.org/10.1038/s41598-020-65663-0>
- Moniot, B., Declosmenil, F., Barrionuevo, F., Scherer, G., Aritake, K., Malki, S., Marzi, L., Cohen-Solal, A., Georg, I., Klattig, J., Englert, C., Kim, Y., Capel, B., Eguchi, N., Urade, Y., Boizet-Bonhoure, B., Poulat, F., 2009. The PGD2 pathway, independently of FGF9, amplifies SOX9 activity in Sertoli cells during male sexual differentiation. Development. 136(11), 1813–1821. <https://doi.org/10.1242/dev.032631>
- Moore, E.C., Roberts, R.B., 2013. Polygenic sex determination. Current Biology. 23(12), R510–R512. <https://doi.org/10.1016/j.cub.2013.04.004>
- Morretti, A., 1999. Manual on hatchery production of seabass and gilthead seabream. Food & Agriculture Org. Volume 1. <https://www.fao.org/3/x3980e/x3980e.pdf>
- Moskalev, E.A., Zavgorodnij, M.G., Majorova, S.P., Vorobjev, I.A., Jandaghi, P., Bure, I.V., Hoheisel, J.D., 2011. Correction of PCR-bias in quantitative DNA methylation studies by means of cubic polynomial regression. Nucleic Acids Research. 39(11), e77. <https://doi.org/10.1093/nar/gkr213>
- Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D.J., Salichos, L., Zhang, J., Weinstock, G.M., Isaacs, F., Rozowsky, J., Gerstein, M., 2016. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biology 17(1), 53. <https://doi.org/10.1186/s13059-016-0917-0>
- Munger, S.C., Aylor, D.L., Syed, H.A., Magwene, P.M., Threadgill, D.W., Capel, B., 2009. Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal. Genes & Development. 23(21), 2521–2536. <https://doi.org/10.1101/gad.1835809>
- Munger, S.C., Natarajan, A., Looger, L.L., Ohler, U., Capel, B., 2013. Fine time course expression analysis identifies cascades of activation and repression and maps a putative regulator of mammalian sex determination. PLoS Genetics. 9(7), e1003630. <https://doi.org/10.1371/journal.pgen.1003630>
- Murozumi, N., Nakashima, R., Hirai, T., Kamei, Y., Ishikawa-Fujiwara, T., Todo, T., Kitano, T., 2014. Loss of follicle-stimulating hormone receptor function causes masculinization and suppression of Ovarian Development in Genetically Female Medaka. Endocrinology. 155(8), 3136–3145. <https://doi.org/10.1210/en.2013-2060>
- Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamaguchi, S., Sakaizumi, M., 2012. Tracing the emergence of a novel sex-determining gene in Medaka, *Oryzias latipes*. Genetics. 191(1), 163–170. <https://doi.org/10.1534/genetics.111.137497>

N

- Nagahama, Y., 1983. 6 The functional morphology of teleost gonads, In: Fish Physiology. Academic Press. 9, 223–275 pp. [https://doi.org/10.1016/S1546-5098\(08\)60290-3](https://doi.org/10.1016/S1546-5098(08)60290-3)
- Nagahama, Y., Chakraborty, T., Paul-Prasanth, B., Ohta, K., Nakamura, M., 2021. Sex determination, gonadal sex differentiation, and plasticity in vertebrate species. *Physiological Reviews*. 101, 1237–1308. <https://doi.org/10.1152/physrev.00044.2019>
- Nakabayashi, K., 2020. Chapter Seven - The Illumina Infinium methylation assay for genome-wide methylation analyses, In: Tollefsbol, T. (Ed.), *Epigenetics Methods, Translational Epigenetics*. Academic Press, 117–140 pp. <https://doi.org/10.1016/B978-0-12-819414-0.00007-0>
- Nakamoto, M., Uchino, T., Koshimizu, E., Kuchiishi, Y., Sekiguchi, R., Wang, L., Sudo, R., Endo, M., Guiguen, Y., Scharl, M., Postlethwait, J.H., Sakamoto, T., 2021. A Y-linked anti-Müllerian hormone type-II receptor is the sex-determining gene in ayu, *Plecoglossus altivelis*. *PLoS Genetics*. 17(8), e1009705. <https://doi.org/10.1371/journal.pgen.1009705>
- Nakamura, M., 2013. Is a sex-determining gene(s) necessary for sex-determination in amphibians? steroid hormones may be the key factor. *Sexual Development*. 7(1-3), 104–114. <https://doi.org/10.1159/000339661>
- Nakato, R., Sakata, T., 2021. Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods*. 187, 44–53. <https://doi.org/10.1016/j.ymeth.2020.03.005>
- Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., Schmid, M., 2002. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proceedings of the National Academy of Sciences*. 99(18), 11778–11783. <https://doi.org/10.1073/pnas.182314699>
- Naue, J., Hoefsloot, H.C.J., Mook, O.R.F., Rijlaarsdam-Hoekstra, L., van der Zwalm, M.C.H., Henneman, P., Kloosterman, A.D., Verschure, P.J., 2017. Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression. *Forensic Science International: Genetics*. 31, 19–28. <https://doi.org/10.1016/j.fsigen.2017.07.015>
- Navarro-Martín, L., Blázquez, M., Piferrer, F., 2009a. Masculinization of the European sea bass (*Dicentrarchus labrax*) by treatment with an androgen or aromatase inhibitor involves different gene expression and has distinct lasting effects on maturation. *General and Comparative Endocrinology*. 160(1), 3–11. <https://doi.org/10.1016/j.ygcen.2008.10.012>
- Navarro-Martín, L., Blázquez, M., Viñas, J., Joly, S., Piferrer, F., 2009b. Balancing the effects of rearing at low temperature during early development on sex ratios, growth and maturation in the European sea bass (*Dicentrarchus labrax*): Limitations and opportunities for the production of highly female-biased stocks. *Aquaculture*. 296(3-4), 347–358. <https://doi.org/10.1016/j.aquaculture.2009.07.022>
- Navarro-Martín, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Croce, L.D., Piferrer, F., 2011. DNA methylation of the gonadal aromatase (*cyp19a*) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLoS Genetics*. 7(12), e1002447. <https://doi.org/10.1371/journal.pgen.1002447>
- Nelson, J.S., Grande, T.C., Wilson, M.V., 2016. *Fishes of the World*. John Wiley & Sons. Hoboken, NJ, 624 pp.
- Nguyen, N.D., Blaby, I.K., Wang, D., 2019. ManiNetCluster: a novel manifold learning approach to reveal the functional links between gene networks. *BMC Genomics*. 20(1003), 1-4. <https://doi.org/10.1186/s12864-019-6329-2>
- Nguyen, T., Chin, W.C., Verdugo, P., 1998. Role of Ca²⁺/K⁺ ion exchange in intracellular storage and release of Ca²⁺. *Nature*. 395(6705), 908–912. <https://doi.org/10.1038/27686>
- Niculescu, M.D., Zeisel, S.H., 2002. Diet, methyl donors and DNA methylation: interactions between dietary folate, methionine and choline. *The Journal of Nutrition*. 132(8), 2333S–2335S. <https://doi.org/10.1093/jn/132.8.2333S>
- Niemira, M., Collin, F., Szalkowska, A., Bielska, A., Chwialkowska, K., Reszec, J., Niklinski, J., Kwasniewski, M., Kretowski, A., 2020. Molecular signature of subtypes of non-small-cell lung cancer by large-scale transcriptional profiling: identification of key modules and genes by weighted gene co-expression network analysis (WGCNA). *Cancers*. 12, 37. <https://doi.org/10.3390/cancers12010037>
- Nivelle, R., Gennotte, V., Kalala, E.J.K., Ngoc, N.B., Muller, M., Mélard, C., Rougeot, C., 2019. Temperature preference of Nile tilapia (*Oreochromis niloticus*) juveniles induces spontaneous sex reversal. *PLoS One*. 14(2), e0212504. <https://doi.org/10.1371/journal.pone.0212504>

O

- O'Dea, R.E., Noble, D.W.A., Johnson, S.L., Hesselson, D., Nakagawa, S., 2016. The role of non-genetic inheritance in evolutionary rescue: epigenetic buffering, heritable bet hedging and epigenetic traps. *Environmental Epigenetics*. 2(1), dvv014. <https://doi.org/10.1093/eeep/dvv014>

- Oike, A., Kodama, M., Yasumasu, S., Yamamoto, T., Nakamura, Y., Ito, E., Nakamura, M., 2017. Participation of androgen and its receptor in sex determination of an amphibian species. *PLoS One*. 12(6), e0178067. <https://doi.org/10.1371/journal.pone.0178067>
- Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., Geschwind, D.H., 2008. Functional organization of the transcriptome in human brain. *Nature Neuroscience*. 11(11), 1271–1282. <https://doi.org/10.1038/nn.2207>
- Ortega-Recalde, O., Hore, T.A., 2019. DNA methylation in the vertebrate germline: balancing memory and erasure. *Essays in Biochemistry*. 63(6), 649–661. <https://doi.org/10.1042/EBC20190038>
- Ortega-Recalde, O., Peat, J.R., Bond, D.M., Hore, T.A., 2021. Estimating global methylation and erasure using low-coverage whole-genome bisulfite sequencing (WGBS). In: *TET Proteins and DNA Demethylation: Methods and Protocols*. New York, Springer US. 272, 29–44 pp. https://doi.org/10.1007/978-1-0716-1294-1_3
- Ospina-Alvarez, N., Piferrer, F., 2008. Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PloS One*. 3(7), e2837. <https://doi.org/10.1371/journal.pone.0002837>
- Ottolenghi, C., Omari, S., Garcia-Ortiz, J.E., Uda, M., Crisponi, L., Forabosco, A., Pilia, G., Schlessinger, D., 2005. *Foxl2* is required for commitment to ovary differentiation. *Human Molecular Genetics*. 14(14), 2053–2062. <https://doi.org/10.1093/hmg/ddi210>
- Ovens, K., Eames, B.F., McQuillan, I., 2021a. Comparative analyses of gene co-expression networks: Implementations and applications in the study of evolution. *Frontiers in Genetics*. 12, 695399. <https://doi.org/10.3389/fgene.2021.695399>
- Ovens, K., Maleki, F., Eames, B.F., McQuillan, I., 2021b. Juxtapose: a gene-embedding approach for comparing co-expression networks. *BMC Bioinformatics* 22, 1–26. <https://doi.org/10.1186/s12859-021-04055-1>

P

- Pan, Q., Anderson, J., Bertho, S., Herpin, A., Wilson, C., Postlethwait, J.H., Scharl, M., Guiguen, Y., 2016. Vertebrate sex-determining genes play musical chairs. *Comptes Rendus Biologies, Trajectories of genetics, 150 years after Mendel / Trajectoire de la génétique, 150 après Mendel* Guest Editors / Rédacteurs en chef invités: Bernard Dujon, Georges Pelletier 339, 258–262. <https://doi.org/10.1016/j.crv.2016.05.010>
- Pan, Q., Feron, R., Jouanno, E., Darras, H., Herpin, A., Koop, B., Rondeau, E., Goetz, F.W., Larson, W.A., Bernatchez, L., Tringali, M., Curran, S.S., Saillant, E., Denys, G.P., von Hippel, F.A., Chen, S., López, J.A., Verreycken, H., Ocalewicz, K., Guyomard, R., Eche, C., Lluch, J., Roques, C., Hu, H., Tabor, R., DeHaan, P., Nichols, K.M., Journot, L., Parrinello, H., Klopp, C., Interesova, E.A., Trifonov, V., Scharl, M., Postlethwait, J., Guiguen, Y., 2021a. The rise and fall of the ancient northern pike master sex-determining gene. *eLife* 10, e62858. <https://doi.org/10.7554/eLife.62858>
- Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., Wen, M., Busnel, J.M., Bobe, J., Concordet, J.P., Parrinello, H., 2019. Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLoS Genetics*. 15(8), e1008013. <https://doi.org/10.1371/journal.pgen.1008013>
- Pan, Q., Kay, T., Depincé, A., Adolphi, M., Scharl, M., Guiguen, Y., Herpin, A., 2021b. Evolution of master sex determiners: TGF- β signalling pathways at regulatory crossroads. *Philosophical Transactions of the Royal Society B*. 376(1832), 20200091. <https://doi.org/10.1098/rstb.2020.0091>
- Pan, S., Zhan, Y., Chen, X., Wu, B., Liu, B., 2019. Identification of biomarkers for controlling cancer stem cell characteristics in bladder cancer by network analysis of transcriptome data stemness indices. *Frontiers in Oncology*. 9, 613. <https://doi.org/10.3389/fonc.2019.00613>
- Papandroulakis, N., Divanach, P., Anastasiadis, P., Kentouri, M., 2001. The pseudo-green water technique for intensive rearing of sea bream (*Sparus aurata*) larvae. *Aquaculture International*. 9, 205–216. <https://doi.org/10.1023/A:1016813623122>
- Park, J., Lee, K., Kim, K., Yi, S.J., 2022. The role of histone modifications: From neurodevelopment to neurodegenerative diseases. *Signal Transduction and Targeted Therapy*. 7(1), 217. <https://doi.org/10.1038/s41392-022-01078-9>
- Parmigiani, G., Garrett, E.S., Irizarry, R.A., Zeger, S.L., 2003. The analysis of gene expression data: An overview of methods and software, In: Parmigiani, G., Garrett, E.S., Irizarry, R.A., Zeger, S.L. (Eds.), *The analysis of gene expression data: methods and software, statistics for biology and health*. Springer, New York, NY, 1–45 pp. https://doi.org/10.1007/0-387-21679-0_1
- Parrott, B.B., Kohno, S., Cloy-McCoy, J.A., Guillelte Jr, L.J., 2014. Differential incubation temperatures result in dimorphic DNA methylation patterning of the *SOX9* and aromatase promoters in gonads of alligator (*Alligator mississippiensis*) embryos. *Biology of reproduction*. 90(1), 2–1. <https://doi.org/10.1095/biolreprod.113.111468>

- Pauly, D., 2021. Why do fish reach first maturity when they do? *Journal of Fish Biology*. 1-9. <https://doi.org/10.1111/jfb.14902>
- Pavlidis, M.A., Mylonas, C.C., (eds.) 2011. *Sparidae: Biology and aquaculture of gilthead sea bream and other species*. John Wiley & Sons. Chichester, XVII+ 390 pp, ISBN: 978-1-4051-9772-4.
- Pen, I., Uller, T., Feldmeyer, B., Harts, A., While, G.M., Wapstra, E., 2010. Climate-driven population divergence in sex-determining systems. *Nature*. 468(7322), 436–438. <https://doi.org/10.1038/nature09512>
- Penman, D.J., Piferrer, F., 2008. Fish gonadogenesis. Part I: genetic and environmental mechanisms of sex determination. *Reviews in Fisheries Science*. 16(1), 16–34. <https://doi.org/10.1080/10641260802324610>
- Pepin, P., 1991. Effect of temperature and size on development, mortality, and survival rates of the pelagic early life history stages of marine fish. *Canadian Journal of Fisheries and Aquatic Sciences*. 48, 503–518. <https://doi.org/10.1139/f91-065>
- Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M.T., Baker, M., Browne, W.J., Clark, A., Cuthill, I.C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S.T., Howells, D.W., Karp, N.A., Lazic, S.E., Lidster, K., MacCallum, C.J., Macleod, M., Pearl, E.J., Petersen, O.H., Rawle, F., Reynolds, P., Rooney, K., Sena, E.S., Silberberg, S.D., Steckler, T., Würbel, H., 2020. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *PLoS Biology*. 18, e3000410. <https://doi.org/10.1371/journal.pbio.3000410>
- Perez, M.F., Lehner, B., 2019. Intergenerational and transgenerational epigenetic inheritance in animals. *Nature Cell Biology*. 21(2), 143–151. <https://doi.org/10.1038/s41556-018-0242-9>
- Pérez-Lluch, S., Blanco, E., Tilgner, H., Curado, J., Ruiz-Romero, M., Corominas, M., Guigó, R., 2015. Absence of canonical marks of active chromatin in developmentally regulated genes. *Nature Genetics*. 47(10), 1158–1167. <https://doi.org/10.1038/ng.3381>
- Perrella, G., Zioutopoulou, A., Headland, L.R., Kaiserli, E., 2020. The impact of light and temperature on chromatin organization and plant adaptation. *Journal of Experimental Botany*. 71(17), 5247–5255. <https://doi.org/10.1093/jxb/eraa154>
- Perrin, N., 2021. Sex-chromosome evolution in frogs: What role for sex-antagonistic genes? *Philosophical Transactions of the Royal Society B*. 376(1832):20200094. <https://doi.org/10.1098/rstb.2020.0094>
- Perrin, N., 2016. Random sex determination: When developmental noise tips the sex balance. *BioEssays*. 38(12), 1218–1226. <https://doi.org/10.1002/bies.201600093>
- Pértille, F., Brantsæter, M., Nordgreen, J., Coutinho, L.L., Janczak, A.M., Jensen, P., Guerrero-Bosagna, C., 2017. DNA methylation profiles in red blood cells of adult hens correlate with their rearing conditions. *Journal of Experimental Biology*. 220, 3579–3587. <https://doi.org/10.1242/jeb.157891>
- Pettersson, M.E., Rochus, C.M., Han, F., Chen, J., Hill, J., Wallerman, O., Fan, G., Hong, X., Xu, Q., Zhang, H., Liu, S., Liu, X., Haggerty, L., Hunt, T., Martin, F.J., Flicek, P., Bunikis, I., Folkvord, A., Andersson, L., 2019. A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. *Genome Research*. 29(11), 1919–1928. <https://doi.org/10.1101/gr.253435.119>
- Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S., Smyth, G.K., 2016. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *The Annals of Applied Statistics*. 10(2), 946–963. <https://doi.org/10.1214/16-AOAS920>
- Pickett, G.D., Pawson, M.G., 1994. *Sea bass: Biology*. Springer Science & Business Media. Fish & Fisheries Series 12, 342. ISBN: 9780412400902.
- Pierron, F., Lorient, S., Hérouin, D., Daffe, G., Etcheverria, B., Cachot, J., Morin, B., Dufour, S., Gonzalez, P., 2021. Transgenerational epigenetic sex determination: Environment experienced by female fish affects offspring sex ratio. *Environmental Pollution*. 277, 116864. <https://doi.org/10.1016/j.envpol.2021.116864>
- Piétu, G., Mariage-Samson, R., Fayein, N.A., Matingou, C., Eveno, E., Houlgatte, R., Decraene, C., Vandenbrouck, Y., Tahi, F., Devignes, M.-D., Wirkner, U., Ansorge, W., Cox, D., Nagase, T., Nomura, N., Auffray, C., 1999. The genexpress IMAGE knowledge base of the human brain transcriptome: a prototype integrated resource for functional and computational genomics. *Genome Research*. 9(2), 195–209. <https://doi.org/10.1101/gr.9.2.195>
- Piferrer, F., 2023. Epigenetics in aquaculture: knowledge gaps, challenges and future prospects. In: Piferrer F., Wang, H.P., eds. 2023. *Epigenetics in Aquaculture*. John Wiley and Sons, Hoboken, New Jersey. 450 pp. ISBN: 9781119821915. <https://doi.org/10.1002/9781119821946.ch20>
- Piferrer, F., 2018a. Epigenetics of sex determination and differentiation in fish. In: Wang, H.P., Piferrer F., Chen, S.L., Shen, Z.G., eds. 2018. *Sex control in aquaculture*. Wiley, New York. 1, 65–83 pp. ISBN: 9781119127260. <https://doi.org/10.1002/9781119127291.ch3>
- Piferrer, F., 2018b. La determinació del sexe. *L'Atzavara*. 28, 39-60. <https://raco.cat/index.php/Atzavara/article/view/332608>
- Piferrer, F., 2016. Altered sex ratios in response to climate change—Who will fall into the (epigenetic) trap? (Comment on DOI 10.1002/bies.201600058). *BioEssays*. 38(10), 939–939. <https://doi.org/10.1002/bies.201600172>
- Piferrer, F., 2013. Epigenetics of sex determination and gonadogenesis. *Developmental Dynamics*. 242(4), 360–370. <https://doi.org/10.1002/dvdy.23924>

- Piferrer, F., Anastasiadi, D., 2023. Age estimation in fishes using epigenetic clocks: Applications to fisheries management and conservation biology. *Frontiers in Marine Science* 10, 1062151. <https://doi.org/10.3389/fmars.2023.1062151>
- Piferrer, F., Anastasiadi, D., 2021. Do the offspring of sex reversals have higher sensitivity to environmental perturbations? *Sexual Development*. 15(1-3), 134–147. <https://doi.org/10.1159/000515192>
- Piferrer, F., Anastasiadi, D., Díaz, N., Ribas, L., 2015. Integrating epigenetics into aquaculture research. *International Symposium on Genetics in Aquaculture XII*. <http://hdl.handle.net/10261/141387>
- Piferrer, F., Anastasiadi, D., Valdivieso, A., Sánchez-Baizán, N., Moraleda-Prados, J., Ribas, L., 2019. The model of the conserved epigenetic regulation of sex. *Frontiers in Genetics*. 10, 857. <https://doi.org/10.3389/fgene.2019.00857>
- Piferrer, F., Blázquez, M., Navarro, L., González, A., 2005. Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (*Dicentrarchus labrax* L.). *General and Comparative Endocrinology*. 142(1-2), 102–110. <https://doi.org/10.1016/j.ygcen.2005.02.011>
- Piferrer, F., Guiguen, Y., 2008. Fish gonadogenesis. Part II: molecular biology and genomics of sex differentiation. *Reviews in Fisheries Science*. 16(1), 35–55. <https://doi.org/10.1080/10641260802324644>
- Piferrer, F., Zanuy, S., Carrillo, M., Solar, I.I., Devlin, R.H., Donaldson, E.M., 1994. Brief treatment with an aromatase inhibitor during sex differentiation causes chromosomally female salmon to develop as normal, functional males. *Journal of Experimental Zoology*. 270(3), 255–262. <https://doi.org/10.1002/jez.1402700304>
- Piprek, R.P., Damulewicz, M., Kloc, M., Kubiak, J.Z., 2018. Transcriptome analysis identifies genes involved in sex determination and development of *Xenopus laevis* gonads. *Differentiation*. 100, 46–56. <https://doi.org/10.1016/j.diff.2018.02.004>
- Piprek, R.P., Kloc, M., Tassan, J.P., Kubiak, J.Z., 2017. Development of *Xenopus laevis* bipotential gonads into testis or ovary is driven by sex-specific cell-cell interactions, proliferation rate, cell migration and deposition of extracellular matrix. *Developmental Biology*. 432(2), 298–310. <https://doi.org/10.1016/j.ydbio.2017.10.020>
- Pizzorusso, T., Tognini, P., 2020. Interplay between metabolism, nutrition and epigenetics in shaping brain DNA methylation, neural function and behavior. *Genes*. 11(7), 742. <https://doi.org/10.3390/genes11070742>
- Pla, S., 2019. Evolutionary transitions, environmental correlates and life-history traits associated with the distribution of the different forms of hermaphroditism in fish. *Universitat Autònoma de Barcelona*.
- Pla, S., Benvenuto, C., Capellini, I., Piferrer, F., 2022. Switches, stability and reversals in the evolutionary history of sexual systems in fish. *Nature Communications*. 13(1), 3029. <https://doi.org/10.1038/s41467-022-30419-z>
- Plongthongkum, N., Eijk, K.R. van, Jong, S. de, Wang, T., Sul, J.H., Boks, M.P.M., Kahn, R.S., Fung, H.L., Ophoff, R.A., Zhang, K., 2014. Characterization of genome-methylome interactions in 22 nuclear pedigrees. *PLoS One*. 9, e99313. <https://doi.org/10.1371/journal.pone.0099313>
- Podgorniak, T., Brockmann, S., Konstantinidis, I., Fernandes, J.M.O., 2019. Differences in the fast muscle methylome provide insight into sex-specific epigenetic regulation of growth in Nile tilapia during early stages of domestication. *Epigenetics*. 14(8), 818–836. <https://doi.org/10.1080/15592294.2019.1618164>
- Polanowski, A.M., Robbins, J., Chandler, D., Jarman, S.N., 2014. Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources*. 14(5), 976–987. <https://doi.org/10.1111/1755-0998.12247>
- Potok, M.E., Nix, D.A., Parnell, T.J., Cairns, B.R., 2013. Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell*. 153(4), 759–772. <https://doi.org/10.1016/j.cell.2013.04.030>
- Poulat, F., 2021. Non-coding genome, transcription factors, and sex determination. *Sexual Development*. 15(5-6), 295–307. <https://doi.org/10.1159/000519725>
- Prado, N.A., Brown, J.L., Zoller, J.A., Haghani, A., Yao, M., Bagryanova, L.R., Campana, M.G., E. Maldonado, J., Raj, K., Schmitt, D., Robeck, T.R., Horvath, S., 2021. Epigenetic clock and methylation studies in elephants. *Aging Cell*. 20(7), e13414. <https://doi.org/10.1111/acel.13414>
- Probst, A.V., Dunleavy, E., Almouzni, G., 2009. Epigenetic inheritance during the cell cycle. *Nature Reviews Molecular Cell Biology*. 10(3), 192–206. <https://doi.org/10.1038/nrm2640>

Q

- Qiu, W., Guo, F., Glass, K., Yuan, G.C., Quackenbush, J., Zhou, X., Tantisira, K.G., 2018. Differential connectivity of gene regulatory networks distinguishes corticosteroid response in asthma. *Journal of Allergy and Clinical Immunology*. 141(4), 1250–1258. <https://doi.org/10.1016/j.jaci.2017.05.052>

R

- R Core Team, 2022. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.

- R Core Team, 2021. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.
- Rahmoun, M., Lavery, R., Laurent-Chaballier, S., Bellora, N., Philip, G.K., Rossitto, M., Symon, A., Pailhoux, E., Cammas, F., Chung, J., Bagheri-Fam, S., Murphy, M., Bardwell, V., Zarkower, D., Boizet-Bonhoure, B., Clair, P., Harley, V.R., Poulat, F., 2017. In mammalian foetal testes, SOX9 regulates expression of its target genes by binding to genomic regions with conserved signatures. *Nucleic Acids Research*. 45(12), 7191–7211. <https://doi.org/10.1093/nar/gkx328>
- Ramakrishnan, N., Pillai, S.R.B., Padinhateeri, R., 2022. High fidelity epigenetic inheritance: Information theoretic model predicts threshold filling of histone modifications post replication. *PLoS Computational Biology*. 18(2), e1009861. <https://doi.org/10.1371/journal.pcbi.1009861>
- Ramalho-Carvalho, J., Henrique, R., Jerónimo, C., 2018. Methylation-Specific PCR. *Methods Molecular Biology, DNA Methylation Protocols*. 1708, 447–472. https://doi.org/10.1007/978-1-4939-7481-8_23
- Rang, F.J., Boonstra, J., 2014. Causes and consequences of age-related changes in DNA methylation: A role for ROS? *Biology*. 3(2), 403–425. <https://doi.org/10.3390/biology3020403>
- Rao, M.S., Van Vleet, T.R., Ciurlionis, R., Buck, W.R., Mittelstadt, S.W., Blomme, E.A.G., Liguori, M.J., 2019. Comparison of RNA-seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies. *Frontiers in Genetics*. 9, 636. <https://doi.org/10.3389/fgene.2018.00636>
- Ratel, D., Ravanat, J.-L., Berger, F., Wion, D., 2006. N6-methyladenine: the other methylated base of DNA. *Bioessays*. 28(3), 309–315. <https://doi.org/10.1002/bies.20342>
- Rauschert, S., Raubenheimer, K., Melton, P.E., Huang, R.C., 2020. Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. *Clinical Epigenetics*. 12(1), 51. <https://doi.org/10.1186/s13148-020-00842-4>
- Ravi, V., Venkatesh, B., 2018. The divergent genomes of teleosts. *Annual review of animal biosciences*. 6, 47–68. <https://doi.org/10.1146/annurev-animal-030117-014821>
- Raymond, C.S., Shamu, C.E., Shen, M.M., Seifert, K.J., Hirsch, B., Hodgkin, J., Zarkower, D., 1998. Evidence for evolutionary conservation of sex-determining genes. *Nature*. 391(6668), 691–695. <https://doi.org/10.1038/35618>
- Recalde, M., Gárate-Rascón, M., Herranz, J.M., Elizalde, M., Azkona, M., Unfried, J.P., Boix, L., Reig, M., Sangro, B., Fernández-Barrena, M.G., Fortes, P., Ávila, M.A., Berasain, C., Arechederra, M., 2022. DNA methylation regulates a set of long non-coding RNAs compromising hepatic identity during hepatocarcinogenesis. *Cancers*. 14(9), 2048. <https://doi.org/10.3390/cancers14092048>
- Reichwald, K., Petzold, A., Koch, P., Downie, B.R., Hartmann, N., Pietsch, S., Baumgart, M., Chalopin, D., Felder, M., Bens, M., Sahm, A., 2015. Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell*. 163(6), 1527–1538. <https://doi.org/10.1016/j.cell.2015.10.071>
- Reik, W., Santos, F., Mitsuya, K., Morgan, H., Dean, W., 2003. Epigenetic asymmetry in the mammalian zygote and early embryo: relationship to lineage commitment? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. 358(1436), 1403–1409. <https://doi.org/10.1098/rstb.2003.1326>
- Reis, B.Y., Butte, A.S., Kohane, I.S., 2001. Extracting knowledge from dynamics in gene expression. *Journal of Biomedical Informatics*. 34(1), 15–27. <https://doi.org/10.1006/jbin.2001.1005>
- Rey, S., Boltana, S., Vargas, R., Roher, N., MacKenzie, S., 2013. Combining animal personalities with transcriptomics resolves individual variation within a wild-type zebrafish population and identifies underpinning molecular differences in brain function. *Molecular Ecology*. 22(24), 6100–6115. <https://doi.org/10.1111/mec.12556>
- Rhen, T., Schroeder, A., 2010. Molecular mechanisms of sex determination in reptiles. *Sexual Development*. 4(1–2), 16–28. <https://doi.org/10.1159/000282495>
- Ribas, L., Crespo, B., Sánchez-Baizán, N., Xavier, D., Kuhl, H., Rodríguez, J.M., Díaz, N., Boltaña, S., MacKenzie, S., Morán, F., Zanuy, S., Gómez, A., Piferrer, F., 2019. Characterization of the European Sea Bass (*Dicentrarchus labrax*) Gonadal Transcriptome During Sexual Development. *Marine Biotechnology*. (NY) 21, 359–373. <https://doi.org/10.1007/s10126-019-09886-x>
- Ribas, L., Valdivieso, A., Díaz, N., Piferrer, F. 2017. Appropriate rearing density in domesticated zebrafish to avoid masculinization: links with the stress response. *Journal of Experimental Biology*. 220(6):1056–64. <https://doi.org/10.1242/jeb.144980>
- Ribas, L., Liew, W.C., Díaz, N., Sreenivasan, R., Orbán, L., Piferrer, F., 2017b. Heat-induced masculinization in domesticated zebrafish is family-specific and yields a set of different gonadal transcriptomes. *Proceedings of the National Academy of Sciences*. 114(6), E941–E950. <https://doi.org/10.1073/pnas.1609411114>
- Ribas, L., Robledo, D., Gómez-Tato, A., Viñas, A., Martínez, P., Piferrer, F., 2016. Comprehensive transcriptomic analysis of the process of gonadal sex differentiation in the turbot (*Scophthalmus maximus*). *Molecular and Cellular Endocrinology*. 422, 132–149. <https://doi.org/10.1016/j.mce.2015.11.006>
- Richards, E.J., 2008. Population epigenetics. *Current Opinion in Genetics & Development*, Chromosomes and expression mechanisms. 18(2), 221–226. <https://doi.org/10.1016/j.gde.2008.01.014>

- Richards, E.J., 2006. Inherited epigenetic variation — revisiting soft inheritance. *Nature Reviews Genetics*. 7(5), 395–401. <https://doi.org/10.1038/nrg1834>
- Risha, M.A., Ali, A., Siengdee, P., Trakooljul, N., Haack, F., Dannenberger, D., Wimmers, K., Ponsuksili, S., 2021. Wnt signaling related transcripts and their relationship to energy metabolism in C2C12 myoblasts under temperature stress. *PeerJ*. 9, e11625. <https://doi.org/10.7717/peerj.11625>
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 43(7), e47. <https://doi.org/10.1093/nar/gkv007>
- Roberts, N.B., Juntti, S.A., Coyle, K.P., Dumont, B.L., Stanley, M.K., Ryan, A.Q., Fernald, R.D., Roberts, R.B., 2016. Polygenic sex determination in the cichlid fish the *Astatotilapia burtoni*. *BMC Genomics*. 17(1), 835. <https://doi.org/10.1186/s12864-016-3177-1>
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Roblin, C., Bruslé, J., 1983. Ontogenèse gonadique et différenciation sexuelle du Loup *Dicentrarchus labrax*, en conditions d'élevage. *Reproduction Nutrition Développement*. 23(1), 115–127. <https://doi.org/10.1051/rnd:19830110>
- Rocha, A., Zanuy, S., Gómez, A., 2016. Conserved anti-müllerian hormone: Anti-müllerian hormone type-2 receptor specific interaction and intracellular signaling in teleosts. *Biology of Reproduction*. 94(69), 141. <https://doi.org/10.1095/biolreprod.115.137547>
- Rodriguez Barreto, D., de Leaniz, C.G., Verspoor, E., Sobolewska, H., Coulson, M., Consuegra, S., 2019. DNA methylation changes in the sperm of captive-reared fish: a route to epigenetic introgression in wild populations. *Molecular Biology and Evolution*. 36(10):2205–11. <https://doi.org/10.1093/molbev/msz135>
- Rodríguez, L., Halm, S., Mylonas, C., Zanuy, S., 2003. Gene expression of estrogen receptors α and β during early sexual differentiation in the European sea bass (*Dicentrarchus labrax*). *Fish Physiology and Biochemistry*. 28, 175–175. <https://doi.org/10.1023/B:FISH.0000030518.51999.5d>
- Rohland, N., Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*. 22(5), 939–946. <https://doi.org/10.1101/gr.128124.111>
- Rojo-Bartolomé, I., Valencia, A., Cancio, I., 2017. Transcription of ribogenesis genes in fish gonads: Applications in the identification of stages of oogenesis and in environmental monitoring of intersex condition. *Marine Pollution Bulletin*. 121(1-2), 292–301. <https://doi.org/10.1016/j.marpolbul.2017.06.027>
- Rönn, T., Volkov, P., Gillberg, L., Kokosar, M., Perfilyev, A., Jacobsen, A.L., Jørgensen, S.W., Brøns, C., Jansson, P.A., Eriksson, K.-F., Pedersen, O., Hansen, T., Groop, L., Stener-Victorin, E., Vaag, A., Nilsson, E., Ling, C., 2015. Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. *Human Molecular Genetics*. 24(13), 3792–3813. <https://doi.org/10.1093/hmg/ddv124>
- Roux, J., Rosikiewicz, M., Robinson-Rechavi, M., 2015. What to compare and how: Comparative transcriptomics for evo-devo. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*. 324(4), 372–382. <https://doi.org/10.1002/jez.b.22618>
- RStudio Team, 2021. RStudio: Integrated development environment for R. RStudio, PBC, Boston, MA.
- Ruan, Y., Lin, Y.F., Feng, Y.C.A., Chen, C.Y., Lam, M., Guo, Z., He, L., Sawa, A., Martin, A.R., Qin, S., Huang, H., Ge, T., 2022. Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*. 54(5), 573–580. <https://doi.org/10.1038/s41588-022-01054-7>
- Ruiz-García, A., Roco, Á.S., Bullejos, M., 2021. Sex differentiation in amphibians: Effect of temperature and its influence on sex reversal. *Sexual Development*. 15(1-3), 157–167. <https://doi.org/10.1159/000515220>
- Russo, V.E., Martienssen, R.A., Riggs, A.D., 1996. Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory Press.
- Ryu, T., Veilleux, H.D., Munday, P.L., Jung, I., Donelson, J.M., Ravasi, T., 2020. An epigenetic signature for within-generational plasticity of a reef fish to ocean warming. *Frontiers in Marine Science*. 7, 284. <https://doi.org/10.3389/fmars.2020.00284>

S

- Sadoul, B., Alfonso, S., Goold, C., Pratlong, M., Rialle, S., Geffroy, B., Bégout, M.L., 2022. Transcriptomic profiles of consistent risk-taking behaviour across time and contexts in European sea bass. *Proceedings of the Royal Society B*. 289(1975), 20220399. <https://doi.org/10.1098/rspb.2022.0399>
- Saillant, E., Chatain, B., Fostier, A., Przybyla, C., Fauvel, C., 2001. Parental influence on early development in the European sea bass. *Journal of Fish Biology*. 58(6), 1585–1600. <https://doi.org/10.1111/j.1095-8649.2001.tb02314.x>

- Saillant, E., Fostier, A., Haffray, P., Menu, B., Chatain, B., 2003a. Saline preferendum for the European sea bass, *Dicentrarchus labrax*, larvae and juveniles: effect of salinity on early development and sex determination. *Journal of Experimental Marine Biology and Ecology*. 287(1), 103–117. [https://doi.org/10.1016/S0022-0981\(02\)00502-6](https://doi.org/10.1016/S0022-0981(02)00502-6)
- Saillant, E., Fostier, A., Haffray, P., Menu, B., Laureau, S., Thimonier, J., Chatain, B., 2003b. Effects of rearing density, size grading and parental factors on sex ratios of the sea bass (*Dicentrarchus labrax* L.) in intensive aquaculture. *Aquaculture*. 221(1-4), 183–206. [https://doi.org/10.1016/S0044-8486\(02\)00539-2](https://doi.org/10.1016/S0044-8486(02)00539-2)
- Saillant, Eric, Fostier, A., Menu, B., Haffray, P., Chatain, B., 2001. Sexual growth dimorphism in sea bass *Dicentrarchus labrax*. *Aquaculture*. 202, 371–387. [https://doi.org/10.1016/S0044-8486\(01\)00786-4](https://doi.org/10.1016/S0044-8486(01)00786-4)
- Sammouth, S., d'Orbcastel, E.R., Gasset, E., Lemarié, G., Breuil, G., Marino, G., Coeurdacier, J.L., Fivelstad, S., Blancheton, J.P., 2009. The effect of density on sea bass (*Dicentrarchus labrax*) performance in a tank-based recirculating system. *Aquacultural Engineering*. 40(2), 72–78. <https://doi.org/10.1016/j.aquaeng.2008.11.004>
- Samudio-Ruiz, S.L., Hudson, L.G., 2012. Increased DNA methyltransferase activity and DNA methylation following epidermal growth factor stimulation in ovarian cancer cells. *Epigenetics*. 7(3), 216–224. <https://doi.org/10.4161/epi.7.3.19273>
- Sánchez-Baizán, N., Ribas, L., Piferrer, F., 2022. Improved biomarker discovery through a plot twist in transcriptomic data analysis. *BMC Biology*. 20(1), 208. <https://doi.org/10.1186/s12915-022-01398-w>
- Santos, F., Dean, W., 2004. Epigenetic reprogramming during early development in mammals. *Reproduction* 127(6), 643–651. <https://doi.org/10.1530/rep.1.00221>
- Santos, F., Hendrich, B., Reik, W., Dean, W., 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Developmental Biology*. 241(1), 172–182. <https://doi.org/10.1006/dbio.2001.0501>
- Sarre, S.D., Ezaz, T., Georges, A., 2011. Transitions between sex-determining systems in reptiles and amphibians. *Annual review of genomics and human genetics* 12, 391–406. <https://doi.org/10.1146/annurev-genom-082410-101518>
- Sarre, S.D., Georges, A., Quinn, A., 2004. The ends of a continuum: genetic and temperature-dependent sex determination in reptiles. *Bioessays*. 26(6), 639–645. <https://doi.org/10.1002/bies.20050>
- Schartl, M., Georges, A., Marshall Graves, J.A., 2023. Polygenic sex determination in vertebrates – is there any such thing? *Trends in Genetics*. 39(4), 242–250. <https://doi.org/10.1016/j.tig.2022.12.002>
- Schlicker, A., Domingues, F.S., Rahnenführer, J., Lengauer, T., 2006. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*. 7, 218. <https://doi.org/10.1186/1471-2105-7-302>
- Schulte, P.M., Healy, T.M., Fangué, N.A., 2011. Thermal performance curves, phenotypic plasticity, and the time scales of temperature exposure. *Integrative and Comparative Biology*. 51(5), 691–702. <https://doi.org/10.1093/icb/ucr097>
- Schumacher, A., Kapranov, P., Kaminsky, Z., Flanagan, J., Assadzadeh, A., Yau, P., Virtanen, C., Winegarden, N., Cheng, J., Gingeras, T., Petronis, A., 2006. Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Research*. 34(2), 528–542. <https://doi.org/10.1093/nar/gkj461>
- Scott, M., De Sario, A., 2020. DNA methylation changes in cystic fibrosis: Cause or consequence? *Clinical Genetics* 98, 3–9. <https://doi.org/10.1111/cge.13731>
- Sentner, C.P., Vos, Y.J., Niezen-Koning, K.N., Mol, B., Smit, G.P.A., 2012. Mutation analysis in glycogen storage disease Type III patients in the Netherlands: Novel genotype-phenotype relationships and five novel mutations in the AGL gene. In: Brown, G., Morava, E., Peters, V., Gibson, K.M., Zschocke, J. (Eds.), *JIMD Reports - Case and Research Reports, 2012/4*, JIMD Reports. Springer Berlin Heidelberg, Berlin, Heidelberg, 19–26 pp. https://doi.org/10.1007/8904_2012_134
- Seo, H.W., Rengaraj, D., Choi, J.W., Ahn, S.E., Song, Y.S., Song, G., Han, J.Y., 2010. Claudin 10 is a glandular epithelial marker in the chicken model as human epithelial ovarian cancer. *International Journal of Gynecologic Cancer*. 20, 9. <https://doi.org/10.1111/IGC.0b013e3181fb18e4>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*. 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Shao, C., Li, Q., Chen, S., Zhang, P., Lian, J., Hu, Q., Sun, B., Jin, L., Liu, S., Wang, Z., Zhao, H., Jin, Z., Liang, Z., Li, Y., Zheng, Q., Zhang, Y., Wang, J., Zhang, G., 2014. Epigenetic modification and inheritance in sexual reversal of fish. *Genome Research*. 24(4), 604–615. <https://doi.org/10.1101/gr.162172.113>
- Shen, Z.G., Yao, H., Guo, L., Li, X.X., Wang, H.P., 2017. Ribosome RNA profiling to quantify ovarian development and identify sex in fish. *Scientific Report*. 7(1), 4196. <https://doi.org/10.1038/s41598-017-04327-y>
- Shi, D.Q., Ali, I., Tang, J., Yang, W.C., 2017. New insights into 5hmC DNA modification: Generation, distribution and function. *Frontiers in Genetics*. 8, 100. <https://doi.org/10.3389/fgene.2017.00100>
- Sidoli, S., Bhanu, N.V., Karch, K.R., Wang, X., Garcia, B.A., 2016. Complete workflow for analysis of histone post-translational modifications using bottom-up mass spectrometry: From histone extraction to data analysis. *JoVE (Journal of Visualized Experiments)* 111, e54112. <https://doi.org/10.3791/54112>

- Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A. M., Lovell-Badge, R., Goodfellow, P.N., 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346(6281), 240–244. <https://doi.org/10.1038/346240a0>
- Skene, P.J., Henikoff, S., 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* 6, e21856. <https://doi.org/10.7554/eLife.21856>
- Slatkin, M., 2009. Epigenetic inheritance and the missing heritability problem. *Genetics*. 182(3), 845–850. <https://doi.org/10.1534/genetics.109.102798>
- Smith, C.A., Roeszler, K.N., Ohnesorg, T., Cummins, D.M., Farlie, P.G., Doran, T.J., Sinclair, A.H., 2009. The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature*. 461(7261), 267–271. <https://doi.org/10.1038/nature08298>
- Socorro, S., Martins, R.S., Deloffre, L., Mylonas, C.C., Canario, A.V.M., 2007. A cDNA for European sea bass (*Dicentrarchus labrax*) 11 beta-hydroxylase: gene expression during the thermosensitive period and gonadogenesis. *General and comparative endocrinology*. 150(1), 164–173. <https://doi.org/10.1016/j.ygcen.2006.07.018>
- Song, C.X., Yin, S., Ma, L., Wheeler, A., Chen, Y., Zhang, Y., Liu, B., Xiong, J., Zhang, W., Hu, J., Zhou, Z., Dong, B., Tian, Z., Jeffrey, S.S., Chua, M.S., So, S., Li, W., Wei, Y., Diao, J., Xie, D., Quake, S.R., 2017. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Research*. 27(10), 1231–1242. <https://doi.org/10.1038/cr.2017.106>
- Song, H., Dong, T., Yan, X., Wang, W., Tian, Z., Sun, A., Dong, Y., Zhu, H., Hu, H., 2023. Genomic selection and its research progress in aquaculture breeding. *Reviews in Aquaculture*. 15(1), 274–291. <https://doi.org/10.1111/raq.12716>
- Song, J., Angel, A., Howard, M., Dean, C., 2012. Vernalization - a cold-induced epigenetic switch. *Journal of Cell Science*. 125(16), 3723–3731. <https://doi.org/10.1242/jcs.084764>
- Song, W., Xie, Y., Sun, M., Li, X., Fitzpatrick, C.K., Vaux, F., O'Malley, K.G., Zhang, Q., Qi, J., He, Y., 2021. A duplicated *amh* is the master sex-determining gene for *Sebastes* rockfish in the Northwest Pacific. *Open Biology*. 11(7), 210063. <https://doi.org/10.1098/rsob.210063>
- Sosa, C.C., Clavijo-Buriticá, D.C., García-Merchán, V.H., López-Rozo, N., Riccio-Rengifo, C., Diaz, M.V., Londoño, D.A., Quimbaya, M.A., 2023. GOCmpare: An R package to compare functional enrichment analysis between two species. *Genomics*. 115(1), 110528. <https://doi.org/10.1016/j.ygeno.2022.110528>
- Spinks, R.K., Bonzi, L.C., Ravasi, T., Munday, P.L., Donelson, J.M., 2021. Sex- and time-specific parental effects of warming on reproduction and offspring quality in a coral reef fish. *Evolutionary Applications*. 14(4), 1145–1158. <https://doi.org/10.1111/eva.13187>
- Staack, A., Donjacour, A.A., Brody, J., Cunha, G.R., Carroll, P., 2003. Mouse urogenital development: A practical approach. *Differentiation*. 71(7), 402–413. <https://doi.org/10.1046/j.1432-0436.2003.7107004.x>
- Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazar, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., Lancet, D., 2016. The GeneCards suite: From gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics*. 54(1), 1.30.1–1.30.33. <https://doi.org/10.1002/cpbi.5>
- Stöck, M., Kratochvíl, L., Kuhl, H., Rovatsos, M., Evans, B.J., Suh, A., Valenzuela, N., Veyrunes, F., Zhou, Q., Gamble, T., Capel, B., Schartl, M., Guiguen, Y., 2021. A brief review of vertebrate sex evolution with a pledge for integrative research: towards ‘sexomics.’ *Philosophical Transactions of the Royal Society B: Biological Sciences*. 376(1832), 20200426. <https://doi.org/10.1098/rstb.2020.0426>
- Strüssmann, C.A., Saito, T., Usui, M., Yamada, H., Takashima, F., 1997. Thermal thresholds and critical period of thermolabile sex determination in two atherinid fishes, *Odontesthes bonariensis* and *Patagonina hatcheri*. *Journal of Experimental Zoology*. 278(3), 167–177. [https://doi.org/10.1002/\(SICI\)1097-010X\(19970615\)278:3<167::AID-JEZ6>3.0.CO;2-M](https://doi.org/10.1002/(SICI)1097-010X(19970615)278:3<167::AID-JEZ6>3.0.CO;2-M)
- Stueve, T.R., Li, W.-Q., Shi, J., Marconett, C.N., Zhang, T., Yang, C., Mullen, D., Yan, C., Wheeler, W., Hua, X., Zhou, B., Borok, Z., Caporaso, N.E., Pesatori, A.C., Duan, J., Laird-Offringa, I.A., Landi, M.T., 2017. Epigenome-wide analysis of DNA methylation in lung tissue shows concordance with blood studies and identifies tobacco smoke-inducible enhancers. *Human molecular genetics*. 26(15), 3014–3027. <https://doi.org/10.1093/hmg/ddx188>
- Su, S., Law, C.W., Ah-Cann, C., Asselin-Labat, M.-L., Blewitt, M.E., Ritchie, M.E., 2017. Glimma: interactive graphics for gene expression analysis. *Bioinformatics*. 33(13), 2050–2052. <https://doi.org/10.1093/bioinformatics/btx094>
- Suelves, M., Carrió, E., Núñez-Álvarez, Y., Peinado, M.A., 2016. DNA methylation dynamics in cellular commitment and differentiation. *Briefings in Functional Genomics*. 15, 443–453. <https://doi.org/10.1093/bfpg/ewl017>
- Sun, B.J., Li, T., Gao, J., Ma, L., Du, W.G., 2015. High incubation temperatures enhance mitochondrial energy metabolism in reptile embryos. *Scientific Reports*. 5(1), 8861. <https://doi.org/10.1038/srep08861>

- Sun, L.X., Wang, Y.Y., Zhao, Y., Wang, H., Li, N., Ji, X.S., 2016. Global DNA methylation changes in Nile tilapia gonads during high temperature-induced masculinization. *PLoS One*. 11(8), e0158483. <https://doi.org/10.1371/journal.pone.0158483>
- Sun, R., Zhu, P., 2022. Advances in measuring DNA methylation. *Blood Science*. 4(1), 8–15. <https://doi.org/10.1097/BS9.0000000000000098>
- Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 6(7), e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Swagers, J., De Cupere, S., Gaens, N., Lancaster, L.T., Carbonell, J.A., Sánchez Guillén, R.A., Stoks, R., 2023. Plasticity and associated epigenetic mechanisms play a role in thermal evolution during range expansion. *Evolution Letters*. qrac007. <https://doi.org/10.1093/evlett/qrac007>
- Szyf, M., Weaver, I.C.G., Champagne, F.A., Diorio, J., Meaney, M.J., 2005. Maternal programming of steroid receptor expression and phenotype through DNA methylation in the rat. *Frontiers in Neuroendocrinology*. 26(3–4), 139–162. <https://doi.org/10.1016/j.yfrne.2005.10.002>

T

- Takahashi, H., 1977. Juvenile hermaphroditism in the zebrafish, *Brachydanio rerio*. *Bulletin of the Faculty of Fisheries Hokkaido University*. 28(2), 57–65. [https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/23605/1/28\(2\)_P57-65.pdf](https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/23605/1/28(2)_P57-65.pdf)
- Takahashi, Y., Valencia, M., Yu, Y., Ouchi, Y., Takahashi, K., Shokhirev, M.N., Lande, K., Williams, A.E., Fresia, C., Kurita, M., Hishida, T., Shojima, K., Hatanaka, F., Nuñez-Delgado, E., Esteban, C.R., Belmonte, J.C.I., 2023. Transgenerational inheritance of acquired epigenetic signatures at CpG islands in mice. *Cell*. 186(4), 715–731.e19. <https://doi.org/10.1016/j.cell.2022.12.047>
- Takehana, Y., Demiyah, D., Naruse, K., Hamaguchi, S., Sakaizumi, M., 2007. Evolution of different Y chromosomes in two medaka species, *Oryzias dancena* and *O. latipes*. *Genetics*. 175(3), 1335–1340. <https://doi.org/10.1534/genetics.106.068247>
- Takehana, Y., Hamaguchi, S., Sakaizumi, M., 2008. Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi*. *Chromosome Research*. 16, 801–811. <https://doi.org/10.1007/s10577-008-1227-5>
- Takehana, Y., Matsuda, M., Myosho, T., Suster, M.L., Kawakami, K., Shin-i, T., Kohara, Y., Kuroki, Y., Toyoda, A., Fujiyama, A., Hamaguchi, S., 2014. Co-option of *Sox3* as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. *Nature communications*. 5(1), 4157. <https://doi.org/10.1038/ncomms5157>
- Tavolga, W.N., 1949. Embryonic development of the platyfish (*Platyfocilus*), the swordtail (*Xiphophorus*), and their hybrids. *Bulletin of the AMNH. New York University*. 94(4), 7320887. <https://www.proquest.com/dissertations-theses/embryonic-development-platyfish-platyfocilus/docview/301824279/se-2>
- Taylor, J.H., 2012. DNA methylation and cellular differentiation. Springer Science & Business Media. Wien, New York. ISBN: 13:9783709187234 <https://doi.org/10.1007/978-3-7091-8721-0>
- Teng, J., Zhao, Y., Chen, H.J., Wang, H., Ji, X.S., 2020. Transcriptome profiling and analysis of genes associated with high temperature-induced masculinization in sex-undifferentiated Nile tilapia gonad. *Marine Biotechnology*. 22, 367–379. <https://doi.org/10.1007/s10126-020-09956-5>
- Tesson, B.M., Breitling, R., Jansen, R.C., 2010. DiffCoEx: A simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics*. 11(1), 497. <https://doi.org/10.1186/1471-2105-11-497>
- Thelie, A., Pascal, G., Angulo, L., Perreau, C., Papillier, P., Dalbès-Tran, R., 2012. An oocyte-preferential histone mRNA stem-loop-binding protein like is expressed in several mammalian species. *Molecular Reproduction and Development*. 79(6), 380–391. <https://doi.org/10.1002/mrd.22040>
- Timmons, J.A., Szkop, K.J., Gallagher, I.J., 2015. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biology*. 16(1), 186. <https://doi.org/10.1186/s13059-015-0761-7>
- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F.A.M., Bargelloni, L., Power, D.M., Bonhomme, F., Canario, A.V.M., Reinhardt, R., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*. 5(1), 5770. <https://doi.org/10.1038/ncomms6770>
- Todd, E.V., Liu, H., Muncaster, S., Gemmell, N.J., 2016. Bending genders: the biology of natural sex change in fish. *Sexual Development*. 10(5–6), 223–241. <https://doi.org/10.1159/000449297>
- Todd, E.V., Ortega-Recalde, O., Liu, H., Lamm, M.S., Rutherford, K.M., Cross, H., Black, M.A., Kardailsky, O., Marshall Graves, J.A., Hore, T.A., Godwin, J.R., Gemmell, N.J., 2019. Stress, novel sex genes, and epigenetic

- reprogramming orchestrate socially controlled sex change. *Science Advances*. 5(7), eaaw7006. <https://doi.org/10.1126/sciadv.aaw7006>
- Trucchi, E., Mazzarella, A.B., Gilfillan, G.D., Lorenzo, M.T., Schönswetter, P., Paun, O., 2016. BsRADseq: Screening DNA methylation in natural populations of non-model species. *Molecular Ecology*. 25(8), 1697–1713. <https://doi.org/10.1111/mec.13550>
- Tsai, C.F., Sung, Y.T., 2020. Ensemble feature selection in high dimension, low sample size datasets: Parallel and serial combination approaches. *Knowledge-Based Systems*. 203, 106097. <https://doi.org/10.1016/j.knosys.2020.106097>
- Tsukazaki, T., Chiang, T.A., Davison, A.F., Attisano, L., Wrana, J.L., 1998. SARA, a FYVE Domain Protein that Recruits Smad2 to the TGF β Receptor. *Cell*. 95(6), 779–791. [https://doi.org/10.1016/S0092-8674\(00\)81701-8](https://doi.org/10.1016/S0092-8674(00)81701-8)
- Tu, J., Huang, Z., Wang, Y., Wang, M., Yin, Z., Mei, X., Li, M., Li, L., 2021. Transcriptome analysis of the progression from chronic pancreatitis to pancreatic cancer and metastatic pancreatic cancer. *Scientific Reports*. 11(1), 3409. <https://doi.org/10.1038/s41598-021-83015-4>
- Tucker, K., Beard, C., Dausmann, J., Jackson-Grusby, L., Laird, P.W., Lei, H., Li, E., Jaenisch, R., 1996. Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes & development*. 10(8), 1008–1020. <https://doi.org/10.1101/gad.10.8.1008>
- Turcotte, H., Hooker, J., Samanfar, B., Parent, J.S., 2022. Can epigenetics guide the production of better adapted cultivars? *Agronomy*. 12(4), 838. <https://doi.org/10.3390/agronomy12040838>
- Turner, B.M., 2005. Reading signals on the nucleosome with a new nomenclature for modified histones. *Nature structural & molecular biology*. 12(2), 110–112. <https://doi.org/10.1038/nsmb0205-110>

U

- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research*. 40(15), e115. <https://doi.org/10.1093/nar/gks596>

V

- Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. *PLoS One*. 14(11), e0224365. <https://doi.org/10.1371/journal.pone.0224365>
- Vainio, S., Heikkilä, M., Kispert, A., Chin, N., McMahon, A.P., 1999. Female development in mammals is regulated by Wnt-4 signalling. *Nature*. 397(6718), 405–409. <https://doi.org/10.1038/17068>
- Vaisvila, R., Ponnaluri, V.K.C., Sun, Z., Langhorst, B.W., Saleh, L., Guan, S., Dai, N., Campbell, M.A., Sexton, B.S., Marks, K., Samaranayake, M., Samuelson, J.C., Church, H.E., Tamanaha, E., Corrêa, I.R., Pradhan, S., Dimalanta, E.T., Evans, T.C., Williams, L., Davis, T.B., 2021. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research*. 31(7), 1280–1289. <https://doi.org/10.1101/gr.266551.120>
- Valdivieso, A., Anastasiadi, D., Ribas, L., Piferrer, F., 2023a. Development of epigenetic biomarkers for the identification of sex and thermal stress in fish using DNA methylation analysis and machine learning procedures. *Molecular Ecology Resources*. 23(2), 453–470. <https://doi.org/10.1111/1755-0998.13725>
- Valdivieso, A., Sánchez-Baizán, N., Mitizakis, N., Papandroulakis, N., Piferrer, F., 2023b. Development of epigenetic biomarkers with diagnostic and prognostic value to assess the lasting effects of early temperature changes in farmed fish. *Aquaculture*. 563, 738918. <https://doi.org/10.1016/j.aquaculture.2022.738918>
- Valente, L.M.P., Moutou, K.A., Conceição, L.E.C., Engrola, S., Fernandes, J.M.O., Johnston, I.A., 2013. What determines growth potential and juvenile quality of farmed fish species? *Reviews in Aquaculture* 5, S168–S193. <https://doi.org/10.1111/raq.12020>
- Valenzuela, N., Lance, V., 2004. Temperature-dependent sex determination in vertebrates. Smithsonian Books Washington, DC. ISBN 1-58834-203-4.
- van Doorn, G.S., Kirkpatrick, M., 2007. Turnover of sex chromosomes induced by sexual conflict. *Nature* 449(7164), 909–912. <https://doi.org/10.1038/nature06178>
- Vandeputte, M., Dupont-Nivet, M., Chavanne, H., Chatain, B., 2007. A Polygenic Hypothesis for Sex Determination in the European Sea Bass *Dicentrarchus labrax*. *Genetics*. 176(2), 1049–1057. <https://doi.org/10.1534/genetics.107.072140>
- Vandeputte, M., Gagnaire, P.A., Allal, F., 2019. The European sea bass: a key marine fish model in the wild and in aquaculture. *Animal Genetics*. 50(3), 195–206. <https://doi.org/10.1111/age.12779>

- Vandeputte, Marc, Piferrer, F., 2019. Genetic and environmental components of sex determination in the European sea bass, In: Wang, H.P., Piferrer, F., Chen, S.L., Shen, Z.-G. (Eds.), Sex Control in Aquaculture. John Wiley & Sons, Ltd, Chichester, UK, 305–325 pp. <https://doi.org/10.1002/9781119127291.ch14>
- Vandeputte, M., Quillet, E., Chatain, B., 2012. Are sex ratios in wild European sea bass (*Dicentrarchus labrax*) populations biased? Aquatic Living Resources. 25(1), 77–81. <https://doi.org/10.1051/alr/2012002>
- Varona, L., Munilla, S., Mouresan, E.F., González-Rodríguez, A., Moreno, C., Altarriba, J., 2015. A bayesian model for the analysis of transgenerational epigenetic variation. G3 Genes|Genomes|Genetics. 5(4), 477–485. <https://doi.org/10.1534/g3.115.016725>
- Vazquez, F.J.S., Munoz-Cueto, J.A. (Eds.), 2014. Biology of European sea bass. CRC Press. 13:9781466599451. <https://doi.org/10.1201/b16043>
- Veillard, A.-C., Datlinger, P., Laczik, M., Squazzo, S., Bock, C., 2016. Diagenode® Premium RRBS technology: cost-effective DNA methylation mapping with superior coverage. Nature Methods. 13, i–ii. <https://doi.org/10.1038/nmeth.f.391>
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E., Hieter, P., Vogelstein, B., Kinzler, K.W., 1997. Characterization of the yeast transcriptome. Cell. 88(2), 243–251. [https://doi.org/10.1016/S0092-8674\(00\)81845-0](https://doi.org/10.1016/S0092-8674(00)81845-0)
- Velten, B., Braunger, J.M., Argelaguet, R., Arnol, D., Wirbel, J., Bredikhin, D., Zeller, G., Stegle, O., 2022. Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. Nature Methods. 19(2):179–86. <https://doi.org/10.1038/s41592-021-01343-9>
- Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, Fourth. ed. Springer, New York. XII, 498 pp. ISBN: 978-0-387-95457-8. <https://doi.org/10.1007/978-0-387-21706-2>
- Villamizar, N., Ribas, L., Piferrer, F., Vera, L.M., Sánchez-Vázquez, F.J. 2012. Impact of daily thermocycles on hatching rhythms, larval performance and sex differentiation of zebrafish. PLoS One. 7(12), e52153. <https://doi.org/10.1371/journal.pone.0052153>
- Vining, B., Ming, Z., Bagheri-Fam, S., Harley, V., 2021. Diverse regulation but conserved function: *SOX9* in vertebrate sex determination. Genes. 12(4), 486. <https://doi.org/10.3390/genes12040486>
- Vogt, G., 2021. Epigenetic variation in animal populations: Sources, extent, phenotypic implications, and ecological and evolutionary relevance. Journal of Biosciences. 46(1), 24. <https://doi.org/10.1007/s12038-021-00138-6>
- Voldoire, E., Brunet, F., Naville, M., Volff, J.N., Galiana, D., 2017. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. PLoS One. 12(7), e0180936. <https://doi.org/10.1371/journal.pone.0180936>
- Volff, J.N., Scharl, M., 2002. Sex determination and sex chromosome evolution in the medaka, *Oryzias latipes*, and the platyfish, *Xiphophorus maculatus*. Cytogenetic and genome research. 99(1-4), 170–177. <https://doi.org/10.1159/000071590>
- Volff, J.N., Scharl, M., 2001. Variability of genetic sex determination in poeciliid fishes. Genetica. 111, 101–110. <https://doi.org/10.1023/a:1013795415808>
- Voshall, A., Moriyama, E.N., 2018. Next-generation transcriptome assembly: strategies and performance analysis, bioinformatics in the era of post genomics and big data. Bioinformatics in the era of post genomics and big data. 15-36. <https://doi.org/10.5772/intechopen.73497>

W

- Waddington, C., 1957. The Strategy of the Genes. Routledge. (Vol. George Allen & Unwin): London.
- Waddington, C.H., 1942. Canalization of development and the inheritance of acquired characters. Nature 150, 563–565. <https://doi.org/10.1038/150563a0>
- Wagner, S., Whiteley, S.L., Castelli, M., Patel, H.R., Deveson, I.W., Blackburn, J., Holleley, C.E., Marshall Graves, J.A., Georges, A., 2023. Gene expression of male pathway genes *sox9* and *amh* during early sex differentiation in a reptile departs from the classical amniote model. BMC Genomics. 24(1), 243. <https://doi.org/10.1186/s12864-023-09334-0>
- Wallis, M.C., Waters, P.D., Graves, J. a. M., 2008. Sex determination in mammals--before and after the evolution of *SRY*. Cellular and molecular life sciences. 65, 3182–95. <https://doi.org/10.1007/s00018-008-8109-z>
- Wang, C.H., Lai, H.C., Liou, T.M., Hsu, K.F., Chou, C.Y., Lee, G.B., 2013. A DNA methylation assay for detection of ovarian cancer cells using a HpaII/MspI digestion-based PCR assay in an integrated microfluidic system. Microfluid Nanofluid. 15, 575–585. <https://doi.org/10.1007/s10404-013-1179-8>
- Wang, G., Yu, J., Yang, Y., Liu, X., Zhao, X., Guo, X., Duan, T., Lu, C., Kang, J., 2020. Whole-transcriptome sequencing uncovers core regulatory modules and gene signatures of human fetal growth restriction. Clinical and Translational Medicine. 9, 1-3. <https://doi.org/10.1186/s40169-020-0259-0>

- Wang, H.Q., Tuominen, L.K., Tsai, C.J., 2011. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics*. 27(2), 225–231. <https://doi.org/10.1093/bioinformatics/btq650>
- Wang, J., Liu, Y., Jiang, S., Li, W., Gui, L., Zhou, T., Zhai, W., Lin, Z., Lu, J., Chen, L., 2019a. Transcriptomic and epigenomic alterations of Nile tilapia gonads sexually reversed by high temperature. *Aquaculture*. 508, 167–177. <https://doi.org/10.1016/j.aquaculture.2019.04.073>
- Wang, J., Tang, C., Wang, Q., Su, J., Ni, T., Yang, W., Wang, Yongsheng, Chen, W., Liu, X., Wang, S., Zhang, J., Song, H., Zhu, J., Wang, Yuan, 2017. NRF1 coordinates with DNA methylation to regulate spermatogenesis. *The FASEB Journal*. 31(11), 4959–4970. <https://doi.org/10.1096/fj.201700093R>
- Wang, J., Tian, G.G., Zheng, Z., Li, B., Xing, Q., Wu, J., 2019b. Comprehensive transcriptomic analysis of mouse gonadal development involving sexual differentiation, meiosis and gametogenesis. *Biology Procedures Online*. 21(1), 20. <https://doi.org/10.1186/s12575-019-0108-y>
- Wang, L., Zhang, Jun, Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, Jing, Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X., Liu, J., 2014. Programming and inheritance of parental DNA methylomes in mammals. *Cell*. 157(4), 979–991. <https://doi.org/10.1016/j.cell.2014.04.017>
- Wang, Q., Liu, K., Feng, B., Zhang, Z., Wang, R., Tang, L., Li, W., Li, Q., Piferrer, F., Shao, C., 2019. Transcriptome of gonads from high temperature induced sex reversal during sex determination and differentiation in Chinese tongue sole, *Cynoglossus semilaevis*. *Frontiers Genetics* 10. <https://doi.org/10.3389/fgene.2019.01128>
- Wang, S.Y., Lau, K., Lai, K.P., Zhang, J.W., Tse, A.C.K., Li, J.W., Tong, Y., Chan, T.F., Wong, C.K.C., Chiu, J.M.Y., Au, D.W.T., Wong, A.S.T., Kong, R.Y.C., Wu, R.S.S., 2016. Hypoxia causes transgenerational impairments in reproduction of fish. *Nature Communications*. 7, 12114. <https://doi.org/10.1038/ncomms12114>
- Wang, T., Ma, J., Hogan, A.N., Fong, S., Licon, K., Tsui, B., Kreisberg, J.F., Adams, P.D., Carvunis, A.-R., Bannasch, D.L., Ostrander, E.A., Ideker, T., 2020. Quantitative translation of dog-to-human aging by conserved remodeling of the DNA methylome. *Cell Systems*. 11(2), 176–185.e6. <https://doi.org/10.1016/j.cels.2020.06.006>
- Wang, X., Bhandari, R.K., 2019. DNA methylation dynamics during epigenetic reprogramming of medaka embryo. *Epigenetics*. 14(6), 611–622. <https://doi.org/10.1080/15592294.2019.1605816>
- Wang, Xiyan, Su, L., Pan, X., Yao, J., Li, Z., Wang, Xuenan, Xu, B., 2015. Effect of BMP-6 on development and maturation of mouse preantral follicles *in vitro*. *Biotechnology & Biotechnological Equipment*. 29(2), 336–344. <https://doi.org/10.1080/13102818.2014.996605>
- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nature reviews genetics*. 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Wang, Z.F., Ingledue, T.C., Dominski, Z., Sanchez, R., Marzluff, W.F., 1999. Two *Xenopus* proteins that bind the 3' end of histone mRNA: implications for translational control of histone synthesis during oogenesis. *Molecular and Cellular Biology*. 19(1), 835–845. <https://doi.org/10.1128/MCB.19.1.835>
- Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B., Galili, T., 2020. gplots: Various R programming tools for plotting data. R package version 3.0.1.2. <https://CRAN.R-project.org/package=gplots>
- Weaver, I.C.G., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., Meaney, M.J., 2004. Epigenetic programming by maternal behavior. *Nature Neuroscience*. 7(8), 847–854. <https://doi.org/10.1038/nn1276>
- Weber, C., Capel, B., 2018. Sex reversal. *Current Biology*. 28(21), R1234–R1236. <https://doi.org/10.1016/j.cub.2018.09.043>
- Weber, C., Zhou, Y., Lee, J.G., Looger, L.L., Qian, G., Ge, C., Capel, B., 2020. Temperature-dependent sex determination is mediated by pSTAT3 repression of *Kdm6b*. *Science*. 368(6488), 303–306. <https://doi.org/10.1126/science.aaz4165>
- Weber, D.N., Fields, A.T., Patterson III, W.F., Barnett, B.K., Hollenbeck, C.M., Portnoy, D.S., 2022. Novel epigenetic age estimation in wild-caught Gulf of Mexico reef fishes. *Canadian Journal of Fisheries and Aquatic Sciences*. 79(1), 1–5. <https://doi.org/10.1139/cjfas-2021-0240>
- Wei, S.H., Balch, C., Paik, H.H., Kim, Y.-S., Baldwin, R.L., Liyanarachchi, S., Li, L., Wang, Z., Wan, J.C., Davuluri, R.V., Karlan, B.Y., Gifford, G., Brown, R., Kim, S., Huang, T.H.-M., Nephew, K.P., 2006. Prognostic DNA methylation biomarkers in ovarian cancer. *Clinical Cancer Research*. 12(9), 2788–2794. <https://doi.org/10.1158/1078-0432.CCR-05-1551>
- Wen, A.Y., You, F., Sun, P., Li, J., Xu, D.D., Wu, Z.H., Ma, D.Y., Zhang, P.J., 2014. CpG methylation of *dmrt1* and *cyp19a* promoters in relation to their sexual dimorphic expression in the Japanese flounder *Paralichthys olivaceus*. *Journal of Fish Biology*. 84(1), 193–205. <https://doi.org/10.1111/jfb.12277>
- Weternan, M.A.J., Barth, P.G., van Spaendonck-Zwarts, K.Y., Aronica, E., Poll-The, B.-T., Brouwer, O.F., van Tintelen, J.P., Qahar, Z., Bradley, E.J., de Wissel, M., Salviati, L., Angelini, C., van den Heuvel, L., Thomasse, Y.E.M., Backx, A.P., Nürnberg, G., Nürnberg, P., Baas, F., 2013. Recessive *MYL2* mutations cause infantile type I muscle fibre disease and cardiomyopathy. *Brain* 136(1), 282–293. <https://doi.org/10.1093/brain/aws293>

- Whiteley, S.L., Holleley, C.E., Wagner, S., Blackburn, J., Deveson, I.W., Graves, J.A.M., Georges, A., 2021. Two transcriptionally distinct pathways drive female development in a reptile with both genetic and temperature dependent sex determination. *PLoS Genetics*. 17(4), e1009465. <https://doi.org/10.1371/journal.pgen.1009465>
- Wibbels, T., Rostal, D., Byles, R., 1998. High pivotal temperature in the sex determination of the olive ridley sea turtle, *Lepidochelys olivacea*, from Playa Nancite, Costa Rica. *Copeia*. 4, 1086–1088. <https://doi.org/10.2307/1447364>
- Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., RStudio, 2021. ggplot2: create elegant data visualisations using the grammar of graphics. R package version: 3.35 and 3.4.2. <https://ggplot2.tidyverse.org>
- Wilkins, A.S., 1995. Moving up the hierarchy: A hypothesis on the evolution of a genetic sex determination pathway. *BioEssays*. 17(1), 71–77. <https://doi.org/10.1002/bies.950170113>
- Williams, T.M., Carroll, S.B., 2009. Genetic and molecular insights into the development and evolution of sexual dimorphism. *Nature Reviews Genetics*. 10(11), 797–804. <https://doi.org/10.1038/nrg2687>
- Wingett, S.W., Andrews, S., 2018. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*. 7, 1338. <https://doi.org/10.12688/f1000research.15931.2>
- Wu, T.P., Wang, T., Seetin, M.G., Lai, Y., Zhu, S., Lin, K., Liu, Y., Byrum, S.D., Mackintosh, S.G., Zhong, M., Tackett, A., Wang, G., Hon, L.S., Fang, G., Swenberg, J.A., Xiao, A.Z., 2016. DNA methylation on *N*⁶-adenine in mammalian embryonic stem cells. *Nature*. 532(7599), 329–333. <https://doi.org/10.1038/nature17640>

X

- Xi, Y., Li, W., 2009. BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10, 232. <https://doi.org/10.1186/1471-2105-10-232>
- Xie, J., Yang, F., Wang, J., Karikomi, M., Yin, Y., Sun, J., Wen, T., Nie, Q., 2020. DNF: A differential network flow method to identify rewiring drivers for gene regulatory networks. *Neurocomputing*. 410, 202–210. <https://doi.org/10.1016/j.neucom.2020.05.028>
- Xu, X., Li, S., Liang, T., Sun, T., 2020. Sample selection-based hierarchical extreme learning machine. *Neurocomputing*. 377, 95–102. <https://doi.org/10.1016/j.neucom.2019.10.013>

Y

- Yang, F., Zheng, J., 2014. High temperature sensitivity is intrinsic to voltage-gated potassium channels. *eLife* 3, e03255. <https://doi.org/10.7554/eLife.03255>
- Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., Cabau, C., Bouchez, O., Fostier, A., Guiguen, Y., 2012. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Current Biology*. 22(15), 1423–1428. <https://doi.org/10.1016/j.cub.2012.05.045>
- Yao, Z.L., Zhao, Y., Chen, H.J., Wang, H., Ji, X.S., 2022. Integrated analysis of DNA methylome and RNA transcriptome during high-temperature-induced masculinization in sex-undifferentiated Nile tilapia gonad. *Aquaculture*. 552, 738043. <https://doi.org/10.1016/j.aquaculture.2022.738043>
- Yi, M., Chen, F., Luo, M., Cheng, Y., Zhao, H., Cheng, H., Zhou, R., 2014. Rapid evolution of piRNA pathway in the teleost fish: implication for an adaptation to transposon diversity. *Genome Biology and Evolution*. 6(6), 1393–1407. <https://doi.org/10.1093/gbe/evu105>
- Yildirim, E., Aksoy, S., Onel, T., Yaba, A., 2020. Gonadal development and sex determination in mouse. *Reproductive Biology*. 20(2), 115–126. <https://doi.org/10.1016/j.repbio.2020.01.007>
- Yin, L., Cai, Z., Zhu, B., Xu, C., 2018. Identification of key pathways and genes in the dynamic progression of HCC based on WGCNA. *Genes*. 9(2), 92. <https://doi.org/10.3390/genes9020092>
- Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T., Ito, M., 2008. A W-linked DM-domain gene, *DM-W*, participates in primary ovary development in *Xenopus laevis*. *Proceedings of the National Academy of Sciences*. 105(7), 2469–2474. <https://doi.org/10.1073/pnas.0712244105>
- Yu, M., Han, D., Hon, G.C., He, C., 2018. Tet-Assisted Bisulfite Sequencing (TAB-seq). *Methods Molecular Biology, DNA Methylation Protocols*. 1708, 645–663. https://doi.org/10.1007/978-1-4939-7481-8_33
- Yu, Y., Zhu, D., Ma, C., Cao, H., Wang, Y., Xu, Y., Zhang, W., Yan, Y., 2016. Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development. *The Crop Journal*. 4(2), 92–106. <https://doi.org/10.1016/j.cj.2016.01.006>
- Yu-Wei, C. 2015. Machine learning with R cookbook. Packt Publishing Ltd, Birmingham. 442. ISBN 978-1-78398-204-2. <https://www.perlego.com/book/117994/machine-learning-with-r-cookbook-pdf>

Z

- Zaratiegui, M., Irvine, D.V., Martienssen, R.A., 2007. Noncoding RNAs and gene silencing. *Cell* 128(4), 763–776. <https://doi.org/10.1016/j.cell.2007.02.016>
- Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*. 4,1. <https://doi.org/10.2202/1544-6115.1128>
- Zhang, B., Zhou, Y., Lin, N., Lowdon, R.F., Hong, C., Nagarajan, R.P., Cheng, J.B., Li, D., Stevens, M., Lee, H.J., Xing, X., Zhou, J., Sundaram, V., Elliott, G., Gu, J., Shi, T., Gascard, P., Sigaroudinia, M., Tlsty, T.D., Kadlecsek, T., Weiss, A., O'Geen, H., Farnham, P.J., Maire, C.L., Ligon, K.L., Madden, P.A.F., Tam, A., Moore, R., Hirst, M., Marra, M.A., Zhang, Baoxue, Costello, J.F., Wang, T., 2013. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Research*. 23(9), 1522–1540. <https://doi.org/10.1101/gr.156539.113>
- Zhang, F., Kaufman, H.L., Deng, Y., Drabier, R., 2013. Recursive SVM biomarker selection for early detection of breast cancer in peripheral blood. *BMC medical genomics*. 6(1), 1–10. <https://doi.org/10.1186/1755-8794-6-S1-S4>
- Zhang, J., Wang, L., Xu, X., Li, X., Guan, W., Meng, T., Xu, G., 2020. Transcriptome-based network analysis unveils eight immune-related genes as molecular signatures in the immunomodulatory subtype of triple-negative breast cancer. *Frontiers in Oncology*. 10,1787. <https://doi.org/10.3389/fonc.2020.01787>
- Zhang, J., Yu, Z., Zhang, W., Shi, H., Tao, M., Chen, J., Liu, Y., Zhao, M., Li, S., Zhang, Y., Lin, H., 2021. Jumonji family histone demethylases in *Epinephelus coioides* gonad. *Reproduction and Breeding*. 1(1), 1–10. <https://doi.org/10.1016/j.repbre.2021.02.002>
- Zhang, S., Cheng, Y., Věchtová, P., Boryshpolets, S., Shazada, N.E., Alavi, S.M.H., Cosson, J., Linhart, O., 2023. Potential implications of sperm DNA methylation functional properties in aquaculture management. *Reviews in Aquaculture*. 15(2), 536–556. <https://doi.org/10.1111/raq.12735>
- Zhang, Z., Lau, S.-W., Zhang, L., Ge, W., 2015. Disruption of zebrafish follicle-stimulating hormone receptor (*fshr*) but not luteinizing hormone receptor (*lhcg*) gene by TALEN leads to failed follicle activation in females followed by sexual reversal to males. *Endocrinology*. 156(10), 3747–3762. <https://doi.org/10.1210/en.2015-1039>
- Zhang, Z., Wang, Y., Wang, S., Liu, J., Warren, W., Mitreva, M., Walter, R.B., 2011. Transcriptome analysis of female and male *Xiphophorus maculatus* Jp 163 A. *PLoS One*. 6(4), e18379. <https://doi.org/10.1371/journal.pone.0018379>
- Zhao, F., Li, R., Xiao, S., Diao, H., El Zowalaty, A.E., Ye, X., 2014. Multigenerational exposure to dietary zearalenone (ZEA), an estrogenic mycotoxin, affects puberty and reproduction in female mice. *Reproductive Toxicology*. 47, 81–88. <https://doi.org/10.1016/j.reprotox.2014.06.005>
- Zhao, H., Du, X., Zhang, K., Liu, Y., Wang, Y., Liu, J., He, Y., Wang, X., Zhang, Q., 2018. Weighted correlation network analysis (WGCNA) of Japanese flounder (*Paralichthys olivaceus*) embryo transcriptome provides crucial gene sets for understanding haploid syndrome and rescue by diploidization. *Journal of Ocean University of China*. 17, 1441–1450. <https://doi.org/10.1007/s11802-018-3656-x>
- Zhao, H., Tang, X., Wu, M., Li, Q., Yi, X., Liu, S., Jiang, J., Wang, S., Sun, X., 2021. Transcriptome characterization of short distance transport stress in beef cattle blood. *Frontiers in Genetics*. 12:616388. <https://doi.org/10.3389/fgene.2021.616388>
- Zhao, S., Fung-Leung, W.P., Bittner, A., Ngo, K., Liu, X., 2014. Comparison of RNA-seq and microarray in transcriptome profiling of activated T cells. *PLoS One*. 9(1), e78644. <https://doi.org/10.1371/journal.pone.0078644>
- Zheng, S., Tao, W., Yang, H., Kocher, T.D., Wang, Z., Peng, Z., Jin, L., Pu, D., Zhang, Y., Wang, D., 2022. Identification of sex chromosome and sex-determining gene of southern catfish (*Silurus meridionalis*) based on XX, XY and YY genome sequencing. *Proceedings of the Royal Society B: Biological Sciences*. 289(1971), 20212645. <https://doi.org/10.1098/rspb.2021.2645>
- Zheng, W., Jin, M., 2022. Improving the performance of feature selection methods with low-sample-size data. *The Computer Journal*. bxac033. <https://doi.org/10.1093/comjnl/bxac033>
- Zheng, X., Li, Z., Wang, G., Wang, H., Zhou, Y., Zhao, X., Cheng, C.Y., Qiao, Y., Sun, F., 2021. Sperm epigenetic alterations contribute to inter- and transgenerational effects of paternal exposure to long-term psychological stress via evading offspring embryonic reprogramming. *Cell Discovery*. 7(1), 1–22. <https://doi.org/10.1038/s41421-021-00343-5>
- Zhi, J., Sun, J., Wang, Z., Ding, W., 2018. Support vector machine classifier for prediction of the metastasis of colorectal cancer. *International journal of molecular medicine* 41(3), 1419–1426. <https://doi.org/10.3892/ijmm.2018.3359>
- Zhou, H., Zhuang, Z.X., Sun, Y.Q., Chen, Q., Zheng, X.Y., Liang, Y.T., Mahboob, S., Wang, Q., Zhang, R., Al-Ghanim, K.A., Shao, C.-W., Li, Y.J., 2019. Changes in DNA methylation during epigenetic-associated sex

- reversal under low temperature in *Takifugu rubripes*. PLoS One. 14, e0221641. <https://doi.org/10.1371/journal.pone.0221641>
- Zhou, R., Shang, R., Gong, D., Xu, X., Liu, S., 2019. Characterization of H3 methylation in regulating oocyte development in cyprinid fish. Science China Life Sciences. 62, 829–837. <https://doi.org/10.1007/s11427-018-9346-6>
- Zhou, Z., Leng, C., Wang, Z., Long, L., Lv, Y., Gao, Z., Wang, Y., Wang, S., Li, P., 2023. The potential regulatory role of the lncRNA-miRNA-mRNA axis in teleost fish. Frontiers in Immunology. 14, 1065357. <https://doi.org/10.3389/fimmu.2023.1065357>
- Zhu, L., Ding, Y., Chen, C.-Y., Wang, L., Huo, Z., Kim, S., Sotiriou, C., Oesterreich, S., Tseng, G.C., 2017. MetaDCN: meta-analysis framework for differential co-expression network detection with an application in breast cancer. Bioinformatics. 33(8), 1121–1129. <https://doi.org/10.1093/bioinformatics/btw788>
- Zhu, R., Lin, W., Tang, L., Hu, Y., 2021. Identification of hub genes associated with adult acute myeloid leukemia progression through weighted gene co-expression network analysis. Aging. (Albany NY) 13(4), 5686–5697. <https://doi.org/10.18632/aging.202493>
- Zou, H., Shi, M., Li, R., Zhang, X., Lu, W., 2021. Comparative transcriptome and methylome analysis of the hindbrain in olive flounder (*Paralichthys olivaceus*) considering individual behavior-type and oxygen metabolism. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics. 38, 100799. <https://doi.org/10.1016/j.cbd.2021.100799>
- Zuo, Z., Shen, J.X., Pan, Y., Pu, J., Li, Y.-G., Shao, X.H., Wang, W.P., 2018. Weighted gene correlation network analysis (WGCNA) detected loss of MAGI2 promotes chronic kidney disease (CKD) by podocyte damage. Cell Physiology Biochemistry. 51(1), 244–261. <https://doi.org/10.1159/000495205>

Annex



Characterization of the European Sea Bass (*Dicentrarchus labrax*) Gonadal Transcriptome During Sexual Development

L. Ribas¹ · B. Crespo^{2,3} · N. Sánchez-Baizán¹ · D. Xavier⁴ · H. Kuhl^{5,6} · J. M. Rodríguez^{7,8} · N. Díaz^{1,9} · S. Boltaña^{10,11} · S. MacKenzie^{10,12} · F. Morán⁴ · S. Zanuy² · A. Gómez² · F. Piferrer¹

Received: 17 December 2018 / Accepted: 18 February 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

The European sea bass is one of the most important cultured fish in Europe and has a marked sexual growth dimorphism in favor of females. It is a gonochoristic species with polygenic sex determination, where a combination between still undifferentiated genetic factors and environmental temperature determines sex ratios. The molecular mechanisms responsible for gonadal sex differentiation are still unknown. Here, we sampled fish during the gonadal developmental period (110 to 350 days post fertilization, dpf), and performed a comprehensive transcriptomic study by using a species-specific microarray. This analysis uncovered sex-specific gonadal transcriptomic profiles at each stage of development, identifying larger number of differentially expressed genes in ovaries when compared to testis. The expression patterns of 54 reproduction-related genes were analyzed. We found that *hsd17b10* is a reliable marker of early ovarian differentiation. Further, three genes, *pdgfb*, *snx1*, and *nfy*, not previously related to fish sex differentiation, were tightly associated with testis development in the sea bass. Regarding signaling pathways, lysine degradation, bladder cancer, and NOD-like receptor signaling were enriched for ovarian development while eight pathways including basal transcription factors and steroid biosynthesis were enriched for testis development. Analysis of the transcription factor abundance showed an earlier increase in females than in males. Our results show that, although many players in the sex differentiation pathways are conserved among species, there are peculiarities in gene expression worth exploring. The genes identified in this study illustrate the diversity of players involved in fish sex differentiation and can become potential biomarkers for the management of sex ratios in the European sea bass and perhaps other cultured species.

Keywords Genomics · Transcriptomics · Reproduction · Sex differentiation · Gonads · Aquaculture

L. Ribas and B. Crespo contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10126-019-09886-x>) contains supplementary material, which is available to authorized users.

✉ A. Gómez
a.gomez@csic.es

✉ F. Piferrer
piferrer@icm.csic.es

¹ Institute of Marine Sciences (ICM), Spanish National Research Council (CSIC), Barcelona, Catalonia, Spain

² Institute of Aquaculture of Torre de la Sal (IATS-CSIC), Ribera de Cabanes s/n. Torre la Sal, 12595 Castellón, Spain

³ Present address: UCL GOS Institute of Child Health, University College London, London, UK

⁴ Department of Biochemistry and Molecular Biology I, Complutense University, Madrid, Spain

⁵ Max Planck Institute for Molecular Genetics, Berlin, Germany

⁶ Present address: Department of Ecophysiology and Aquaculture, Leibniz Institute for Freshwater Ecology and Inland Fisheries, Berlin, Germany

⁷ Spanish National Bioinformatics Institute, Madrid, Spain

⁸ Present address: Cardiovascular Proteomics Laboratory, Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain

⁹ Present address: Max Planck Institute for Molecular Biomedicine, Muenster, Germany

¹⁰ Autonomous University of Barcelona, Barcelona, Spain

¹¹ Present address: Interdisciplinary Center for Aquaculture Research (INCAR), Department of Oceanography, Biotechnology Center, University of Concepción, Concepción, Chile

¹² Present address: Institute of Aquaculture, University of Stirling, Stirling, Scotland, UK

Introduction

Many fish species exhibit sexual dimorphic growth where one sex, in many cases the females, grows more than the other. Mixed sex rearing constitutes a problem for the aquaculture industry because it means producing at suboptimal capacity when compared to monosex culture. In addition, farming conditions result in many cases in masculinization, which further aggravates the problem if the desired sex is the females. Thus, the control of sex ratios is of major importance for many farmed species (Piferrer 2001; Budd et al. 2015; Wang et al. 2019). Deciphering the molecular mechanisms involved in gonadal development and the establishment of population sex ratios has then not only basic interest but also is of practical importance for modern fish farming.

Transcriptomic studies on gonadal tissues have been conducted in a relatively large number of fish species, both cultured and not. For example, in the channel catfish (*Ictalurus punctatus*), the genes differentially expressed (DEG) were described during testis development (Zeng et al. 2016). In fugu (*Takifugu rubripes*) and in the spotted knifejaw (*Oplegnathus punctatus*), a larger number of DEG were found in adult testes when compared to ovaries (Du et al. 2017; Wang et al. 2017). However, the number of studies focusing on sex-related differences precisely during sexual development and with a genomics approach is much more limited, and includes those carried out in Nile tilapia (*Oreochromis niloticus*) (Tao et al. 2013), turbot (*Scophthalmus maximus*) (Ribas et al. 2016), zebrafish (*Danio rerio*) (Ribas et al. 2017), and yellow river carp (*Cyprinus carpio*) (Jia et al. 2018).

The European sea bass (*Dicentrarchus labrax*) stands as the third most important cultured species of marine fish in Europe with a production of ~180 thousand tons per year (Food and Agriculture Organization of the United Nations 2005). Its production has benefited from advances in reproduction and the implementation of breeding programs (reviewed in Felip and Piferrer 2018; Wang et al. 2019). In this species, females grow about 30% more than males (Saillant et al. 2001). However, temperature during early development can affect sex ratios, favoring a higher number of males in the populations (reviewed in Vandeputte and Piferrer 2019). This problem can be even more aggravated if males undergo precocious maturation as it slows down their growth (Carrillo et al. 2015). Thus, there is interest in producing monosex female stocks.

The European sea bass is one of the richest species in terms of genomic resources among cultured fish, which include the availability of the genome and single nucleotide polymorphism (SNP) markers. The European sea bass has a small genome size (675 Mb) with a total of 26,719 annotated genes (Tine et al. 2014; Chaves-Pozo et al. 2017) and has 24 haploid chromosome pairs (Aref'yev 1989). It is a gonochoristic species with a polygenic sex determination system with both

genetic and environmental influences (Piferrer et al. 2005; Vandeputte et al. 2007; Vandeputte and Piferrer 2019). Although the genetic factors are still not known, efforts have been done towards identifying sex-determinant markers to aid in genetic selection programs. Studies using ~6700 SNP markers showed that there are at least three loci linked to sex (Palaiokostas et al. 2015) but these may be family-specific. Studies on the effects of food supply during early development evidenced changes in energy balance during testicular development (Díaz et al. 2014). In a similar manner, transcriptomic profiles of differentiating gonads subjected to different temperatures or estrogen treatments identified the involvement of steroidogenic- and epigenetic-related genes (Díaz and Piferrer 2015, 2017). Recently, the epigenetic regulation of key sex-related genes has been reported and a method based on the analysis of epimutations to predict sex in the European sea bass devised (Anastasiadi et al. 2018). Nevertheless, a study of the transcriptomic changes that occur during sexual development had never been carried out.

The objective of this study was to fill this gap. To do so, we transcriptomically analyzed sexually undifferentiated, differentiating, and differentiated gonads. Gene expression levels were evaluated by using a homologous custom-made microarray enriched with reproduction-related genes. Studying transcriptomes by using a microarray platform provides accuracy and reproducibility of the performed analysis and allows studying a broad range, if not all, of the transcripts of the genome (Shi et al. 2006). In particular, our custom microarray contained 78.5% of the annotated coding genes of the European sea bass genome, thus providing a powerful molecular tool to study gene expression patterns of this species.

Materials and Methods

European Sea Bass Gonad Sampling

In order to obtain the widest possible range of expressed transcripts, gonads were dissected from fish (5.2 ± 0.5 cm of standard length, SD) at 110 days post fertilization (dpf), when they are still morphologically undifferentiated but can be sexed by analyzing the expression of sex markers (i.e., *cyp19a1a* (Blázquez et al. 2008), when differentiating at 250 dpf (12.7 ± 5.7 cm of SD and 11.2 ± 0.6 cm of SD in females and males, respectively) and when differentiated at 350 dpf (16 ± 1.3 cm of SD and 14.8 ± 1.1 cm of SD, in females and males, respectively). The set of samples consisted of 11 gonads at 110 dpf, 12 gonads (six testes and six ovaries) at 250 dpf, and 18 gonads (nine testes and nine ovaries) at 350 dpf. When possible, gonads were fully isolated from fish at 250 and 350 dpf and thus gonadal tissue was devoid of any other tissue. However, a clean gonad isolation was not always feasible in

younger fish (110 dpf) due to their extremely small size, and thus some epithelial contamination could not be ruled out.

Microarray Platform

The microarray platform used in the present study consists of 1417 Agilent control probes and a total of 43,803 transcript probes that represent 20,978 genes of which 20,028 have two probe copies each while the rest have between 1 and 6 copies per gene (Supplementary Table S1). Genes with known reproduction-related functions had at least four copies. Microarray was based on sequences obtained from two 454 FLX Titanium runs on the European sea bass gonad tissues at different ages (from 40 dpf up to 6 years), a former custom European sea bass microarray platform (GPL13443) available in our laboratory (Díaz et al. 2014; Díaz and Piferrer 2015), a previously published European sea bass microarray (Ferraresso et al. 2010), and a battery of selected reproduction key genes. Only non-redundant and annotated sequences were selected. This microarray was submitted to the Gene Expression Omnibus (GEO) database (Edgar et al. 2002) with the platform number GPL16767 and its functionality was reported in a previous work using the European sea bass larvae (Schaeck et al. 2017). In the present work, this microarray has been re-annotated by using the European sea bass genome (Tine et al. 2014) and used to study the gonadal development in the European sea bass. Microarray data of the present study were submitted to GEO and are accessible through GEO series accession number GSE115841. For a complete list of gene names and abbreviations, see Dataset 1.

RNA Isolation and Microarray Hybridization

Total RNAs was extracted from 41 gonad samples using RNeasy Mini Kit (Qiagen, Germany) following the manufacturer's instructions. Quantity was determined by a Nanodrop spectrophotometer (Nanodrop Technologies, USA) and quality (RNA integrity number, RIN) measured with a Bioanalyzer (Agilent Technologies, USA). Only RNA samples with a mean RIN ≥ 8.4 were further processed for microarray analysis. RNA was labeled using the Low Input Quick Amp Labelling Kit, One-Color (Cy3; Agilent Technologies), and cRNA was hybridized overnight with the corresponding buffers during 17 h at 65 °C and washed on the following day. Samples were hybridized individually in the European sea bass custom 4x44K Agilent microarray described above at the Parc de Recerca Biomèdica de Barcelona (PRBB) and slides were scanned using an Agilent G2565B microarray scanner (Agilent Technologies, USA). Agilent software was used to avoid saturation and the extraction feature generated the raw data for further pre-processing.

Data Analysis

Statistical analyses were performed with R software (2.13.1 version; www.R-project.org). Array normalization was implemented using the Quantile method in the Linear Models for Microarray Analysis (Limma) R package (Wettenhall and Smyth 2004; Ritchie et al. 2015). Potential batch effects were removed by ComBat correction, a bioinformatic tool based on empirical Bayes algorithms (Chen et al. 2011). Data visualization, principal component analysis of the variance, and identification of clusters and outliers (two samples at 350 dpf were detected as outliers and excluded from further analysis) were performed using R. TIGR Multiexperiment Viewer version 4.9 (TMeV) software (Saeed et al. 2003) was used to determine the number of differentially expressed genes (DEG) between sexes at a given stage of development or between stages of development within the same sex. Significance was assessed by significant analysis of microarrays (SAM) statistical test with a false discovery rate (FDR) and adjusted *P* values < 0.01 and < 0.001 were applied to identify genes with statistically significant differences in expression. The above-mentioned analysis generated lists of DEG at each stage (110, 250, and 350 dpf) in the same sex or in comparison to the other sex, including the $\log_2$ transformation of fluorescence intensity measured for each gene.

Gene Ontology Terms and KEGG Pathway Analysis

The over-represented Gene Ontology (GO) functional categories of the DEG between females and males at each stage were obtained by GO term enrichment analysis using GO.db and topGO packages from the Bioconductor Project (Gentleman et al. 2004; Alexa and Rahnenfuhrer 2016; Carlson 2017) in R software (R Core 2017). The graphs and heatmaps were produced using gplots and ggplot2 packages (Wickham 2009; Warnes et al. 2016).

We used the Gene Set Variation Analysis (GSVA) from Pathway Processor 2.0 to study the signaling pathways involved in gonadal development. GSVA transforms the gene expression values into a normalized expression matrix with enrichment scores of differentially regulated pathways (DRP) with the corrected *P* value between males and females at each developmental stage (Beltrame et al. 2013).

Four pathways involved in sex differentiation: fanconi anemia and wnt signaling pathways, associated with female differentiation (Rodríguez-Marí and Postlethwait 2011; Sreenivasan et al. 2014), p53 signaling and cytokine-cytokine interaction receptor pathways, associated with male differentiation (Yasuda et al. 2012; Ribas et al. 2017), were also studied. The lists of genes making-up these pathways were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG), using zebrafish as background. The numbers of DEG from these pathways as well as expression values

(in log₂ fold change, FC) were plotted together, upregulated in male vs. upregulated in female over time.

Transcription Factor Analysis

The transcription factors (TFs) present in the microarray were identified with the aid of the TF checkpoint database, a list of TFs compiled from nine databases (Chawla et al. 2013). Using this list as background, the percentage of TFs present in the DEG between sexes and differentially expressed at each age was calculated.

Validation of the Microarray

Microarray results were validated by quantitative real-time polymerase chain reaction (qPCR) analyzing the expression of 12 genes selected with a wide range of FC values and equal amount of upregulated and downregulated genes when ovaries and testes were compared. Two house-keeping genes were chosen as reference: elongation factor-1 alpha (*ef-1α*) and 40S ribosomal protein (*fau*) that were previously validated in the European sea bass (Mitter et al. 2009). One hundred nanograms of total RNA were reverse transcribed into cDNA using Superscript III (Invitrogen) and 100 ng of random hexamer primers (Sigma) following the manufacturer's instructions. The reaction was carried out with SYBR Green chemistry (Power SYBR Green PCR Master Mix; Applied Biosystems). qPCR reactions contained 1X SYBR green master mix (Applied Biosystems), 10 pmol of each primer and 1 µl of the RT reaction. Samples were run individually and in triplicate in optically clear 384-well plates in Applied Biosystems 7900 machine. Cycling parameters were as follows: 50 °C for 2 min, 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 1 min. Finally, a temperature-determining dissociation step was performed at 95 °C for 15 s, 60 °C for 15 s, and 95 °C for 15 s at the end of the amplification phase. qPCR data were collected by SDS 2.3 and RQ Manager 1.2 software and relative quantity (RQ) values for each reaction replicate were calculated by the 2^{ΔΔCT} method (Schmittgen and Livak 2008). Primer sequences used for gene expression study are shown in Supplementary Table S2.

Results

Microarray Platform and Validation

Hybridization repeatability and consistency of results was verified in 55 genes related to reproduction and previously selected in turbot (Ribas et al. 2016) and zebrafish (Ribas et al. 2017) to study sex differentiation and reproduction in fish (Supplementary Table S3). Probe copy tendency for 52 of these genes was the same for all the copies and only three of

them (representing ~5% of the 52 tested probes) showed different probe tendency in at least one of the copies. Some examples of upregulated (Supplementary Fig. S1 a, c, e, g), downregulated (Supplementary Fig. S1 b, d, f, h), or variable (i) gene expression are shown. Since most probes had two or four copies (Supplementary Fig. 1j) to further evaluate the hybridization accuracy, we determined the magnitude of variation between technical replicates. The mean of the standard deviations for all 52 probes ranged between 0.205 and 0.347 (Supplementary Fig. 1j). Thus, given the low standard error among probe copies of the same gene, the average FC value of all probe copies was used for each gene.

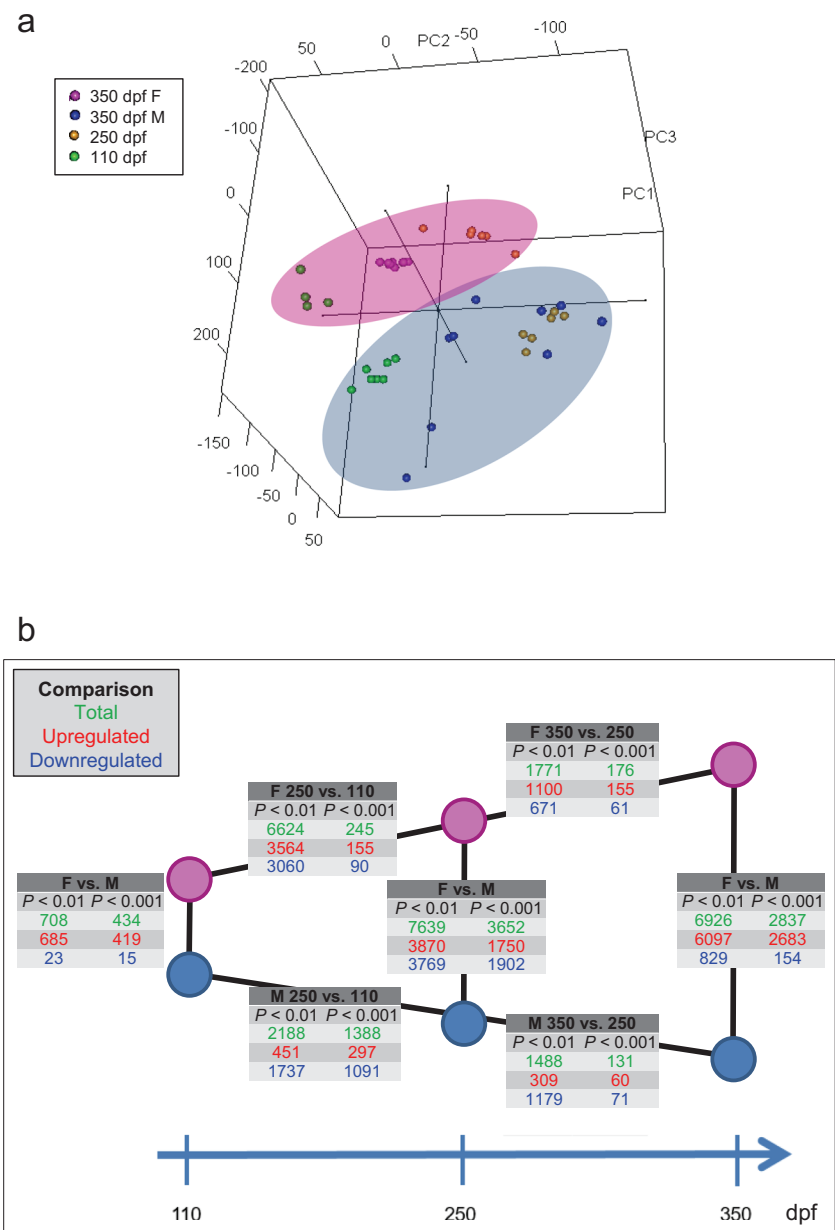
Microarray validation by qPCR for 12 DEGs showed a good correlation between the results obtained either by microarray or qPCR techniques ($R^2 = 0.748$, $P = 0.0003$; Supplementary Fig. S2a). Additionally, gene expression values of aromatase (*cyp19a1a*) from samples at different stages of gonadal development (110, 250, and 350 dpf) determined either by using microarray or by qPCR, further validated our results (Supplementary Fig. S2b).

Gonadal Transcriptomes Overview

Samples clustered in two distinctive groups according to sex and within each group samples tended to group according to age, except for males at 350 dpf, which showed more variation in the PCA (Fig. 1a). The component 1 of the PCA alone explained 87.14% of the variance, while components 2 and 3 contributed to 8.35% and 1.98% of the total variance, respectively. Thus, the first three components together explained 97.47% of variance.

A total of seven comparisons were analyzed with the SAM statistical test with an adjusted P value ≤ 0.01 : three between sexes, two within males, and two within females at the three different ages (Fig. 1b). Among the 20,978 genes included in the microarray, 64.93% were differentially expressed at one or several of these comparisons. The number of DEGs between ovaries and testes was 708, 7639 and 6926, at 110, 250 and 350 dpf, respectively (Fig. 1b, Dataset 1). A larger number of genes were upregulated in females when compared to M: 685 vs. 23 at 110 dpf; 3870 vs. 3769 at 250 dpf, and 6097 vs. 829 at 350 dpf. Between 110–250 and 250–350 dpf, the number of upregulated genes was 3564 and 1100 in females while 451 and 309 in males, respectively. The number of downregulated genes between 110–250 and 250–350 dpf was 3060 and 671 in females while 1737 and 1179 in males, respectively. Among the upregulated genes in the developing ovaries, higher FC values were found in 110 dpf and 350 dpf, while at 250 dpf, FC values were higher in testes when compared to ovaries (Supplementary Table S4). The highest FC values were observed at 250 dpf ($P < 0.001$) which was 7.5 and 6.7 Log₂ in males and females, respectively (Dataset 1), indicating

Fig. 1 Overview of transcriptomic changes during the European sea bass sex differentiation. **a** Principal component analysis of microarray results at three developmental stages: 110, 250, and 350 days post fertilization (dpf). Samples cluster together by gender: females (pink ellipse), males (blue ellipse). **b** Number of differentially expressed genes found along gonadal development (110, 250, and 350 dpf) for the seven comparisons analyzed using the SAM test



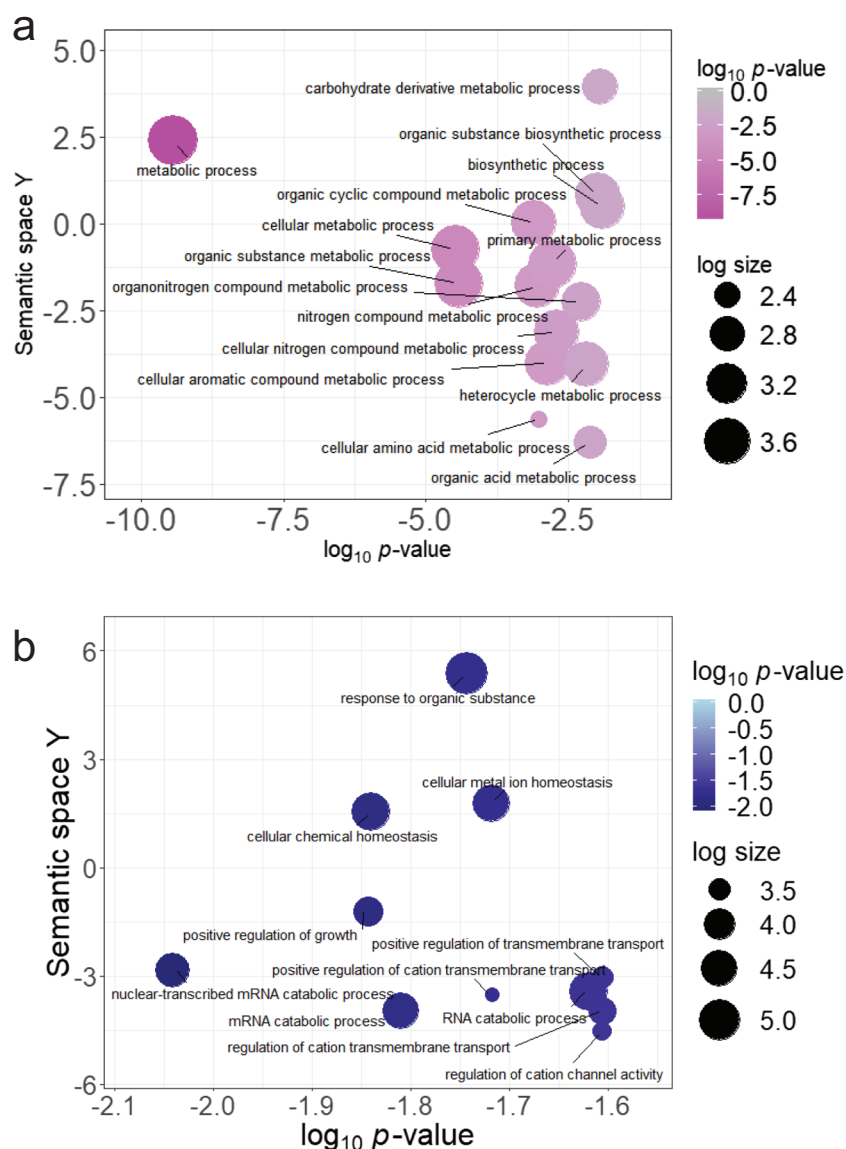
important sex-related differences in expression levels (Supplementary Table S4).

Gene Ontology and Gene Pathway Enrichments Along Gonadal Development

The GO term enrichment analysis of the DEGs between females and males revealed several categories related to biological processes (BP), molecular function (MF), and cellular component (CC) throughout development (Dataset 2). A total of 39 GO terms in the three categories were enriched during ovarian formation. The 15 GO terms significantly enriched in BP common at 110, 250, and 350 dpf ($P < 0.01$) are shown in Fig. 2a. The three developmental stages were enriched in GO terms related to

metabolic processes (GO:0008152, GO:0071704, GO:0044237), catalytic activity (GO:0003824), oxidoreductase activity (GO:0016491), coenzyme and cofactor binding (GO:0050662; GO:0048037), and biosynthetic processes (GO:0009058, GO:1901576). Among the GO terms enriched for testis formation, there were a total of 52 significantly enriched terms that were common at 250 and 350 dpf, but none at 110 dpf which were related to catabolic processes (GO:0000956, GO:0006402, GO:0006401), regulation of ion transmembrane activity (GO:1904427, GO:0032414, GO:0034767), regulation of calcium ion (GO:0010524, GO:0050850, GO:0051281, GO:0060316, GO:1901021, GO:1904427), and positive regulation of growth (GO:0045927). Figure 2 b shows the enriched GO terms found for testis formation in BP category ($P < 0.02$).

Fig. 2 Common biological processes Gene Ontology terms at 110, 250, and 350 days post fertilization (dpf) of differentially expressed genes during European sea bass sex differentiation. **a** Female-related genes ($P < 0.01$). **b** Male-related genes ($P < 0.02$)



Then, we determined the significantly differentially regulated cellular pathways (DRP) between males and females along gonadal development. A total of 41, 151, and 106 DRP were found between males and females at 110, 250, and 350, respectively (Dataset 3, $P < 0.05$). Some of these DRP were related to sex differentiation, for example, at early gonadal development (110 dpf): p53 signaling pathway, steroid hormone biosynthesis, or *erbB* signaling pathway; at 250 dpf: wnt signaling pathway, oocyte meiosis, or steroid biosynthesis; and at 350 dpf: MAPK signaling pathway or cytokine-cytokine receptor interaction among others. A total of 16 DPR were consistently differentially regulated at the three developmental stages in both sexes (Table 1). The lysine degradation, bladder cancer, and the nucleotide-binding oligomerization (NOD)-like receptor signaling pathways were upregulated in females at the three gonadal developmental stages when

compared to males. The dorso-ventral axis formation pathway was significantly downregulated in females at 110 dpf but upregulated at 250 and 350 dpf. The p53 signaling pathway and the Chagas disease (American trypanosomiasis) pathways showed significance, being upregulated at 110 and 350 dpf in females but downregulated at 250 dpf when compared to males. There were two pathways, phosphatidylinositol signaling system and the *ErbB* signaling pathway that were upregulated in females at 110 dpf but at 250 and 350 dpf in males. Finally, there were up to eight pathways upregulated in males when compared to females throughout the studied period: butirosin and neomycin biosynthesis, basal transcription factors, amino sugar and nucleotide sugar metabolism, type II diabetes mellitus, glycine, serine, and threonine metabolism, steroid biosynthesis and ribosome and folate biosynthesis pathways.

Table 1 Sixteen common differentially regulated pathways identified from the microarrays data along gonadal development ($P < 0.05$). Pink color indicates pathways upregulated in females, whereas blue indicates the upregulated pathways in males

Pathway description	110 dpf	250 dpf	350 dpf
Lysine degradation			
Bladder cancer			
NOD-like receptor signaling pathway			
Dorso-ventral axis formation			
p53 signaling pathway			
Chagas disease (American trypanosomiasis)			
Phosphatidylinositol signaling system			
ErbB signaling pathway			
Butirosin and neomycin biosynthesis			
Basal transcription factors			
Amino sugar and nucleotide sugar metabolism			
Type II diabetes mellitus			
Glycine, serine, and threonine metabolism			
Steroid biosynthesis			
Ribosome			
Folate biosynthesis			

Expression of Key Genes and Pathways Related to Sex Differentiation

Of the 54 key genes known to be relevant for reproduction and sex differentiation in fish according to the primary literature, 49 of them had sex-related significant differences in at least one of the three ages studied. The majority (80.5%) were DEG at 250 dpf (Supplementary Table S3). Of the 49 just mentioned above, 25 were pro-female and 24 pro-male genes. Hierarchical clustering analysis and the corresponding heatmaps of the 25 pro-female genes (Fig. 3a) and the 24 pro-male genes (Fig. 3b) showed that gene expression results mostly matched according to their phenotypic gender. The expression profiles of 12 key genes are shown in Fig. 4. Six of them are related to gonadal development (*cyp19a1a*, *hsd17b10*, *hsd3b*, *cyp11b*, *arα*, *fshr*, Fig. 4a–f) while the other six genes are TFs related to sex differentiation (*foxl2*, *sox3*, *figlα*, *nr5a1a*, *sox9b*, and *dmrt1*, Fig. 4g–l). The genes *cyp11b1*, *arα*, *hsd3b*, and *fshr* were upregulated in males when compared to females at 250 dpf and onwards while *hsd17b10* was upregulated in females already at 110 dpf. In all these genes, sex-specific significant differences in expression were observed at least in one of the three sampling ages. All genes except *hsd3b* had maximal sex-related expression differences at 250 dpf. Regarding the six canonical TFs, they were upregulated as expected according to sex: *foxl2*, *sox3*, and *figlα* in females while *nr5a1a*, *sox9b*, and *dmrt1* in males. All of them were differentially expressed between sexes at least in 250 dpf and some also at 350 dpf (*sox3*, *figlα*, *sfla*, and *sox9b*).

Next, we looked specifically at the four signaling pathways known to be associated with sex differentiation from previous studies (see [Materials and methods](#)). Among genes that constitute these signaling pathways, we looked at the number of

DEG and the magnitude of the gene expression values. The two selected pathways related to ovarian development, fanconi anemia (Fig. 5a, b) and wnt (Fig. 5c, d) signaling pathways, had a larger number of DEG and a higher gene expression (FC) values in females. Similarly, when looking among pathways related to testis development, p53 signaling pathway (Fig. 5e, f) and cytokine-cytokine interaction receptor pathway (Fig. 5g, h), they had a larger number of DEG and a higher gene expression (FC) values in males.

The Role of Transcription Factors During Gonadal Development

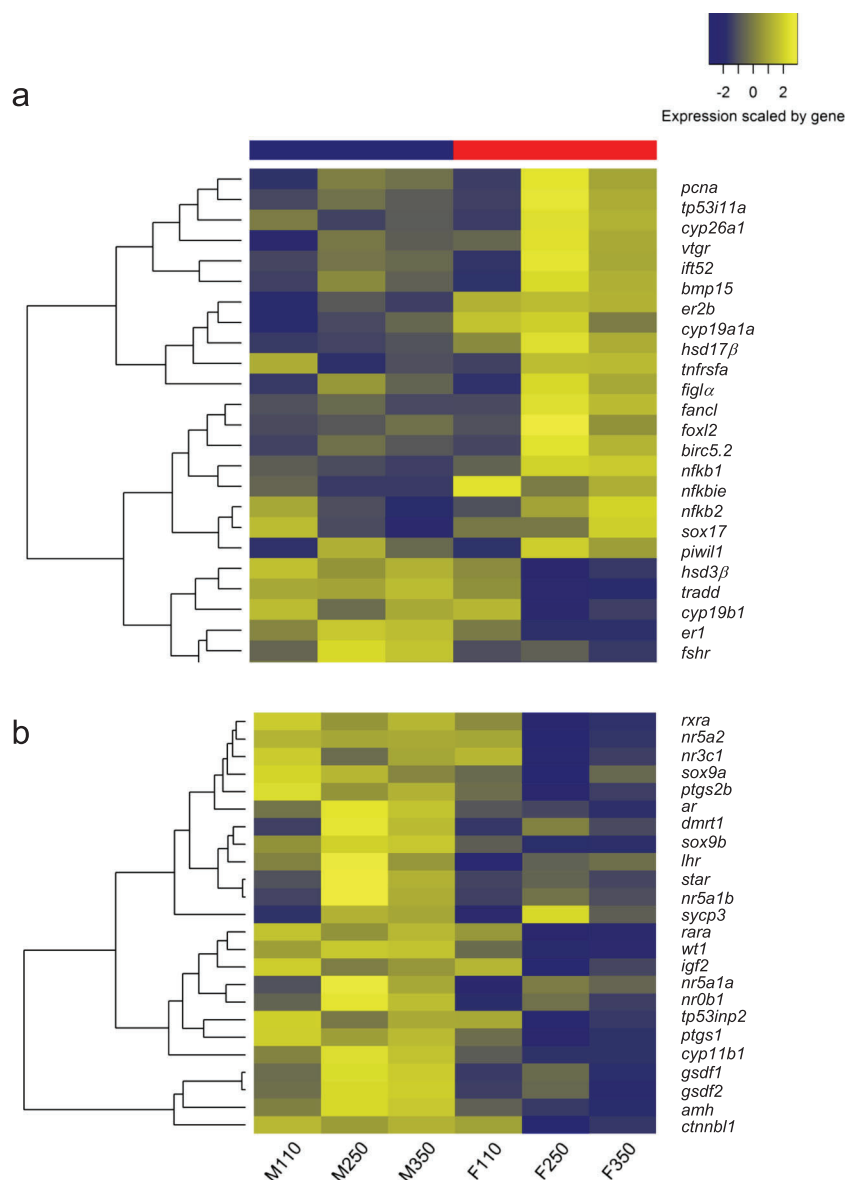
Our microarray included 2822 TFs in total, i.e., 13.5% of all the probes. The proportion of differentially expressed TFs was variable between sexes and across time (Fig. 6). At 110 dpf, 8.0% of the DEG were upregulated TFs in ovaries when compared to testes, while no TFs were upregulated in testes. At 250 dpf, the percentage of differentially upregulated TFs increased up to 13.0% in ovaries and 9.6% in testis. Up to 25.6% of the DEG at 350 dpf were identified as TFs; 14.5% upregulated in ovaries and 11.1% were upregulated in testis.

Discussion

Robustness of the Microarray

This study provides a comprehensive transcriptomic analysis of gonad differentiation in the European sea bass using a custom species-specific microarray (Schaeck et al. 2017) that has been here further validated. First, the microarray was completely re-annotated and includes almost 80% of the genes identified in the

Fig. 3 Heatmap of the microarray expression data for 49 out of 5 reproduction-related key genes: **a** 25 upregulated and **b** 24 down-regulated genes in females. Each row represents a gene and each column represents a group of fish by age and sex: M110 = 8 males at 110 days post fertilization (dpf), M250 = 6 males at 250 dpf, and M350 = 9 males at 350 dpf; F110 = 4 females at 110 dpf; F250 = 6 females at 250 dpf and F350 = 9 females at 350 dpf. The key color represents the level of expression scaled by gene (yellow: high expression and blue: low expression). The dendrograms inform of the similarity between genes and between the different samples. Notice that most genes were grouped as pro-female and pro-male as expected from studies in other species. See Dataset 1 for a complete list of gene names and abbreviations

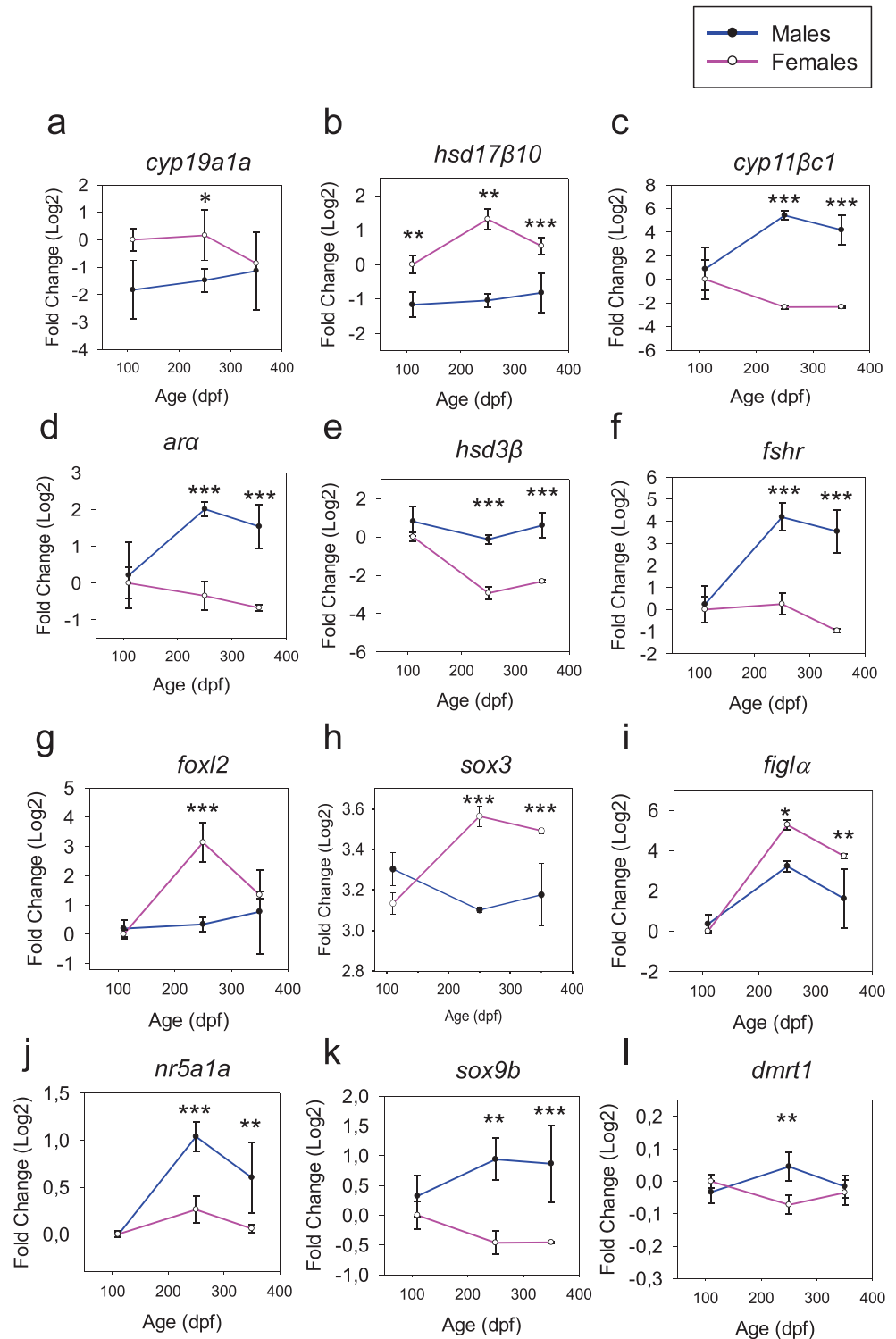


European sea bass genome. Quality control showed that it had high reproducibility and accuracy. Transcript expression values were very robust as the standard deviations of probe replicates were very low (average 0.276 for 20,029 of the duplicated probes), confirming the high reproducibility of RNA analysis using the Agilent oligo-array (Shi et al. 2006). To date, microarray analyses have been very useful in the study of fish transcriptomes, (e.g., Millan et al. 2010; Jantzen et al. 2011; Tingaud-Sequeira et al. 2013; Schaeck et al. 2017). Although in the last years RNA sequencing have gained favor over array platforms, analysis of the same samples with the two different techniques gives similar results (Zhao et al. 2014). Here, we provide a validated, fast, and cost-effective tool for aquaculture research to study the expression patterns of genes, including all major reproduction-related genes, in the European sea bass (Schaeck et al. 2017).

Transcriptomic Differences Between Females and Males During Gonadal Development

PCA classified individuals in well-defined and separated clusters according to sex and stage. At 110 dpf, when gonads were still histologically undifferentiated, transcriptomic analysis was already capable of classifying samples according to phenotypic sex. Statistical analysis showed that the highest number of DEG was found in differentiated females when compared to differentiated males, in particular at 250 dpf, probably explained by the fact that sex differentiation in the European sea bass starts earlier in females (Piferrer et al. 2005) as in many other fish species (Piferrer 2001; Devlin and Nagahama 2002; Wang et al. 2019). This contrasts with results found in other fish species such as zebrafish (Small et al. 2009), tilapia (Tao et al. 2013), or turbot (Ribas et al. 2016),

Fig. 4 Changes in expression of a set of key genes related to sex differentiation and reproduction in female and male gonads during European sea bass sex differentiation. **a–f** Key genes related to sex differentiation. **g–l** Key transcription factors. Abbreviations: dpf, days post fertilization; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$



where male-related genes were enriched with respect to female-related genes. Testis development implied downregulation of genes in a certain stage when compared to the previous developmental stages. This tendency was also observed in turbot (Ribas et al. 2016), supporting the importance of active gene repression for testis development. This is in accordance

with the current view stating that positive and negative regulatory loops are required for sex differentiation in vertebrates (Munger et al. 2013; Capel 2017).

There were 15 enriched GO terms in upregulated genes during ovarian development (from 110 dpf to 350 dpf), all of them related to metabolic functions. In this regard, the

metabolic process category was the most enriched one found in the differentiating ovary of the protogynous ricefield eel (*Monopterus albus*) (Cai et al. 2017) and was described as well for ovarian development in turbot (Ribas et al. 2016). In contrast, we did not find any common GO term in the three developmental stages studied in testis. Nevertheless, 30 common GO terms were found enriched between 250 and 350 dpf in developing males and were related to catabolic processes, regulation, and positive regulation of growth, among others.

Three pathways (i.e., lysine degradation, bladder cancer, and NOD-like receptor signaling) showed a consistent upregulation in the ovaries in all stages when compared to testes. Lysine is an essential amino acid required for protein synthesis, enzyme catalysis, and L-carnitine biosynthesis and thus essential for energy metabolism in all body tissues, including ovaries (Ramseyer and Garling 1994; Hallen et al. 2013). The bladder cancer pathway includes genes involved in gonadal development such as the tumor protein tp53 (Mitra et al. 2006), a gene present in germ cells and that induces apoptosis and atresia in oocytes (Rodríguez-Marí et al. 2010; Sayed et al. 2018). The NOD-like receptor signaling pathway is activated in response to host defense and inflammatory disease response (Caruso et al. 2014) and it is upregulated in human polycystic ovaries (Wang et al. 2014). On the other hand, during testicular development, eight pathways were identified, including pathways related to metabolism (e.g., butirosin and neomycin biosynthesis, amino sugar, and nucleotide sugar metabolism), to genetic information processing (e.g., basal transcription factors, type II diabetes mellitus ribosome) and to steroid biosynthesis. A sex-biased representation of these molecular pathways might also be species-specific. Thus, for example, in Japanese flounder (*Paralichthys olivaceus*), the upregulation of metabolic-related pathways was found in ovaries rather than in testes (Fan et al. 2014). Also, in the tilapia, steroidogenic pathways were more expressed in females than in males, particularly at early stages of development (Tao et al. 2013).

Sex-Biased Expression of Genes During Sex Differentiation

We selected 54 pro-female or pro-male genes to study their expression along gonadal development. We found 49 DEGs whose expression matched the expected sex bias described in previous studies of other fish species (see [Materials and Methods](#)). However, six of the genes upregulated in male European sea bass had been previously described as upregulated in female zebrafish (Ribas et al. 2017), namely *hsd3 β* , *cyp19b1*, *tradd*, *er1*, *fshr*, and *er2b*.

In females, the steroidogenic gene *hsd17 β 10* was upregulated at 110 dpf. The expression of *hsd17 β* , *cyp19a1a*, and *star* was downregulated in the European sea bass at 170 dpf in fish previously exposed to high temperature, due to the

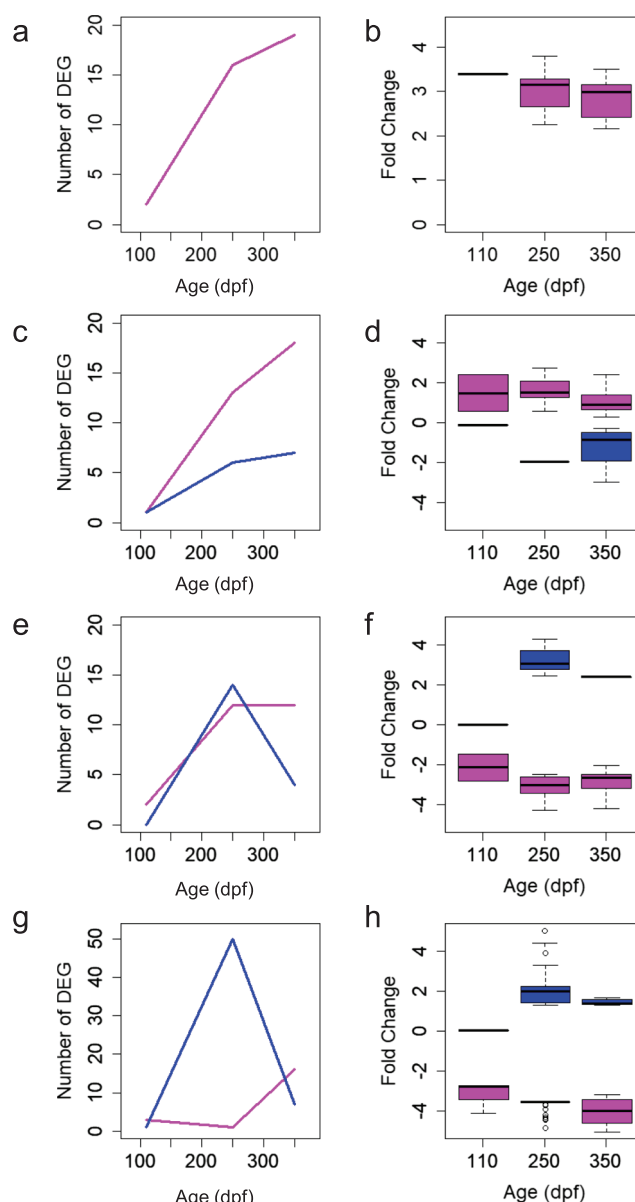


Fig. 5 Number of differentially expressed genes (DEG; left panels) and fold change of reproduction-related pathways (right panels) during European sea bass sex differentiation (pink, females; blue, males): **a, b** Fanconi anemia signaling pathway, **c, d** Wnt signaling pathway, **e, f** p53 signaling pathway, **g, h** Cytokine-cytokine interaction receptor signaling pathway. In the left panels data is expressed as the total number of genes (absolute values) differentially expressed at each time of development, in the right panels data is expressed as fold change using male values at 110 dpf as control group set at 0

masculinization of the ovary (Díaz and Piferrer 2017). The *cyp19a1a* gene, a key enzyme responsible for converting androgens into estrogens (Guiguen et al. 2009), is considered an early marker of ovarian differentiation in several fish species, including sea bass (Blázquez et al. 2008) and also Atlantic halibut (*Hippoglossus hippoglossus*) (Matsuoka et al. 2006) and turbot (Ribas et al. 2016). In the present study, *cyp19a1a* showed differential expression at 110 dpf but differences were

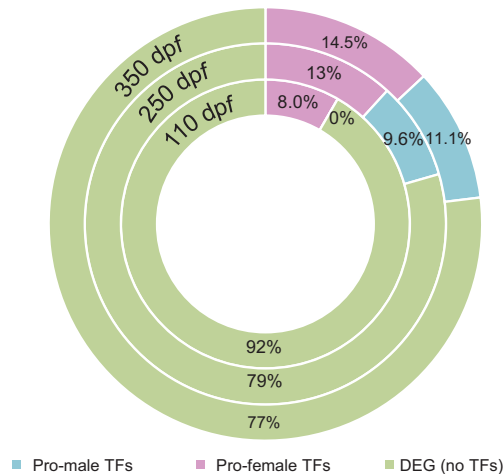


Fig. 6 Diagram showing at three ages during the European sea bass sex differentiation the percentage of pro-male and pro-female transcription factors (TFs) differentially expressed at each developmental stage of female-related genes, in pink or male-related genes, in blue

not significant until 250 dpf. *Hsd17β10* is a mitochondrial enzyme involved in multiple cellular functions, which include fatty acid oxidation, amino acid degradation, and steroid metabolism (Yang et al. 2007; Zschocke 2012). In humans, *hsd17β10* is related to neurodegenerative diseases such as Parkinson or Alzheimer and has been fully documented (Zschocke 2012; Yang et al. 2014). In fish, there is a lack of information about this isoform as it has been only described in amphioxus (*Branchiostoma belcheri*) (Zhang et al. 2008) and zebrafish (He and Yang 2009), but no biological functions have been yet ascribed. However, information does exist for other genes of the same family such as *hsd17β1* and *hsd17β3*, which are involved in sex steroid biosynthesis: *hsd17β1* is responsible to convert inactive estrone to active estradiol and leads to female sex differentiation, while *hsd17β3* is required for 11-ketotestosterone synthesis (Tokarz et al. 2015). *Hsd17β1* has been identified in some fish species such as Nile tilapia (Zhou et al. 2005), Atlantic cod (*Gadus morhua*) (Breton and Berlinsky 2014), and olive flounder (Fan et al. 2014), while *hsd17β3* only in zebrafish and in medaka (*Oryzias latipes*) (Mindnich et al. 2004; Kim et al. 2014). *Hsd17β1* was already detected at early stages of development in pre-differentiated fathead minnow (*Pimephales promelas*) embryos, although its expression was not correlated to any sex in particular (Wood et al. 2015). Recently, *hsd17β1* has been suggested as the sex-determining gene in the California yellowtail (*Seriola dorsalis*), which has a putative ZW sex determination system (Purcell et al. 2018). In our data, we did not find any differential expression for *hsd17β1* and *hsd17β3* genes, but we found for *hsd17β10*. To our knowledge, this is the first time that *hsd17β10* is described in the European sea bass and it is detected early in the ovaries. Therefore, it is a candidate to be considered as a novel early ovarian marker in

this species, although research on its functional role during ovarian differentiation needs further attention.

In males, the first signs of sex-biased expression of canonical reproduction-related genes were detected at 250 dpf onwards and not earlier, indicating that their expression starts somewhere between 110 and 250 dpf as previously showed in fish subjected to high temperature treatments at 170 dpf (Díaz and Piferrer 2015). In the present study, genes involved in the steroidogenic pathway such as *hsd3β* or *cyp11β1*, in androgen action such as *arα*, or in gonadotropin signaling, *fshr* and *lhr*, were differentially expressed during testis development, as previously reported (Blázquez and Piferrer 2005; Mazón et al. 2014). In contrast, the gonadotropin subunits *fshb* and *lhb*, detected in the gonads of the sea bass in this study, as also described in other fish species (Wong and Zohar 2004; von Schalburg et al. 2005; Levavi-Sivan et al. 2010), were not differentially expressed.

When looking for DEG ($P < 0.01$) in testes compared to ovaries at 110 dpf, a total of 15 genes were found, although none could be considered as canonical reproduction-related genes. Among them, we identified three genes that were previously described to be expressed in the reproductive system: platelet-derived growth factor beta polypeptide (*pdgfb*), sorting nexin 1 (*snx1*), and nuclear transcription factor Y beta (*nfy*), although there are few data on the role of these genes in testis, not only in fish, but also in mammals. For example, *pdgfb* is involved in the regulation of many biological processes including embryonic development and sexual phenotype, since alteration of this gene generated male and female infertility in several species, including humans (Donnem et al. 2010). *Snx* is involved in cellular endocytosis functions and its role in oogenesis was described in the gibel carp (*Carassius gibelio*) (Wen et al. 2003) while *nfy* is a pleiotropic transcription factor that participates in multiple processes such as cell proliferation and development (Li et al. 2018). For example, it has been detected in breast cancer cells (Lagadec et al. 2014) and it is involved in spermatogenesis (Vanwert et al. 2008), but its specific role in transcriptional regulation is not fully understood although several functional studies have been published so far. It is known that it can bind to the *piwil1* promoter in the germ cells (Chang et al. 2015), but also binds, together with the orphan nuclear receptor steroidogenic factor-1 (Nr5a1), to the promoter of *fshb* gene (Jacobs et al. 2003).

Next, we looked for canonical KEGG pathways involved in sex differentiation and previously described in some but few fish species. This was the case of two pathways required for ovarian development: the fanconi anemia pathway, identified in zebrafish (Rodríguez-Marí and Postlethwait 2011) and in common carp (*Cyprinus carpio*) (Jia et al. 2018), and the wnt signaling pathway, identified in zebrafish (Sreenivasan et al. 2014) and in rainbow trout (*Oncorhynchus mykiss*) (Nicol and Guiguen 2011). In the European sea bass, we found that the number of genes differentially expressed ascribed to these two pathways increased at 250 dpf and

onwards in the developing ovaries. In testes, we studied the apoptotic pathway p53 previously described in zebrafish (Rodríguez-Marí et al. 2010), medaka (Yasuda et al. 2012), and spotted knifejaw testes (*Oplegnathus punctatus*) (Du et al. 2017), and the cytokine-cytokine interaction pathway identified in Japanese flounder (Zhang et al. 2015) and in zebrafish (Ribas et al. 2017) gonads. In the European sea bass, we found an increase in the number of genes differentially expressed and associated with these pathways at 250 dpf that then decreased. Thus, these results confirm that, as occurs in other fish species, these four pathways are also involved in gonad development in the European sea bass.

Transcription Factors During Gonadal Development

TFs tightly control gene expression in a large number of processes including gonadal development (Migeon and Wisniewski 2000) and so, in the last years, many studies have revealed their importance in fish sex differentiation (Herpin and Scharlt 2011; Nakamura et al. 2011; Shen and Wang 2014; Tanaka 2016). Consequently, with the aim of deciphering the involvement of TFs in gonadal development in the European sea bass, we studied the expression of TFs already known to be sexually dimorphic. These included *foxl2* (Yamaguchi et al. 2007) and *figla* (Kanamori et al. 2008), related to ovarian development, and *sox9b* (Bagheri-Fam et al. 2010), *nr5a1a* (Crespo et al. 2013), and *dmrt1* (Deloffre et al. 2009), related to testis development, the expression of some being in accordance with what was previously described in the European sea bass gonads fish at 170 dpf and subjected to high temperatures (Díaz and Piferrer 2015). *Sox3* was considered as a male-determining gene in ricefish (*Oryzias latipes*) (Takehana et al. 2014) although its expression was related to both oocyte and testis development in other fish species. This is the case found in grouper (*Epinephelus coioides*) (Yao et al. 2007) and in Japanese flounder (Jeng et al. 2018) with an expression bias towards female development. In the present study, *sox3* clearly showed a female bias in the European sea bass gonads.

Then, we explored the presence of TFs at each specific stage of gonadal development. The number of DE TFs increased as the gonadal development progressed. Thus, at 350 dpf, when gonads were fully differentiated, the largest number of DE TFs was detected. In all stages, there was a larger number of DE TFs in females than in males that is in concordance with the larger number of DEGs found in females in this study. This skewed number towards females was evident already at 110 dpf. DE TFs were only detected in ovaries, probably due to sex differentiation starting earlier in females (Piferrer et al. 2005) and to the increased activity of the tissue by ovary formation and meiotic division actions (D'Cotta et al. 2001).

Conclusions

A species-specific microarray enriched for reproduction-related genes was used to study gene expression during the European sea bass gonadal development. In contrast to what had been described in other species, a larger number of DEG and DE TFs were observed in ovaries when compared to testis. The expression profiles of 54 genes previously associated to sex differentiation in other species were examined and the steroidogenic gene *hsd17b10* is described as a promising ovarian marker capable of identifying females as early as 110 dpf. Also, three genes: *pdgfb*, *snx1*, and *nfy*, were identified as potential markers for male development. Further, three and eight pathways that are consistently enriched along gonadal development in ovary or testis, respectively, were also identified. Taken together, these results contribute to our understanding of gene expression during sexual development in an economically important species in particular and in non-mammalian vertebrates in general, and emphasize the great diversity, also at the molecular level, of fish sexual development.

Acknowledgments We would like to thank Dr. Richard Reinhardt for his work on the 454 sequencing of the European sea bass gonads performed at Max Planck Institute for Molecular Genetics, Berlin (Germany).

Funding Information This work was supported by Spanish Government Consolider-Ingenio 2010 grant “Aquagenomics” (ref. CDS2007-0002) to SMK, FM, SZ and FP. BC was supported by an Aquagenomics contract; LR was supported by Aquagenomics postdoctoral contract; ND was supported by Aquagenomics predoctoral contract. NSB was supported by AGL2016-78710-R.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no competing interests.

References

- Alexa A, Rahnenführer J (2016) topGO: enrichment analysis for gene ontology. R package version 2.30.0. Available from: <https://bioconductor.org/packages/release/bioc/html/topGO.html>. Accessed 25 Apr 2018.
- Anastasiadi D, Vandeputte M, Sanchez-Baizan N, Allal F, Piferrer F (2018) Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a fish with mixed genetic and environmental sex determination. *Epigenetics* 13(9):988–1011
- Aref'yev VA (1989) Cytogenetic analysis and nuclear organization of the sea bass *Dicentrarchus labrax*. *J Ichthyol* 29:1–12
- Bagheri-Fam S, Sinclair AH, Koopman P, Harley VR (2010) Conserved regulatory modules in the Sox9 testis-specific enhancer predict roles for SOX, TCF/LEF, Forkhead, DMRT, and GATA proteins in vertebrate sex determination. *Int J Biochem Cell Biol* 42(3):472–477
- Beltrame L, Bianco L, Fontana P, Cavalieri D (2013) Pathway processor 2.0: a web resource for pathway-based analysis of high-throughput data. *Bioinformatics* 29(14):1825–1826

- Blázquez M, Piferrer F (2005) Sea bass (*Dicentrarchus labrax*) androgen receptor: cDNA cloning, tissue-specific expression, and mRNA levels during early development and sex differentiation. *Mol Cell Endocrinol* 237(1–2):37–48
- Blázquez M, González A, Papadaki M, Mylonas C, Piferrer F (2008) Sex-related changes in estrogen receptors and aromatase gene expression and enzymatic activity during early development and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *Gen Comp Endocrinol* 158(1):95–101
- Breton TS, Berlinsky DL (2014) Characterizing ovarian gene expression during oocyte growth in Atlantic cod (*Gadus morhua*). *Comp Biochem Physiol Part D Genomics Proteomics* 9:1–10
- Budd A, Banh Q, Domingos J, Jerry D (2015) Sex control in fish: approaches, challenges and opportunities for aquaculture. *J Mar Sci Eng* 3(2):329–355
- Cai J, Yang W, Chen D, Zhang Y, He Z, Zhang W, Zhang L (2017) Transcriptomic analysis of the differentiating ovary of the protogynous ricefield eel *Monopterus albus*. *BMC Genomics* 18(1):573
- Capel B (2017) Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat Rev Genet* 18(11):675–689
- Carlson M (2017) GO.db: a set of annotation maps describing the entire Gene Ontology. R package version 3.5.0. Available from: <https://bioconductor.org/packages/release/data/annotation/html/GO.db.html>. Accessed 7 Jul 2018
- Carrillo M, Espigares F, Felip A, Escobar S, Moles G, Rodríguez R, Victoria Alvarado M, Gómez A, Zanuy S (2015) Updating control of puberty in male European sea bass: a holistic approach. *Gen Comp Endocrinol* 221:42–53
- Caruso R, Warner N, Inohara N, Nunez G (2014) NOD1 and NOD2: signaling, host defense, and inflammatory disease. *Immunity* 41(6):898–908
- Chang GB, Chen R, Xu L, Ma T, Wang HZ, Chen J, Zhang Y, Li ZT, Wan F, Guo XM, Xu Q, Zhao WM, Chen GH (2015) DNA methylation and NF-Y regulate Piwi1 expression during chicken spermatogenesis. *Anim Reprod Sci* 162:95–103
- Chaves-Pozo E, Valero Y, Esteve-Codina A, Gomez-Garrido J, Dabad M, Alioto T, Meseguer J, Esteban MA, Cuesta A (2017) Innate cell-mediated cytotoxic activity of European sea bass leucocytes against nodavirus-infected cells: a functional and RNA-seq study. *Sci Rep* 7(1):15396
- Chawla K, Tripathi S, Thommesen L, Laegreid A, Kuiper M (2013) TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics* 29(19):2519–2520
- Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, Liu C (2011) Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS One* 6(2):e17238
- Crespo B, Gómez A, Mazón MJ, Carrillo M, Zanuy S (2013) Isolation and characterization of Ff1 and Gsdf family genes in European sea bass and identification of early gonadal markers of precocious puberty in males. *Gen Comp Endocrinol* 191(0):155–167
- D’Cotta H, Fostier A, Guiguen Y, Govoroun M, Baroiller JF (2001) Search for genes involved in the temperature-induced gonadal sex differentiation in the tilapia, *Oreochromis niloticus*. *J Exp Zool* 290(6):574–585
- Deloffre L, Martins R, Mylonas C, Canario A (2009) Alternative transcripts of DMRT1 in the European sea bass: expression during gonadal differentiation. *Aquaculture* 293(1–2):89–99
- Devlin RH, Nagahama Y (2002) Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208(3–4):191–364
- Díaz N, Piferrer F (2015) Lasting effects of early exposure to temperature on the gonadal transcriptome at the time of sex differentiation in the European sea bass, a fish with mixed genetic and environmental sex determination. *BMC Genomics* 16:679
- Díaz N, Piferrer F (2017) Estrogen exposure overrides the masculinizing effect of elevated temperature by a downregulation of the key genes implicated in sexual differentiation in a fish with mixed genetic and environmental sex determination. *BMC Genomics* 18:973
- Díaz N, Ribas L, Piferrer F (2014) Effects of changes in food supply at the time of sex differentiation on the gonadal transcriptome of juvenile fish. Implications for natural and farmed populations. *PLoS One* 9(10):e111304
- Donnem T, Al-Saad S, Al-Shibli K, Busund LT, Bremnes RM (2010) Co-expression of PDGF-B and VEGFR-3 strongly correlates with lymph node metastasis and poor survival in non-small-cell lung cancer. *Ann Oncol* 21(2):223–231
- Du X, Wang B, Liu X, Liu X, He Y, Zhang Q, Wang X (2017) Comparative transcriptome analysis of ovary and testis reveals potential sex-related genes and pathways in spotted knifejaw *Oplegnathus punctatus*. *Gene* 637:203–210
- Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1):207–210
- Fan ZF, You F, Wang LJ, Weng SD, Wu ZH, Hu JW, Zou YX, Tan XG, Zhang PJ (2014) Gonadal transcriptome analysis of male and female Olive flounder (*Paralichthys olivaceus*). *Biomed Res Int* 2014:1–10
- Felip A, Piferrer F (2018) State of culture and breeding of European sea bass, *Dicentrarchus labrax* L. In: Liang, X.F., H.P Wang, H. Liu and R.W. Hardy, eds. *World Perch and Bass Culture: Innovation and Industrialization*. China Science Press, Beijing, pp. 332–351
- Ferraresso S, Milan M, Pellizzari C, Vitulo N, Reinhardt R, Canario AVM, Patarnello T, Bargelloni L (2010) Development of an oligo DNA microarray for the European sea bass and its application to expression profiling of jaw deformity. *BMC Genomics* 11:354
- © FAO 2005–2019. Cultured Aquatic Species Information Programme. *Dicentrarchus labrax*. Cultured Aquatic Species Information Programme. Text by Bagni, M. In: FAO Fisheries and Aquaculture Department [online]. Rome. Updated 18 February 2005. Available from: http://www.fao.org/fishery/culturedspecies/Dicentrarchus_labrax/en#tcNA00EA. Accessed 5 May 2017
- Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge YC, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang JH (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5(10):R80
- Guiguen Y, Fostier A, Piferrer F, Chang CF (2009) Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish. *Gen Comp Endocrinol* 165(3):352–366
- Hallen A, Jamie JF, Cooper AJ (2013) Lysine metabolism in mammalian brain: an update on the importance of recent discoveries. *Amino Acids* 45(6):1249–1272
- He X, Yang S (2009) Comments on ‘Significance of developmental expression of amphioxus *Branchiostoma belcheri* and zebrafish *Danio rerio* Hsd17b10 in biological and medical research’. *J Fish Biol* 74(8):1689–1692
- Herpin A, Scharl M (2011) Dmrt1 genes at the crossroads: a widespread and central class of sexual development factors in fish. *FEBS J* 278(7):1010–1019
- Jacobs SBR, Coss D, McGillivray SM, Mellon PL (2003) Nuclear factor Y and steroidogenic factor 1 physically and functionally interact to contribute to cell-specific expression of the mouse follicle-stimulating hormone-beta gene. *Mol Endocrinol* 17(8):1470–1483
- Jantzen SG, Sanderson DS, von Schalburg KR, Yasuike M, Marass F, Koop BF (2011) A 44K microarray dataset of the changing transcriptome in developing Atlantic salmon (*Salmo salar* L.). *BMC Res Notes* 4:88
- Jeng SR, Wu GC, Yueh WS, Kuo SF, Dufour S, Chang CF (2018) Gonadal development and expression of sex-specific genes during

- sex differentiation in the Japanese eel. *Gen Comp Endocrinol* 257: 74–85
- Jia Y, Nan P, Zhang W, Wang F, Zhang R, Liang T, Ji X, Du Q, Chang Z (2018) Transcriptome analysis of three critical periods of ovarian development in Yellow River carp (*Cyprinus carpio*). *Theriogenology* 105:15–26
- Kanamori A, Toyama K, Kitagawa S, Kamehara A, Higuchi T, Kamachi Y, Kinoshita M, Hori H (2008) Comparative genomics approach to the expression of fig alpha, one of the earliest marker genes of oocyte differentiation in medaka (*Oryzias latipes*). *Gene* 423(2): 180–187
- Kim S, Jung D, Kho Y, Choi K (2014) Effects of benzophenone-3 exposure on endocrine disruption and reproduction of Japanese medaka (*Oryzias latipes*)-a two generation exposure study. *Aquat Toxicol* 155:244–252
- Lagadec C, Vlashi E, Frohnen P, Alhiyari Y, Chan M, Pajonk F (2014) The RNA-binding protein Musashi-1 regulates proteasome subunit expression in breast cancer and glioma-initiating cells. *Stem Cells* 32(1):135–144
- Levavi-Sivan B, Bogerd J, Mananos EL, Gomez A, Lareyre JJ (2010) Perspectives on fish gonadotropins and their receptors. *Gen Comp Endocrinol* 165(3):412–437
- Li GL, Zhao H, Wang LJ, Wang Y, Guo XQ, Xu BH (2018) The animal nuclear factor Y: an enigmatic and important heterotrimeric transcription factor. *Am J Cancer Res* 8(7):1106–1125
- Matsuoka MP, van Nes S, Andersen O, Benfey TJ, Reith M (2006) Real-time PCR analysis of ovary- and brain-type aromatase gene expression during Atlantic halibut (*Hippoglossus hippoglossus*) development. *Comp Biochem Physiol B Biochem Mol Biol* 144(1):128–135
- Mazón MJ, Gómez A, Yilmaz O, Carrillo M, Zanuy S (2014) Administration of follicle-stimulating hormone in vivo triggers testicular recrudescence of juvenile European sea bass (*Dicentrarchus labrax*). *Biol Reprod* 90(1):10
- Migeon CJ, Wisniewski AB (2000) Human sex differentiation: from transcription factors to gender. *Horm Res* 53(3):111–119
- Millan A, Gomez-Tato A, Fernandez C, Pardo BG, Alvarez-Dios JA, Calaza M, Bouza C, Vazquez M, Cabaleiro S, Martinez P (2010) Design and performance of a turbot (*Scophthalmus maximus*) oligo-microarray based on ESTs from immune tissues. *Mar Biotechnol* 12(4):452–465
- Mindnich R, Deluca D, Adamski J (2004) Identification and characterization of 17 beta-hydroxysteroid dehydrogenases in the zebrafish, *Danio rerio*. *Mol Cell Endocrinol* 215(1–2):19–30
- Mitra AP, Datar RH, Cote RJ (2006) Molecular pathways in invasive bladder cancer: new insights into mechanisms, progression, and target identification. *J Clin Oncol* 24(35):5552–5564
- Mitter K, Kotoulas G, Magoulas A, Mulero V, Sepulcre P, Figueras A, Novoa B, Sarropoulou E (2009) Evaluation of candidate reference genes for QPCR during ontogenesis and of immune-relevant tissues of European seabass (*Dicentrarchus labrax*). *Comp Biochem Physiol B Biochem Mol Biol* 153(4):340–347
- Munger SC, Natarajan A, Looger LL, Ohler U, Capel B (2013) Fine time course expression analysis identifies cascades of activation and repression and maps a putative regulator of mammalian sex determination. *PLoS Genet* 9(7):e1003630
- Nakamura S, Kobayashi K, Nishimura T, Tanaka M (2011) Ovarian germline stem cells in the teleost fish, medaka (*Oryzias latipes*). *Int J Biol Sci* 7(4):403–409
- Nicol B, Guiguen Y (2011) Expression profiling of wnt signaling genes during gonadal differentiation and gametogenesis in rainbow trout. *Sex Dev* 5(6):318–329
- Palaikostas C, Bekaert M, Taggart JB, Gharbi K, McAndrew BJ, Chatain B, Penman DJ, Vandeputte M (2015) A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*). *Genet Sel Evol* 47:68
- Piferrer F (2001) Endocrine sex control strategies for the feminization of teleost fish. *Aquaculture* 197(1–4):229–281
- Piferrer F, Blázquez M, Navarro L, González A (2005) Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (*Dicentrarchus labrax* L.). *Gen Comp Endocrinol* 142(1–2):102–110
- Purcell CM, Seetharam AS, Snodgrass O, Ortega-Garcia S, Hyde JR, Severin AJ (2018) Insights into teleost sex determination from the *Seriola dorsalis* genome assembly. *BMC Genomics* 19(1):31
- R Core Team (2017) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.R-project.org>. Accessed 15 Nov 2017
- Ramseyer LJ, Garling DL (1994) Amino acid composition of the ovaries, muscle, and whole body of yellow perch. *Prog Fish Cult* 56(3):175–179
- Ribas L, Robledo D, Gómez-Tato A, Viñas A, Martínez P, Piferrer F (2016) Comprehensive transcriptomic analysis of the process of gonadal sex differentiation in the turbot (*Scophthalmus maximus*). *Mol Cell Endocrinol* 422(C):132–149
- Ribas L, Liew WC, Díaz N, Sreenivasan R, Orbán L, Piferrer F (2017) Heat-induced masculinization in domesticated zebrafish is family-specific and yields a set of different gonadal transcriptomes. *PNAS* 114(6):E941–E950
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acid Res* 43(7):e47
- Rodríguez-Marí A, Postlethwait JH (2011). The role of Fanconi anemia/BRCA genes in zebrafish sex determination. *Methods Cell Biol* 105: 461–490
- Rodríguez-Marí A, Canestro C, BreMiller RA, Nguyen-Johnson A, Asakawa K, Kawakami K, Postlethwait JH (2010) Sex reversal in zebrafish fanci mutants is caused by Tp53-mediated germ cell apoptosis. *PLoS Genet* 6(7):e1001034
- Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezaiansev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34(2):374–378
- Saillant E, Fostier A, Menu B, Haffray P, Chatain B (2001) Sexual growth dimorphism in sea bass *Dicentrarchus labrax*. *Aquaculture* 202(3–4):371–387
- Sayed AEH, Ismail RF, Mitani H (2018) Oocyte atresia in WT (HdR) and P53 (-/-) medaka (*Oryzias latipes*) exposed to UVA. *J Photochem Photobiol B Biol* 183:57–63
- Schaeck M, Reyes-Lopez FE, Vallejos-Vidal E, Van Cleemput J, Duchateau L, Van den Broeck W, Tort L, Decostere A (2017) Cellular and transcriptomic response to treatment with the probiotic candidate *Vibrio lentus* in gnotobiotic sea bass (*Dicentrarchus labrax*) larvae. *Fish Shellfish Immunol* 63:147–156
- Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* 3(6):1101–1108
- Shen Z-G, Wang H-P (2014) Molecular players involved in temperature-dependent sex determination and sex differentiation in Teleost fish. *Genet Sel Evol* 46:26
- Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, Herman D, Jensen RV, Johnson CD, Lobenhofer EK, Puri RK, Scherf U, Thierry-Mieg J, Wang C, Wilson M, Wolber PK, Zhang L, Amur S, Bao W, Barbacioru CC, Lucas AB, Bertholet V, Boysen C, Bromley B, Brown D, Brunner A, Canales R, Cao XM, Cebula TA, Chen JJ, Cheng J, Chu T-M, Chudin E, Corson J, Corton JC, Croner LJ, Davies C, Davison TS, Delenstarr G, Deng X, Dorris D, Eklund AC, X-h F, Fang H, Fulmer-Smentek S, Fuscoe JC, Gallagher K, Ge W, Guo L, Guo X, Hager J, Haje PK, Han J, Han T, Harbottle HC, Harris SC, Hatchwell E, Hauser CA, Hester S, Hong H, Hurban P, Jackson SA, Ji H, Knight CR, Kuo WP, LeClerc JE, Levy S, Li Q-Z, Liu C, Liu Y, Lombardi MJ, Ma Y, Magnuson

- SR, Maqsodi B, McDaniel T, Mei N, Myklebost O, Ning B, Novoradovskaya N, Orr MS, Osborn TW, Papallo A, Patterson TA, Perkins RG, Peters EH, Peterson R, Philips KL, Pine PS, Pusztai L, Qian F, Ren H, Rosen M, Rosenzweig BA, Samaha RR, Schena M, Schroth GP, Shchegrova S, Smith DD, Staedtler F, Su Z, Sun H, Szallasi Z, Tezak Z, Thierry-Mieg D, Thompson KL, Tikhonova I, Turpaz Y, Vallanat B, Van C, Walker SJ, Wang SJ, Wang Y, Wolfinger R, Wong A, Wu J, Xiao C, Xie Q, Xu J, Yang W, Zhang L, Zhong S, Zong Y, Slikker W Jr, Consortium M (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24(9):1151–1161
- Small CM, Carney GE, Mo Q, Vannucci M, Jones AG (2009) A microarray analysis of sex- and gonad-biased gene expression in the zebrafish: evidence for masculinization of the transcriptome. *BMC Genomics* 10:579
- Sreenivasan R, Jiang JH, Wang XG, Bartfai R, Kwan HY, Christoffels A, Orban L (2014) Gonad differentiation in zebrafish is regulated by the canonical wnt signaling pathway. *Biol Reprod* 90(2):45:41–10
- Takehana Y, Matsuda M, Myosho T, Suster ML, Kawakami K, Shin-I T, Kohara Y, Kuroki Y, Toyoda A, Fujiyama A, Hamaguchi S, Sakaizumi M, Naruse K (2014) Co-option of *Sox3* as the male-determining factor on the Y chromosome in the fish *Oryzias latipes*. *Nat Commun* 5:4157
- Tanaka M (2016) Germline stem cells are critical for sexual fate decision of germ cells. *Bioessays* 38(12):1227–1233
- Tao WJ, Yuan J, Zhou LY, Sun LN, Sun YL, Yang SJ, Li MH, Zeng S, Huang BF, Wang DH (2013) Characterization of gonadal transcriptomes from Nile tilapia (*Oreochromis niloticus*) reveals differentially expressed genes. *PLoS One* 8(5):e63604
- Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RST, Hecht J, Knaust F, Belkhir K, Klages S, Dieterich R, Stueber K, Piferrer F, Guinand B, Bierre N, Volckaert FAM, Bargelloni L, Power DM, Bonhomme F, Canario AVM, Reinhardt R (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat Commun* 5:5770
- Tingaud-Sequeira A, Lozano JJ, Zapater C, Otero D, Kube M, Reinhardt R, Cerda J (2013) A rapid transcriptome response is associated with desiccation resistance in aerially-exposed killifish embryos. *PLoS One* 8(5):e64410
- Tokarz J, Moeller G, de Angelis MH, Adamski J (2015) Steroids in teleost fishes: a functional point of view. *Steroids* 103:123–144
- Vandeputte M, Piferrer F (2019) Genetic and environmental components of sex determination in the European sea bass. In: Wang HP, Piferrer F, Chen SL (eds) *Sex Control in Aquaculture*. Wiley-Blackwell, Hoboken, pp 307–325
- Vandeputte M, Dupont-Nivet M, Chavanne H, Chatain B (2007) A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. *Genetics* 176(2):1049–1057
- Vanwert JM, Wolfe SA, Grimes SR (2008) Binding of RFX2 and NF-Y to the testis-specific histone H1t promoter may be required for transcriptional activation in primary spermatocytes. *J Cell Biochem* 104(3):1087–1101
- von Schalburg KR, Rise ML, Brown GD, Davidson WS, Koop BF (2005) A comprehensive survey of the genes involved in maturation and development of the rainbow trout ovary. *Biol Reprod* 72(3):687–699
- Wang XX, Wei JZ, Jiao J, Jiang SY, Yu DH, Li D (2014) Genome-wide DNA methylation and gene expression patterns provide insight into polycystic ovary syndrome development. *Oncotarget* 5(16):6603–6610
- Wang Z, Qiu X, Kong D, Zhou X, Guo Z, Gao C, Ma S, Hao W, Jiang Z, Liu S, Zhang T, Meng X, Wang X (2017) Comparative RNA-Seq analysis of differentially expressed genes in the testis and ovary of *Takifugu rubripes*. *Comp Biochem Physiol Part D Genomics Proteomics* 22:50–57
- Wang H-P, Piferrer F, Chen SL, Shen ZG (2019) *Sex Control in Aquaculture*. Wiley, Hoboken
- Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M and Venables B (2016) *gplots: various R programming tools for plotting data*. R package version 3.0.1. Available from: <https://CRAN.R-project.org/package=gplots>. Accessed 4 Nov 2018
- Wen HJ, Xie J, Gui JF (2003) cDNA cloning and characterization of a novel SNX gene differentially expressed in previtellogenic oocytes of gibel carp. *Comp Biochem Physiol B Biochem Mol Biol* 136(3):451–461
- Wettenhall JM, Smyth GK (2004) limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* 20(18):3705–3706
- Wickham H (2009) *ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York
- Wong TT, Zohar Y (2004) Novel expression of gonadotropin subunit genes in oocytes of the gilthead seabream (*Sparus aurata*). *Endocrinology* 145(11):5210–5220
- Wood RK, Seidel JS, Martyniuk CJ (2015) Transcripts involved in steroid biosynthesis and steroid receptor signaling are expressed early in development in the fathead minnow (*Pimephales promelas*). *Comp Biochem Physiol B Biochem Mol Biol* 182:64–72
- Yamaguchi T, Yamaguchi S, Hirai T, Kitano T (2007) Follicle-stimulating hormone signaling and Foxl2 are involved in transcriptional regulation of aromatase gene during gonadal sex differentiation in Japanese flounder, *Paralichthys olivaceus*. *Biochem Biophys Res Commun* 359(4):935–940
- Yang S-Y, He X-Y, Miller D (2007) HSD17B10: a gene involved in cognitive function through metabolism of isoleucine and neuroactive steroids. *Mol Genet Metab* 92(1–2):36–42
- Yang S-Y, He XY, Isaacs C, Dobkin C, Miller D, Philipp M (2014) Roles of 17 beta-hydroxysteroid dehydrogenase type 10 in neurodegenerative disorders. *J Steroid Biochem Mol Biol* 143:460–472
- Yao B, Zhou L, Wang Y, Xia W, Gui J-F (2007) Differential expression and dynamic changes of SOX3 during gametogenesis and sex reversal in protogynous hermaphroditic fish. *J Exp Zool A Ecol Genet Physiol* 307A(4):207–219
- Yasuda T, Oda S, Li Z, Kimori Y, Kamei Y, Ishikawa T, Todo T, Mitani H (2012) Gamma-ray irradiation promotes premature meiosis of spontaneously differentiating testis-ova in the testis of p53-deficient medaka (*Oryzias latipes*). *Cell Death Dis* 3:e395
- Zeng QF, Liu SK, Yao J, Zhang Y, Yuan ZH, Jiang C, Chen AL, Fu Q, Su BF, Dunham R, Liu ZJ (2016) Transcriptome display during testicular differentiation of Channel catfish (*Ictalurus punctatus*) as revealed by RNA-seq analysis. *Biol Reprod* 95(1):19
- Zhang Y, Wang L, Shao M, Li J, Li B, Zhang H (2008) A comparative study on the developmental expression of *hadh2* in amphioxus and zebrafish. *J Fish Biol* 72(5):1215–1222
- Zhang XY, Hou JL, Wang GX, Jiang HB, Wang YF, Sun ZH, Jiang XF, Yu QH, Liu HJ (2015) Gonadal transcriptome analysis in sterile double haploid Japanese flounder. *PLoS One* 10(11):e0143204
- Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 9(1):e78644
- Zhou LY, Wang DS, Senthilkumaran B, Yoshikuni M, Shibata Y, Kobayashi T, Sudhakumari CC, Nagahama Y (2005) Cloning, expression and characterization of three types of 17 beta-hydroxysteroid dehydrogenases from the Nile tilapia, *Oreochromis niloticus*. *J Mol Endocrinol* 35(1):103–116
- Zschocke J (2012) HSD10 disease: clinical consequences of mutations in the HSD17B10 gene. *J Inher Metab Dis* 35(1):81–89



The Model of the Conserved Epigenetic Regulation of Sex

Francesc Piferrer*, Dafni Anastasiadi, Alejandro Valdivieso, Núria Sánchez-Baizán, Javier Moraleda-Prados and Laia Ribas

Institut de Ciències del Mar (ICM), Spanish National Research Council (CSIC), Barcelona, Spain

OPEN ACCESS

Edited by:

Peng Xu,
Xiamen University,
China

Reviewed by:

Zhigang Shen,
Huazhong Agricultural University,
China
Eveline M. Ibeagha-Awemu,
Agriculture and Agri-Food Canada
(AAFC), Canada
Tao Zhou,
Auburn University,
United States

*Correspondence:

Francesc Piferrer
piferrer@icm.csic.es

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 27 March 2019

Accepted: 16 August 2019

Published: 26 September 2019

Citation:

Piferrer F, Anastasiadi D,
Valdivieso A, Sánchez-Baizán N,
Moraleda-Prados J and Ribas L
(2019) The Model of the Conserved
Epigenetic Regulation of Sex.
Front. Genet. 10:857.
doi: 10.3389/fgene.2019.00857

Epigenetics integrates genomic and environmental information to produce a given phenotype. Here, the model of Conserved Epigenetic Regulation of Sex (CERS) is discussed. This model is based on our knowledge on genes involved in sexual development and on epigenetic regulation of gene expression activation and silencing. This model was recently postulated to be applied to the sexual development of fish, and it states that epigenetic and gene expression patterns are more associated with the development of a particular gonadal phenotype, e.g., testis differentiation, rather than with the intrinsic or extrinsic causes that lead to the development of this phenotype. This requires the existence of genes with different epigenetic modifications, for example, changes in DNA methylation levels associated with the development of a particular sex. Focusing on DNA methylation, the identification of CpGs, the methylation of which is linked to sex, constitutes the basis for the identification of Essential Epigenetic Marks (EEM). EEMs are defined as the number and identity of informative epigenetic marks that are strictly necessary, albeit perhaps not sufficient, to bring about a specific, measurable, phenotype of interest. Here, we provide a summary of the genes where DNA methylation has been investigated so far, focusing on fish. We found that *cyp19a1a* and *dmrt1*, two key genes for ovary and testis development, respectively, consistently show an inverse relationship between their DNA methylation and expression levels, thus following CERS predictions. However, in *foxl2a*, a pro-female gene, and *amh*, a pro-male gene, such relationship is not clear. The available data of other genes related to sexual development such as *sox9*, *gsdf*, and *amhr2* are also discussed. Next, we discuss the use of CERS to make testable predictions of how sex is epigenetically regulated and to better understand sexual development, as well as the use of EEMs as tools for the diagnosis and prognosis of sex. We argue that CERS can aid in focusing research on the epigenetic regulation of sexual development not only in fish but also in vertebrates in general, particularly in reptiles with temperature sex-determination, and can be the basis for possible practical applications including sex control in aquaculture and also in conservation biology.

Keywords: conserved epigenetic regulation of sex, essential epigenetic marks, DNA methylation, sex determination, sex differentiation, sex control, environmental sex determination

INTRODUCTION

Background on Epigenetics

The origin of the term “epigenetics” and its implications are continuously subjected to revision. Here, we will use the definition proposed by Deans and Maggert (2015): “the study of phenomena and mechanisms that cause chromosome-bound, heritable changes to gene expression that are not dependent on changes to DNA sequence.” These epigenetic changes or epimutations can be inherited not only during mitosis from mother to daughter cells but also through meiosis from parents to offspring (Dupont et al., 2009). Epigenetics has emerged as a powerful discipline in the study of the integration of genomic and environmental information to bring about a specific phenotype (Turner, 2009; Vogt, 2017).

Fish sex is remarkably plastic when compared with the situation in other vertebrates since it can be determined genetically, environmentally, or by a combination of both types of influences (see Wang et al., 2019 and articles therein). Fish present three major sexual patterns: gonochorism, hermaphroditism, and unisexuality. Thus, the phenotypic sex is, in many fish, a clear example of phenotypic plasticity not only because, in hermaphrodites, the same genotype is capable of producing two different phenotypes but also because, under certain environmental conditions, e.g., unusually warm temperatures, some gonochoristic species may develop a phenotypic sex different from its genotypic sex (Ospina-Alvarez and Piferrer, 2008; Baroiller and D’Cotta, 2016; Ribas et al., 2017).

During sexual differentiation, cells of the germ and somatic lines acquire identity and, in this process, changes in gene expression patterns play a central role. Thus, sexual differentiation involves a certain antagonism between male and female pathways as well as multiple feedback loops that reinforce the effects of the primary effector, be genetic or environmental (Munger and Capel, 2012). Gene networks, involved in testis or ovarian differentiation, consist of genes the expression of which is activated or suppressed in a tight spatial and temporal fashion (Capel, 2017). We now know that in this type of regulation, epigenetic mechanisms such as DNA methylation, histone modification, and noncoding RNAs (Berger et al., 2009) play a role, and hence, in the last years, the contribution of epigenetics to sex determination and differentiation across taxa has emerged (reviewed in Piferrer, 2013). In the rest of this paper, we will use the term “sexual development” when collectively referring to sex determination and sex differentiation.

The Model of Conserved Epigenetic Regulation of Sex

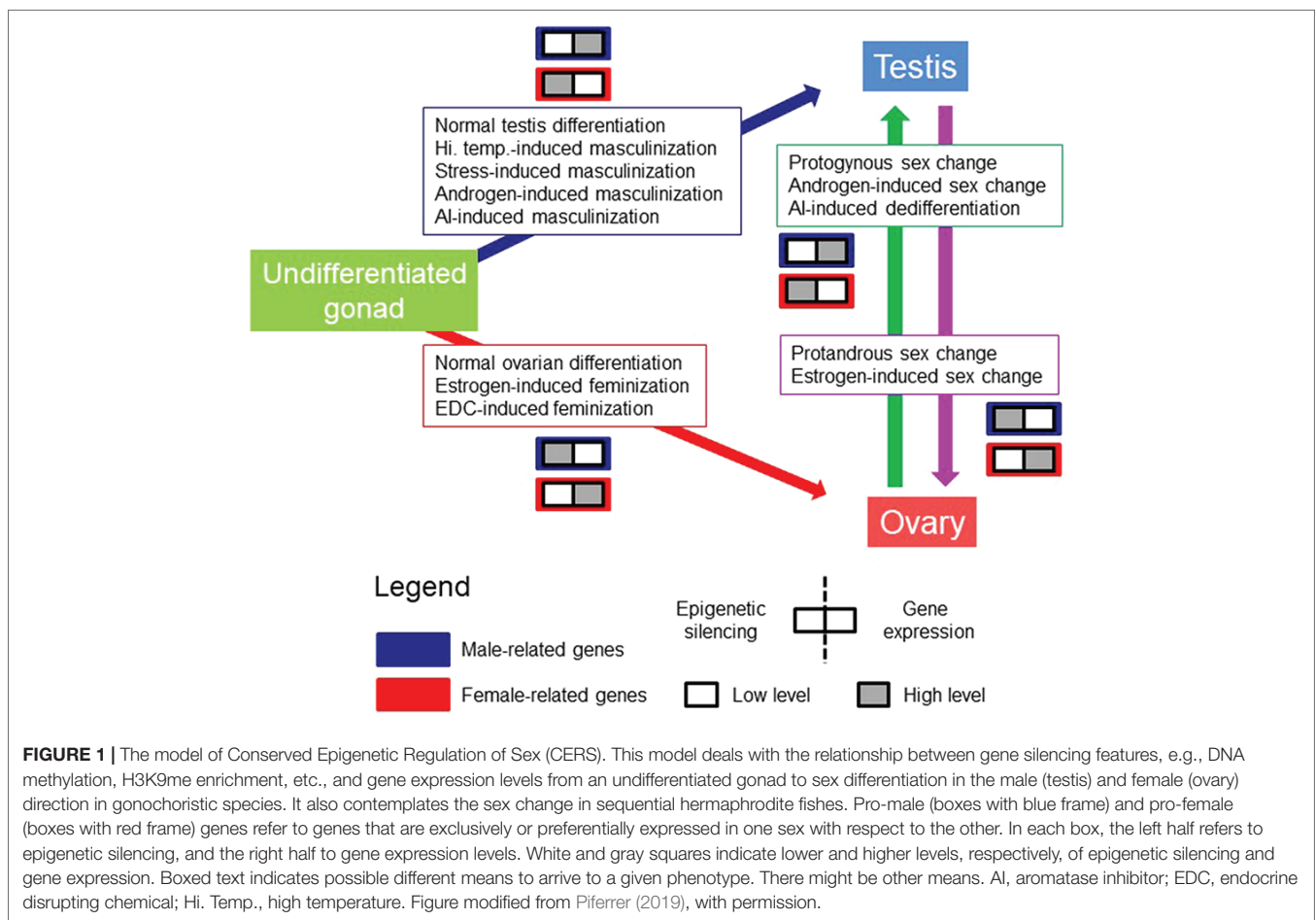
Recently, the concept of Essential Epigenetic Marks (EEM), defined as “the number and identity of informative epigenetic marks that are strictly necessary, albeit perhaps not sufficient, to bring about a specific, measurable, phenotype of interest,” was proposed (Piferrer, 2019). The model of Conserved Epigenetic Regulation of Sex (CERS) was also proposed (Piferrer, 2019) in regards to the regulation of gene expression during the emergence of the sexual phenotype. This model is based on the assumptions that there are “pro-male” and “pro-female” genes and that there is an inverse relationship between epigenetic

silencing and expression of the genes. The terms “pro-male” and “pro-female” genes refer to the exclusive or preferential expression of these genes in one sex rather than in the other. Specifically, the model applies to sex differentiation in gonochoristic species and sex change in hermaphroditic species regardless of the underlying sex-determining mechanism. The CERS model postulates that, for a given sex-related gene, the association between DNA methylation and expression levels with a particular gonadal phenotype is stronger than the means by which this phenotype is obtained (Piferrer, 2019). This implies that, in females, DNA methylation of pro-female genes will be low while expression of these genes will be high and that, in contrast, DNA methylation of pro-male genes will be high while their expression will be low. Conversely, in males, DNA methylation of pro-male genes will be low, while expression of these genes will be high and, in contrast, DNA methylation of pro-female genes will be high, while their expression will be low. Notice that “low” and “high” rather than absolute values indicate values of one sex relative to the other sex. The regulation of gene expression levels by changes in DNA methylation constitute one of the main molecular mechanisms of CERS (the other two would be regulation of gene expression by histone modifications or variants and abundance and activity of miRNAs).

Regarding the causation of differentially methylation levels of “pro-male” and “pro-female” genes, currently, there is debate on whether epigenetic changes are a cause or a consequence of changes in gene expression (probably both things are correct). Allele-specific effects have been found in the half-smooth tongue sole, *Cynoglossus semilaevis*, neomales (ZW females sex reversed into males) with Z chromosomes inherited from high-temperature-exposed sires (Shao et al., 2014). In the European sea bass, *Dicentrarchus labrax*, we found genes with methylation levels that resembled those of oocytes, while other genes had methylation levels resembling those of the sperm, suggesting female- and male-specific inheritance, respectively (Anastasiadi et al., 2018b).

Testis development, at least in fish, where sex can be labile, can be achieved as a consequence of normal male sex differentiation, protogynous sex change, or as masculinization induced by high temperature, stress, aromatase inhibitors, or androgens (Blazquez et al., 2001; Navarro-Martin et al., 2009; Piferrer, 2019). Conversely, ovarian development can be achieved as a consequence of normal female sex differentiation, protandrous sex change, or feminization induced by estrogens or endocrine disrupting chemicals. The model is called conserved because the underlying mechanisms are thought to be shared across species even if they have different reproductive strategies (Figure 1). It should be noted that DNA methylation patterns may differ depending on the cell type within the same gonad. Thus, DNA methylation values reported until now in the gonads represent the combined values of the different cell types.

In the first inception of this model, the following aspects were discussed (Piferrer, 2019): 1) What species are more fruitful to study and why; 2) Which are the best developmental stages to target; 3) Whether there are other organs than the gonads worth targeting; 4) The links with ecotoxicology; and 5) The added comparative value of these studies. In this review, the concept CERS will be further developed. Thus, here we will: 1) Discuss some general considerations about epigenetic marks to put CERS and the concept of EEM in a broader perspective; 2) Since, in the



last 2–3 years, several studies have provided information on DNA methylation levels and given the extraordinary diversity of fishes, we will attempt to summarize the available data on the epigenetic regulation of sex and hence test CERS. This will allow drawing conclusions that can be used not only to establish an appropriate framework but also to help to focus future studies; and 3) Make the suggestion that the CERS can be also applied to other vertebrates regardless of the sex-determining system, whether is genetic or environmental. In fact, even in plants, there is evidence of the involvement of epigenetic regulatory mechanisms in sex determination. This is the case of the *Populus balsamifera* tree, where the *pbr9* gene showed sex-specific patterns of DNA methylation (mostly male-biased) in the putative promoter and in the first intron (Bräutigam et al., 2017).

EPIGENETIC BIOMARKERS

General Concepts

Biomarkers have been developed mostly in the context of human health (e.g., Liu et al., 2019). According to the Biomarkers Definitions Working Group, a biomarker is defined as “a characteristic that is objectively measured and evaluated as indicator of normal biological processes, pathogenic processes

or pharmacologic responses to a therapeutic intervention” (Atkinson et al., 2001). Biomarkers can be proteins, levels of mRNA transcripts, or epigenetic modifications and can mainly be used for diagnosis and prognosis, e.g., to predict responses to therapy in cancer (Costa-Pinheiro et al., 2015; Prensner et al., 2012). Proper biomarkers have to be harmless and characterized by high sensitivity, specificity, and reproducibility (Atkinson et al., 2001; Costa-Pinheiro et al., 2015; García-Giménez et al., 2017).

Also in the context of human health, epigenetic alterations including DNA methylation, histone modifications, and noncoding RNAs have been suggested as good candidates for becoming cancer biomarkers because they can be stable, frequent, abundant, and accessible (Costa-Pinheiro et al., 2015). Nevertheless, the most frequently studied epigenetic modification as potential biomarker is DNA methylation, mainly because of its stability and relative ease of measurement by the available technologies (Bock, 2009; Van Neste et al., 2012). Thus, DNA methylation biomarkers are thought to be extremely promising in the context of human health (Van Neste et al., 2012; Costa-Pinheiro et al., 2015). However, other biomarkers such as microRNAs have been identified also as good candidates for human diseases (Navickas et al., 2016) and, to a lesser extent, as an aid in animal breeding programs (Ibeagha-Awemu and Zhao, 2015).

Biomarker Development

A systematic approach to develop epigenetic biomarkers based on DNA methylation has been suggested by Bock (2009) in the context of clinical applications, where different steps have to be completed. Here, we modify this approach for the development of epigenetic biomarkers to test the CERS (Figure 2). In the first step, a whole-genome or genome-wide method should be used in order to simultaneously assess hundreds or thousands of candidate sites. For DNA methylation biomarkers, whole-genome bisulfite sequencing (WGBS), reduced representation bisulfite sequencing (RRBS) (Gu et al., 2011), or bisulfite RAD-seq (Trucchi et al., 2016) could be employed. These techniques allow to measure the actual DNA methylation levels present in those cytosines located in a CpG context in vertebrate genomes and should lead to the identification of candidate EEMs. These can include differentially methylated cytosines (DMCs) or differentially methylated regions (DMRs) between sexes. In the second step, selected biomarkers are tested using targeted approaches in a large number of independent samples. Here, appropriate approaches include, but are not limited to, multiplex bisulfite sequencing (MBS) (Masser et al., 2013; Anastasiadi et al., 2018b), enrichment bisulfite sequencing (Diep et al., 2012;

Paul et al., 2014), pyrosequencing, or mass spectrometric analysis of DNA methylation (Coolen et al., 2007; Bock et al., 2016). Computational and statistical, machine learning procedures involving regression (best subsets regression, penalized regression, principal components-based regression analysis) or classification analysis should be used. In the third step, from all those EEMs that are strongly correlated to the trait of interest, a handful of them that allow an optimal trait association and/or prediction are validated and a targeted assay is developed (array-type, MeDIP-qPCR or MBS) (Bøvelstad et al., 2007; James et al., 2013; Anastasiadi et al., 2018b).

Switching the perspective from clinical research to ecology and animal production, biomarkers in vertebrates have been used as indicators of environmental pollution (Monserrat et al., 2007) and animal health, including endocrine, immune, nutritional, and metabolic processes (Warne et al., 2015). Epigenetic biomarkers have already been used to predict age and sex in vertebrates. Thus, after the discovery of an epigenetic clock in humans, i.e., a panel of DNA methylation biomarkers as diagnostic of biological age (Horvath, 2013), epigenetic clocks have been constructed in other vertebrates, such as mice, *Mus musculus* (Han et al., 2018), chimpanzees, *Pan troglodytes* (Horvath, 2013), humpback

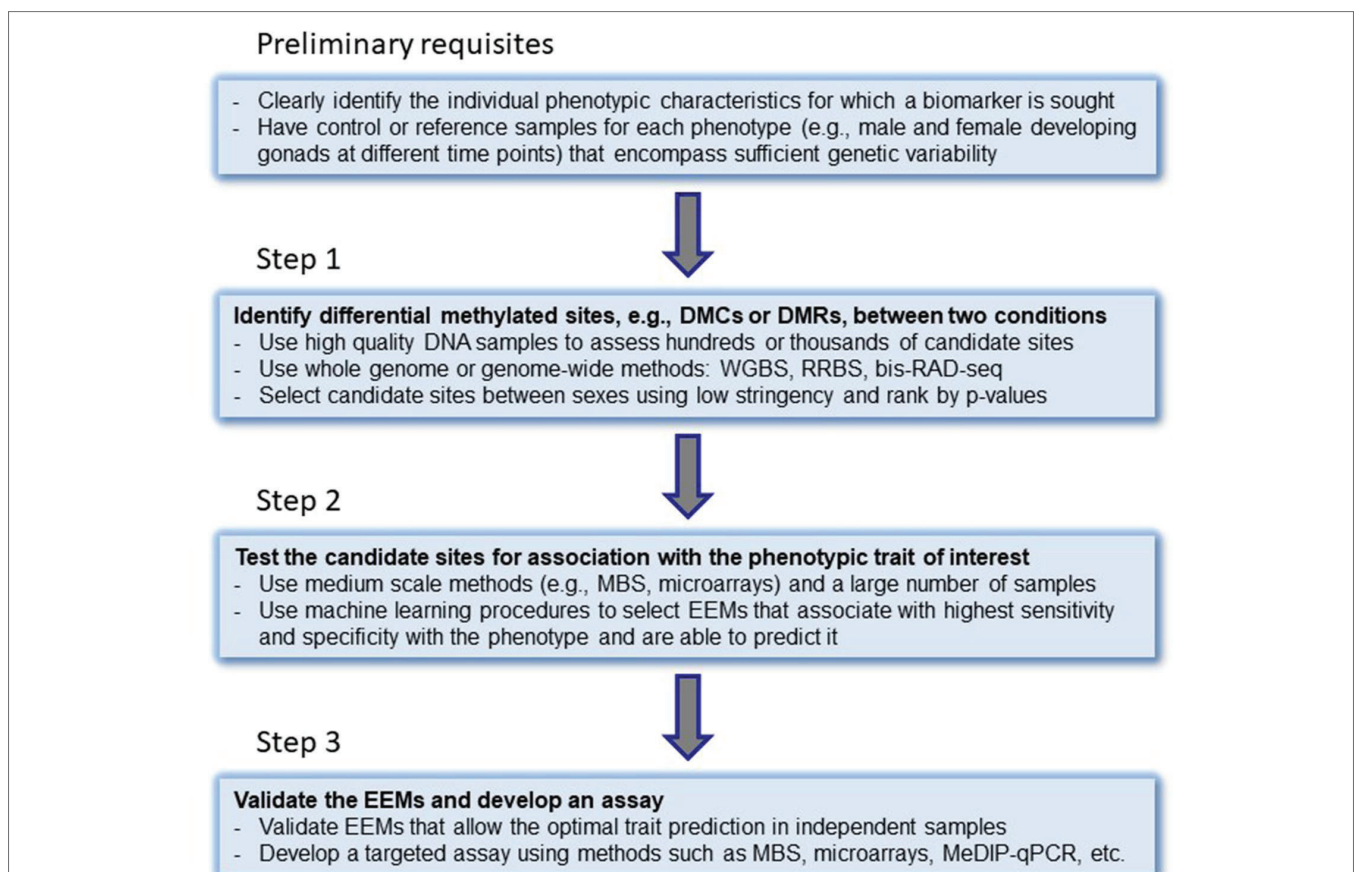


FIGURE 2 | Step-by-step approach for the development of epigenetic biomarkers. DMCs, differentially methylated cytosines; DMRs, differentially methylated regions; WGBS, whole-genome bisulfite sequencing; RRBS, reduced representation bisulfite sequencing; bis-RAD-seq, bisulfite randomly amplified DNA sequencing; MBS, multiplex bisulfite sequencing; EEMs, essential epigenetic marks. MeDIP-seq, methylated DNA immunoprecipitation sequencing. The original idea of steps to epigenetic biomarker development was laid down by Bock (2009).

whales, *Megaptera novaeangliae* (Polanowski et al., 2014), in a long-lived seabird, *Ardena tenuirostris* (De Paoli-Iseppi et al., 2019) and in the European sea bass (Anastasiadi et al., 2019).

Development of Biomarkers of Sex

In livestock and animal production, epigenetic biomarkers have been suggested recently as candidates with extreme potential to predict the phenotypic outcome, as well as to improve production traits (Ibeagha-Awemu and Zhao, 2015; Moghadam et al., 2015). This need was first described in a report by the Food and Agriculture Organization of the United Nations in 2015 stating that the knowledge on epigenetics will offer new opportunities for animal breeding (Scherf and Pilling, 2015). In fish, using the European sea bass, as a model, a carefully selected panel of CpGs in three genes constitute an example of EEMs that were capable to predict the sex phenotype of the gonad with ~90% accuracy (Anastasiadi et al., 2018b). To our knowledge, this is so far the first and only method to predict sex based on EEMs. Currently, sex prediction using EEMs is lethal and is not cost-effective. However, we are testing the possible existence of correlations between DNA methylation in predictor CpGs in the gonads with equally predictive CpGs in other tissues. On the other hand, the own development of biomarkers involves, in the last step, the use of CpG in an array-type approach or in multiplexing (MBS) that, along with the continued decrease of next-generation sequencing costs, should make the cost of screening per sample affordable.

TESTING THE MODEL OF THE CONSERVED EPIGENETIC REGULATION OF SEX

Epigenetic regulation of gene expression is involved in the sexual development of gonochoristic fish with different types of sex-determining mechanisms, as well as in driving the process of sex change in different types of hermaphrodites.

Here, we searched the published literature in fish and collected information on the DNA methylation of genes related to sexual development. A WGBS was used in the half-smooth tongue sole, (Shao et al., 2014), while a MBS was used in the European sea bass (Anastasiadi et al., 2018b). These are the exceptions because in the rest of studies carried so far, which concern around 15 different species, just one or two genes have been analyzed in each case (Table 1). DNA methylation at a single CpG is of a binary nature, since a given CpG can be either methylated or unmethylated. However, mean percent DNA methylation can, theoretically, fall in any value between 0 and 100%. This applies regardless of whether one considers the promoter or the first intron (Anastasiadi et al., 2018a) or other genomic features in a predefined window of a given length. Information drawn from the primary literature shows that DNA methylation levels are more or less evenly distributed across five arbitrarily defined methylation classes (0–20%, 21–40%, 41–60%, 61–80%, and 81–100%), perhaps with a higher preponderance in the 0–20% class, regardless of other considerations such as method of analysis, targeted genomic feature, sex, species, etc. Thus, these preliminary data indicate that there are no preferred or typical DNA

methylation values for the sex-related genes as a whole (Figure 3A). Again, it should be remembered that DNA methylation values represent the combined values resulting from the different cell types making up the gonads. Thus, the correlation with gene expression, if present, should take this into account.

Gonadal aromatase (*cyp19a1a*) was the first gene shown to be under epigenetic regulation during sexual development in a vertebrate, the European sea bass (Navarro-Martín et al., 2011). This is not surprising because it is the only steroidogenic enzyme responsible for the balance between androgens and estrogens and because estrogens are needed for ovarian differentiation in all nonmammalian vertebrates (Guiguen et al., 2010). Since then, the DNA methylation of only few genes has been studied in more than two species: *cyp19a1a* just cited earlier (studied in 10 species), doublesex- and mab-3-related transcription factor 1 (*dmrt1*) (6 species), anti-Müllerian hormone or Müllerian-inhibiting hormone (*amh*) (4 species), and the member of the winged helix/forkhead group (*foxl2a*) (3 species) (Figure 3B). From the analysis of the published data and our own unpublished data, we found that mean DNA methylation levels of *cyp19a1a* were typically <50% in ovaries (mean: 46.2%, sd: 15.98) and >75% in testes (mean: 77.0%, sd: 24.89) (*t*-test: -4.0439; df = 28, *p* = 0.00037) in a fairly consistent manner across species (see list of species in Table 1). This finding was in accordance with the constitutive higher expression of *cyp19a1a* in ovaries when compared with testes (Piferrer and Blázquez, 2005; Guiguen et al., 2010). Likewise, mean DNA methylation levels of *dmrt1* were ~30% in ovaries (mean: 32.54%, sd: 15.98) and <10% in testes (mean: 5.54%, sd: 5.21) (*t*-test: 4.54; df = 14, *p* = 0.00046), also in accordance with the higher constitutive expression of *dmrt1* in testes when compared with ovaries (Herpin and Scharl, 2011) (Figure 3B). Therefore, these two important genes for sex differentiation, which have been used as sex markers in some fish species, e.g., turbot, *Scophthalmus maximus* (Ribas et al., 2016), and medaka, *Oryzias latipes* (Herpin and Scharl, 2011), do indeed conform to the CERS predictions, since there is an inverse relationship between DNA methylation and gene expression with clear sex-specific differences.

This inverse relationship does not seem apparent when two other well-known genes with sex-biased expression in fish are considered: *amh* and *foxl2a* (Figure 3B). *Amh* is a member of the TGF- β superfamily of growth and differentiation factors involved in sex differentiation from mammals to fish (Piferrer and Guiguen, 2008). Relatively low and equal levels of *amh* expression are detected in gonads prior to the appearance of sex-specific differences. However, once sex differentiation is underway, higher *amh* levels are typically associated with testis differentiation in several species analyzed (reviewed in Pfennig et al., 2015). Here, we found that mean DNA methylation levels of *amh* were 54.05% (sd: 25.26) in ovaries and 80.24% (sd: 12.74) in testes, a difference that did not reach statistical significance with the data available so far (*t*-test: -2.07; df = 8, *p* = 0.07211). In the same way, *foxl2a* is expressed at higher levels in the ovary when compared with the testis (reviewed in Bertho et al., 2016), like *cyp19a1a*. On the other hand, *foxl2a* is actually one of the earliest transcriptional activators of *cyp19a1a* that co-localizes in the granulosa cells (Wang et al., 2004). However, DNA methylation levels were clearly not different (mean = 3.08% and sd = 3.88 in ovaries and mean = 2.59% and sd = 3.3 in testes)

TABLE 1 | Studies involving fish where DNA methylation of genes associated with sexual development has been measured

Sex determination Species	Common name	Genes	References
Gonochorism			
<i>Dicentrarchus labrax</i> (Polygenic)	European sea bass	<i>cyp19a1a</i> <i>amhr2</i> , <i>cyp19a1a</i> , <i>dmrt1</i> , <i>foxl2a</i> , <i>fshr</i> , <i>erβ2</i> , <i>nr3c1</i> <i>amh</i> , <i>cyp11a1</i> , <i>hsd3b2</i> , <i>sox9a</i> , <i>vasa</i> (*)	Navarro-Martin et al., 2011 Anastasiadi et al., 2018b Anastasiadi and Piferrer, 2019 (submitted)
<i>Danio rerio</i> (Polygenic)	Zebrafish	<i>amh</i> <i>amh</i> , <i>cyp11a1</i> , <i>cyp11c1</i> , <i>cyp19a1</i> , <i>dmrt1</i> , <i>foxl2a</i> , <i>hsd11β2</i> , <i>hsd17β1</i> , <i>hsd3b2</i> , <i>nr3c1</i> <i>cyp11c1</i> , <i>cyp19a1a</i> , <i>dmrt1</i> , <i>hsd11β2</i> , <i>hsd17β1</i>	Laing et al., 2018 Valdivieso et al., (unpublished)
<i>Cynoglossus semilaevis</i> (ZW/ZZ)	Half-smooth tongue sole	<i>amh</i> , <i>amhr2</i> , <i>arx</i> , <i>cxcr4a</i> , <i>cyp19a1a</i> , <i>daz1</i> , <i>dmrt1</i> , <i>emx2</i> , <i>figla</i> , <i>gsdf</i> , <i>lhx9</i> , <i>pdgfrb</i> , <i>sdf1a</i> , <i>vasa</i> , <i>wt1a</i> , <i>wt1b</i> (**) <i>gata4</i>	Moraleda-Prados et al., (unpublished) Shao et al., 2014
<i>Scophthalmus maximus</i> (ZW/ZZ)	Turbot	<i>wnt4a</i> <i>rson1</i> <i>smiwi12</i>	Hu et al., 2014 Liu et al., 2018 Wang et al., 2017a
<i>Gobiocypris rarus</i> (XX/XY)	Chinese rare minnow	<i>cyp19a1a</i>	Liu et al., 2014
<i>Paralichthys olivaceus</i> (XX/XY)	Olive flounder	<i>cyp19a1a</i> , <i>dmrt1</i>	Wen et al., 2014
<i>Oreochromis niloticus</i> (XX/XY)	Nile tilapia	<i>cyp19a1a</i> <i>cyp19a1a</i> <i>cyp19a1a</i>	Chen et al., 2017 Wang et al., 2017b Chen et al., 2018b
<i>Lateolabrax maculatus</i> (XX/XY)	Chinese sea perch	<i>cyp19a1a</i>	Chen et al., 2018a
<i>Culter alburnus</i> (???)	Topmouth culter	<i>dmrt1</i>	Jia et al., 2019
Hermaphroditism			
<i>Acanthopagrus schlegelii</i> (Protandry)	Black porgy	<i>cyp19a1a</i>	Wu et al., 2016
<i>Lates calcarifer</i> (Protandry)	Barramundi	<i>amh</i> , <i>cyp19a1a</i> , <i>dmrt1</i> , <i>foxl2a</i> , <i>sox8</i> , <i>sox9a</i>	Domingos et al., 2018
<i>Monopterus albus</i> (Protogyny)	Ricefield eel	<i>cyp19a1a</i>	Zhang et al., 2013
<i>Kryptolebias marmoratus</i> (Simultaneous)	Mangrove killfish	<i>cyp19a1a</i>	Ellison et al., 2015

(*) This is from a multiplex bisulfite sequencing analysis with a larger panel of genes. Here, a subset of the most sex-related genes is shown.

(**) This subset of genes showed differential methylation level between ovaries and testes and are taken from Supplementary Table 8 in Shao et al. (2014), where whole-genome bisulfite sequencing was used.

(*t*-test: 0.1656; *df* = 4, *p* = 0.8765). Therefore, unlike *cyp19a1a* and *dmrt1*, and with the information available so far, data suggest that *amh* and *foxl2a* do not seem to conform to CERS predictions or that, in these genes, the relationship between DNA methylation and gene expression is positive (Figure 3B), although, clearly, further research is needed.

There are other genes related to sex differentiation at different degrees for which it may be premature to attempt any sort of generalizations. These genes include *amhr2*, *cyp11a*, *hsd3b2*, *nr3c1*, *sox9*, *vasa*, and *gsdf* (Figure 3C). Here, it is worth noting that allelic diversification of *amhr2* in *Takigugu rubripes* results in a dominant master sex-determining gene, while allelic diversification of *gsdf* has given rise to the sex-determining gene in some fish species, including *Oryzias luzonensis* and *Anoplopoma fimbria* (reviewed in Piferrer, 2018; Guiguen et al., 2019). DNA methylation levels of *amhr2* in the European sea bass were ~50% without sex-related differences (Anastasiadi et al., 2018b), while in the half-smooth tongue sole DNA methylation levels are higher in females (Shao et al., 2014). Similarly, in the latter species, the only species where *gsdf* DNA methylation values have

been determined, these values are clearly lower in males, in accordance with the higher expression of *gsdf* in males (Shao et al., 2014) (Figure 3C).

Except in the half-smooth tongue sole (Shao et al., 2014), where WGBS was used, in the rest of the studies reported in Table 1 and used to draw Figure 3, targeted approaches were utilized to query the DNA methylation status of the target genes. For these studies, an average of ~9 late juvenile or adult fish per sex was used. Typically, amplicons spawn ~450 bp and usually include ~15 CpG located around the transcription start site, although the latter figure may vary considerably among species. It is interesting to note that while sex-specific differences involve change in DNA methylation of several CpGs in some genes, in contrast, in other genes, sex-differences involve only a low number of CpGs (Figure 4). A more comprehensive picture will emerge when genome-wide DNA methylation techniques such as WGBS or RRBS will be employed in lieu of the targeted approaches used so far in most studies.

For the rest of the genes, *cyp11a*, *hsd3b2*, and *nr3c1*, there are only preliminary data gathered in our lab with the European

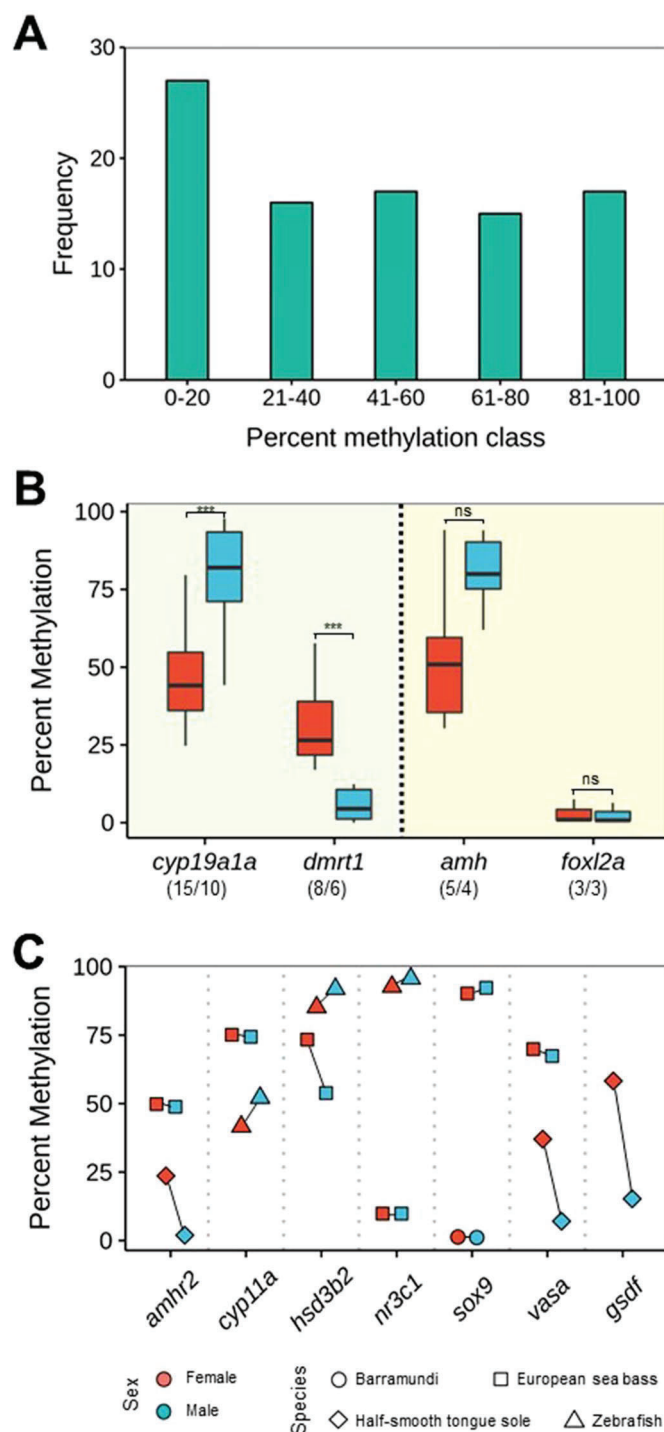


FIGURE 3 | DNA methylation levels of some key genes (see **Table 1**) involved in sexual development. **(A)** Histogram of overall methylation levels for the genes discussed in this paper. Frequency refers to number of DNA methylation values obtained by combining published data and unpublished research performed in our lab. **(B)** Boxplot of DNA methylation levels of *cyp19a1a* and *dmrt1*, which conform to CERS postulates (left side), and *amh* and *foxl2a*, which do not conform to CERS postulates (right side). The lower and upper hinges display the distribution of values between the first and third quartiles, the upper whisker extends to the maximum value up to 1.5 * interquartile range (IQR), the lower whisker extends to the minimum value up to 1.5 * IQR, while the black line indicates the median of the distribution. One outlier outside the end of the whiskers has been excluded. Numbers between parentheses indicate number of datapoints/species. If the first number is bigger than the second, it indicates that there are species for which there is more than one datapoint. Significant differences were assessed with the *t*-test. ****P* < 0.001; ns, not significant. **(C)** DNA methylation levels of *amhr2*, *cyp11a*, *hsd3b2*, *nr3c1*, *sox9*, *vasa*, and *gsdf* in different species. For easier visualization, lines connect datapoints of the same species. In all genes except *gsdf*, there is data for at least two different species. In addition, in **(B)** and **(C)**, data are also color-coded according to sex.

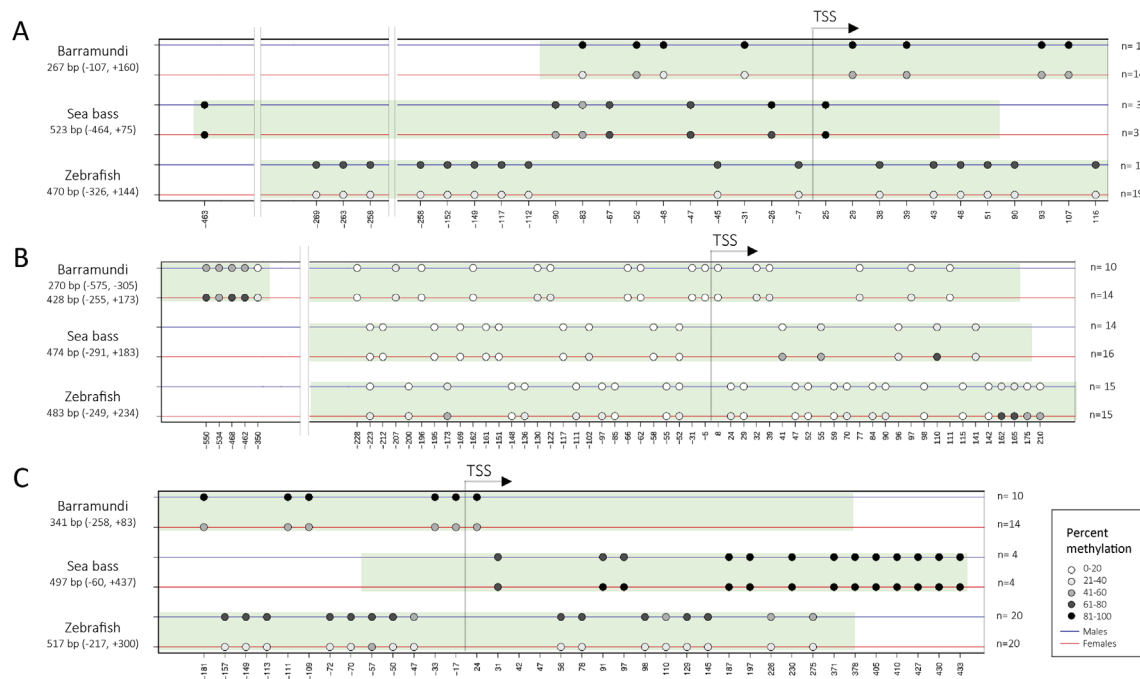


FIGURE 4 | Representation of sex-specific differences in the DNA methylation of CpGs, indicated by circles, around the transcription start site of *cyp19a1* (A), *dmrt1* (B), and *amh* (C) in three different species: Barramundi (Domingos et al., 2018), European sea bass (Anastasiadi et al., 2018b and our own unpublished data), and zebrafish (Valdivieso et al. unpubl. data). The shaded green area indicates the region targeted by the amplicon. Percent methylation is indicated by a gray scale. n = sample size. The graph was built with “Methylation plotter,” developed by Mallona et al. (2014) and available from http://maplab.imppc.org/methylation_plotter/

sea bass (Anastasiadi et al., unpublished) and zebrafish (*Danio rerio*) (Valdivieso et al., unpublished; Moraleda-Prados et al., unpublished). In the European sea bass, methylation values of *hsd3b2* are higher in females. This is in agreement with the expression of this gene that is male-skewed in the developing gonads of Nile tilapia, *Oreochromis niloticus* (Ijiri et al., 2008). On the other hand, DNA methylation values of *nr3c1* and *sox9* were quite different between the two species.

In many species, sex determination has an environmental component. Hence, it is worth mentioning that an environmental factor such as temperature or population density may be connected to sex through epigenetic mechanisms. DNA methylation changes in sex-related genes is the type of epigenetic modification most commonly studied so far. Temperature can affect DNA methylation of many genes, as shown by MeDIP-seq in the Nile tilapia (Sun et al., 2016), although the exact mechanism is not known yet. In the European sea bass, elevated temperature induces hypermethylation in the promoter of *cyp19a1a*, and this prevents the binding of *cyp19a1a* transcriptional activators such as *sfl* and *foxl2a* (Navarro-Martín et al., 2011). Other epigenetic modifications can also be involved in the connection between environmental factors and sex. Thus, temperature increases the transcription of lysine-specific demethylase 6B (*kdm6b*), a chromatin modifier gene in the red-eared slider turtle, *Trachemys scripta*. *Kdm6b* eliminates the trimethylation of H3K27 in the promoter of *dmrt1*, leading to upregulation of its expression and male development (Ge et al., 2018; Georges and Holleley, 2018).

We would like to mention three considerations for further testing the CERS model. First, what species are worth testing?

Obviously, fish, due to their great diversity of sexual systems and sex determining systems, which can vary even in closely related species. Reptiles can also provide very relevant information. Many reptiles possess temperature-dependent sex determination and thus offer the opportunity to test whether DNA methylation in key genes do correlate with gene expression and phenotypic sex under different incubation temperatures during the thermosensitive period. Thus, in the red-eared slider turtle *cyp19a1* DNA methylation levels conformed to CERS predictions (Matsumoto et al., 2013). The same is true in the alligator, *Alligator mississippiensis*, for *cyp19a1* and *sox9* (Parrott et al., 2014) and in the sea turtle, *Lepidochelys olivacea*, for *sox9* (Venegas et al., 2016). In birds and mammals, sexual development is strongly canalized (Capel, 2017), and therefore there is little or no room for sexual plasticity. Nevertheless, in such canalized systems, it would also be interesting to determine to what extent DNA methylation of key genes correlates with expression and whether this is established before the completion of gonadal differentiation. In any case, and regardless of the species of choice, testing the role of epigenetic regulation on the expression of key sex-related genes during the process of sex differentiation should involve, in our opinion, the analysis of at least three different time points. The first one, ideally, should be prior to any morphological sign of sex differentiation, the second around the middle of the process, and the third towards the end or after the completion of sex differentiation.

Second, what other genes can be targeted? In our view, the genes to be tested should include at least the ones that consistently follow or not the predictions of the CERS model, namely, *cyp19a1a*, *dmrt1*, *amh*, and *foxl2a*, as shown in this

TABLE 2 | Genes related to sexual development in mammals, birds, reptiles, and fish (Ge et al., 2018; Capel, 2017; Valenzuela et al., 2019; Todd et al., 2019) where its epigenetic regulation would be worth studying

Gene symbol	Gene description	Gene symbol	Gene description
<i>amh</i> *	Anti-Müllerian Hormone	<i>foxl2</i> *	Forkhead Box L2
<i>amhr2</i> *	Anti-Müllerian Hormone Receptor Type 2	<i>fst</i>	Follistatin
<i>crebp</i>	cAMP-Response Element-Binding Protein	<i>gata4</i> *	GATA Binding Protein 4
<i>ctnnb</i>	Beta-Catenin	<i>gsdf</i> *	Gonadal Soma Derived Factor
<i>cyp11a</i> *	Cytochrome P450 Family 11 Subfamily A Member 1	<i>hsd11b2</i> *	Hydroxysteroid 11-Beta Dehydrogenase 2
<i>cyp11c1</i> *	Cytochrome P450 Family 11 Subfamily B Member 1	<i>kdm6</i>	Lysine-specific Demethylase 6B
<i>cyp19a1a</i> *	Cytochrome P450 Family 19 Subfamily A Member 1	<i>nr3c1</i> *	Nuclear Receptor Subfamily 3 Group C Member 1
<i>dax1</i>	Nuclear Receptor Subfamily 0 Group B Member 1	<i>rspo1</i> *	R-Spondin 1
<i>ddx4</i> (<i>vasa</i>)	DEAD-Box Helicase 4	<i>sf1</i>	Splicing Factor 1
<i>dmrt1</i> *	Doublesex And Mab-3 Related Transcription Factor 1	<i>sox17</i>	SRY-Box 17
<i>erb2</i> *	Estrogen Receptor Beta 2	<i>sox9</i> *	SRY-Box 9
<i>fgf9</i>	Fibroblast Growth Factor 9	<i>sry</i>	Sex Determining Region Y
<i>figla</i> *	Folliculogenesis Specific BHLH Transcription Factor	<i>wnt4</i> *	Wnt Family Member 4
<i>fog2</i>	Zinc Finger Protein, FOG Family Member 2	<i>wt1</i> *	Wilms Tumor Protein 1

(*) Genes for which there is data on DNA methylation during sex differentiation, as detailed in Table 1.

paper. However, other genes with known functions in sexual development in vertebrates, including mammals, birds, and reptiles, should also be studied. We propose here a list of some of the most relevant genes found in the literature (Table 2). Information on DNA methylation of additional genes during gonadal differentiation and any possible sex-related differences will help to better understand the epigenetic regulation of sexual development.

Third, what other approaches can be used? Gene-editing techniques such as CRISPR/Cas9 or the more recently developed technique to edit the methylome in the mammalian genome by Liu et al. (2016b) can be very useful. To date, knockout mutants of sex-related genes in fish have been mostly developed for some model species, e.g., in zebrafish: *cyp19a1a* (Lau et al., 2016), *amh*, and *dmrt1* (Lin et al., 2017), and in medaka: estrogen receptor 1 (*esr1*) (Tohyama et al., 2017), *gnrh* family genes (Marvel et al., 2018), and *cyp19a1a* knockout (Nakamoto et al., 2018). Lau et al. (2016) found that all knockout mutants of *cyp19a1a* were males, supporting the view that aromatase plays an essential role in ovarian differentiation and development. Yet, Lin et al. (2017) found that *dmrt1* and *amh* knockout zebrafish mutants displayed female-biased sex ratios, but the development of abnormal testes was still possible. *Dmrt1* was suggested to be necessary for the maintenance, self-renewal, and differentiation of male germ cells, and *amh* was proposed to control the balance between proliferation and differentiation of these cells. Therefore, it would be interesting to analyze the DNA methylation of *dmrt1* and other male-biased genes in *amh* knockout mutants and vice versa, the DNA methylation of *amh* and other male-biased genes of the network in *dmrt1* knockout mutants.

GAPS IN KNOWLEDGE AND FUTURE PROSPECTS

There are some aspects worth discussing regarding future studies of the involvement of DNA methylation on the

regulation of sexual development. First, one aspect concerns the genomic feature on which one should focus when the goal is to associate DNA methylation with gene expression levels. Determination of the expression should be accurate and consistent for each gene assuming that the method of measurement, e.g., qPCR, is properly employed according to the appropriate standards (primers, reference genes, etc.). In contrast, DNA methylation levels can vary across different genomic features of the same gene. In most studies, the promoter region has been typically targeted. However, methylation of other genomic regions has been found to be equally or even better associated with gene silencing. Indeed, it was shown that the first exon is tightly linked to transcriptional silencing (Brenet et al., 2011). Furthermore, in a systematic study aimed at addressing this question, it was found that the first intron, more than the promoter and the first exon, is tightly related to gene silencing. This seems to be conserved across vertebrate species since it was observed in fish (*Japanese puffer*, *Takifugu rubripes*, and the European sea bass), frog (*Xenopus*), and humans (Anastasiadi et al., 2018a). Thus, for the epigenetic regulation of sex, as well as for sex-related development of biomarkers, it is better to focus around the transcription start site and to prioritize the CpGs localized in the first intron, first exon, and promoter regions, in the order mentioned. Furthermore, gene expression can also be positively correlated to tissue-specific DNA methylation, and this should be kept in mind (Lokk et al., 2014; Wan et al., 2015; Anastasiadi et al., 2018a).

Another aspect concerns the possible effect of genetic variation on DNA methylation levels and how to account for it in the data analysis (Lea et al., 2017; Anastasiadi et al., 2018b). This is related to the number of samples to be analyzed per treatment in studies of DNA methylation, which has been discussed elsewhere (Bock et al., 2016). Also, it would be desirable to overcome the noise induced by the cell heterogeneity of the gonadal tissue. In this regard, recent technological advances allow to determine the epigenome of single cells (Farlik et al., 2015). Efforts toward such type of measurements would definitively help in obtaining more robust measurements of DNA methylation.

Furthermore, DNA methylation and gene expression levels discussed throughout this paper refer to the gonads. DNA methylation is known to be tissue-specific. However, it cannot be ruled out that the methylation patterns of the gonads could be replicated in other tissues. This could be the case of tissues involved in the control of reproduction (e.g., the hypothalamus) or that present sex dimorphism (e.g., secondary sexual characters) because they are under the control of hormonal steroids. To the best of our knowledge, this information does not still exist despite increasing evidence of sex-related differences in DNA methylation for many genes in nonreproductive tissues, such as the muscle or the liver (Davegårdh et al., 2019; Grimm et al., 2019).

Another major challenge will be to determine the sexual phenotype just by the DNA methylation levels of selected EEMs before it can be determined by other means (e.g., by analyzing transcriptomic or histological changes). This would be achievable if demonstrated that the epigenetic modifications precede changes in gene expression. In this case, the EEMs and the CERS model can be foreseen as having potentially useful applications. For example, a defined set of EEMs could be used to predict the sexual phenotype in species with marked sexual growth dimorphism (Parker, 1992; Wang et al., 2019). EEMs could allow to predict the sex ratio in a subsample of a clutch before gonadal differentiation. This would aid in the stock management and in the selection of future broodstock. The same principle could be applied in ornamental fish culture, where the secondary sexual characteristics of males make them usually more desirable than females (Piferrer and Lim, 1997). Another case would be to aid in selection of broodstock fish with a certain epigenetic profile that is suitable to withstand, for example, a masculinization environment due to elevated density or temperature. In reptiles, the use of EEMs combined with temperature manipulations could aid in the research toward our understanding of the underlying molecular mechanism of temperature-dependent sex determination.

Finally, epigenetic modifications can recapitulate past environmental influences (Turner, 2009; Vogt, 2017). Taking advantage of this, EEMs could help to determine whether animals in the wild were exposed to altered environmental conditions such as, for example, exposure to pollutants or elevated temperatures. These EEMs could therefore be useful in conservation programs aimed at determining the environmental hazards to which natural populations may have been previously exposed. As an example along these lines, Guillette et al. (2016) identified epigenetic biomarkers to assess the environmental exposures and health impacts on populations of alligators from lakes contaminated with endocrine-disrupting compounds. The effects of endocrine-disrupting compounds on DNA methylation in the field of aquatic toxicology and biodiversity conservation have recently been reviewed by Tubbs and McDonough, (2018). This approach would allow determining whether a wild population was subjected to a sex-altering condition in the past. To our knowledge, this type of applications has not been fully developed to date, but several studies have started to identify biomarkers with this aim in mind.

CONCLUSIONS

1. There are genes such as *cyp19a1a* and *dmrt1* where DNA methylation and gene expression in developing gonads are not only sex-specific but also inversely correlated. Thus, these genes conform to CERS predictions. Although certainly more research is needed, it is tempting to speculate that similar results will be found when new species are tested.
2. There are genes such as *amh* where DNA methylation and gene expression in the developing gonads could be positively correlated, and this relationship seems to be fairly conserved across species. Thus, in principle, these genes do not conform to CERS predictions. Undoubtedly, research in additional species is needed, but results probably will be similar to what has been found until now.
3. The clear sex-related differences in DNA methylation observed for *cyp19a1a* and *dmrt1* (and also for *amh* if more data confirm the trend observed here) suggest that their combined values could be used as EEMs in a test to predict gonadal sex in fish. Such a test is already available for the European sea bass and species with close sequence similarity for these genes (Anastasiadi et al., 2018b).
4. There are other genes for which there is less information available on DNA methylation, but the data collected so far suggest that perhaps *amhr2*, *hsd3b2*, *gsdf*, and *vasa* could conform to CERS predictions. These and other genes involved in sexual development should be examined in additional species.
5. The CERS model can become a useful tool to better focus research in other species and can contribute to our understanding of the role of epigenetic modifications in the regulation of gene expression during sexual development. However, a major question to be answered is whether sex-related differences in DNA methylation (either positively or negatively correlated) are a cause or a consequence of concomitant sex-related differences in gene expression. Based on the properties of epigenetic modifications, it is tempting to speculate that the role of epigenetic modifications is first to regulate and then to stabilize gene expression during sexual development. In fact, during this process, both somatic and germ cells differentiate and acquire identity in response to interactions with each other or with the environment, and epigenetic regulatory mechanisms are involved in differentiation and acquisition of cell identity.

DATA AVAILABILITY

All datasets analyzed for this study are included in the manuscript and the supplementary files.

AUTHOR CONTRIBUTIONS

FP conceived the study, supervised the collection of data, coined the concepts of EEM and CERS, and wrote the paper. DA developed the technique of MBS and wrote the paper. AV, NS-B, JM-P, and LR collected data on different species

and wrote the paper. All authors approved the final version of the manuscript.

FUNDING

This study was supported by the Spanish Ministry of Science grants AGL2016–787107-R “Epimark” to FP and AGL2015-73864-JIN “Ambisex” to LR. DA was supported by an Epimark contract, AV and NS-B were supported by Spanish government scholarships

(BES-2014-069051 and BES-2017-079744, respectively); LR and JM-P were supported by Ambisex contracts.

ACKNOWLEDGMENTS

We would like to thank the editors, Drs. Peng Xu, Lior David, Paulino Martínez, and Gen Hua Yue, for allowing us to prepare this paper.

REFERENCES

- Anastasiadi, D., Esteve-Codina, A., and Piferrer, F. (2018a). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenet. Chromatin* 11 (1), 37. doi: 10.1186/s13072-018-0205-1
- Anastasiadi, D., and Piferrer, F. (2019). A clockwork fish. Age-prediction using DNA methylation-based biomarkers in the European sea bass. *Mol. Ecol. Res.* (submitted).
- Anastasiadi, D., Vandeputte, M., Sánchez-Baizán, N., Allal, F., and Piferrer, F. (2018b). Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a fish with mixed genetic and environmental sex determination. *Epigenetics* 13 (9), 988–1011. doi: 10.1080/15592294.2018.1529504
- Atkinson, A. J. Jr., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., and Spilker, B. A. (2001). Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* 69 (3), 89–95. doi: 10.1067/mcp.2011.113989
- Baroiller, J. F., and D'Cotta, H. (2016). The reversible sex of gonochoristic fish. *Sex. Dev.* 10 (5–6), 242–266. doi: 10.1159/000452362
- Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009). An operational definition of epigenetics. *Genes Dev.* 23 (7), 781–783. doi: 10.1101/gad.1787609
- Bertho, S., Pasquier, J., Pan, Q., Le Trionnaire, G., Bobe, J., Postlethwait, J. H., et al. (2016). Foxl2 and its relatives are evolutionary conserved players in gonadal sex differentiation. *Sex. Dev.* 10 (3), 111–129. doi: 10.1159/000447611
- Blazquez, M., Felip, A., Zanuy, S., Carrillo, M., and Piferrer, F. (2001). Critical period of androgen-inducible sex differentiation in a teleost fish, the European sea bass. *J. Fish Biol.* 58, 342–358. doi: 10.1111/j.1095-8649.2001.tb02257.x
- Bock, C. (2009). Epigenetic biomarker development. *Epigenomics* 1 (1), 99–110. doi: 10.2217/epi.09.6
- Bock, C., Halbritter, F., Carmona, F. J., Tierling, S., Datlinger, P., Assenov, Y., et al. (2016). Quantitative comparison of DNA methylation assays for biomarker development and clinical applications. *Nat. Biotechnol.* 34 (7), 726. doi: 10.1038/nbt.3605
- Bøvelstad, H. M., Nygård, S., Størvold, H. L., Aldrin, M., Borgan, Ø., Frigessi, A., et al. (2007). Predicting survival from microarray data—a comparative study. *Bioinformatics* 23 (16), 2080–2087. doi: 10.1093/bioinformatics/btm305
- Bräutigam, K., Soolanayakanahally, R., Champigny, M., Mansfield, S., Douglas, C., Campbell, M. M., et al. (2017). Sexual epigenetics: gender-specific methylation of a gene in the sex determining region of *Populus balsamifera*. *Sci. Rep.* 7, 45388. doi: 10.1038/srep45388
- Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A. J., Socci, N. D., et al. (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. *PLoS ONE* 6 (1), e14524. doi: 10.1371/journal.pone.0014524
- Capel, B. (2017). Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet.* 18 (11), 675. doi: 10.1038/nrg.2017.60
- Chen, X., Wang, Z., Tang, S., Zhao, Y., and Zhao, J. (2017). Genome-wide mapping of DNA methylation in Nile Tilapia. *Hydrobiologia* 791 (1), 247–257. doi: 10.1007/s10750-016-2823-6
- Chen, X., He, Y., Wang, Z., and Li, J. (2018a). Expression and DNA methylation analysis of *cyp19a1a* in Chinese sea perch *Lateolabrax maculatus*. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 226, 85–90. doi: 10.1016/j.cbpb.2018.07.008
- Chen, X., Zhao, Y., He, Y., and Zhao, J. (2018b). Methylation pattern polymorphism of *cyp19a* in Nile tilapia and hybrids. *Open Life Sci.* 13 (1), 327–334. doi: 10.1515/biol-2018-0040
- Coolen, M. W., Statham, A. L., Gardiner-Garden, M., and Clark, S. J. (2007). Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic Acids Res.* 35 (18), e119. doi: 10.1093/nar/gkm662
- Costa-Pinheiro, P., Montezuma, D., Henrique, R., and Jeronimo, C. (2015). Diagnostic and prognostic epigenetic biomarkers in cancer. *Epigenomics* 7 (6), 1003–1015. doi: 10.2217/epi.15.56
- Davegårdh, C., Wedin, E. H., Broholm, C., Henriksen, T. I., Pedersen, M., Pedersen, B. K., et al. (2019). Sex influences DNA methylation and gene expression in human skeletal muscle myoblasts and myotubes. *Stem Cell Res. Ther.* 10 (1), 26. doi: 10.1186/s13287-018-1118-4
- Deans, C., and Maggert, K. A. (2015). What do you mean, “Epigenetic”? *Genetics* 199, 887–896. doi: 10.1534/genetics.114.173492
- De Paoli-Iseppi, R., Deagle, B. E., Polanowski, A. M., McMahon, C. R., Dickinson, J. L., Hindell, M. A., et al. (2019). Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Mol. Ecol. Resour.* 19 (2), 411–425. doi: 10.1111/1755-0998.12981
- Diep, D., Plongthongkum, N., Gore, A., Fung, H. L., Shoemaker, R., and Zhang, K. (2012). Library-free methylation sequencing with bisulfite padlock probes. *Nat. Methods* 9 (3), 270. doi: 10.1038/nmeth.1871
- Domingos, J. A., Budd, A. M., Banh, Q. Q., Goldsbury, J. A., Zenger, K. R., and Jerry, D. R. (2018). Sex-specific *dmrt1* and *cyp19a1* methylation and alternative splicing in gonads of the protandrous hermaphrodite barramundi. *PLoS One* 13 (9), e0204182. doi: 10.1371/journal.pone.0204182
- Dupont, C., Armant, D. R., and Brenner, C. A. (2009). Epigenetics: definition, mechanisms and clinical perspective. *Semin. Reprod. Med.* 27 (05), 351–357. doi: 10.1055/s-0029-1237423
- Ellison, A., Rodriguez Lopez, C. M., Moran, P., Breen, J., Swain, M., Megias, M., et al. (2015). Epigenetic regulation of sex ratios may explain natural variation in self-fertilization rates. *Proc. R. Soc. Lond. B Biol. Sci.* 282 (1819), 20151900. doi: 10.1098/rspb.2015.1900
- Farlik, M., Sheffield, N. C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., et al. (2015). Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep.* 10 (8), 1386–1397. doi: 10.1016/j.celrep.2015.02.001
- García-Giménez, J. L., Seco-Cervera, M., Tollefsbol, T. O., Romá-Mateo, C., Peiró-Chova, L., Lapunzina, P., et al. (2017). Epigenetic biomarkers: current strategies and future challenges for their use in the clinical laboratory. *Crit. Rev. Clin. Lab. Sci.* 54 (7–8), 529–550. doi: 10.1080/10408363.2017.1410520
- Ge, C., Ye, J., Weber, C., Sun, W., Zhang, H., Zhou, Y., et al. (2018). The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science* 360 (6389), 645–648. doi: 10.1126/science.aap8328
- Georges, A., and Holleley, C. E. (2018). How does temperature determine sex? *Science* 360 (6389), 601–602. doi: 10.1126/science.aat5993
- Grimm, S. A., Shimbo, T., Takaku, M., Thomas, J. W., Auerbach, S., Bennett, B. D., et al. (2019). DNA methylation in mice is influenced by genetics as well as sex and life experience. *Nat. Commun.* 10 (1), 305. doi: 10.1038/s41467-018-08067-z
- Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat. Protoc.* 6 (4), 468. doi: 10.1038/nprot.2010.190

- Guiguen, Y., Fostier, A., Piferrer, F., and Chang, C. F. (2010). Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish. *Gen. Comp. Endocrinol.* 165 (3), 352–366. doi: 10.1016/j.ygcen.2009.03.002
- Guiguen, Y., Fostier, A., and Herpin, A. (2019). “Sex determination and differentiation in fish: genetic, genomic and endocrine aspects,” in *Sex Control in Aquaculture*. Eds. Wang, H. P., Piferrer, F., Chen, and S. L. (Hoboken, NJ: John Wiley & Sons), 35–63. doi: 10.1002/9781119127291.ch2
- Guillette, L. J., Jr., Parrott, B. B., Nilsson, E., Haque, M., and Skinner, M. K. (2016). Epigenetic programming alterations in alligators from environmentally contaminated lakes. *Gen. Comp. Endocrinol.* 238, 4–12. doi: 10.1016/j.ygcen.2016.04.012
- Han, Y., Eipel, M., Franzen, J., Sakk, V., Dethmers-Ausema, B., Yndriago, L., et al. (2018). Epigenetic age-predictor for mice based on three CpG sites. *Elife* 7, e37462. doi: 10.7554/eLife.37462.001
- Herpin, A., and Scharl, M. (2011). *Dmrt1* genes at the crossroads: a widespread and central class of sexual development factors in fish. *FEBS J.* 278 (7), 1010–1019. doi: 10.1111/j.1742-4658.2011.08030.x
- Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol.* 14 (10), 3156. doi: 10.1186/gb-2013-14-10-r115
- Hu, Q., Zhu, Y., Liu, Y., Wang, N., and Chen, S. (2014). Cloning and characterization of *wnt4a* gene and evidence for positive selection in half-smooth tongue sole (*Cynoglossus semilaevis*). *Sci. Rep.* 4, 7167. doi: 10.1038/srep07167
- Ibeagha-Awemu, E. M., and Zhao, X. (2015). Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Front. Genet.* 6, 302. doi: 10.3389/fgene.2015.00302
- Ijiri, S., Kaneko, H., Kobayashi, T., Wang, D. S., Sakai, F., Paul-Prasanth, B., et al. (2008). Sexual dimorphic expression of genes in gonads during early differentiation of a teleost fish, the Nile tilapia *Oreochromis niloticus*. *Biol. Reprod.* 78 (2), 333–341. doi: 10.1095/biolreprod.107.064246
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An introduction to statistical learning. (New York: Springer). 112, 8
- Jia, Y., Zheng, J., Chi, M., Liu, S., Jiang, W., Cheng, S., and Chen, L. (2019). Molecular identification of *dmrt1* and its promoter CpG methylation in correlation with gene expression during gonad development in Culter alburnus. *Fish Physiol. Biochem.* 45 (1), 245–252. doi: 10.1007/s10695-018-0558-1
- Laing, L., Viana, J., Dempster, E., Webster, T. U., van Aerle, R., Mill, J., et al. (2018). Sex-specific transcription and DNA methylation profiles of reproductive and epigenetic associated genes in the gonads and livers of breeding zebrafish. *Comp. Biochem. Physiol., Part A Mol. Integr. Physiol.* 222, 16–25. doi: 10.1016/j.cbpa.2018.04.004
- Lau, E. S. W., Zhang, Z., Qin, M., and Ge, W. (2016). Knockout of zebrafish ovarian aromatase gene (*cyp19a1a*) by TALEN and CRISPR/Cas9 leads to all-male offspring due to failed ovarian differentiation. *Sci. Rep.* 6, 37357. doi: 10.1038/srep37357
- Lea, A. J., Vilgalys, T. P., Durst, P. A., and Tung, J. (2017). Maximizing ecological and evolutionary insight in bisulfite sequencing data sets. *Nat. Ecol. Evol.* 1 (8), 1074. doi: 10.1038/s41559-017-0229-0
- Lin, Q., Mei, J., Li, Z., Zhang, X., Zhou, L., and Gui, J. F. (2017). Distinct and cooperative roles of *amh* and *dmrt1* in self-renewal and differentiation of male germ cells in zebrafish. *Genetics* 207 (3), 1007–1022. doi: 10.1534/genetics.117.300274
- Liu, J., Liu, T., Niu, J., Wu, X., Zhai, J., Zhang, Q., et al. (2018). Expression pattern and functional analysis of R-spondin1 in tongue sole *Cynoglossus semilaevis*. *Gene* 642, 453–460. doi: 10.1016/j.gene.2017.11.047
- Liu, J., Zhang, W., Du, X., Jiang, J., Wang, C., Wang, X., et al. (2016a). Molecular characterization and functional analysis of the *GATA4* in tongue sole (*Cynoglossus semilaevis*). *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* 193, 1–8. doi: 10.1016/j.cbpb.2015.12.001
- Liu, W. J., Cui, Y., and Ren, W. (2019). Epigenetic biomarker screening by FLIM-FRET for combination therapy in ER plus breast cancer. *Clinical Epigenet.* 11, 16. doi: 10.1186/s13148-019-0620-6
- Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., et al. (2016b). Editing DNA methylation in the mammalian genome. *Cell* 167 (1), 233–247. e217. doi: 10.1016/j.cell.2016.08.056
- Liu, Y., Yuan, C., Chen, S., Zheng, Y., Zhang, Y., Gao, J., et al. (2014). Global and *cyp19a1a* gene specific DNA methylation in gonads of adult rare minnow *Gobiocypris rarus* under bisphenol A exposure. *Aquat. Toxicol.* 156, 10–16. doi: 10.1016/j.aquatox.2014.07.017
- Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., et al. (2014). DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* 15 (4), 3248. doi: 10.1186/gb-2014-15-4-r54
- Mallona, I., Díez-Villanueva, A., and Peinado, M. A. (2014). Methylation plotter: a web tool for dynamic visualization of DNA methylation data. *Sour. Code Biol. Med.* 9 (1), 11. doi: 10.1186/1751-0473-9-11
- Marvel, M., Spicer, O. S., Wong, T. T., Zmora, N., and Zohar, Y. (2018). Knockout of the *Gnrh* genes in zebrafish: effects on reproduction and potential compensation by reproductive and feeding-related neuropeptides. *Biol. Reprod.* 99 (3), 565–577. doi: 10.1093/biolre/iy078
- Masser, D. R., Berg, A. S., and Freeman, W. M. (2013). Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenet. Chromatin* 6 (1), 33. doi: 10.1186/1756-8935-6-33
- Matsumoto, Y., Buemio, A., Chu, R., Vafaee, M., and Crews, D. (2013). Epigenetic control of gonadal aromatase (*cyp19a1*) in temperature-dependent sex determination of red-eared slider turtles. *PLoS ONE* 8 (6), e63599. doi: 10.1371/journal.pone.0063599
- Moghadam, H., Mørkøre, T., and Robinson, N. (2015). Epigenetics—potential for programming fish for aquaculture? *J. Mar. Sci. Eng.* 3 (2), 175–192. doi: 10.3390/jmse3020175
- Monserat, J. M., Martínez, P. E., Geracitano, L. A., Amado, L. L., Martins, C. M. G., Pinho, G. L. L., et al. (2007). Pollution biomarkers in estuarine animals: critical review and new perspectives. *Comp. Biochem. Physiol. C, Pharmacol. Toxicol. Endocrinol.* 146 (1–2), 221–234. doi: 10.1016/j.cbpc.2006.08.012
- Munger, S. C., and Capel, B. (2012). Sex and the circuitry: progress toward a systems-level understanding of vertebrate sex determination. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 4 (4), 401–412. doi: 10.1002/wsbm.1172
- Nakamoto, M., Shibata, Y., Ohno, K., Usami, T., Kamei, Y., Taniguchi, Y., et al. (2018). Ovarian aromatase loss-of-function mutant medaka undergo ovary degeneration and partial female-to-male sex reversal after puberty. *Mol. Cell. Endocrinol.* 460, 104–122. doi: 10.1016/j.mce.2017.07.013
- Navarro-Martin, L., Blázquez, M., and Piferrer, F. (2009). Masculinization of the European sea bass (*Dicentrarchus labrax*) by treatment with an androgen or aromatase inhibitor involves different gene expression and has distinct lasting effects on male maturation. *Gen. Comp. Endocrinol.* 160, 3–11. doi: 10.1016/j.ygcen.2008.10.012
- Navarro-Martin, L., Viñas, J., Ribas, L., Díaz, N., Gutiérrez, A., Di Croce, L., et al. (2011). DNA methylation of the gonadal aromatase (*cyp19a*) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLoS Genet.* 7 (12), e1002447. doi: 10.1371/journal.pgen.1002447
- Navickas, R., Gal, D., Laucevičius, A., Taparuskaitė, A., Zdanytė, M., and Holvoet, P. (2016). Identifying circulating microRNAs as biomarkers of cardiovascular disease: a systematic review. *Cardiovasc. Res.* 111 (4), 322–337. doi: 10.1093/cvr/cvw174
- Ospina-Alvarez, N., and Piferrer, F. (2008). Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS ONE* 3 (7), e2837. doi: 10.1371/journal.pone.0002837.g001
- Parker, G. (1992). The evolution of sexual size dimorphism in fish. *J. Fish Biol.* 41, 1–20. doi: 10.1111/j.1095-8649.1992.tb03864.x
- Parrott, B. B., Kohno, S., Cloy-McCoy, J. A., and Guillette, Jr., L. J. (2014). Differential incubation temperatures result in dimorphic DNA methylation patterning of the *SOX9* and aromatase promoters in gonads of alligator (*Alligator mississippiensis*) embryos. *Biol. Reprod.* 90 (1) 2, 1–11. doi: 10.1095/biolreprod.113.111468
- Paul, D. S., Guilhamon, P., Karpathakis, A., Butcher, L. M., Thirlwell, C., Feber, A., et al. (2014). Assessment of RainDrop BS-seq as a method for large-scale, targeted bisulfite sequencing. *Epigenetics* 9 (5), 678–684. doi: 10.4161/epi.28041
- Pfennig, F., Standke, A., and Gutzeit, H. O. (2015). The role of *Amh* signaling in teleost fish—multiple functions not restricted to the gonads. *Gen. Comp. Endocrinol.* 223, 87–107. doi: 10.1016/j.ygcen.2015.09.025
- Piferrer, F. (2013). Epigenetics of sex determination and gonadogenesis. *Dev. Dyn.* 242 (4), 360–370. doi: 10.1002/dvdy.23924
- Piferrer, F. (2018). La determinació del sexe. *L'Atzavara* 28, 39–60.
- Piferrer, F. (2019). “Epigenetics of sex determination and differentiation in fish,” in *Sex Control in Aquaculture*. Eds. Wang, H. P., Piferrer, F., Chen, and S. L. (Hoboken, NJ: John Wiley & Sons), 65–83. doi: 10.1002/9781119127291.ch3

- Piferrer, F., and Blázquez, M. (2005). Aromatase distribution and regulation in fish. *Fish Physiol. Biochem.* 31 (2), 215. doi: 10.1007/s10695-006-0027-0
- Piferrer, F., and Guiguen, Y. (2008). Fish gonadogenesis. Part II: molecular biology and genomics of sex differentiation. *Rev. Fish. Sci.* 16 Suppl. 1, 35–55. doi: 10.1080/10641260802324644
- Piferrer, F., and Lim, L. C. (1997). Application of sex reversal technology in ornamental fish culture. *Aquarium Sci. Conserv.* 1, 113–118. doi: 10.1023/A:1018391702814
- Polanowski, A. M., Robbins, J., Chandler, D., and Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales. *Mol. Ecol. Resour.* 14 (5), 976–987. doi: 10.1111/1755-0998.12247
- Prensner, J. R., Rubin, M. A., Wei, J. T., and Chinnaiyan, A. M. (2012). Beyond PSA: the next generation of prostate cancer biomarkers. *Sci. Transl. Med.* 4 (127), 127rv123–127rv123. doi: 10.1126/scitranslmed.3003180
- Ribas, L., Robledo, D., Gómez-Tato, A., Viñas, A. M., Martínez, P., and Piferrer, F. (2016). Comprehensive transcriptomic analysis of the process of gonadal sex differentiation in the turbot (*Scophthalmus maximus*). *Mol. Cell Endocrinol.* 422, 132–149. doi: 10.1016/j.mce.2015.11.006
- Ribas, L., Liew, W. C., Díaz, N., Sreenivasan, R., Orbán, L., and Piferrer, F. (2017). Heat-induced masculinization in domesticated zebrafish is family-specific and yields a set of different gonadal transcriptomes. *Proc. Natl. Acad. Sci.* 114 (6), E941–E950. doi: 10.1073/pnas.1609411114
- Scherf, B. D., and Pilling, D. (2015). *The second report on the state of the world's animal genetic resources for food and agriculture*. Rome: FAO Commission on Genetic Resources for Food and Agriculture Assessments.
- Shao, C., Li, Q., Chen, S., Zhang, P., Lian, J., Hu, Q., et al. (2014). Epigenetic modification and inheritance in sexual reversal of fish. *Genome Res.* 24 (4), 604–615. doi: 10.1101/gr.162172.113
- Sun, L. X., Wang, Y. Y., Zhao, Y., Wang, H., Li, N., and Ji, X. S. (2016). Global DNA methylation changes in Nile tilapia gonads during high temperature-induced masculinization. *PLoS ONE* 11(8), e0158483. doi:10.1371/journal.pone.0158483
- Todd, E. V., Ortega-Recalde, O., Liu, H., Lamm, M. S., Rutherford, K. M., Cross, H., et al. (2019). Stress, novel sex genes, and epigenetic reprogramming orchestrate socially controlled sex change. *Sci. Adv.* 5(7), eaaw7006. doi: 10.1126/sciadv.aaw7006
- Tohyama, S., Ogino, Y., Lange, A., Myosho, T., Kobayashi, T., Hirano, Y., et al. (2017). Establishment of estrogen receptor 1 (ESR1)-knockout medaka: ESR1 is dispensable for sexual development and reproduction in medaka, *Oryzias latipes*. *Dev. Growth Differ.* 59 (6), 552–561. doi: 10.1111/dgd.12386
- Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönschetter, P., and Paun, O. (2016). Bs RAD seq: screening DNA methylation in natural populations of non-model species. *Mol. Ecol.* 25 (8), 1697–1713. doi: 10.1111/mec.13550
- Tubbs, C. W., and McDonough, C. E. (2018). Reproductive impacts of endocrine-disrupting chemicals on wildlife species: implications for conservation of endangered species. *Annu. Rev. Anim. Biosci.* 6, 287–304. doi: 10.1146/annurev-animal-030117-014547
- Turner, B. M. (2009). Epigenetic responses to environmental change and their evolutionary implications. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 364 (1534), 3403–3418. doi: 10.1098/rstb.2009.0125
- Valenzuela, N., Litterman, J. L., Neuwald, B., Mizoguchi, J. B., Iverson, J. L., and Riley, and Litzgus, J. D. (2019). Extreme thermal fluctuations from climate change unexpectedly accelerate demographic collapse of vertebrates with temperature-dependent sex determination. *Sci. Rep.* 9, 11. doi: 10.1038/s41598-019-40597-4
- Van Neste, L., Herman, J. G., Otto, G., Bigley, J. W., Epstein, J. I., and Van Criekinge, W. (2012). The epigenetic promise for prostate cancer diagnosis. *Prostate* 72 (11), 1248–1261. doi: 10.1002/pros.22459
- Venegas, D., Marmolejo-Valencia, A., Valdes-Quezada, C., Govenzensky, T., Recillas-Targa, F., and Merchant-Larios, H. (2016). Dimorphic DNA methylation during temperature-dependent sex determination in the sea turtle *Lepidochelys olivacea*. *Gen. Comp. Endocrinol.* 236, 35–41. doi: 10.1016/j.ygcen.2016.06.026
- Vogt, G. (2017). Facilitation of environmental adaptation and evolution by epigenetic phenotype variation: insights from clonal, invasive, polyploid, and domesticated animals. *Environ. Epigenet.* 3 (1). doi: 10.1093/eep/dvx002
- Wan, J., Oliver, V. F., Wang, G., Zhu, H., Zack, D. J., Merbs, S. L., et al. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics* 16(1), 49. doi:10.1186/s12864-015-1271-4
- Wang, D. S., Kobayashi, T., Zhou, L. Y., and Nagahama, Y. (2004). Molecular cloning and gene expression of Foxl2 in the Nile tilapia, *Oreochromis niloticus*. *Biochem. Biophys. Res. Commun.* 320 (1), 83–89. doi: 10.1016/j.bbrc.2004.05.133
- Wang, H., Wang, B., Liu, X., Liu, Y., Du, X., Zhang, Q., et al. (2017a). Identification and expression of piwil2 in turbot *Scophthalmus maximus*, with implications of the involvement in embryonic and gonadal development. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* 208, 84–93. doi: 10.1016/j.cbpb.2017.04.007
- Wang, H. P., Piferrer, F., and Chen, S. L. (2019). *Sex Control in Aquaculture*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9781119127291
- Wang, Y. Y., Sun, L. X., Zhu, J. J., Zhao, Y., Wang, H., Liu, H. J., et al. (2017b). Epigenetic control of *cyp19a1a* expression is critical for high temperature induced Nile tilapia masculinization. *J. Therm. Biol.* 69, 76–84. doi: 10.1016/j.jtherbio.2017.06.006
- Warne, R. W., Proudfoot, G. A., and Crespi, E. J. (2015). Biomarkers of animal health: integrating nutritional ecology, endocrine ecophysiology, ecoimmunology, and geospatial ecology. *Ecol. Evol.* 5 (3), 557–566. doi: 10.1002/ece3.1360
- Wen, A., You, F., Sun, P., Li, J., Xu, D., Wu, Z., et al. (2014). CpG methylation of *dmrt1* and *cyp19a* promoters in relation to their sexual dimorphic expression in the Japanese flounder *Paralichthys olivaceus*. *J. Fish Biol.* 84 (1), 193–205. doi: 10.1111/jfb.12277
- Wu, G. C., Li, H. W., Huang, C. H., Lin, H. J., Lin, C. J., and Chang, C. F. (2016). The testis is a primary factor that contributes to epigenetic modifications in the ovaries of the protandrous black porgy, *Acanthopagrus schlegelii*. *Biol. Reprod.* 94 (6132), 131–113. doi: 10.1095/biolreprod.115.137463
- Zhang, Y., Zhang, S., Liu, Z., Zhang, L., and Zhang, W. (2013). Epigenetic modifications during sex change repress gonadotropin stimulation of *cyp19a1a* in a teleost ricefield eel (*Monopterus albus*). *Endocrinology* 154 (8), 2881–2890. doi: 10.1210/en.2012-2220

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Piferrer, Anastasiadi, Valdivieso, Sánchez-Baizán, Moraleda-Prados and Ribas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Unraveling the genotype by environment interaction in a thermosensitive fish with a polygenic sex determination system

Benjamin Geffroy^{a,1} , Mathieu Besson^{b,c}, Núria Sánchez-Baizán^d , Frederic Clota^{a,c}, Alexander Goikoetxea^a , Bastien Sadoul^{a,e}, François Ruelle^f, Marie-Odile Blanc^f, Hugues Parrinello^g, Sophie Hermet^h , Eva Blondeau-Bidet^h , Marine Pratlong^g, Francesc Piferrer^d , Marc Vandeputte^{a,c} , and François Allal^a

^aMARBEQ Université de Montpellier, CNRS, Ifremer, IRD, Palavas-les-Flots, France; ^bSysAAF, Station LPGP/INRAE, 35042 Rennes, France; ^cUniversité Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France; ^dInstitut de Ciències del Mar, Spanish National Research Council, Barcelona, Spain; ^eESE, Ecology and Ecosystem Health, Institut Agro, INRAE, Rennes, France; ^fLaboratoire Service d'Expérimentations Aquicoles, Ifremer, Palavas-les-Flots, France; ^gMGX, BCM, Université de Montpellier, CNRS, INSERM, Montpellier, France; and ^hMARBEQ Université de Montpellier, CNRS, Ifremer, IRD, Montpellier, France

Edited by James Bull, Department of Biological Sciences, University of Idaho, Moscow, ID; received July 21, 2021; accepted October 28, 2021

In most animals, sex determination occurs at conception, when sex chromosomes are segregated following Mendelian laws. However, in multiple reptiles and fishes, this genetic sex can be overridden by external factors after fertilization or birth. In some species, the genetic sex may also be governed by multiple genes, further limiting our understanding of sex determination in such species. We used the European sea bass (*Dicentrarchus labrax*) as a model and combined genomic (using a single nucleotide polymorphism chip) and transcriptomic (RNA-Sequencing) approaches to thoroughly depict this polygenic sex determination system and its interaction with temperature. We estimated genetic sex tendency (eGST), defined as the estimated genetic liability to become a given sex under a liability threshold model for sex determination, which accurately predicts the future phenotypic sex. We found evidence that energetic pathways, concerning the regulation of lipids and glucose, are involved in sex determination and could explain why females tend to exhibit higher energy levels and improved growth compared to males. Besides, early exposure to high-temperature up-regulated *sox3*, followed by *sox9a* in individuals with intermediate eGST, but not in individuals showing highly female-biased eGST, providing the most parsimonious explanation for temperature-induced masculinization. This gonadal state was maintained likely by DNA methylation and the up-regulation of several genes involved in histone modifications, including *jmjd1c*. Overall, we describe a sex determination system resulting from continuous genetic and environmental influences in an animal. Our results provide significant progress in our understanding of the mechanisms underlying temperature-induced masculinization in fish.

sex determination | genomics | temperature | fish | epigenetic

Sex determination is a central biological process with consequences relevant for natural population dynamics and livestock production. A plethora of systems, from purely genetic sex determination (GSD) to environmental sex determination (ESD), have been described in the animal kingdom (1). Sex determination involves the interaction of promale and profemale genetic pathways in birds and mammals (1), but interestingly, those pathways are often impacted by various environmental factors in reptiles and fish (2, 3). Undoubtedly, fishes represent the taxon exhibiting the widest diversity of sex determination systems (4), in which biotic (e.g., density) (5) or abiotic factors (e.g., pH and temperature) (6, 7) can interact with, or even override, the genetic background of sex.

These external factors affecting the sex of individuals are then transduced at the physiological level in different manners, depending on the species. Yet, two main routes have been identified in fishes, one involving the stress-axis pathway (8) and another one involving epigenetic mechanisms (differential

methylation or histone modification) (9). Interestingly, the latter seems more conserved in reptiles (10, 11) when compared to the former (12). Temperature, the most studied environmental factor affecting fish sex, has been shown to either increase cortisol production (the main stress hormone) with cascading effects on sex (13) or to change methylation profiles in the promoters of key genes mostly involved in sex differentiation (9).

Temperature-dependent sex determination (TSD) has been detected in various fish species including the Atlantic silverside (*Menidia menidia*) (3), the Nile tilapia (*Oreochromis niloticus*) (14), the olive flounder (*Paralichthys olivaceus*) (15), the African spiny catfish (*Clarias gariepinus*) (16), the pejerrey (*Odontesthes hatcheri*), or the cobaltcap silverside (*Hypoaetherina tsurugaen*) (17). In these species, natural temperatures within the thermal range of what fish usually encounter in the wild can impact sexual fate. Moreover, even in species with a supposedly strong GSD, extreme water temperatures outside the natural thermal range can sometimes override their sex determination pathway (18–20). In most of the above-mentioned species, sex reversal (usually from female to male) induced by temperature fluctuations is relatively easy to detect since the

Significance

Traditionally, fish sex determination was considered to be governed by genetic or environmental factors. However, many teleost species defy this dichotomy. We combined genomic and transcriptomic approaches to characterize the temperature-dependent polygenic sex determination of European sea bass. We observed that the estimated genetic sex tendency (eGST) provides an accurate estimation of the phenotypic sex. Our data support the hypothesis that sexually dimorphic growth is the consequence rather than the cause of sex determination. We also showed that temperature-induced masculinization involves the up-regulation of *sox3* and *sox9a* for individuals in the middle of the eGST distribution. We depict a sex determination system influenced by both continuous genetic and environmental variation that results in variable proportions of males and females.

Author contributions: B.G., F.P., M.V., and F.A. designed research; B.G., M.B., F.C., A.G., B.S., F.R., M.-O.B., H.P., S.H., and E.B.-B. performed research; B.G., M.B., N.S.-B., E.B.-B., M.P., F.P., and F.A. analyzed data; and B.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

¹To whom correspondence may be addressed. Email: bgeffroy@ifremer.fr.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2112660118/-DCSupplemental>.

Published December 8, 2021.

genetic sex can be identified either at the gene (21–25) or at the chromosome level (26, 27), which enables the investigation of the underlying physiological mechanisms of sex reversal on an individual basis. Identifying cases of sex reversal becomes much more complicated when species exhibit a polygenic sex determination system (28). In such instances, each individual presents a specific combination of promale and profemale genes involved in sex determination resulting in a genetic sex tendency (GST), defined as the genetic liability to become a male or a female under a liability threshold model (29) for sex determination. This GST is by definition continuous, as opposed to the dichotomic pattern found in species with a master sex-determining gene, in which the presence or the absence of such gene governs sex at conception.

The European sea bass (*Dicentrarchus labrax*) is a gonochoristic species that possesses a GST whereby the genetic architecture (likely involving many genes) interacts with temperature during a labile period where sex can be altered before the sexual fate of the gonad is definitively fixed (30–32). The labile period encompasses the larval and the juvenile stages (SI Appendix, Fig. S1). Thus, as it occurs in many other fish species, exposure to relatively high temperature (HT) (>17°C) during the larval stage promotes male differentiation (30–32). However, long-term exposure to relatively low temperature (LT) (<17°C) before gonadal sex differentiation is complete (i.e., during the juvenile stage) can also trigger masculinization (31, 33). In the European sea bass, future females are already bigger when compared to future males (34), and it has been shown that sex-related differences in growth are established well before the appearance of the first currently known molecular markers of sex (34). However, whether being female induces enhanced growth rate or, conversely, if high early growth rate promotes feminization needs to be further studied.

The fact that this species possesses a polygenic sex determination system, where temperature also influences sex determination, has complicated the studies aiming at deciphering the underpinning mechanisms. Indeed, environmental effects have commonly been detected at the group level (35–37) or after the labile period (36), and genetic effects are deduced from the propensity of specific parents to produce a biased sex ratio (32, 38). While these earlier studies have improved our knowledge of the potential mechanisms involved, they did not allow the identification of the earliest molecular signs of environmental effects, when the gonad is not yet differentiated, even at the molecular level.

Here, we took advantage of the recently developed 57K single nucleotide polymorphism (SNP) chip in the European sea bass (39) to determine the estimated genetic sex tendency (eGST) of individuals, exposed to either high or low temperatures. A prediction equation for eGST can be obtained by combining multilocus SNP genotypes and sex phenotypes in a training population and then this equation can be used to estimate the eGST of fish for which only the SNP genotype is available, in a genomic evaluation framework (40). We predicted that the sex of individuals at both extremes of the eGST distributions would not be impacted by temperature, while those exhibiting intermediate values would be sensitive to temperature. To test this hypothesis, individually genotyped fish were sampled at four key time points during the labile period (SI Appendix, Fig. S1). We then used RNA-Sequencing (RNA-Seq) approaches, both at the whole-body and at the gonadal level, during gonadal sex differentiation. Based on the transcriptomic analysis of individually genotyped fish, we combined gene expression data with sex prediction through eGST data, enabling the investigation of the pathways involved in sex determination of the European sea bass at an early stage. Furthermore, our experimental design allowed us to test the

overlooked hypothesis that masculinization by elevated temperature may result from sex-specific mortality of females rather than from induced female-to-male sex reversal, which could not be tested previously, as mortality in larval temperature treatments occurs before any phenotypic or molecular difference between sexes is visible.

Results

Sex Ratio Analysis of Control and HT-Exposed 1-y-Old Fish Validates the eGST Model. In individuals reared at 21°C (HT, $n = 493$) from 8 to 390 d post hatching (dph), the sex ratio was highly biased toward males ($75 \pm 8.5\%$ males). On the contrary, those fish that were kept at 16°C (LT, $n = 537$) from 2 to 59 dph before being switched to 21°C had a more balanced sex ratio ($46.5 \pm 1.4\%$ males), confirming a strong effect of early exposure to HT on the sex of European sea bass (z -value = -6.3 , P value < 0.001). When combining genomic relationships with the phenotypic sex in a single trait threshold model, where low and high temperatures were considered a fixed effect, sex was found to be highly heritable, with a heritability estimate of $h^2 = 0.56 \pm 0.06$. In the multitrait analysis, where sex in each temperature treatment was considered a specific trait, both sex at LT (sex_{LT}) and at HT (sex_{HT}) were also found highly heritable: $h^2_{\text{sex}_{\text{LT}}} = 0.65 \pm 0.06$ and $h^2_{\text{sex}_{\text{HT}}} = 0.51 \pm 0.08$, with a strong genetic correlation between both temperature-specific sex traits, $r_G = 0.91 \pm 0.09$. From a leave-one-out cross-validation approach, where we predicted the sex of an individual based on a genomic prediction equation established without providing information on its phenotype, the genomic prediction successfully classified animals in 74.5% of the cases for the fish reared at LT and 72.5% for the fish at HT, based on the area under the curve values of the receiver operator characteristic curves (SI Appendix, Fig. S2). Importantly, we did not detect any skew in the distribution of eGST over time (randomly sampled at the four time points, SI Appendix, Fig. S1) when comparing one temperature to the other (SI Appendix, Fig. S3), emphasizing that the skew in sex ratio observed at high temperatures is due to sex reversal rather than to genotype-specific mortality. Based on the observed sex ratio of 1-y-old fish, we predicted that about 25% of the whole population had sex reversed at HT and that this would likely concern those individuals with an eGST lower than 0.5 (i.e., “weak” genetic females exhibiting male sex differentiation at HT, reference SI Appendix, Fig. S4).

A genome-wide association study identified five genomic regions explaining more than 2% of the total genetic variance of the GST (SI Appendix, Fig. S5), which we considered as putative quantitative trait loci (QTL). The region with the highest association with sex was in LG7, between positions 6.3 and 6.8 Mb (QTL_{LG7}) with a 8% variance explained. Two other regions in LG19 in the region 6.6 to 7.3 Mb (QTL_{LG19a}) and in the region 17.0 to 17.9 Mb (QTL_{LG19b}) explained 3.2 and 2.7% of the genetic variance, respectively. In LG13, the region 24.0 to 24.6 Mb (QTL_{LG13}) explained 2.3% of the genetic variance. A last region in LG1A (25.7 to 26.3 Mb, QTL_{LG1A}) explained 2.1% of the genetic variance for GST.

Sox Genes and Genes Linked to Histone Modification Processes Are Affected by HT in Fish with Intermediate eGST Values at the “Flexion” Stage. Whole-body RNA-Seq was performed at the “flexion” stage (25 to 40 dph), coinciding with the early stages of the labile period for sex determination (SI Appendix, Fig. S1). At this stage, only one germ cell per future gonad is observable in transversal fish sections (41). A total of 10 individuals (5 per treatment) were selected for their intermediate, though positive, eGST. The DESeq2 analysis highlighted 341 differentially expressed genes (P value < 0.05) between HT and

LT individuals. The gene ontologies (GOs) response to steroid hormone and cellular response to steroid hormone stimulus were among the biological processes up-regulated at HT (SI Appendix, Fig. S6). Specifically, we found that two *sox* (Sry-related HMG box) genes classically involved in sex determination and differentiation, *sox3* and *sox9b*, were up-regulated at HT (Fig. 1 A and B). Seven genes were linked to histone modification processes, two of them up-regulated at HT (*ncoab* and *lpin1*) and five up-regulated at LT: *ube2a*, *zbtb7b*, *setd5*, *suz12*, and *aunts2*.

Sox Genes and Genes Linked to Energy Regulation Correlate with eGST at the "All Fins" Stage. Whole-body RNA-Seq was performed at the "all fins" stage (53 to 78 dph period) on 68 individuals (29 LT and 39 HT). This period coincides with rapid primordial germ cell proliferation (41). Among the 17,303 genes that respected our inclusion criteria (>30 reads per gene), we detected 584 genes for which expression was correlated with the eGST (eGST P value < 0.05; eGST $\times$ T ($^{\circ}$ C) > 0.5). In total, 12 and 2 genes, respectively (SI Appendix, Table S1), were part of the GOs sex differentiation and sex determination, with *sox9a*

and *sox3* among them (P value = 0.052) (Fig. 1 E and F). For both these genes, we also detected a strong temperature effect on transcript number (P < 0.001), with a higher number of transcripts at HT compared to LT (Fig. 1 C and D).

In total, 16 genes were involved in the GO lipid biosynthetic process and 19 in the GO regulation of growth (SI Appendix, Table S1). The gene encoding the growth hormone (*gh*) was one of these genes and was positively and significantly correlated with the eGST. Three other genes (*prkca*, *gf11b*, and *eya2*) that are in close vicinity of the three previously detected QTL (*QTL_LG7*, *QTL_LG19b*, and *QTL_LG1A*) exhibited a significant correlation with the eGST, though their expression was independent of their SNP genotype of the QTLs (AA, AB, or BB).

The "response to glucose" was among the biological processes presenting a positive correlation with the eGST and thus more expressed in females (SI Appendix, Fig. S7). A total of 11 genes involved in the GO histone modification were also significantly correlated to the eGST (SI Appendix, Table S1). Interestingly, the genes from the GO "histone H3-K27 methylation" and "histone H3-K4 methylation" were negatively correlated with the eGST (thus more expressed in males; SI Appendix, Fig. S7).

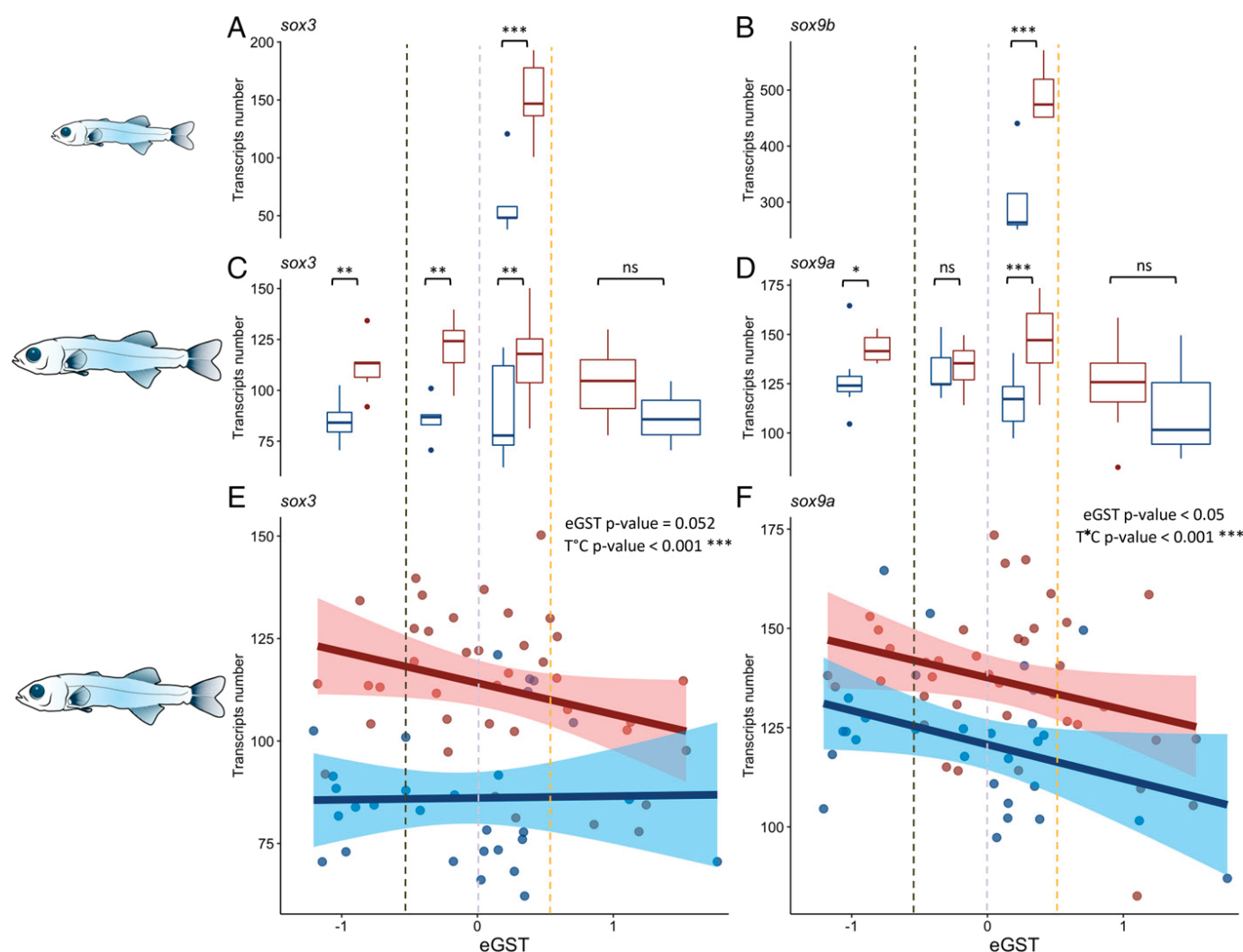


Fig. 1. Both (A) *sox3* and (B) *sox9b* were differentially expressed following the DeSeq2 analysis between neutral individuals (0 < eGST < 0.5) at the "flexion" stage kept at HT (in red) and LT (in blue). The number of RNA-Seq transcripts of (C) *sox3* and (D) *sox9a* differed according to the temperature between groups of eGST < -0.5, -0.5 < eGST < 0, 0 < eGST < 0.5, and eGST > 0.5 in fish sampled at the "all fins" stage following the DeSeq2 analysis. The RNA-Seq analysis at the all fins stage revealed an overall negative and significant correlation between (E) *sox3* and eGST and (F) *sox9a* and eGST, as well as a temperature effect. Abbreviations: *** = P value < 0.001; ** = P value < 0.01, * = P value < 0.05; ns = not significant.

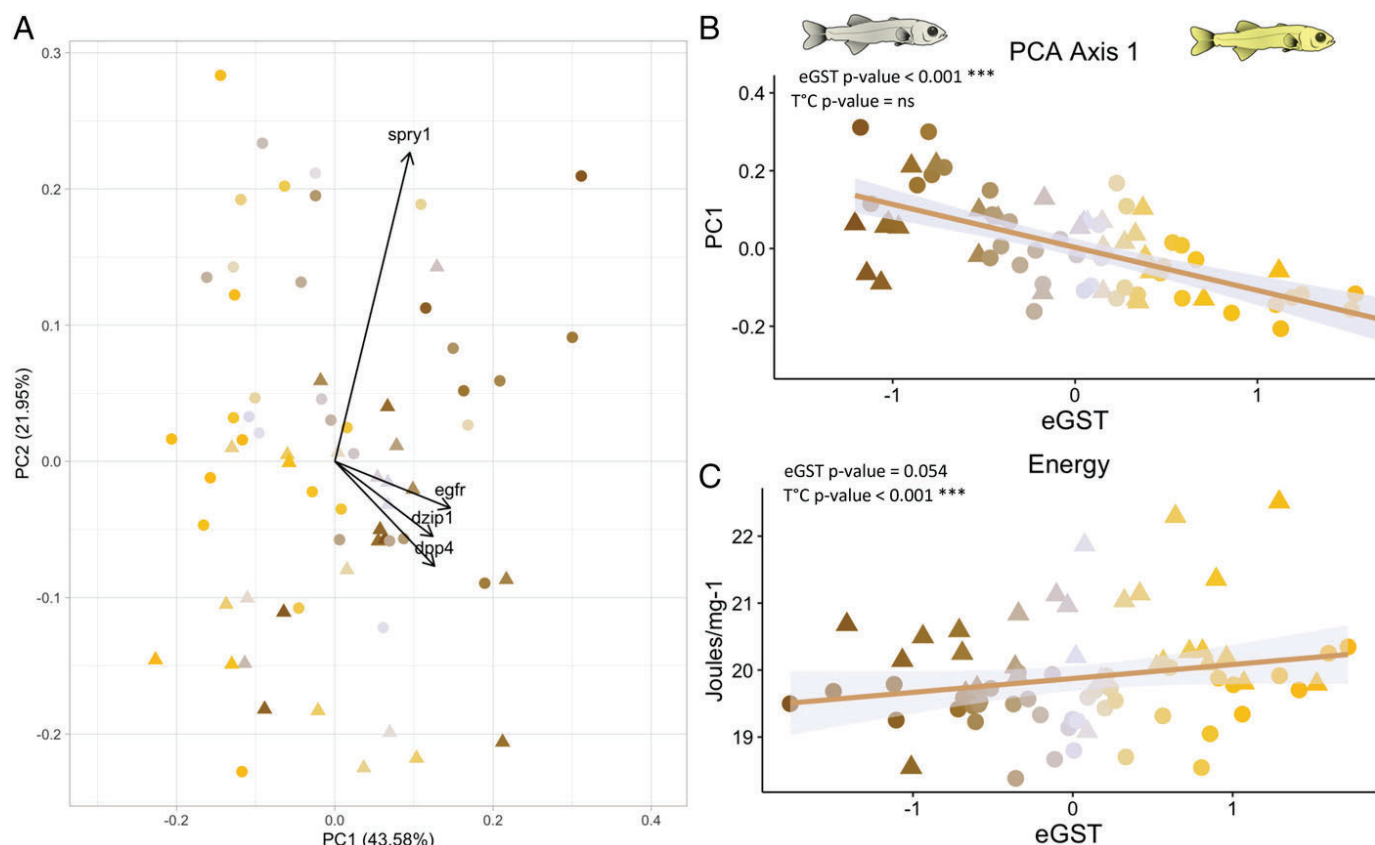


Fig. 2. (A) PCA of four genes (*dzip1*, *dpp4*, *egfr*, and *spry1*) having a highly significant (P value < 0.001) and linear correlation with the sex tendency (eGST) and detected from the RNA-Seq analysis of whole individuals at the “all fins” stage. (B) The first component axis strongly correlated to the eGST. (C) The energy content (joules · mg⁻¹ of tissue) of fish sampled at the “all fins” stage correlated positively with eGST so that genetic females displayed slightly higher energy content than males. Fish kept at low temperatures also displayed higher energy content than those kept at HT. Individuals are represented with a color gradient, from maroon to yellow, representing their eGST. Circles represent fish kept at HT (HT = 21 °C; n = 39) and triangles those kept at LT (LT = 16 °C; n = 29). ns = not significant.

Using a more stringent significance threshold (P value < 0.001), four genes (*spry1*, *egfr*, *dpp4*, and *dzip1*) were correlated to the eGST. Only the gene encoding the Daz interacting protein 1 (*dzip1*), involved in spermatogenesis, showed a clear dimorphic expression higher for individuals with negative eGST. The three other genes have a role in growth rate (Epidermal growth factor receptor, *Egfr*; Sprouty rtk signaling antagonist 1, *Spry1*) and glucose (The dipeptidyl-peptidase IV, *Dpp4*) regulation (based on their GO). The first axis of a Principal Component Analysis (PCA), representing these four genes (Fig. 2A), was highly correlated to the eGST (Fig. 2B). With a quadratic model, only seven genes (*SI Appendix, Table S1*) showed both an overall linear relationship with eGST (P < 0.01) and significant interaction with temperature, revealed by a temperature-specific quadratic component (P value < 0.05). Three of these genes were involved in epigenetic processes: *sgsm2*, *entpd2*, and *map3k3*.

The Juvenile Gonadal Transcriptome Faithfully Reflects the Underlying eGST Independently of Temperature Influences. RNA-Seq was performed on total RNA extracted from the gonads of 42 individuals (21 HT and 21 LT) sampled at the juvenile stage (117 to 124 dph), before the first signs of morphological sex differentiation (*SI Appendix, Fig. S1*). Among the 15,724 genes that respected our inclusion criteria, 1,297 showed a significant (P value < 0.01) linear correlation, either positive or negative, between their expression level and the eGST, independently of the initial temperature treatment (HT versus LT). Among those genes, 19 and 6 genes (*SI Appendix, Table S3*) were within the

GOs of sex differentiation and sex determination, respectively, including *cyp19a1a* (gonadal aromatase), *foxl2* (forkhead box l2), *dmrt1* (doublesex and mab-3 related transcription factor 1), *gsdf* (gonadal soma derived factor), *amh* (anti-Müllerian hormone), *sox9a* (sry-related HMG box 9a), and *insr* (insulin receptor). Those genes, well described to be involved in sexual development, allowed to distinguish two groups on the first axis of the PCA: the differentiating males as opposed to the differentiating females (Fig. 3A). As expected, the correlation was positive for genes involved in ovarian development and negative for those involved in testis development (Fig. 3B and *SI Appendix*). This was confirmed by genes involved in the GO steroids metabolic process, namely, *hsd17b1*, *cyp26a*, and *3β-hsd* (*SI Appendix, Fig. S8*).

Overall, this allowed ascertaining the high relevance of the GST estimated with the Gibbs model (eGST), especially for individuals at both extremes of the distribution, independently of the temperature. Our results were further validated at the group level (eGST > 0 = genetic females versus eGST < 0 = genetic males) with DESeq2 on the GO of sex determination (*SI Appendix, Fig. S9*). A total of 52 genes involved in the GO histone modification were also significantly correlated to the eGST (*SI Appendix, Table S2*), which was confirmed with the “without *a priori* approach” showing that genes involved in histone methylation and acetylation were also up- or down-regulated in differentiating gonads (*SI Appendix, Fig. S10*). Interestingly, other epigenetic processes such as those involved in the microRNA (miRNA) production, were negatively correlated with the eGST and thus positively with maleness (*SI*

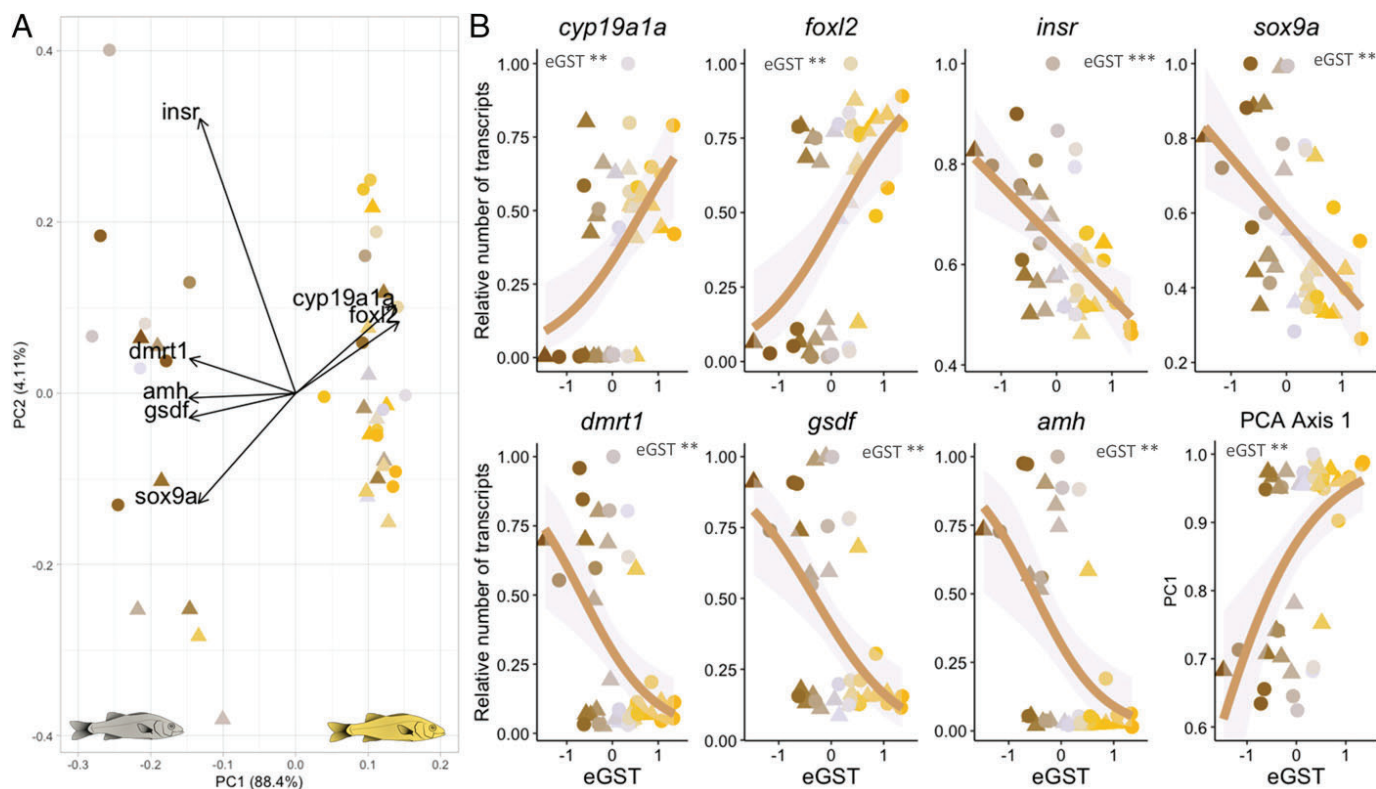


Fig. 3. (A) PCA of seven genes involved in sex determination and differentiation. Data are from the RNA-Seq analysis of the gonads of fish at the juvenile stage ($n = 42$). The PC1 separated the sex horizontally and explained 88.4% of the variance. The PC2 separated the variables vertically and explained 4.1% of the variance. The contribution of the variables (genes) are represented by the arrows. (B) Significant ($P < 0.01$) linear correlation between the eGST and both *insr* and *sox9a* (relative number of transcripts on the y-axis). For five genes, *cyp19a1a*, *foxl2*, *dmrt1*, *amh*, *gsdf*, and the PC1 axis, a dichotomic distribution was observed and modeled with a “quasibinomial” function. Circles represent fish kept at HT (HT = 21 °C; $n = 21$) and triangles those kept at LT (LT = 16 °C; $n = 21$). Individuals are represented with a color gradient, from maroon to yellow, representing their lower or higher eGST. Abbreviations: *** = P value < 0.001 ; ** = P value < 0.01 .

Appendix, Fig. S10). With the quadratic model used for detecting changes linked to the temperature in the middle of the eGST distribution, only seven genes (*thop1*, *paxip1*, *sik3*, *jmjd1c*, *bcor*, *wiz*, and *auts2*) showed both an overall linear correlation with eGST ($P < 0.01$) and a significant interaction with temperature for the quadratic term (P value < 0.05). Four of these genes are involved in epigenetic processes: *jmjd1c*, *bcor*, *wiz*, and *auts2*. The expression of these four genes increased in individuals with an eGST in the middle of the distribution and that were reared at HT, which are the ones with a weak GSD that are expected to be more influenced by the environment (Fig. 4).

DNA Methylation Levels of 1-y-Old Fish Gonads. Reduced representation bisulfite sequencing (RRBS) was conducted at the 1-y-old fish stage using gonadal tissue from 65 males and 42 females. The statistical analysis of methylation data showed several differentially methylated cytosines (DMCs) between fish reared at LT and HT in *sox3* and *sox9a* genes (Fig. 5). For *sox3*, there was a decrease of methylation levels at HT in both males ($P = 0.01844$) and females ($P = 0.0001636$). In males, this gene showed seven hypomethylated DMCs in the first exon, close (< 200 base pair) from the transcription start site (TSS; Fig. 5A). In the females, the same positions were hypomethylated in the first exon, with a total of up to 15 DMCs detected, among which two of them, found around 600 bp from the TSS, were hypermethylated (Fig. 5B). The methylation levels of *sox9a* showed an increase at HT in males (P value = 0.01069) but no significant difference in females (P value = 0.1866) between LT and HT (Fig. 5C). In males, there were three hypermethylated

DMCs toward the end of the gene body (Fig. 5C). However, three out of the five DMCs identified in this region were hypomethylated in females at HT (Fig. 5D).

Gonadal Histology. The sampling at the juvenile stage (117 to 124 dph; $n = 10$ fish per temperature) confirmed that gonads were still not morphologically differentiated. Nevertheless, some oocytes were sparsely observable, but it was impossible to conclude with confidence on the actual phenotypic sex of individuals based on histological analyses alone. Furthermore, the number of oocytes was not correlated to the eGST (data not shown).

Relationship between Energy Content, Body Size, and eGST. The energy content ($\text{joules} \cdot \text{mg}^{-1}$) was not significantly correlated to the eGST at any stage, though it almost reached significance ($P = 0.054$) at the “all fins” stage, with genetic females tending to have higher values than males regardless of their size (Fig. 2C). Individuals from LT presented significantly (P value < 0.001) higher energy content than those from HT at the “all fins” stages (Fig. 2C), likely because LT fish were older. The size of fish was not correlated to the eGST (length: t -value = 1.56, P value = 0.13; wet weight: t -value = 1.3, P value = 0.2), while there was an effect of temperature (length: t -value = 4.6, P value < 0.001 ; wet weight: t -value = 4.75, P value < 0.001).

Discussion

Our analysis combining genomic and transcriptomic data allowed us to shed light on the mechanisms involved in temperature-induced masculinization of the European sea

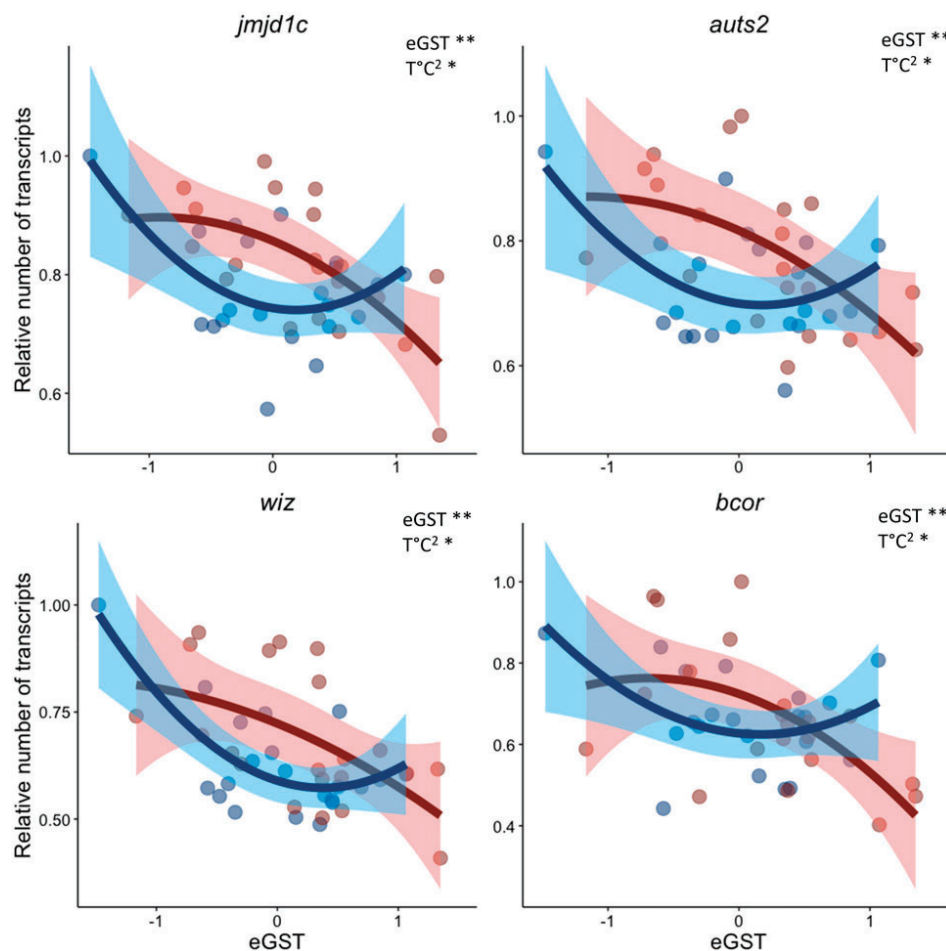


Fig. 4. Quadratic correlation between the eGST and genes involved in histone modification, detected from the RNA-Seq analysis of the gonads of fish at the juvenile stage. The four genes exhibit a significant (**= P value < 0.01) linear correlation with the eGST, plus a significant (* = P value < 0.05) interaction with the temperature for the quadratic term ($T^{\circ}C^2$). Red and blue points represent respectively fish kept at HT (HT = 21 °C; n = 21) or LT (LT = 16 °C; n = 21).

bass. The results confirm the high heritability of the GST. Furthermore, the high genetic correlation between sex_LT and sex_HT suggested low genotype-by-temperature interaction. In other words, the ranking of animals based on their eGST remains very similar at least across the two tested thermal environments. Hence, the effect of larval rearing temperature on sex was mostly additive. This appears to contradict previous results where such interaction occurred (31, 38). However, these previous results were obtained using between-family variation, while we used a genomic relationship matrix for the present study, which is expected to accurately estimate the true genetic parameters (42). Note that the low genotype-by-environment interaction for GST between the two temperature treatments accounts for what happens globally. But it does not impede local $G \times E$ interaction occurring at the genes level or the existence of $G \times E$ interactions in other populations and/or under other environmental circumstances. The prediction equation for GST, which allowed us to estimate the eGST of genotyped individuals, was established on phenotypically sexed individuals at 1 y of age and predicted their sex with a 72.5 to 74.5% success. The relevance of eGST was further confirmed by transcriptomic analysis of gonads at 117 and 124 dph, that is, when the first signs of molecular differentiation can be identified (43). Indeed, we detected a strong correlation between key genes involved in sex differentiation in the gonad and the eGST at this juvenile stage. This good match between phenotypic and genetic sex allowed us to determine the eGST of 1- to 3-mo-old individuals, when the temperature is known to act on the sex of European

sea bass (30). Five putative QTLs, explaining a low but significant part of the variance of GST, were identified in four different chromosomes (LG7, LG19, LG13, and LG1). Faggion et al. (44) already identified *QTL LG7* in Northern Atlantic and Mediterranean populations of European sea bass and the two LG19 QTLs in Mediterranean populations only (origin of the present population). The minor QTLs found in LG13 and LG1A were, however, specific of this study. None of the QTLs previously found in LG6, LG11, and LG18-21 (45) were detected in the present study. Overall, our results are consistent with those of previous studies, pinpointing a GST strongly driven by polygenic variation (~90% of the variance) with a low contribution of minor QTLs (~10%).

These QTLs however participate to the accuracy of the model, which appeared to be strong (100%) for individuals at both extremes of the eGST distribution. However, some mismatches were detected in the middle of the distribution, with some individuals with low negative eGST value that likely were phenotypic females (15% at HT and 29% at LT) based on the dichotomic expression of *cyp19a1a*, *foxl2*, *dmrt1*, *gsdf*, and *amh* and some individuals with positive eGST values that were likely phenotypic males (25% at HT and 14% at LT). The proportion of fish that are supposedly genetic females, but that exhibit a male phenotype (25%), could well be explained by precocious and relatively long-term exposure to HT (31, 46, 47). It also corresponds well to the supposed percentage of masculinized genetic females observed at the end of the experiment: 25%, a figure within the range of masculinization typically observed in European sea bass exposed to HT (48).

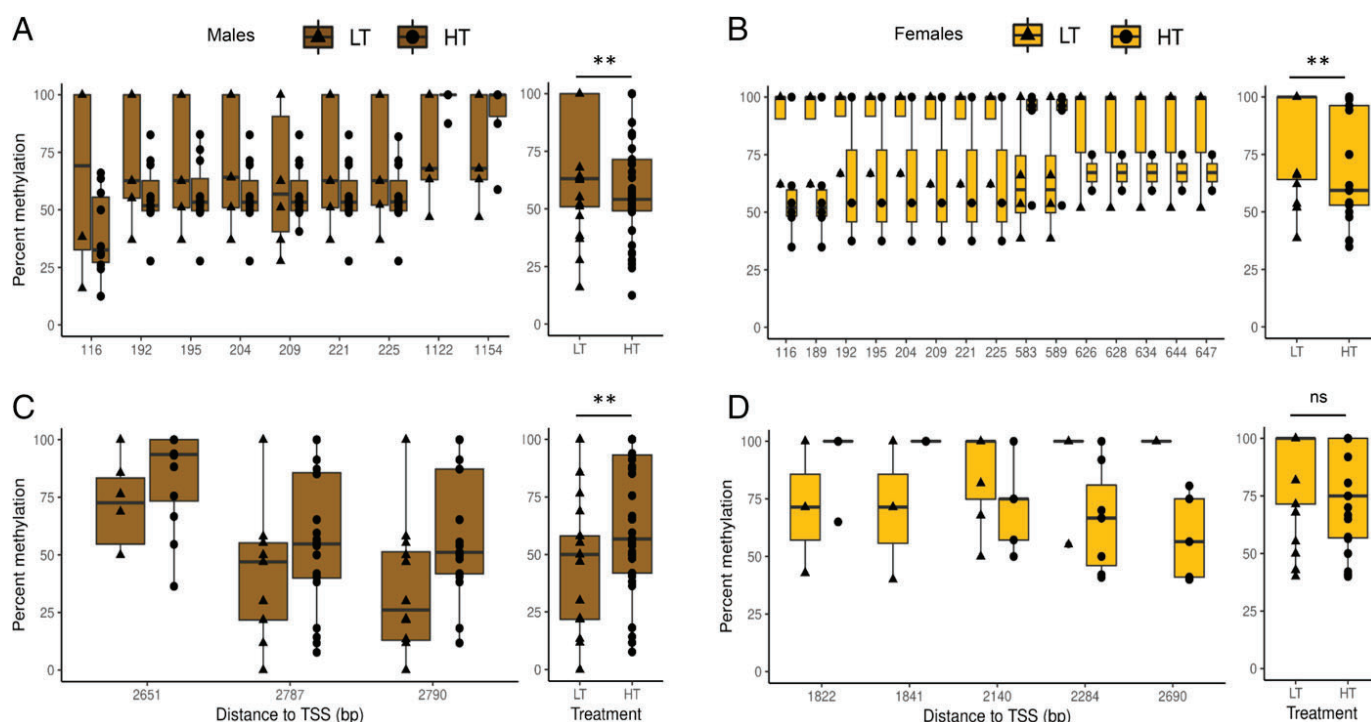


Fig. 5. Box plots of DNA methylation levels of *sox3* (A and B) and *sox9a* (C and D) in 1-y-old fish testes (maroon) and ovaries (yellow), respectively. Individual DMCs identified within the gene region (Left), and average methylation levels of the gene body $\pm 2,000$ bp (Right). The black line within the box indicates the median of the distribution, and the lower and upper hinges display the distribution of values between the first and third quartiles. The upper whisker extends to the maximum value ($1.5 \times$ interquartile range [IQR]), and the lower whisker extends to the minimum value ($1.5 \times$ IQR). Individual DMCs are defined as CpGs with methylation differences $>15\%$ and q -value <0.01 , while significant differences between average data were assessed with the t test. Abbreviations: ** = P value <0.01 ; ns = not significant. Circles represent fish kept at HT (HT = 21°C ; $n = 65$) and triangles those kept at LT (LT = 16°C ; $n = 42$).

The mismatch occurring at LT (14%) could also be due to the masculinization of fish kept too long at this temperature, since there is what could be regarded as a second period of sensitivity to adverse environmental conditions, including prolonged exposure to LT, in European sea bass (31, 37, 46). However, temperature itself might not explain the pattern detected for individuals that are supposedly genetic males. This could come from the fact that some errors occurred in the estimation of eGST (as detected in 1-y-old fish, with the leave-one-out approach), which is expected with a polygenic trait that typically has incomplete penetrance, when heritability is lower than 1, which is the case here (49, 50). It could also be that phenological events linked to gonad development are involved. In the European sea bass, the sex is considered “fixed” once animals reach a size of 8 to 10 cm according to some studies (41) but at a size of 4 to 6 cm according to others (34). It is thus possible that individuals in the middle of the distribution can still develop a phenotypic sex not predicted by their eGST, according to the polygenic nature of sex determination in this species, and that their observed transcript values of sex-related genes remain transitory at this stage/age (7.2 cm and 4.5 g). In this sense, the histological analysis did not permit to unambiguously identify males and females at the early juvenile stage, but some intratesticular oocytes were detected in some individuals, showing that they probably had undergone sex reversal (51). In European eels (*Anguilla anguilla*), individuals presenting intratesticular oocytes showed higher levels of *cyp19a1a* than males (52). It is thus possible that the presence of both tissues in the same gonad drives the mismatch observed between transcript values of sex-related genes and eGST for intermediate individuals. A last plausible explanation involves the sampling design at this stage for transcriptomic analysis. To ensure having sufficient tissue quantity (>100 ng total RNA), we sampled the biggest fish. Since early growth rate is known to impact sex, it is possible that some of the genetic males (eGST <0) developed as phenotypic females at this stage.

Interestingly, the goodness of the linear fit between genetic and phenotypic sex was confirmed by the gene expression of *sox9a* and *insr*, even when considering individuals in the middle of the eGST distribution and at both temperatures. These two genes are within the GO of sex determination and play a key role in male sex differentiation (1, 53–55). None of them were pinpointed as essential for sex differentiation in previous studies on European sea bass (35, 43, 56). However, both genes are known to be overexpressed at high temperatures in TSD reptile species during sex determination (57–59). The fact that both genes present a very linear correlation (as opposed to the dichotomic expression of *cyp19a1a*, *foxl2*, *dmrt1*, *gsdf*, and *amh*) with the eGST suggests that they are involved early in the process of sex determination, while the other sex-related genes just transduce the future state of the gonad (male or female). This was confirmed at the “all fins” stage with *sox9a* exhibiting a negative correlation with the eGST, with higher expression at high temperatures for neutral fish. This gene was also shown to play a key role in the ovary-to-testis transition in the zebrafish (60), where it was strongly expressed in pre-Sertoli cells prior to oocyte apoptosis and degeneration. The DNA methylation levels of *sox9a* increased with high temperatures in testes of 1-y-old fish. This suggests that HT could affect *sox9a* gene expression already at the “flexion” stage and maintain its state through to adulthood in males. However, the higher expression of this gene found at HT for neutral eGST fish and the hypermethylation detected at HT in males would not match the standard association of hypermethylation with down-regulation of gene repression (61). This could be explained by the fact that the DMCs were identified toward the end of the gene body, while it is the methylation level of the first intron, and to a lesser extent, the first exon, which was shown to play an important role and inverse association with gene expression regardless of tissue and species (62).

Another gene that was also reported as a marker of germ cells and supporting cells that preferentially develop into testes in zebrafish juvenile ovary-to-testis transformation (60) was *dzip1*, which is concordant with our results where *dzip1* shows a clear dimorphic expression at the “all fins” stage, predominantly at HT. Three other genes were strongly correlated with the eGST at this stage, namely, *egfr*, *dpp4*, and *spry1*. The specific function of these genes has not been described in fish, and much of our knowledge comes from murine models and humans. Sprouty1 (*Spry1* gene product) is involved in fat and bone development (63). *Spry1* gene knockout in mice adipocytes results in decreased bone mass and increased body fat (63), while adipose tissue-specific expression of the *Spry1* gene in mice protects against fat accumulation and bone loss (64). EGFR is involved in protein kinase activity, and the inhibition of such protein was shown to improve glucose tolerance and favor insulin action (65). Interestingly, its expression was shown to be regulated by *spry1* (66). DPP4 is also involved in glucose homeostasis, and targeted inactivation of this gene in mice yielded individuals with enhanced insulin secretion and improved glucose tolerance (67). All those studies advocate for genetic males having less capacity to produce fat and display appropriate glucose levels. Although this could be viewed as only a conjunction of facts, this hypothesis is enforced by the tendency of genetic females (individuals with high eGST values) to display higher energy content in their tissue (in joules · mg⁻¹) when compared to males (individuals with low eGST values) at the “all fins” stage. At this stage (53 to 78 dph), no correlation was found between size and eGST, and the earliest sexual size dimorphism was found at 103 dph in another experiment (68). Overall, all these results support the hypothesis that early growth differences are the consequence rather than the cause of sex differentiation. In that scheme, the polygenic sex determination of the European sea bass provides information that is later transduced at the phenotypic level, as exemplified by the positive correlation between eGST and the *gh* gene at the “all fins” stage.

Regarding the role of temperature, at both the “flexion” and the “all fins” stages we detected enhanced expression of *sox3* at an HT compared to an LT, while this was not detected for genetic females (eGST > 0.5) at the “all fins” stage following transcriptomic analysis. *Sox3* is the evolutionary precursor of Sry (sex determining region Y) in mammals (69) and a major master-sex determination gene in three medaka fish species (70). Interestingly, ectopic expression of *Sox3* in the developing XX gonads resulted in the complete sex reversal of XX females to males in mice (*Mus musculus*) (71), and loss-of-function of *sox3* caused sex reversal of XY males in the Indian rice fish (*Oryzias dancena*) (72). The expression of *sox3* also gradually increased during the protogynous sex change (female-to-male) of the hermaphroditic fish (*Epinephelus coioides*) (73). This gene is essential in the up-regulation of downstream key-related genes for testicular differentiation, *gsdf* in the Indian rice fish (72) and *sox9* in mice (71). Here, we also detected that up-regulation of *sox3* appeared chronologically before the up-regulation of *sox9a* suggesting similar mechanisms. *Sox3* was strongly hypomethylated at HT in both males and females. These methylation levels were very close (<200 bp) from the TSS and within the first exon. The methylation and gene expression data together suggest that HT could affect *sox3* expression at the flexion stage through DNA hypomethylation-mediated unlocking of gene expression and continuing this state to adults through mitosis. The hypomethylation levels found in this gene at 1-y-old gonads and the higher levels of expression of this gene at HT found at earlier stages match the standard negative correlation between DNA methylation and gene expression (61). It is highly difficult to assess whether the observed methylation changes are the cause or the

consequence. However, DNA methylation is known to contribute to the acquisition and maintenance of cell identity, making it possible that changes in DNA methylation during sex differentiation contribute to stabilize the gene expression program of each sex. Regarding the exact function of *sox9a* and *sox3*, only a proper experiment involving the knockout of these genes would allow to fully understand their role in temperature-induced masculinization.

Regarding other epigenetic mechanisms, genes from the “histone modifications” GO term were always differentially expressed in all our analyses, confirming their implication in the sex-specific response to temperature (74). However, the specific up-regulation of DNA methyltransferases, which are key in the regulation of DNA methylation, was never detected. This was unexpected owing to the previous demonstration of specific methylation of promoters of both *cyp19a1* and *dmrt1* in males and females of European sea bass, respectively (9, 36). Indeed, epigenetic reprogramming mediated by changes in sexually dimorphic DNA methylation has been suggested to be a key mechanism in the determination of sexual fate in sexually isogenic species (75, 76). Furthermore, we found that key Jumonji family (Jmj) gene *jmjd1c*, involved in histone demethylation at the H3K9 site, was up-regulated in male gonads compared to female gonads. Additionally, *jmjd1c* expression was highly impacted by temperature in individuals in the middle of the eGST distribution, with up-regulation in the gonads of HT-exposed individuals and down-regulation in those exposed to LT. Interestingly, this signal was still detectable in gonads of LT fish even after the LT treatment had ended (e.g., fish from the LT treatment were at 21 °C from day 60). It has been hypothesized that temperature may reset epigenetic marks thus redirecting sexual fate (1), supporting our observed results. In certain reptilian species, some specific and related histone demethylases (JMJD3, JARID2, and KDM6B) were shown to be crucial in the shaping of the gonadal phenotype during temperature-induced sex determination (10, 77). Furthermore, although the results presented here are not sufficient to infer the functionality of genes *bcor*, *wiz*, and *auts2*, this study suggests that these genes could be involved in the transduction of temperature signals influencing sexual fate in teleosts, which warrants further examination.

To conclude, we found evidence that sex reversal rather than genotype-specific mortality was the cause of some mismatches between the sex predicted by the eGST and the actual phenotypic sex. We did not find any evidence that stress-axis activation was involved in masculinization. Rather, our data support the involvement of conserved sex-related pathways, epigenetic and energetic processes as the key to understanding temperature-induced masculinization in the European sea bass. We propose a model where the GST of individuals drives the specific expression of genes involved in lipid and glucose metabolism independently of temperature (Fig. 6). In that scheme, individuals presenting higher energy content (transduced by higher transcript levels of genes involved in lipids and glucose production) would become females and those with lower energy would become males (Fig. 6). This may explain the early sexually dimorphic differences in growth usually observed in European sea bass and why domestication (and selection for growth) leads to an increased proportion of females (78). Once the sex is fixed, a classical cascade of genes involved in sex differentiation is activated, starting by *sox9a*. However, if fish are exposed very early to high temperatures, this first triggers an up-regulation of *sox3* that is then followed by an up-regulation of *sox9a*, determining the sex of individuals in the middle of the eGST distribution (Fig. 6). This signal is then maintained, likely thanks to epigenetic processes (DNA methylation and histone modifications, e.g., through *jmjd1c*), which leads to individuals with intermediate eGST values developing as males.

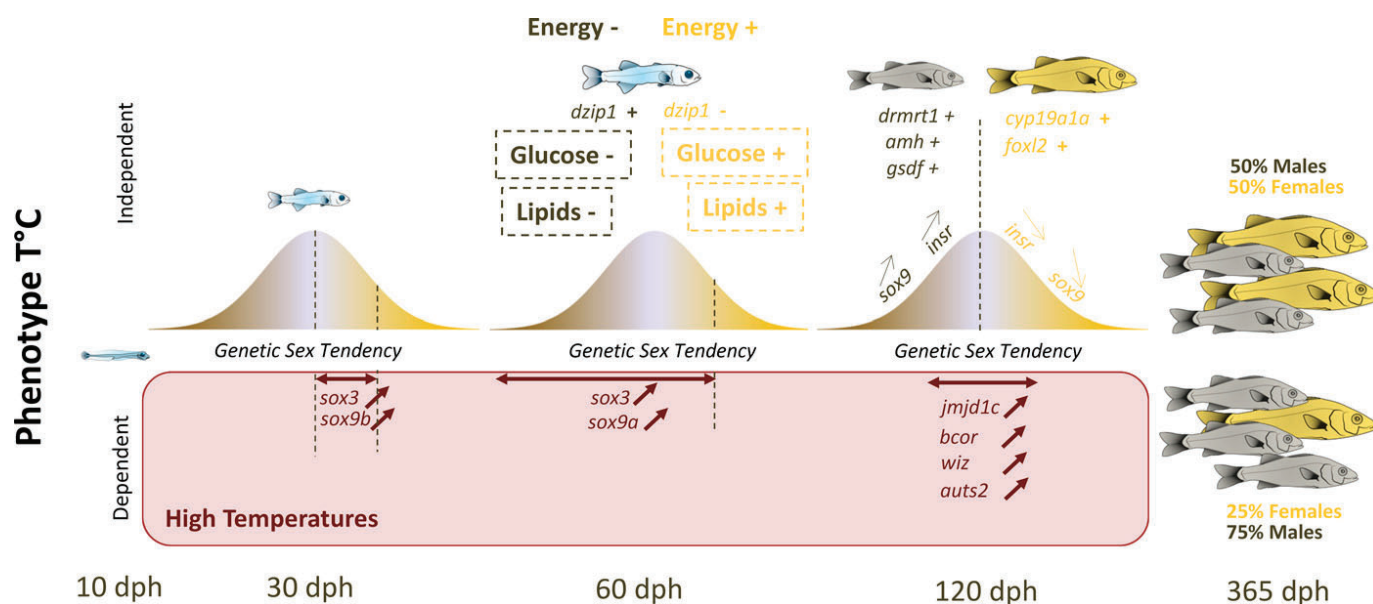


Fig. 6. Summary of the polygenic sex determination system of the European sea bass. In the *Top* panel (temperature-independent sexual phenotype), all fish have their sexual genotype eGST within a Gaussian-like distribution, with those that present a male tendency (maroon) in the left extreme and those that present a female tendency (yellow) in the right extreme. The genes involved in lipid and glucose (in dashed square) regulation are correlated to the eGST, which likely explains the difference in energy content between future males and future females at the “all fins” stage (around 60 dph). This “energetic” information likely drives sex determination, starting with the overexpression of *sox9a*. At 120 dph, the gonad is undergoing molecular sex differentiation involving classical sex pathways. Those in the middle of the distribution can still change sex. Now, if European sea bass are kept at high temperature, this results in the overexpression of *sox3* and *sox9b* for individuals with a relatively “low” female eGST at 30 dph and that is conserved at 60 dph, followed by an increase in the expression of *sox9a*. This sexual phenotype is then likely maintained thanks to the overexpression of key genes involved in histone modification at high, but not low, temperatures for individuals in the middle of the eGST distribution. Following these processes, 25% of the population is then likely masculinized at HT. The double-headed horizontal brown arrows depict individuals with a specific eGST for which we observe increased gene expression at high temperature.

From an adaptive and evolutionary point of view, the conditions favoring the emergence of ESD over GSD have been extensively discussed (28, 79). But species where both strategies coexist, with additive effects, provide new challenges for evolutionary scientists. Understanding how external factors can override genetic information, affecting an essential trait such as phenotypic sex is indeed of major importance in a global warming context. Furthermore, the insights provided by this study on European sea bass can help to illuminate other systems where temperature, stress, or other environmental cues can override a GSD system, as reported in vertebrates, from fish to mammals (80). The present study, therefore, helps pave the way for our understanding of such mechanisms; the genetic architecture likely provides information linked to the regulation of energy and growth, constituting a strong example in support of the hypothesis linking metabolism and sex determination (81). Any environmental cues that might affect this relationship (e.g., temperature influencing metabolism) will trigger a specific expression of a cascade of genes that will, in turn, affect the phenotypic sex of sensitive genotypes.

Materials and Methods

SI Appendix provides a detailed description of the materials and methods used in this study.

Fish Production and Rearing. The fish population used was the result of a mating design including eight males and one female from a West Mediterranean Sea strain of European sea bass, performed by artificial fertilization (March 22, 2017). Fertilized eggs were incubated at 14 °C until 48 h postfertilization. Eggs were then evenly dispatched in six tanks of 500 L each, and the temperature was gradually increased from 14 °C to 16 °C within 1 d. Following hatching, fish density was of 50 larvae per liter. Larvae were then exposed to 21 °C (HT) in triplicates or kept at 16 °C (LT) in triplicates, as described in ref. 82. HT treatment consisted in gradually increasing temperature to reach 21 °C, from 3 dph to 8 dph. From 10 dph onwards, fish were fed *Artemia nauplii* for 40 d, then

weaned on a commercial sea bass diet (Pro Start and Pro Wean, BioMar). For the LT treatment, the temperature was also increased from day 59 (1 °C/day) to day 64, to reach 21 °C. Fish were reared at the Ifremer Plateforme Expérimentale d'Aquaculture (Palavas-les-Flots, France), accredited to use and breed laboratory animals (no. C341926), and the project was approved by the Animal Care Committee no. 36 under project authorization no. APAFIS 19676.

Sampling. Fish were evenly sampled in the triplicates of each temperature treatment, at four different developmental stages. Since fish growth is favored at HT, the development of fish generally greatly differs between thermal treatments. Hence, to enable data comparison between individuals kept at different temperatures, the first three samplings were carried out at the same sum of degree-day (base 10 °C, DD_{10°C}), a procedure previously used to allow standardized measurement of growth in fishes (83). Thus, larvae were not collected at the same date but at the same DD_{10°C}: 77 DD_{10°C}, 242 DD_{10°C}, and 550 DD_{10°C} (reference *SI Appendix, Table S4* for details regarding stage, age, and size). The fourth and last sampling of juveniles, aiming at obtaining developing gonads, was, however, based on size rather than age, since growth difference between treatments declined with age (*SI Appendix, Fig. S1 and Table S4*). For this reason, gonads were sampled on juveniles at 117 dph (HT) or 124 dph (LT). At the end of the experiment, after 1 y (at 390 dph), all fish were euthanized using benzocaine (150 mg/L), measured, weighed, and sexed by *in situ* gonad examination. All samples were genotyped with an SNP chip (see *Genotyping*), but only subsamples of those performed at 242 DD_{10°C}, 550 DD_{10°C}, and 117 to 124 dph were used for molecular analysis (*SI Appendix, Table S4*).

Genotyping. We genotyped three generations of European sea bass using the Thermofisher DlabChip European sea bass array of 57,000 SNP markers (39). Generation 0 includes the parents of the eight sires. Generation 1 corresponds to the parents (the dam and the eight sires). Generation 2 includes the 2,030 offspring, composed of the larvae sampled at 77 DD_{10°C} (*n* = 192), at 242 DD_{10°C} (*n* = 300), and at 550 DD_{10°C} (*n* = 280), the juveniles collected at 117 dph (HT, *n* = 70) and 124 dph (LT, *n* = 70), and fish sexed at 390 dph (*n* = 1,118). Fish were sampled at random at each stage, and the number of fish per family was known a posteriori following genotyping and parentage assignment (*SI Appendix, Table S4*). Complete details of the genotyping analysis can be found in *SI Appendix, Supplementary Materials and Methods*.

Heritability, eGST Prediction, and Genome-Wide Association Scan. We predicted the sex using a liability threshold model, which has been used in a large variety of settings, often with diseases (29, 84) but also for sex determination (28, 32). With this model, the binary phenotype is the realization of an underlying continuous phenotype, the liability trait, here called phenotypic sex tendency (PST). When PST exceeds a given threshold, animals differentiate as females, while they differentiate as males when PST remains below the threshold (*SI Appendix, Fig. S4*). The PST is itself the addition of a genetic and an environmental sex tendency. In a polygenic sex determination system, the PST is considered a quantitative trait, influenced by many genes with small effects plus environmental effects (28). The GST, which cannot be measured directly, is the genetic part of this PST, positive for individuals more likely to develop as females in a neutral environment and negative for individuals more likely to develop as males (*SI Appendix, Fig. S4*). Variance components and heritability were estimated for PST, the underlying liability of the binary sex, with a threshold model using THRGIBBS1F90 (85). Complete details for the establishment of eGST, QTL presence, and heritability assessment can be found in *SI Appendix, Supplementary Materials and Methods*.

Transcriptomics. The transcriptomic approach aimed to detect genes whose expression would be correlated to the eGST. Since we expected to have fewer females at HT versus LT, we collected more individuals in the HT than in the LT treatment to ensure having a sufficient number of females to analyze. Overall, 70 larvae, 40 from the HT group and 30 from the LT group, were randomly collected at the “flexion” (242 DD_{10°C}) and the “all fins” stages (550 DD_{10°C}), euthanized (benzocaine 150 mg/L), and snap frozen in liquid nitrogen. At these stages, the RNA extraction was performed on entire individuals because of the impossibility to neatly dissect the gonads of such small individuals. This was justified from two points of view: 1) it has been previously shown that the chance of underestimating gene expression of sex-related genes using body trunks was negligible in sea bass (86) and 2) we aimed to have a whole picture of the physiological processes associated with temperature sex determination (TSD). At the juvenile stage (117 to 124 dpf), the gonads of 35 euthanized (benzocaine 150 mg/L) individuals from each temperature treatment were collected and snap frozen in liquid nitrogen. At this stage, the biggest individuals were collected to ensure the sampling/collection of enough gonadal tissue for transcriptomic analysis. All samples were stored at -80°C until RNA extraction. Complete details for RNA extraction, RNA-Seq, and RNA-Seq data analysis can be found in *SI Appendix, Supplementary Materials and Methods*. The raw data of RNA-seq can be downloaded from ref. 87.

Histology. At the juvenile stage, 20 gonads (10 HT and 10 LT) were fixed in Bouin's fluid for 6 to 8 h, rinsed in clear water for 1 h, and stocked in a 70% alcohol solution. Each gonad was stained with eosin and then placed in agarose (to improve detection) before being dehydrated and embedded in paraffin. Sections of 5- to 6-mm thickness were stained with Trichrome de Masson, Haematoxylin Groat, Fuschine Ponceau, and Aniline blue using an automated device.

Elemental Analysis. Elemental analysis was performed to estimate the energy content of fish with a known eGe. Elemental composition was determined using a microVario Elemental analyzer (Elementar), allowing to obtain the percentage of carbon, hydrogen, nitrogen, sulfur, and oxygen (CHNSO) per milligram of dry weight. In total, 70 sampled fish were weighted and measured before being euthanized (benzocaine 150 mg/L) and kept at -20°C . Then, fish at each stage (“flexion,” “all fins,” and juveniles) were lyophilized simultaneously for 24 h. Dried fish were then individually homogenized using ball mills. About 1 mg was used for CHNS analyses. Another 1 mg of the same samples was used for oxygen analysis. The analysis was performed by the laboratory of Physical measures (<https://imp.edu.umontpellier.fr/elem>) of Montpellier (certified Association Française de Normalisation, Iso 9001). Individual elemental composition was then transformed to a relative energetic content ($\text{cal} \cdot \text{mg}^{-1}$ dry weight) using the Given (88) formula, as recommended by the analyzer program:

$$\begin{aligned} \text{Relative energetic content (cal.mg}^{-1}\text{)} \\ = ((78.3\text{C} + 339.1\text{H} - 330 + 22\text{S}) + 152)/1000, \end{aligned}$$

where C, H, O, and S refer to the percentages of carbon, hydrogen, oxygen, and sulfur in a sample. It was then divided by 4.184 to obtain the energetic content in $\text{joules} \cdot \text{mg}^{-1}$.

RRBS Library Preparation and Analysis. Genomic DNA was isolated from ~25 mg frozen gonadal tissue from the 1-y-old fish with the Qiagen Blood and Cell Culture kit (catalog no. 13323). Genome-wide profiling of DNA methylation levels was performed by RRBS using the Premium RRBS Kit (catalog no. C02030033; Diagenode) according to the manufacturer's instructions.

Complete details for RRBS sequencing and preliminary analysis can be found in *SI Appendix, Supplementary Materials and Methods*.

Data Analysis. Final sex ratio was analyzed using a binomial mixed model, with “tank” added as a random factor and “initial temperature treatment” as a fixed factor. For the three transcriptomes datasets of fish at the “flexion” stage, “all fins” stage, and juvenile stage, we adopted a traditional approach based on group comparison, where differentially expressed genes are detected using the Bioconductor (89) package DESeq2 version 1.18.1 (90) in R (91) and data normalized using the default method of DESeq2. Regarding the 10 transcriptomes at the “flexion” stage, group comparisons were performed between LT ($n = 5$) and HT ($n = 5$) fish with a positive but not extreme eGST (i.e., $0 < \text{eGST} < 0.5$). Regarding the 68 transcriptomes at the “all fins” stage, individuals were grouped based on their eGST, within each initial temperature treatment (note that they were collected at the same temperature [T, $^{\circ}\text{C}$]). Those at each extreme of the eGST distribution were considered as extremes genetic males (eM, $\text{eGST} < -0.5$) or extreme genetic females (eF, $\text{eGST} > 0.5$), respectively, and those in the middle of the distribution were considered neutrals males (nM, $-0.5 < \text{eGST} < 0$) or neutrals females (nF, $0 < \text{eGST} < 0.5$). Group comparisons between temperature treatments with DESeq2 were thus performed between eight groups, eM-HT ($n = 6$), nM-HT ($n = 10$), nF-HT ($n = 12$), eF-HT ($n = 11$), eM-LT ($n = 8$), nM-LT ($n = 6$), nF-LT ($n = 13$), and eF-LT ($n = 3$). For juveniles' gonads, we considered only two groups, those with an $\text{eGST} > 0$ (genetic females, $n = 26$) and those with an $\text{eGST} < 0$ (genetic males, $n = 17$) to detect differentially expressed genes according to the genetic sex. For all comparisons, genes with an adjusted P value less than 5% (according to the false discovery rate method from Benjamini-Hochberg) were declared differentially expressed.

For the two transcriptomic datasets with the highest number of samples (43, 68), we also considered another approach using our linear predictor, the eGST, as a fixed value in linear models. Using normalized (DESeq2) and filtered data (keeping those with more than 30 reads in average), we ran a loop in R (R version 4.0.4) to detect all genes that are significantly correlated to the eGST in a linear and a quadratic model where the temperature treatment (HT or LT) is added as an interaction term with the eGST ($T \times \text{eGST}$). For the linear model, we considered only the genes without significant interaction with temperature. Once detected, we also ran a linear model with a quasibinomial distribution (link = Logit) for those interesting genes displaying a dichotomic pattern of expression. For the quadratic model, we considered either the genes with a significant correlation with eGST plus a significant quadratic term for the interaction ($(\text{eGST}^2) \times T$) or those where the quadratic term ((eGST^2)) plus the interaction between the temperature treatment and the quadratic term ($(\text{eGST}^2) \times T$) are significant. One outlier (LT) was removed since it was at the extreme for eGST with unusual intermediate values for all sex-related genes, though keeping this outlier resulted in similar outcomes for those genes (*SI Appendix, Fig. S11*). Linear models were also used to assess the effect of eGST on energy content ($\text{joules} \cdot \text{mg}^{-1}$) and the size of the fishes.

For all transcriptomic analyses, the GO dataset of the mouse was used as it gives a more complete overview of the genes involved in each pathway. We adopted two strategies: 1) with a priori, where GOs were selected based on our hypotheses (*SI Appendix, Supplementary Materials and Methods*); and 2) without a priori, where GOs were automatically detected from a functional enrichment analysis. For this second approach, we used a functional enrichment analysis. The clusterProfiler version 3.16.1 (92), an R package, was used to analyze function profiles of genes to identify major biological functions of genes. GO terms were predicted based on differentially expressed genes (from the group comparisons and linear models), including biological process and cellular component categories. Enrichment analysis was performed and a P value < 0.05 was considered to indicate a statistically significant difference. PCA was performed using gene expression levels and illustrated using the ggplot2 package. The same package was used for representing raw values of genes for linear, quadratic regressions and selected genes from group comparisons. The “lme4” package was used for mixed models. Heatmaps of standardized expressions (mean subtraction followed SD division) were created using the pheatmap package. MethyKit and Genomic Ranges version 1.44.0 (93) packages were used to filter percent methylation data for the target regions: gene body $\pm 2,000$ bp. DMCs were defined as cytosines with a methylation difference of $\geq 15\%$ between temperature treatments and a significance threshold of q -value < 0.01 .

Data Availability. RNA-Seq and .xls tables data have been deposited in a publicly accessible database: SEXTANT (<https://sextant.ifremer.fr/record/5cb1485f-385e-46e9-9cc1-4b4d77802183/>). All other study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. The study was supported by the European Maritime and Fisheries Fund (35, Seabass Sex and Stress, Grant No. 4320175237) allocated to B.G. Production of the fish benefited from AQUAEXCEL²⁰²⁰ Transnational access grant "Transsexbass" to F.P. and M.V. Research at the F.P. laboratory was supported by Spanish Ministry of Science Grant No.

1. B. Capel, Vertebrate sex determination: Evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet.* **18**, 675–689 (2017).
2. M. Charnier, Action of temperature on the sex ratio in the Agama agama (Agamidae, Lacertilia) embryo. *C. R. Seances Soc. Biol. Fil.* **160**, 620–622 (1966).
3. D. O. Conover, B. E. Kynard, Environmental sex determination: Interaction of temperature and genotype in a fish. *Science* **213**, 577–579 (1981).
4. R. H. Devlin, Y. Nagahama, Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture* **208**, 191–364 (2002).
5. B. Geffroy, A. Bardonnet, Sex differentiation and sex determination in eels: Consequences for management. *Fish Fish.* **17**, 375–398 (2016).
6. N. Ospina-Álvarez, F. Piferrer, Temperature-dependent sex determination in fish revisited: Prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS One* **3**, e2837 (2008).
7. U. Römer, W. Beisenherz, Environmental determination of sex in Apistogramma (Cichlidae) and two other freshwater fishes (Teleostei). *J. Fish Biol.* **48**, 714–725 (1996).
8. R. S. Hattori, D. C. Castañeda-Cortés, L. F. Arias Padilla, P. H. Strobl-Mazzulla, J. I. Fernandino, Activation of stress response axis as a key process in environment-induced sex plasticity in fish. *Cell. Mol. Life Sci.* **77**, 4223–4236 (2020).
9. F. Piferrer *et al.*, The model of the conserved epigenetic regulation of sex. *Front. Genet.* **10**, 857 (2019).
10. I. W. Deveson *et al.*, Differential intron retention in *Jumonji* chromatin modifier genes is implicated in reptile temperature-dependent sex determination. *Sci. Adv.* **3**, e1700731 (2017).
11. A. Georges, C. E. Holleley, How does temperature determine sex? *Science* **360**, 601–602 (2018).
12. B. Geffroy, M. Douhard, The adaptive sex in stressful environments. *Trends Ecol. Evol.* **34**, 628–640 (2019).
13. B. Geffroy, C. Wedekind, Effects of global warming on sex ratios in fishes. *J. Fish Biol.* **97**, 596–606 (2020).
14. J. F. Baroiller, H. D'Cotta, E. Bezault, S. Wessels, G. Hoerstgen-Schwark, Tilapia sex determination: Where temperature and genetics meet. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **153**, 30–38 (2009).
15. K. Tabata, Reduction of female proportion in lower growing fish separated from normal and feminized seedlings of hiram *Paralichthys olivaceus*. *Fish. Sci.* **61**, 199–201 (1995).
16. S. Santi *et al.*, Thermosensitivity of the sex differentiation process in the African catfish, *Clarias gariepinus*: Determination of the thermosensitive period. *Aquaculture* **455**, 73–80 (2016).
17. R. S. Hattori *et al.*, The duplicated Y-specific *amhy* gene is conserved and linked to maleness in silversides of the genus *Odontesthes*. *Genes (Basel)* **10**, 679 (2019).
18. T. Kitano, Y. Hayashi, E. Shiraishi, Y. Kamei, Estrogen rescues masculinization of genetically female medaka by exposure to cortisol or high temperature. *Mol. Reprod. Dev.* **79**, 719–726 (2012).
19. K. Valdivia *et al.*, High temperature increases the masculinization rate of the all-female (XX) rainbow trout "Mal" population. *PLoS One* **9**, e113355 (2014).
20. H. Zhou *et al.*, Temperature-control-induced masculinization in tiger puffer *Takifugu rubripes*. *J. Oceanol. Limnol.* **37**, 1125–1135 (2019).
21. R. S. Hattori *et al.*, Cortisol-induced masculinization: Does thermal stress affect gonadal fate in pejerrey, a teleost fish with temperature-dependent sex determination? *PLoS One* **4**, e6548 (2009).
22. Y. Hayashi *et al.*, High temperature causes masculinization of genetically female medaka by elevation of cortisol. *Mol. Reprod. Dev.* **77**, 679–686 (2010).
23. T. Kamiya *et al.*, A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLoS Genet.* **8**, e1002798 (2012).
24. K. Miyoshi, R. S. Hattori, C. A. Strüßmann, M. Yokota, Y. Yamamoto, Phenotypic/genotypic sex mismatches and temperature-dependent sex determination in a wild population of an Old World atherinid, the cobaltcap silverside *Hypoatherina tsurugae*. *Mol. Ecol.* **29**, 2349–2358 (2020).
25. A. Yano *et al.*, The sexually dimorphic on the Y-chromosome gene (*sdY*) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evol. Appl.* **6**, 486–496 (2013).
26. J. F. Baroiller, D. Chourrou, A. Fostier, B. Jalabert, Temperature and sex chromosomes govern sex ratios of the mouthbrooding Cichlid fish *Oreochromis niloticus*. *J. Exp. Zool.* **273**, 216–223 (1995).
27. X. Wang *et al.*, High temperature causes masculinization of genetically female olive flounder (*Paralichthys olivaceus*) accompanied by primordial germ cell proliferation detention. *Aquaculture* **479**, 808–816 (2017).
28. M. G. Bulmer, J. J. Bull, Models of polygenic sex determination and sex ratio control. *Evolution* **36**, 13–26 (1982).
29. M. L. A. Hujoel, S. Gazal, P.-R. Loh, N. Patterson, A. L. Price, Liability threshold modeling of case-control status and family history of disease increases association power. *Nat. Genet.* **52**, 541–547 (2020).
30. F. Piferrer, M. Blázquez, L. Navarro, A. González, Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (*Dicentrarchus labrax* L.). *Gen. Comp. Endocrinol.* **142**, 102–110 (2005).
31. E. Saillant *et al.*, Temperature effects and genotype-temperature interactions on sex determination in the European sea bass (*Dicentrarchus labrax* L.). *J. Exp. Zool.* **292**, 494–505 (2002).
32. M. Vandeputte, M. Dupont-Nivet, H. Chavanne, B. Chatain, A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. *Genetics* **176**, 1049–1057 (2007).
33. M. Vandeputte *et al.*, Low temperature has opposite effects on sex determination in a marine fish at the larval/postlarval and juvenile stages. *Ecol. Evol.* **10**, 13825–13835 (2020).
34. N. Díaz, L. Ribas, F. Piferrer, The relationship between growth and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *Aquaculture* **408–409**, 191–202 (2013).
35. N. Díaz, F. Piferrer, Lasting effects of early exposure to temperature on the gonadal transcriptome at the time of sex differentiation in the European sea bass, a fish with mixed genetic and environmental sex determination. *BMC Genomics* **16**, 679 (2015).
36. L. Navarro-Martin *et al.*, DNA methylation of the gonadal aromatase (*cyp19a*) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLoS Genet.* **7**, e1002447 (2011).
37. M. Vandeputte, *et al.*, Low temperature has opposite effects on sex determination in a marine fish at the larval/postlarval and juvenile stages. *Ecol. Evol.* **10**, 13825–13835 (2020).
38. D. Anastasiadi, M. Vandeputte, N. Sánchez-Baizán, F. Allal, F. Piferrer, Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a fish with mixed genetic and environmental sex determination. *Epigenetics* **13**, 988–1011 (2018).
39. R. Griot *et al.*, Genome-wide association studies for resistance to viral nervous necrosis in three populations of European sea bass (*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip. *Aquaculture* **530**, 735930 (2021).
40. T. H. Meuwissen, B. J. Hayes, M. E. Goddard, Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
41. C. Roblin, J. Bruslé, Ontogenèse gonadique et différenciation sexuelle du loup *Dicentrarchus labrax*, en conditions d'élevage. *Reprod. Nutr. Dev.* **23**, 115–127 (1983).
42. J. Odegård, T. H. Meuwissen, Estimation of heritability from limited family data using genome-wide identity-by-descent sharing. *Genet. Sel. Evol.* **44**, 16 (2012).
43. L. Ribas *et al.*, Characterization of the European sea bass (*Dicentrarchus labrax*) gonadal transcriptome during sexual development. *Mar. Biotechnol. (NY)* **21**, 359–373 (2019).
44. S. Faggion, M. Vandeputte, B. Chatain, P.-A. Gagnaire, F. Allal, Population-specific variations of the genetic architecture of sex determination in wild European sea bass *Dicentrarchus labrax* L. *Heredity* **122**, 612–621 (2019).
45. C. Palaiokostas *et al.*, A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*). *Genet. Sel. Evol.* **47**, 68 (2015).
46. M. Blázquez, S. Zanuy, M. Carillo, F. Piferrer, Effects of rearing temperature on sex differentiation in the European sea bass (*Dicentrarchus labrax* L.). *J. Exp. Zool.* **281**, 207–216 (1998).
47. M. Pavlidis *et al.*, Evidence of temperature-dependent sex determination in the European sea bass (*Dicentrarchus labrax* L.). *J. Exp. Zool.* **287**, 225–232 (2000).
48. L. Navarro-Martin, M. Blázquez, J. Vinas, S. Joly, F. Piferrer, Balancing the effects of rearing at low temperature during early development on sex ratios, growth and maturation in the European sea bass (*Dicentrarchus labrax*): Limitations and opportunities for the production of highly female-biased stocks. *Aquaculture* **296**, 347–358 (2009).
49. C. A. Alper, Z. Awdeh, Incomplete penetrance of MHC susceptibility genes: Prospective analysis of polygenic MHC-determined traits. *Tissue Antigens* **56**, 199–206 (2000).
50. D. N. Cooper, M. Krawczak, C. Polychronakos, C. Tyler-Smith, H. Kehrer-Sawatzki, Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132**, 1077–1130 (2013).
51. E. Saillant *et al.*, Sexual differentiation and juvenile intersexuality in the European sea bass (*Dicentrarchus labrax*). *J. Zool. (Lond.)* **260**, 53–63 (2003).
52. B. Geffroy, Y. Guiguen, A. Fostier, A. Bardonnet, New insights regarding gonad development in European eel: Evidence for a direct ovarian differentiation. *Fish Physiol. Biochem.* **39**, 1129–1140 (2013).
53. J. Kent, S. C. Wheatley, J. E. Andrews, A. H. Sinclair, P. Koopman, A male-specific role for SOX9 in vertebrate sex determination. *Development* **122**, 2813–2822 (1996).
54. S. Nef *et al.*, Testis determination requires insulin receptor family function in mice. *Nature* **426**, 291–295 (2003).
55. I. Stévant, S. Nef, Genetic control of gonadal sex determination and development. *Trends Genet.* **35**, 346–358 (2019).
56. M. Blázquez, L. Navarro-Martin, F. Piferrer, Expression profiles of sex differentiation-related genes during ontogenesis in the European sea bass acclimated to two different temperatures. *J. Exp. Zool. B Mol. Dev. Evol.* **312**, 686–700 (2009).

57. N. Moreno-Mendoza, V. R. Harley, H. Merchant-Larios, Temperature regulates SOX9 expression in cultured gonads of *Lepidochelys olivacea*, a species with temperature sex determination. *Dev. Biol.* **229**, 319–326 (2001).
58. S. Radhakrishnan, R. Litterman, J. Neuwald, A. Severin, N. Valenzuela, Transcriptomic responses to environmental temperature by turtles with temperature-dependent and genotypic sex determination assessed by RNAseq inform the genetic architecture of embryonic gonadal development. *PLoS One* **12**, e0172044 (2017).
59. C. Shoemaker, M. Ramsey, J. Queen, D. Crews, Expression of Sox9, Mis, and Dmrt1 in the gonad of a species with temperature-dependent sex determination. *Dev. Dyn.* **236**, 1055–1063 (2007).
60. D. Sun et al., Sox9-related signaling controls zebrafish juvenile ovary-testis transformation. *Cell Death Dis.* **4**, e930 (2013).
61. L. D. Moore, T. Le, G. Fan, DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).
62. D. Anastasiadi, A. Esteve-Codina, F. Piferrer, Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* **11**, 37 (2018).
63. S. Urs et al., Sprouty1 is a critical regulatory switch of mesenchymal stem cell lineage allocation. *FASEB J.* **24**, 3264–3273 (2010).
64. S. Urs, T. Henderson, P. Le, C. J. Rosen, L. Liaw, Tissue-specific expression of Sprouty1 in mice protects against high-fat diet-induced fat accumulation, bone loss and metabolic dysfunction. *Br. J. Nutr.* **108**, 1025–1033 (2012).
65. P. O. Prada et al., EGFR tyrosine kinase inhibitor (PD153035) improves glucose tolerance and insulin action in high-fat diet-fed mice. *Diabetes* **58**, 2910–2919 (2009).
66. Z. Koledova et al., SPRY1 regulates mammary epithelial morphogenesis by modulating EGFR-dependent stromal paracrine signaling and ECM remodeling. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5731–E5740 (2016).
67. D. Marguet et al., Enhanced insulin secretion and improved glucose tolerance in mice lacking CD26. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6874–6879 (2000).
68. S. Faggion et al., Sex dimorphism in European sea bass (*Dicentrarchus labrax* L.): New insights into sex-related growth patterns during very early life stages. *PLoS One* **16**, e0239791 (2021).
69. A. Herpin, M. Scharf, Plasticity of gene-regulatory networks controlling sex determination: Of masters, slaves, usual suspects, newcomers, and usurpaters. *EMBO Rep.* **16**, 1260–1274 (2015).
70. T. Myosho, Y. Takehana, S. Hamaguchi, M. Sakaizumi, Turnover of sex chromosomes in celebensis group medaka fishes. *G3 (Bethesda)* **5**, 2685–2691 (2015).
71. E. Sutton et al., Identification of SOX3 as an XX male sex reversal gene in mice and humans. *J. Clin. Invest.* **121**, 328–341 (2011).
72. Y. Takehana et al., Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias latipes*. *Nat. Commun.* **5**, 4157 (2014).
73. B. Yao, L. Zhou, Y. Wang, W. Xia, J.-F. Gui, Differential expression and dynamic changes of SOX3 during gametogenesis and sex reversal in protogynous hermaphroditic fish. *J. Exp. Zool. A Ecol. Genet. Physiol.* **307**, 207–219 (2007).
74. F. Piferrer, Epigenetics of sex determination and gonadogenesis. *Dev. Dyn.* **242**, 360–370 (2013).
75. N. J. Gemmell, E. V. Todd, A. Goikoetxea, O. Ortega-Recalde, T. A. Hore, “Natural sex change in fish” in *Current Topics in Developmental Biology, Sex Determination in Vertebrates*, B. Capel, Ed. (Academic Press, 2019), pp. 71–117.
76. A. Tsakogiannis et al., The gene toolkit implicated in functional sex in sparidae hermaphrodites: Inferences from comparative transcriptomics. *Front. Genet.* **9**, 749 (2019).
77. C. Ge et al., The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science* **360**, 645–648 (2018).
78. B. Geffroy et al., Parental selection for growth and early-life low stocking density increase the female-to-male ratio in European sea bass. *Sci. Rep.* **11**, 13620 (2021).
79. E. L. Charnov, J. Bull, When is sex environmentally determined? *Nature* **266**, 828–830 (1977).
80. F. Piferrer, D. Anastasiadi, Do the offspring of sex reversals have higher sensitivity to environmental perturbations? *Sex Dev.* **15**, 134–147 (2021).
81. U. Mittwoch, The elusive action of sex-determining genes: Mitochondria to the rescue? *J. Theor. Biol.* **228**, 359–365 (2004).
82. A. Goikoetxea et al., Genetic pathways underpinning hormonal stress responses in fish exposed to short- and long-term warm ocean temperatures. *Ecol. Indic.* **120**, 106937 (2021).
83. K. A. Chezik, N. P. Lester, P. A. Venturelli, Fish growth and degree-days I: Selecting a base temperature for a within-population study. *Can. J. Fish. Aquat. Sci.* **71**, 47–55 (2014).
84. O. Weissbrod, C. Lippert, D. Geiger, D. Heckerman, Accurate liability estimation improves power in ascertained case-control studies. *Nat. Methods* **12**, 332–334 (2015).
85. S. Tsuruta, I. Misztal, “THRGIBBS1F90 for estimation of variance components with threshold and linear models” in *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production* (Instituto Prociencia, 2006), pp. 27–31.
86. M. Blázquez, A. González, M. Papadaki, C. Mylonas, F. Piferrer, Sex-related changes in estrogen receptors and aromatase gene expression and enzymatic activity during early development and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *Gen. Comp. Endocrinol.* **158**, 95–101 (2008).
87. B. Geffroy, RNA-Seq juveniles de bar: Projet 3S (Seabass, Sex and Stress). Sextant. <https://doi.org/10.12770/5cb1485f-385e-46e9-9cc1-4b4d77802183>. Deposited 9 March 2018.
88. P. H. Given, D. Weldon, J. H. Zoeller, Calculation of calorific values of coals from ultimate analyses: theoretical basis and geochemical implications. *Fuel* **65**, 849–854 (1986).
89. R. C. Gentleman et al., Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
90. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
91. R Core Team, *R: A Language and Environment for Statistical Computing* (R Found. Stat. Comput. Vienna, Austria, 2021).
92. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
93. M. Lawrence et al., Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).

RESEARCH ARTICLE

Open Access



Improved biomarker discovery through a plot twist in transcriptomic data analysis

Núria Sánchez-Baizán¹, Laia Ribas¹ and Francesc Piferrer^{1*}

Abstract

Background: Transcriptomic analysis is crucial for understanding the functional elements of the genome, with the classic method consisting of screening transcriptomics datasets for differentially expressed genes (DEGs). Additionally, since 2005, weighted gene co-expression network analysis (WGCNA) has emerged as a powerful method to explore relationships between genes. However, an approach combining both methods, i.e., filtering the transcriptome dataset by DEGs or other criteria, followed by WGCNA (DEGs + WGCNA), has become common. This is of concern because such approach can affect the resulting underlying architecture of the network under analysis and lead to wrong conclusions. Here, we explore a plot twist to transcriptome data analysis: applying WGCNA to exploit entire datasets without affecting the topology of the network, followed with the strength and relative simplicity of DEG analysis (WGCNA + DEGs). We tested WGCNA + DEGs against DEGs + WGCNA to publicly available transcriptomics data in one of the most transcriptomically complex tissues and delicate processes: vertebrate gonads undergoing sex differentiation. We further validate the general applicability of our approach through analysis of datasets from three distinct model systems: European sea bass, mouse, and human.

Results: In all cases, WGCNA + DEGs clearly outperformed DEGs + WGCNA. First, the network model fit and node connectivity measures and other network statistics improved. The gene lists filtered by each method were different, the number of modules associated with the trait of interest and key genes retained increased, and GO terms of biological processes provided a more nuanced representation of the biological question under consideration. Lastly, WGCNA + DEGs facilitated biomarker discovery.

Conclusions: We propose that building a co-expression network from an entire dataset, and only thereafter filtering by DEGs, should be the method to use in transcriptomic studies, regardless of biological system, species, or question being considered.

Keywords: Gene expression analysis, Gene networks, Weighted gene co-expression network analysis (WGCNA), Sex determination and differentiation, Gonadal development, Biomarker discovery

Background

The transcriptome is defined as the entire set of messenger RNA (mRNA or transcripts) expressed by a cell or tissue type of an organism of a given genotype under certain internal and external influences. The total amount and types of transcripts vary depending on multiple factors such as stage of development, physiological state, or environmental conditions [1].

*Correspondence: piferrer@icm.csic.es

Institut de Ciències del Mar (ICM), Spanish National Research Council (CSIC), Barcelona, Passeig Marítim, 37-49, 08003 Barcelona, Spain



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

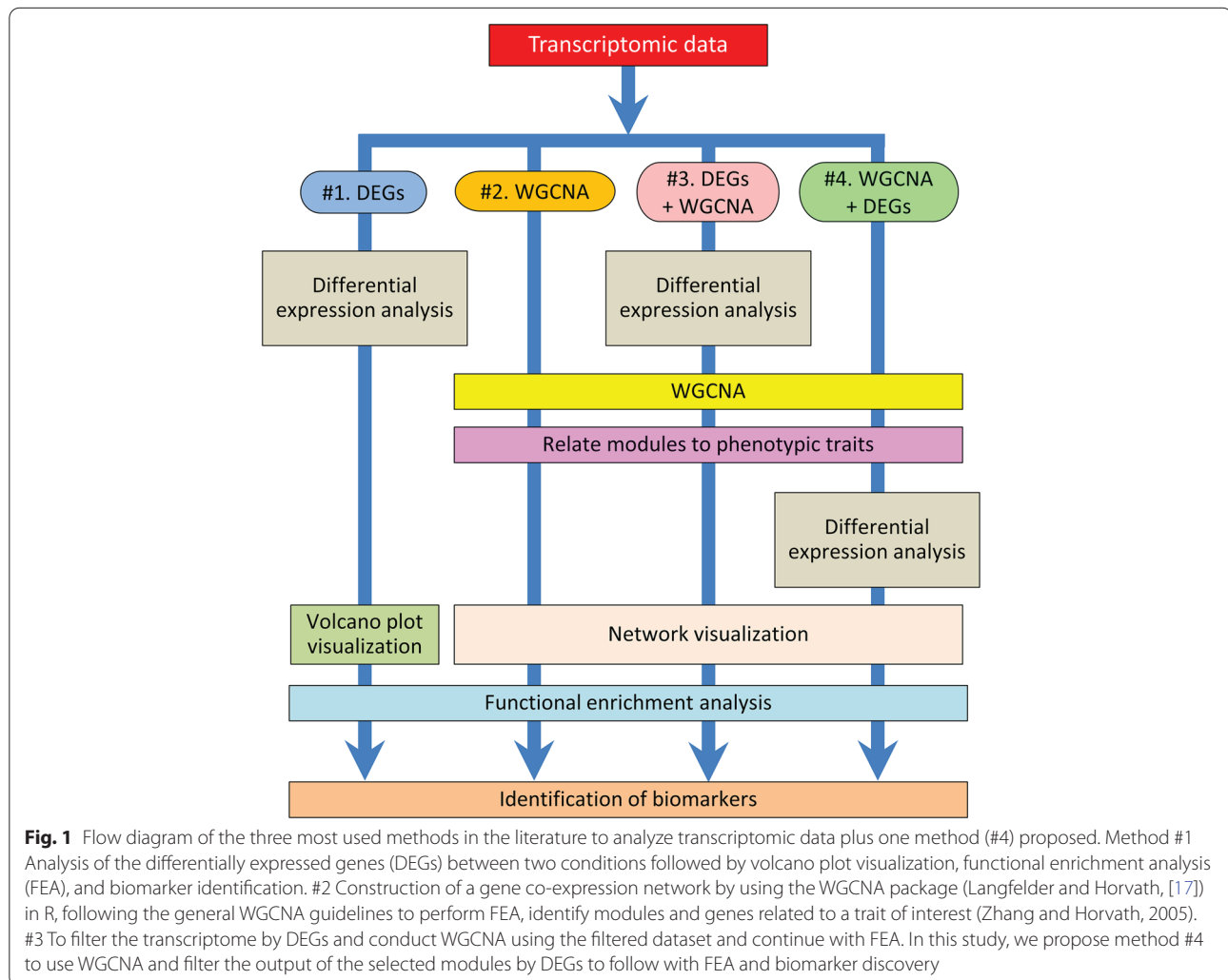
Transcriptomics, the analysis of the whole transcriptome, started using complementary DNA (cDNA) clones to generate expressed sequence tags (ESTs), and the development of the first microarrays in the early 90s [2–4]. An accurate comprehension of the transcriptome is essential to reveal the molecular constituents of cells and tissues, to understand the functional elements of the genome, and also to gain a better understanding of development and disease [5]. With the advent of high-throughput technologies to analyze the transcriptome, several bioinformatic challenges appeared in storing, pre-processing, and analyzing the large datasets produced. As a consequence, a variety of bioinformatic pipelines arose to overcome such challenges [6, 7]. To date, the most commonly used method in various fields of biology and medicine is screening for differentially expressed genes (DEGs), which compares the mean expression levels for individual genes between two or more groups of samples [8–15].

In 2005, Zhang and Horvath proposed a bioinformatic application called weighted gene co-expression network analysis (WGCNA) to gain a deeper understanding of the transcriptome and to elucidate the underlying cellular processes based on the coordinated co-expression of genes encoding the interacting proteins [16]. In contrast to DEG analysis, WGCNA is a gene screening method that takes advantage of the inherent variability in gene expression among different biological samples to illuminate higher-order relationships among genes. WGCNA generates clusters (called modules) of strongly correlated genes based on the Pearson correlation as a measure of their functional relatedness and assigns a different color to each module for easy identification purposes. WGCNA is a powerful method and has many advantages over DEG analysis since it also allows evaluating the association of modules with phenotypic sample traits using network properties. Furthermore, WGCNA facilitates the identification of candidate biomarkers and hub genes relevant to the process under study. The weighted co-expression network consists of an adjacency matrix reporting the connection strength between gene pairs [16, 17]. Since its development, its use has become exponentially widespread and allows integrating network parameters with genetic information from microarray datasets and, more recently, from RNA sequencing experiments [18–20]. It should be noted that variations or methods other than WGCNA have also been developed, such as Differential Co-expression Analysis or metaDCN [21], THD-Module Extractor [22], Diffcoex [23], and module differential analysis for weighted gene co-expression network (MODA) [24]. However, WGCNA remains by far the most commonly used in numerous research fields.

In reviewing the literature extensively, we found three main strategies to analyze the transcriptome (Fig. 1) that can be broadly classified as (1) screening for differentially expressed genes between two conditions (from now on referred to as method #1 or simply as DEGs), (2) WGCNA and derivatives (from now on referred as method #2 or simply as WGCNA), and (3) a method that has become also very popular that can be viewed as a combination of the two former and consists in filtering the transcriptome dataset by DEGs or other criteria, e.g., considering only the most expressed genes, or the top 25% genes with more expression variance, etc., and only then applying WGCNA to the filtered dataset (from now on referred to as method #3 or DEGs + WGCNA). Another strategy that we have found in reviewing the literature is the independent use of methods #1 (DEGs) and #2 (WGCNA) in the same study but then using only one of the two for the subsequent downstream analysis on the data [13, 25–32]. Some examples of studies following each one of the three broad methods defined above are shown in Table 1. Of note, regardless of the method used, usually, these analyses are followed by data visualization and functional analysis (gene ontology and/or pathways enrichment analysis).

The emergence of method #3 (DEGs + WGCNA) apparently seems a logical forward step to take: by first filtering DEGs, the analysis is limited to a set of previously selected genes, based on a statistical or quantitative threshold and, thus, it requires less computational power than the required by WGCNA of the entire dataset. However, and this is very important for the proper analysis of gene interactions, in a scale-free network (i.e., a network whose characteristics are independent of the size or number of nodes making up the network) such as the gene networks, the fraction of nodes with degree k , where k is connectivity (i.e., the sum of connections of a node), follows a power-law $k^{-\alpha}$ distribution, where α is some exponential. In this manner, the network topology is dominated by a few highly connected nodes, called hubs, linked to the rest of less connected nodes [16]. Therefore, applying a filtering step such as DEGs previous to WGCNA, as done in method #3, might eliminate many of the less connected genes that, if not filtered, would contribute to some nodes being identified as hubs. This is of great concern because it thus could affect the core architecture of the network and lead to biased results and interpretations. In contrast, construction of a network with the unfiltered dataset allows drawing the complete map of the network.

As we have just seen, method #3 has the main advantage that one ends up dealing with a much-reduced gene list, albeit all genes are DEGs, but it has serious dangers that can affect the architecture of the networks and lead



to false conclusions. Furthermore, the validity of method #3 was never established and, in fact, the developers of method #2 (WGCNA) recommend against the application of such filter before WGCNA [61]. Thus, to combine the benefits from WGCNA, taking advantage of the whole dataset and not affecting the topology of the network, with the strength and relative simplicity of DEG analysis, in the present study we propose a sort of plot twist in transcriptome data analysis. We propose a new strategy (referred to as method #4): to perform WGCNA using the entire transcriptome dataset and, only subsequently, filtering by DEGs (Fig. 1). To validate this new method, we compared it with the performance of method #3, which is, as mentioned above, nowadays much spread in the literature.

For the purpose of this study, we selected one of the most transcriptomically complex tissues, the gonads [62] and one of the most complicated developmental processes that involve many genes interacting in a delicate

spatio-temporal and orchestrated manner: vertebrate gonadal sex differentiation. At the bi-potential stage, the vertebrate gonad expresses genes from the pro-male (leading to testis differentiation) and pro-female (leading to ovarian differentiation) pathways with opposing effects until the time when, depending on the species, a combination of genetic and/or environmental influences switch the balance towards the differentiation of one sexual phenotype and the repression of the other [63]. Thus, vertebrate gonadal sex differentiation provides an excellent opportunity to compare different methods to analyze large amounts of gene expression data. Furthermore, we used data from the European sea bass, *Dicentrarchus labrax* (from now on sea bass), a modern teleost with a polygenic sex-determining (PSD) mechanism without sex chromosomes [64], the mouse model, where gonadal sex differentiation has been extensively studied, and the human, a species where despite the limited availability of normal fetal gonads, it has been extensively studied for

Table 1 Examples of some transcriptomic studies classified according to the three different methods described (#1, DEGs; #2, WGCNA; # 3, DEGs + WGCNA; # 1 + 2 indicates the use of both methods, without combining them, as in #3) to analyze transcriptomic data since the development of WGCNA in 2005

Method #	Species	Cell type/ tissue	Technology	Reference
1	<i>Rattus norvegicus</i>	Gonads	Microarray	[9]
1	<i>Setaria italica</i> L	Plant seeds	SSH	[11]
1	<i>Gallus gallus</i>	Gonads	RNA-seq	[33]
1	<i>Trachemys scripta</i>	Embryos	RNA-seq	[15]
1	<i>Triticum aestivum</i> L	Plant seeds	Microarray	[14]
1	<i>Homo sapiens</i>	Breast	RNA-seq	[12]
1	<i>Xenopus laevis</i>	Gonads	Microarray	[34]
1	<i>Rattus norvegicus</i>	Brain	RNA-seq	[10]
1	<i>Dicentrarchus labrax</i>	Gonads	Microarray	[35]
2	<i>Homo sapiens</i>	Brain	Microarray	[36]
2	<i>Homo sapiens</i>	Brain	Microarray	[37]
2	<i>Homo sapiens</i>	Brain	RNA-seq	[19]
2	<i>Homo sapiens</i>	Bone	Microarray	[38]
2	<i>Homo sapiens</i>	Liver	Microarray	[39]
2	<i>Mus musculus</i>	Placenta	Microarray	[40]
2	<i>Homo sapiens</i>	Brain	Microarray	[41]
2	<i>Oplegnathus fasciatus</i>	Spleen	RNA-seq	[42]
2	<i>Homo sapiens</i>	Blood	RNA-seq	[43]
2	<i>Homo sapiens</i>	Lung	Microarray	[44]
2	<i>Homo sapiens</i>	Bladder	RNA-seq	[20]
3	<i>Homo sapiens</i>	Bladder	Microarray	[45]
3	<i>Homo sapiens</i>	Colon mucose	Microarray	[46]
3	<i>Homo sapiens</i>	Podocyte cells	Microarray	[47]
3	<i>Bubalus bubalis</i>	Mammary gland	RNA-seq	[48]
3	<i>Homo sapiens</i>	Bladder	Microarray	[49]
3	<i>Homo sapiens</i>	Lung	Microarray	[50]
3	<i>Homo sapiens</i>	Bladder	RNA-seq	[51]
3	<i>Bubalus bubalis</i>	Blood	RNA-seq	[52]
3	<i>Homo sapiens</i>	Brain	RNA-seq	[53]
3	<i>Homo sapiens</i>	Liver	RNA-seq	[54]
3	<i>Scophthalmus maximus</i>	Gill	RNA-seq	[55]
3	<i>Scophthalmus maximus</i>	Kidney	RNA-seq	[56]
3	<i>Homo sapiens</i>	Breast	Microarray	[57]
3	<i>Rattus norvegicus</i>	Spinal cord	RNA-seq	[58]
3	<i>Homo sapiens</i>	Stem cells	RNA-seq	[59]
3	<i>Bos taurus</i>	Blood	RNA-seq	[60]
1 + 2	<i>Scophthalmus maximus</i>	Gonads	Microarray	[13]
1 + 2	<i>Paralichthys olivaceus</i>	Embryo	RNA-seq	[25]
1 + 2	<i>Homo sapiens</i>	Brain	RNA-seq	[26]
1 + 2	<i>Gallus gallus</i>	Lung	RNA-seq	[27]
1 + 2	<i>Homo sapiens</i>	Organoids	RNA-seq	[28]
1 + 2	<i>Homo sapiens</i>	Placenta	Microarray	[29]
1 + 2	<i>Mus musculus</i>	Gonads	RNA-seq	[30]
1 + 2	<i>Homo sapiens</i>	Lung	RNA-seq	[31]
1 + 2	<i>Mus musculus</i>	Pancreas	Microarray	[32]

better understanding of the complex pathways related to disorders of sexual development (DSDs) [65, 66]. As with most mammals, mouse and human have an XX/XY sex-determining system where SRY is the master male sex-determining gene [67]. These three species with different sex-determining mechanism and differentiation dynamics were used to show that our approach was not only working on a complex process but also to show that the results are consistent regardless of the underlying genetic architecture. In all cases, we show that method #4 provides a more realistic and nuanced picture of the complex gene interactions taking place over time during gonadal sex differentiation and propose this method for future transcriptomic studies regardless of the biological system or question being considered.

Results

Construction of WGCNA using different approaches

WGCNA was conducted on male and female samples at two developmental stages (sea bass and mouse) or three developmental stages (human) to investigate the gene network operating during gonadal development in the sea bass, mouse, and human (Supplementary Figure 1).

We applied method #3 to the sea bass transcriptome, composed of a total of 20,978 genes, and identified 8434 DEGs between males and females at 250 days post fertilization (dpf) ($P < 0.05$) (Supplementary Figure 2). The normalized intensities of the 8434 DEGs (Additional file 1) were then used for WGCNA to build the network. The selected soft threshold used for the adjacency function closest to meet scale-free topology criterion and additional considerations was $\beta = 9$, leading to signed $R^2 = 0.7$ (Fig. 2A). Regarding connectivity, the second parameter considered for the selection of the soft threshold indices, the mean value was < 1000 (Fig. 2B) while the slope of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$ was -0.47 ($R^2 = 0.73$) (Fig. 2C). In contrast, when the entire transcriptome was used (methods #2. WGCNA and #4. WGCNA + DEGs), the selected soft threshold obtained was $\beta = 5$, which resulted in a better fit, signed $R^2 = 0.87$ (Fig. 2D), and more than doubled the mean connectivity to ~ 2000 (Fig. 2E) while also improving the slope $-\gamma = -0.51$ ($R^2 = 0.82$) (Fig. 2F).

The same difference between the two methods was observed in the mouse. Analysis of the gonadal transcriptome between males and females at 16.5 days post coitum (dpc) ($P < 0.05$) identified 8109 DEGs (Additional file 2). Using method #3, the selected soft threshold was $\beta = 14$ which resulted in a signed $R^2 = 0.55$ (Supplementary Figure 3 A), a mean connectivity value < 500 (Supplementary Figure 3 B), and slope $-\gamma = -0.79$ ($R^2 = 0.51$) (Supplementary Figure 3 C). However, when the entire gene expression dataset (14,088 genes) were used for

the construction of the network (methods #2. WGCNA and #4. WGCNA + DEGs), the soft threshold was $\beta = 12$, resulting in signed $R^2 = 0.7$ (Supplementary Figure 3 D), the mean connectivity ~ 500 (Supplementary Figure 3 E), and the slope improved to $-\gamma = -0.88$ ($R^2 = 0.55$) (Supplementary Figure 3 F).

The discrepancies between the methods were replicated using human data as well, where using method #3 the selected soft threshold was $\beta = 8$, resulting in signed $R^2 = 0.67$ (Supplementary Figure 4 A), a mean connectivity value < 50 (Supplementary Figure 4 B), and slope $-\gamma = -1.02$ ($R^2 = 0.63$) (Supplementary Figure 4 C). With methods #2 WGCNA and #4 WGCNA + DEGs, the soft threshold was $\beta = 8$, resulting in signed $R^2 = 0.89$ (Supplementary Figure 4 D), the mean connectivity > 200 (Supplementary Figure 4 E), and the slope improved to $-\gamma = -1.38$ ($R^2 = 0.87$) (Supplementary Figure 4 F).

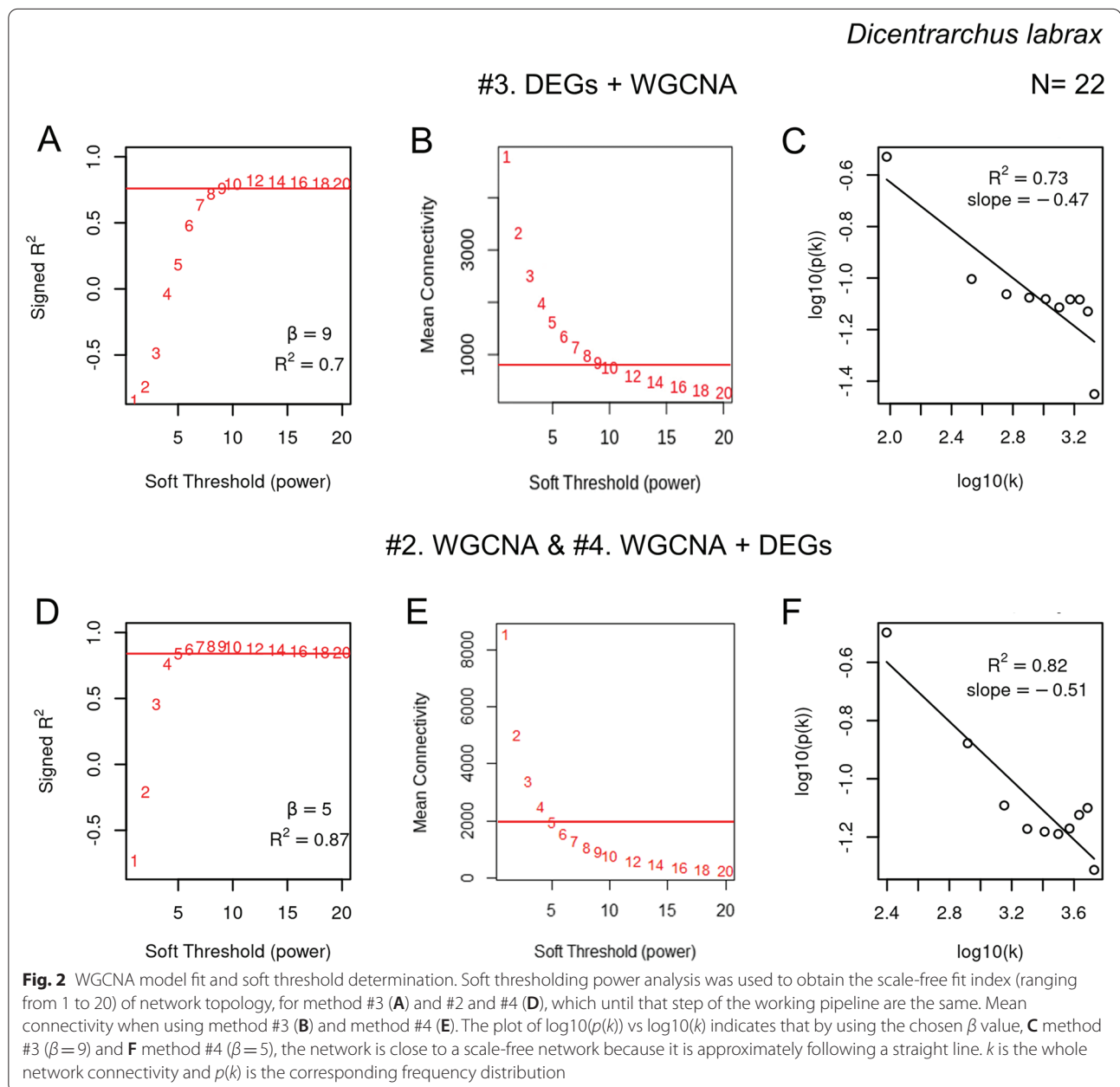
To ensure that the improvement in the model fit were not caused by the selection of a particular soft threshold, results were calculated for a range of thresholds (Table 2). In the three species, method #4 outperformed method #3 when it referred to scale-free topology model fit, mean connectivity, and the slope of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$ being closer to -1 , regardless of the soft threshold selected.

Selection of modules related to sex

Because method #3 uses a smaller dataset than method #4, a different number of modules were produced using the average linkage hierarchical clustering algorithm in WGCNA. Thus, in the sea bass the 8434 DEGs retained in method #3 were grouped into eight modules, of which two were strongly associated with the trait of interest, sex (red module: $R^2 = 0.93$, $P = 7e-10$; blue module: $R^2 = -0.84$, $P = 1e-06$) (Fig. 3A, C). On the other hand, the 20,978 genes of methods #2 and #4 were grouped into 29 modules, of which four were strongly associated with sex (pink, $R^2 = 0.95$, $P = 7e-12$; green, $R^2 = -0.79$, $P = 1e-05$; blue sky, $R^2 = -0.8$, $P = 7e-06$; magenta $R^2 = -0.72$, $P = 2e-04$) (Fig. 3B, D).

In the mouse, the 8109 DEGs retained in method #3 were clustered into six modules, of which two were associated with sex (blue: $R^2 = 0.69$, $P = 0.01$; turquoise: $R^2 = -0.8$, $P = 0.002$). Using the entire transcriptome in methods #2 and #4, a total of 21 modules were obtained, of which two were associated with sex (red: $R^2 = 0.98$, $P = 1e-07$; brown: $R^2 = -0.75$, $P = 0.008$) (Supplementary Figure 5).

The 1479 differentially expressed transcripts at 6 post-conceptional weeks (PCW) in human were grouped into eight modules, among which three of them were positively associated with sex (red: $R^2 = 0.64$, $P = 9e-05$; blue: $R^2 = 0.8$, $P = 3e-08$; yellow: $R^2 = 0.87$, $P = 2e-10$),



while one module was negatively associated with sex (turquoise: $R^2 = -0.91$, $P = 8e-13$). In contrast, the analysis of 35,194 transcripts using method #4 resulted in 36 modules, of which seven modules were associated with sex (light-yellow: $R^2 = -0.6$, $P = 3e-04$; black: $R^2 = -0.85$, $P = 5e-10$; dark olive-green: $R^2 = -0.57$, $P = 6e-04$; red: $R^2 = -0.72$, $P = 4e-06$; dark red: $R^2 = -0.66$, $P = 3e-05$; green: $R^2 = 0.72$, $P = 4e-06$; green-yellow: $R^2 = 0.78$, $P = 2e-07$) (Supplementary Figure 6).

Next, considering only the genes from the modules that significantly associated with sex, we selected those that, in addition, showed a significant correlation between modular membership and the gene significance for sex. All the modules associated with sex in the sea bass using methods #3 (Fig. 4A), #2, and #4 (Fig. 4B) showed a significant positive correlation and were kept for further analysis. However, in the weighted network obtained from the mouse data using method #3 the blue module was discarded because it did not pass the established criteria

Table 2 Comparison of model fit parameters under different soft thresholds using methods #3 (DEGs + WGCNA) and #4 (WGCNA + DEGs) on the gonadal transcriptome during early sex differentiation in three different species: sea bass, mouse, and human. The selected threshold and results obtained with the selected threshold are shown in bold

	Method	Threshold	Signed R^2 (model fit)	Mean connectivity	Slope (scale-free topology)	R^2
<i>Dicentrarchus labrax</i>	#3 DEGs + WGCNA	5	0.19	1580	− 0.12	0.08
		7	0.63	1100	− 0.33	0.59
		9	0.7	800	− 0.47	0.73
		11	0.8	650	− 0.56	0.79
		13	0.8	450	− 0.64	0.81
	#4 WGCNA + DEGs	5	0.89	2050	− 0.51	0.82
		7	0.9	1400	− 0.63	0.87
		9	0.91	900	− 0.71	0.88
		11	0.9	700	− 0.78	0.87
		13	0.9	500	− 0.83	0.86
<i>Mus musculus</i>	#3 DEGs + WGCNA	8	0.13	540	0.02	− 0.3
		10	0.33	375	− 0.5	0.24
		12	0.47	300	− 0.65	0.4
		14	0.55	230	− 0.79	0.51
		16	0.62	180	− 0.89	0.57
	#4 WGCNA + DEGs	8	0.62	495	− 0.62	0.4
		10	0.67	325	− 0.77	0.49
		12	0.7	220	− 0.88	0.55
		14	0.7	200	− 0.96	0.59
		16	0.73	160	− 1.02	0.62
<i>Homo sapiens</i>	#3 DEGs + WGCNA	4	0.45	105	− 0.74	0.38
		6	0.61	50	− 0.93	0.56
		8	0.67	25	− 1.02	0.63
		10	0.68	19	− 1.1	0.64
		12	0.69	15	− 1.17	0.65
	#4 WGCNA + DEGs	4	0.7	950	− 1.3	0.66
		6	0.79	410	− 1.36	0.77
		8	0.89	210	− 1.38	0.87
		10	0.93	100	− 1.42	0.92
		12	0.95	85	− 1.45	0.95

($\text{cor}=0.25$, $P=1.8\text{e}-28$), while the turquoise module was kept for further analysis ($\text{cor}=0.64$, $P<1\text{e}-200$) (Supplementary Figure 7 A). Also the two modules associated with sex obtained from mouse data using the methods #2 (WGCNA) and #4 (WGCNA + DEGs) were kept for further analysis given their positive significant correlation (red: 0.96, $P<1\text{e}-200$; brown: 0.67, $P<1\text{e}-200$) (Supplementary Figure 7 B). In human,

four modules were retained using method #3 (red: 0.75, $P=5.3\text{e}-99$; blue: 0.81, $P=4.6\text{e}-120$; yellow: 0.87, $P=6.6\text{e}-27$; turquoise: 0.84, $P=1.7\text{e}-145$) (Supplementary Figure 8 A), and seven modules were kept for further analysis using method #4 (light-yellow: 0.74, $P<1\text{e}-200$; black: 0.86, $P<1\text{e}-200$; dark green: 0.8, $P=1.4\text{e}-13$; red: 0.77, $P<1\text{e}-200$; dark red: 0.81, $P=1.3\text{e}-56$; green: 0.74, $P<1\text{e}-200$; yellow-green: 0.82,

(See figure on next page.)

Fig. 3 Identification of gene modules associated with sex. Gene hierarchical cluster analysis using method #3 (A) and #4 (B) using the sea bass gonadal transcriptome. Heat map of the correlation of sex with module eigengene distances using method #3 (C) and #4 (D). Each color represents a module in the constructed gene co-expression network by WGCNA. The heat map is colored from red (1, positive) to blue (− 1, negative) to indicate the level of correlation of each module with the trait of interest. The red boxes highlight the selected modules for further analysis associated with sex ($P<0.01$) (color in the online version). Genes not assigned to any of the previous modules are included in the gray module

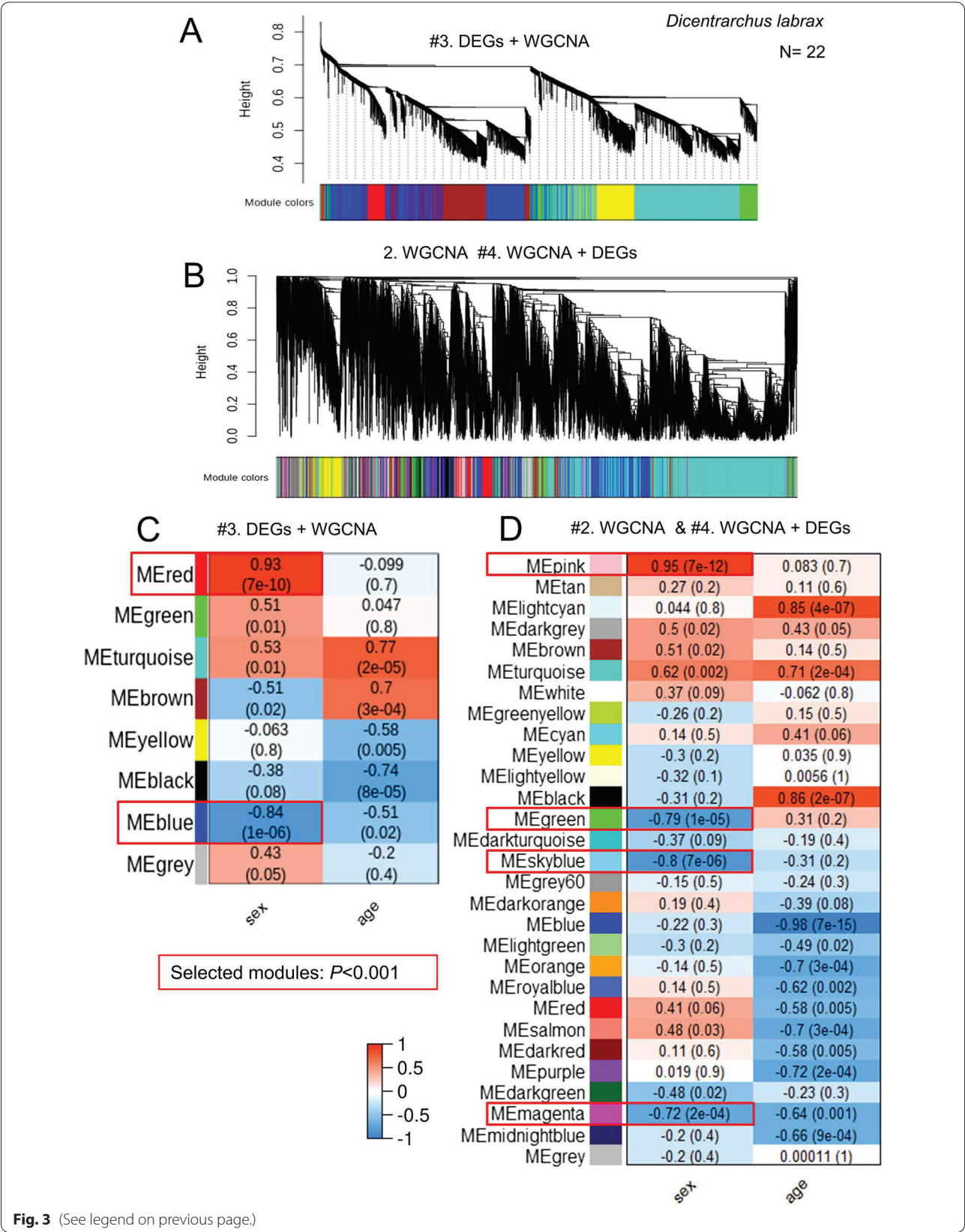


Fig. 3 (See legend on previous page.)

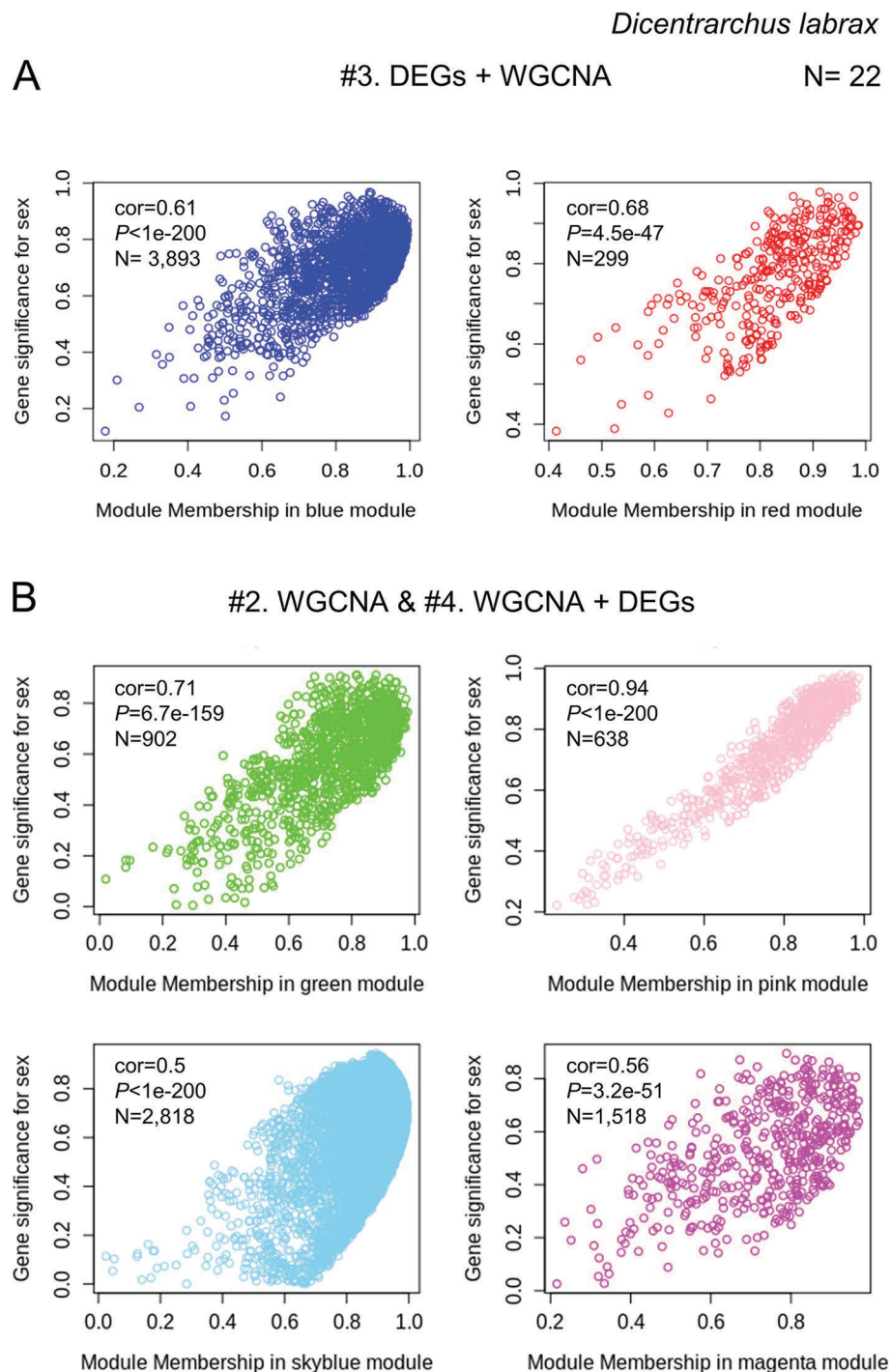


Fig. 4 Filtering of genes based on network properties in the sea bass. Scatterplots of correlation between GS vs module membership of each module which were used for determination of interesting modules for sex trait in the gonadal transcriptome of sea bass using method #3 (**A**) and #2 and/or #4 (**B**). The modules were selected when GS was positively and significantly correlated to modular membership. Gene significance is defined as $-\log$ of the p -value of association of the gene with the trait. Module membership (k_{ME}) measures how correlated each gene is to a particular module eigengene [68]

Table 3 Network statistics results when using methods #2, #3, and #4 in sea bass, mouse, and human transcriptomic data

	Network statistics	Method #2 WGCNA	Method #3 DEGs + WGCNA	Method #4 WGCNA + DEGs
<i>Dicentrarchus labrax</i>	No. of nodes	638	291	365
	No. of edges	180,759	6059	98,783
	Average no. of neighbors	480.7	42.3	263.42
	Characteristic path length	1.37	2.35	1.66
	Network heterogeneity	0.34	0.68	0.77
	Mean degree	480.7	41.6	263.42
	Max degree	720	135	349
	Max betweenness	0.003	0.07	0.03
<i>Mus musculus</i>	No. of nodes	2257	2697	1661
	No. of edges	1,016,156	377,169	798,757
	Average no. of neighbors	903,648	280.21	964.1
	Characteristic path length	1.66	2.39	1.43
	Network heterogeneity	0.563	0.865	0.365
	Mean degree	900.4	105.2	964.1
	Max degree	1867	327	1556
	Max betweenness	0.003	0.02	0.002
<i>Homo sapiens</i>	No. of nodes	558	48	481
	No. of edges	5632	382	3375
	Average no. of neighbors	20.2	15.9	14.0
	Characteristic path length	2.4	1.7	2.2
	Network heterogeneity	1.88	0.72	2.3
	Mean degree	20.2	15.9	14.03
	Max degree	309	42	309
	Max betweenness	0.25	0.09	0.39

$P < 1e - 200$) (Supplementary Figure 8 B). Thus, a total of 5876 genes from the four selected modules in sea bass, 3069 genes obtained from the three chosen modules in the mouse, and 8452 transcripts within the seven selected modules from human were kept for further analysis with method #4 (Supplementary Figure 2).

Co-expressed DEGs involved in gonadal differentiation

After the selection of interesting modules, we further filtered the genes within those modules by its gene significance (GS) for sex ($|GS| > 0.2$) and intra-modular membership ($|k_{IM}| > 0.8$). Furthermore, to implement the method #4 (WGCNA + DEGs), we further filtered the gene lists with the DEGs between males and females at the second stage of development, i.e., 250 dpf in the sea bass, 16.5 dpc in the mouse, and 6 PCW in the human, to identify the most relevant genes of each module. The number of genes co-expressed and, further, that are DEGs at 250 dpf was 3782 in sea bass (pink modules, $n = 365$; green module, $n = 621$; magenta module, $n = 930$; sky blue module, $n = 1866$ genes) (Additional file 3 and Supplementary Figure 2). In the mouse, there were 2205 co-expressed DEGs at 16.5 dpc (red module,

$n = 544$; brown module, $n = 1661$) (Additional file 4), and in the human there were a total of 1023 differentially co-expressed transcripts (Supplementary Figure 2).

Comparison of methods using network properties and statistics

We calculated several topological parameters to compare the performance of methods #2 (WGCNA), #3 (DEGs + WGCNA), and #4 (WGCNA + DEGs). Some of the most relevant parameters are shown in Table 3, including (1) the number of nodes, (2) number of edges, (3) characteristic path length, (4) average number of neighbors, (5) heterogeneity, (6) betweenness centrality, (7) mean degree, and (8) maximum degree. The detailed definitions of topological parameters were published by the developer of the *NetworkAnalyzer* [69]. In the three species, method #4 showed a lower number of nodes than method #2 since only co-DEGs were retained; however, method #4 resulted in a much higher number of nodes and edges than method #3 in the sea bass and human. The average number of neighbors, which indicates the average connectivity of a node, showed higher (better) results when using method #4

than method #3, in proportion to the number of nodes obtained from each method (average number of nodes/number of nodes) $\times 100$, in sea bass: method #2=75%, method #3=14%, method #4=72%, and mouse: method #2=40%, method #3=10%, method #4=58%, but not in human: method #2=3.3%, method #3=37%, method #4=4.2%. Regarding the characteristic path length, the expected distance between two connected nodes (the shortest, the more compacted is a network), method #4 showed better results than method #3, except in human where the network of the module from method #3 was much smaller and, hence, resulted in shorter characteristic path length. Network heterogeneity is an important parameter for biomarker discovery as it reflects the tendency of a network to contain hub nodes [70]. Such parameter was higher when using method #4 in the sea bass and the human. The maximum betweenness centrality was obtained from method #3 in the sea bass and the mouse, as well as from the human using method #4. The maximum degree was higher in the three species when using method #4. Taken together, the topological parameters of the different networks produced using the three different methods showed that the most robust networks were achieved using methods #2 and #4. Besides, method #4 allows further filtering of the interesting genes without altering main network properties and without removing too many genes that could be importantly related to the trait of interest. Of note, when method #3 resulted in better parameter values, this was always in the mouse, incidentally the species in which the minimum number of samples recommended for network construction was not reached ($n < 15$), leading to less robust conclusions. Considering this and that the construction of the network in the mouse required the highest soft threshold ($\beta = 12$ in the mouse vs $\beta = 5$ in the sea bass and $\beta = 8$ in the human) while achieving the poorer model fit using method #4, the results support the original recommendation of at least 15 samples to construct networks with method #2 (WGCNA) and #4 (WGCNA + DEGs).

Proportion of genes related to sexual development identified by the different approaches in the sea bass and the mouse

We found that the proportion of retained genes and, within those, of key genes for gonadal sex differentiation among the four methods tested were similar in the sea bass and mouse (compare Fig. 5A and Supplementary Figure 9 A). When using method #1, filtering the transcriptome by DEGs, a large proportion of the transcriptome was kept. Up to 40.2% in sea bass (Fig. 5A) and 57.6% in mouse (Supplementary Figure 9 A) of the transcriptomes were differentially expressed. Among them, 78.6% of the key genes previously known to be involved

in sexual development were found in the sea bass, and 78% of them were identified in the mouse when using method #1.

When using method #2 for the same datasets, we obtained smaller lists of genes being related to sex. In sea bass, we found that 28% of the transcriptome was involved, of which 64% of the key genes were identified. Similarly, in mouse, 22% of the transcriptome was co-expressed and associated with sex and 76% of the key control genes were detected. The third method, the application of DEG filter previous to network construction, leads to slightly smaller gene lists: 19.9% in the sea bass, and 20% of the mouse transcriptome. However, the proportion of key genes detected by this method was much smaller in the sea bass (34%, 1 gene in the red module and 8 genes in the blue module, Fig. 5B) and in the mouse (50%, 25 genes, Supplementary Figure 9 B). Finally, when using the proposed method #4, to perform WGCNA first and to apply the DEG filter afterwards, it provided us with the smaller proportions of potential novel genes related to sex trait (18% in sea bass and 15.6% of the transcriptome in the mouse) without harming the capability to detect the key genes previously known to be involved in sex development. In sea bass, 64% of the key genes were identified (Fig. 5A, C), and in mouse, 58% of the key genes were captured (Supplementary Figure 9 A and C).

It is important to note that the lists of genes produced by methods #3 and #4 differ. Method #3 yielded a total of 4192 co-expressed genes in sea bass, 1870 of which were uniquely detected using this method. On the other hand, method #4 produced a total of 3782 genes, of which only 877 were unique to that method (Fig. 6A). In the mouse dataset, method #3 detected a total of 2835, of which 1023 were unique to that method, while method #4 detected 2205, of which 393 genes were uniquely detected by this method (Supplementary Figure 9 D). Furthermore, the Jaccard index revealed low similarity between the gene lists obtained from methods #3 and #4 (45.02% in sea bass, and 56.13% in mouse).

Network visualization and gene expression profiles of the selected genes

The data presented so far indicates that our proposed approach (method #4) gives more meaningful results of the biological process being studied than the currently used method #3. To further explore its advantages, we used the gene lists obtained with method #4 in sea bass for further network visualization and plotted the mean gene expression profiles of the chosen modules. Interestingly, the average gene expression profile of the genes making up the four significant modules showed clear differences in their temporal dynamics during sea bass sex differentiation between 110 and 250 dpf. Thus,

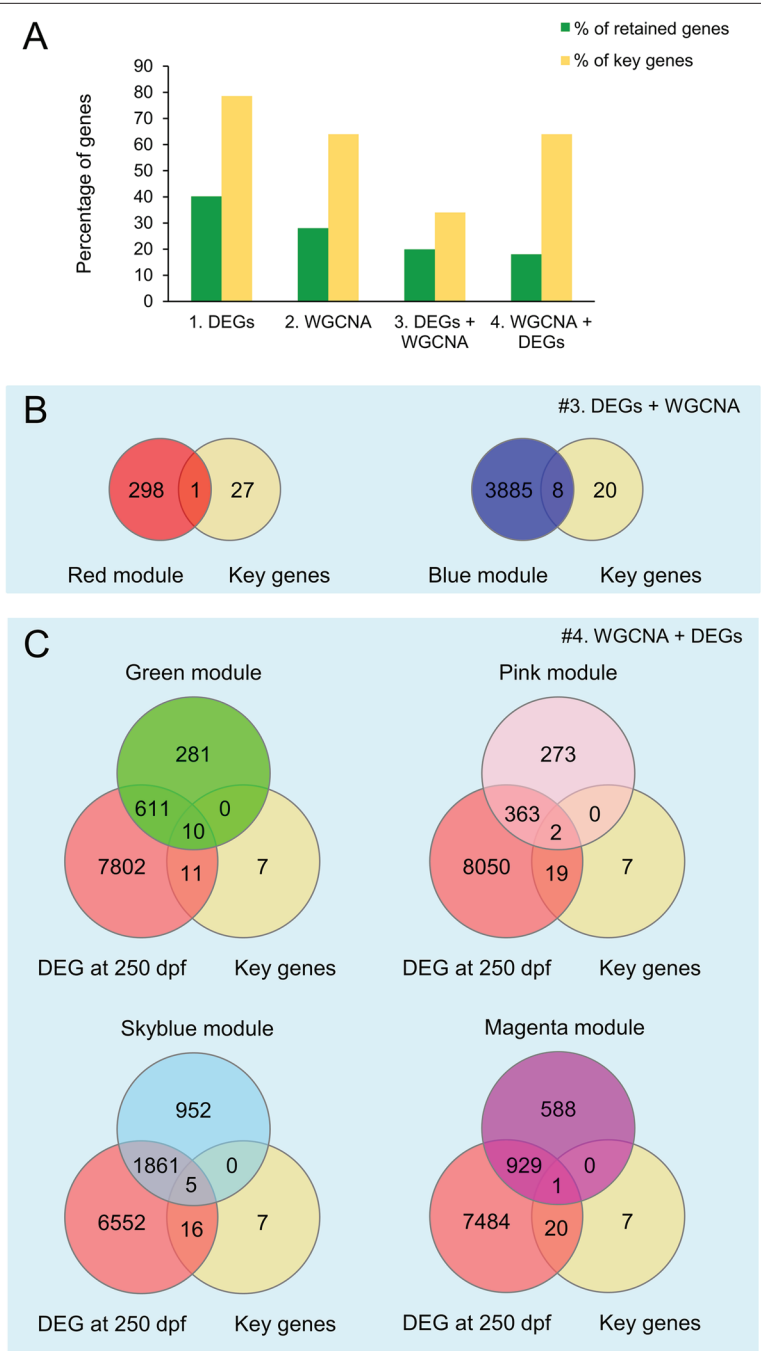


Fig. 5 Identification of key genes involved in sex differentiation in the sea bass when using the different methods to analyze transcriptomic data. **A** Percentage of genes and key genes retained according to the methods mentioned in this study. **B** Venn diagram of the key genes found in each module using method #3 and **C** in method #4

co-expressed genes of the green module were upregulated in males and downregulated in females, reaching very disparate expression levels (Fig. 6B). Genes of the sky blue module did not change expression in males but were actively downregulated in females (Fig. 6C) while

genes of the magenta modules were downregulated in both sexes (Fig. 6D). Finally, genes of the pink module started upregulated in females but were strongly downregulated by 250 dpf, while in the males their expression increased slightly (Fig. 6E).

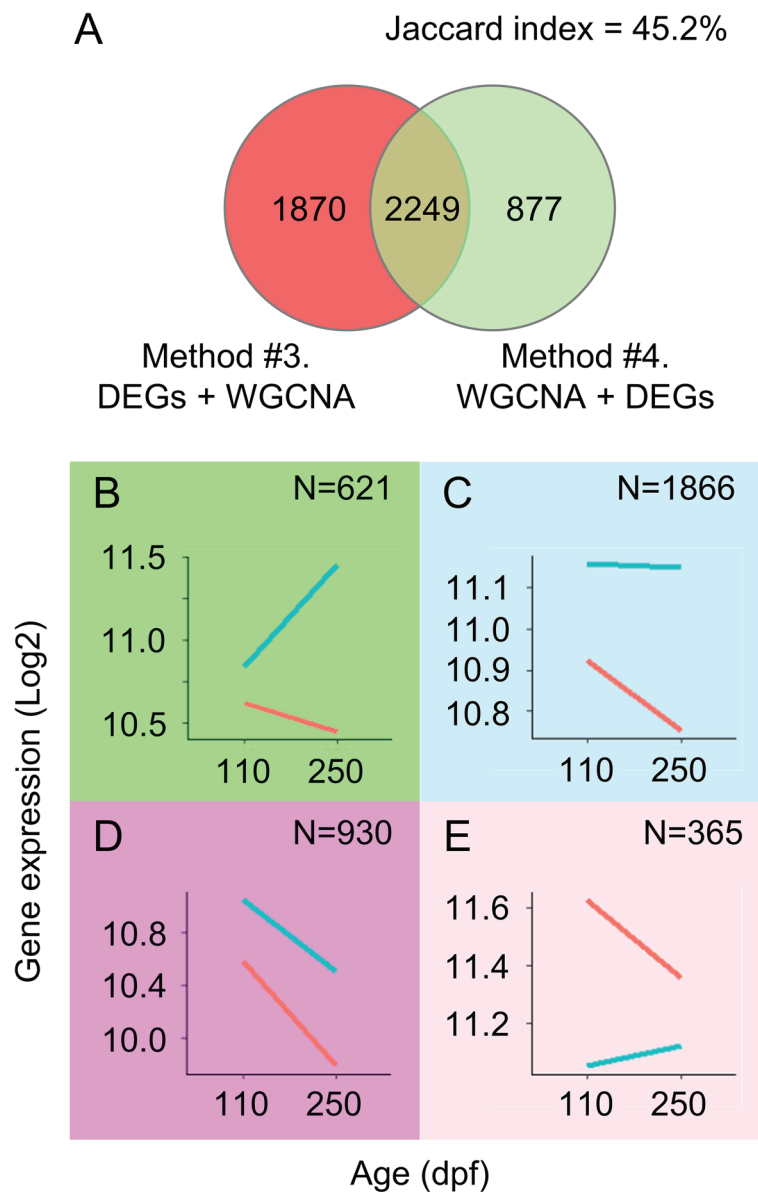
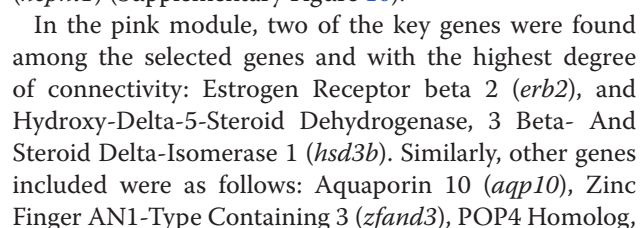


Fig. 6 **A** Comparison of the gene lists obtained from the sea bass dataset when using method #3 and #4. The Jaccard index indicates low similarity between gene lists. Gene expression profile of the co-expressed DEGs from the four modules associated with sex and filtered by method #4. The background color of the plots indicates the color of the modules they belong to: **B** green, **C** sky blue, **D** magenta, and **E** pink module. **G** Gene expression of *sox9a* in males (blue) and females (red) at 110 and 250 dpf

Genes making up the green module ($n=621$) were those whose expression levels differed the most over time. For representation purposes, Fig. 7 shows only a subset of these genes ($n=102$) having the highest number of connections (width > 0.3). From the inside to the outside, the four concentric circles show the genes with a higher degree (number of connections), including Cytochrome C Oxidase Copper Chaperone COX17 (*cox17*), Cyclin A2 (*ccna*), Cytochrome P450 Family 11

Subfamily A Member 1 (*cyp11a1*), Ajuba LIM Protein (*ajuba*), and Proteasome 20S Subunit Alpha 6 (*psma6b*). Importantly, 18 of the 28 key genes selected for their role in sea bass gonadal sex differentiation were represented in the different modules. Nine of them were in the green module with the highest degree (from higher to lower): Anti-Müllerian Hormone (*amh*), Cytochrome P450 Family 17 Subfamily A Member 1 (*cyp17a1*), Follicle Stimulating Hormone Receptor (*fshr*), Steroidogenic



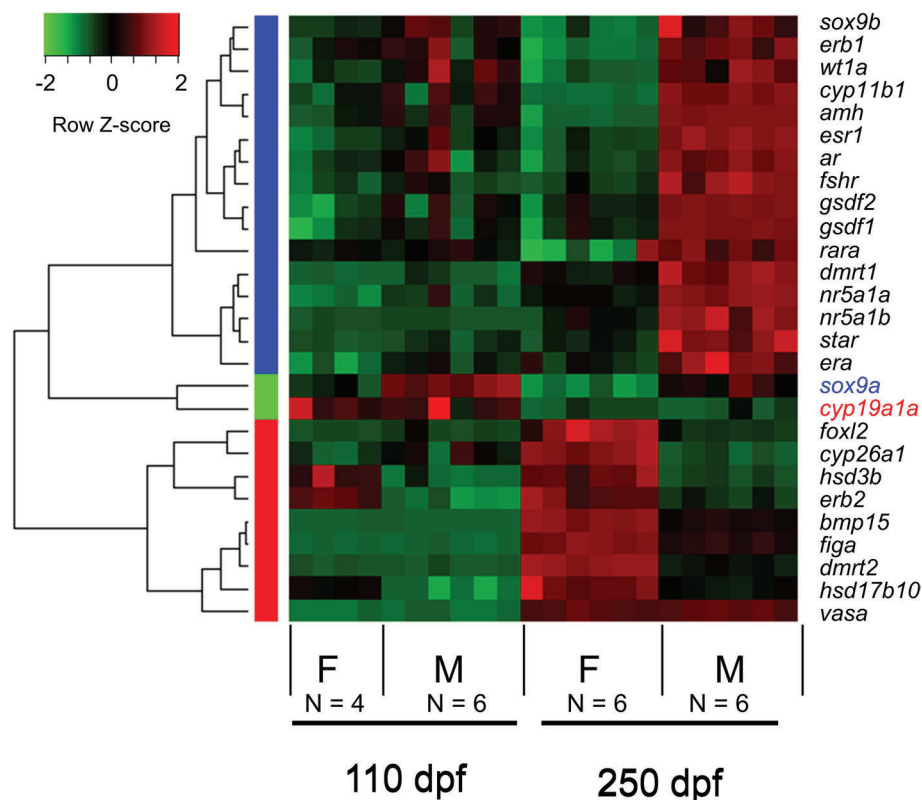


Fig. 8 Hierarchical clustering analysis and heat map of the key genes in the sea bass. Upregulation is indicated in red and downregulation is shown in green. The left bars indicate the three main gene expression patterns found. The blue bar includes the genes upregulated in males at 250 dpf, the red bar includes the genes upregulated in females at 250 dpf while the two genes included in the green bar are those identified as early biomarkers, which show a different gene expression pattern already at 110 dpf between males and females

Ribonuclease P/MRP Subunit (*pop4*), Tubulin Alpha 8 (*tuba8*), Cytochrome C Oxidase Subunit 5A (*cox5a*), and Acyl-CoA Oxidase 3, Pristanoyl (*acox3*) (Supplementary Figure 11). Lastly, with the chosen degree threshold for better network visualization, fewer genes were shown in the sky blue network, including Ubiquitin-Specific Peptidase 5 (*usp5*), Keratin 8 (*krt8*), and Schwannomin Interacting Protein 1 (*schip1*) (Supplementary Figure 12).

Prediction of biomarkers of early gonadal differentiation in the European sea bass

Since the sea bass has PSD, and thus there are no genetic markers for sex, we wanted to focus on genes involved in the process of gonadal sex differentiation to see if we could at least identify robust markers of the early stages of this process. To do so, we used the gene lists obtained from WGCNA built from the entire transcriptome and filtered the output with DEGs only at 110 dpf. We obtained a total of 2089 genes which were strongly correlated to sex. Among these genes, two of the key genes found were *hsd17b10*, in the magenta module, and the ortholog of SRY-Box Transcription Factor 9 a (*sox9a*),

which was detected in the sky blue module. Furthermore, hierarchical cluster analysis of the 28 key genes during the sea bass gonadal sex differentiation showed three main gene expression patterns: (1) genes upregulated in males at 250 dpf, (2) genes upregulated in females at 250 dpf, and (3) genes upregulated in males or females already at 110 that increase their expression by 250 dpf (Fig. 8). *sox9a* was clustered together with *cyp19a1*, the already described biomarker in previous studies for ovarian differentiation. They were clustered together according to the third gene expression pattern mentioned above. Altogether, this suggested that *sox9a* expression could be used as early marker of male sex differentiation in the sea bass.

Prediction of biomarkers of early gonadal differentiation in the mouse

In the case of the mouse, there is no need for biomarkers for sex because they have genetic sex determination (GSD) with sexual chromosomes, which allow identifying the sex as early as the zygote stage. However, we used methods #2 and #4 to test their performance in

biomarker identification using a well-known model. For this objective, method #1 was not used because it generated lists of several thousand candidate genes from which it was impossible to choose a few candidates. Similarly, method #3 was discarded because of the observed bias in the results. However, when using method #2 in the brown module, for example, we found a total of eight key genes. When using method #4, the number of potential biomarkers of sex at the first stage was five genes: Wnt Family Member 4 (*wnt4*), Fibroblast Growth Factor Receptor 2 (*fgfr2*), Bone Morphogenetic Protein 2 (*bmp2*), Follistatin (*fst*), and GATA Binding Protein 6 (*gata6*). The three genes filtered out by method #4 were not DEG at the E11.0–12.0 stage between males and females but were associated with sex by WGCNA. Such genes are relevant for sex development in the mouse but not among the first to show differential expression between males and females.

Therefore, the filter by DEGs in method #4 is essential to increase the probability of finding genes that could show differential gene expression when compared to method #2. This allows the prediction of the sex on other individuals based on the gene expression measurement of a biomarker. Hence, method #4 was the approach that allowed obtaining the shorter lists of candidates meeting all the criteria from WGCNA and, on top, they were among the first to show differential expression at the stage of interest.

Comparison of the functional enrichment analysis from two different methods in the mouse dataset

Next, we performed gene ontology (GO) enrichment analysis of the filtered gene lists produced with methods #3 and #4 in mouse (the system we used that had the most curated annotation). The results showed that GO terms directly related to biological processes associated with gonadal development were more enriched and fell under a more robust significance threshold when using method #4 than when using method #3 (Additional file 5). Secondary alcohol biosynthetic process (GO:1,902,653), cholesterol biosynthetic process (GO:0,006,695), sterol biosynthetic process (GO:0,016,126), and steroid biosynthetic process are a few examples (GO:0,006,694). Furthermore, only method #4 captured the term GO cell morphogenesis involved in differentiation (GO:0,000,904).

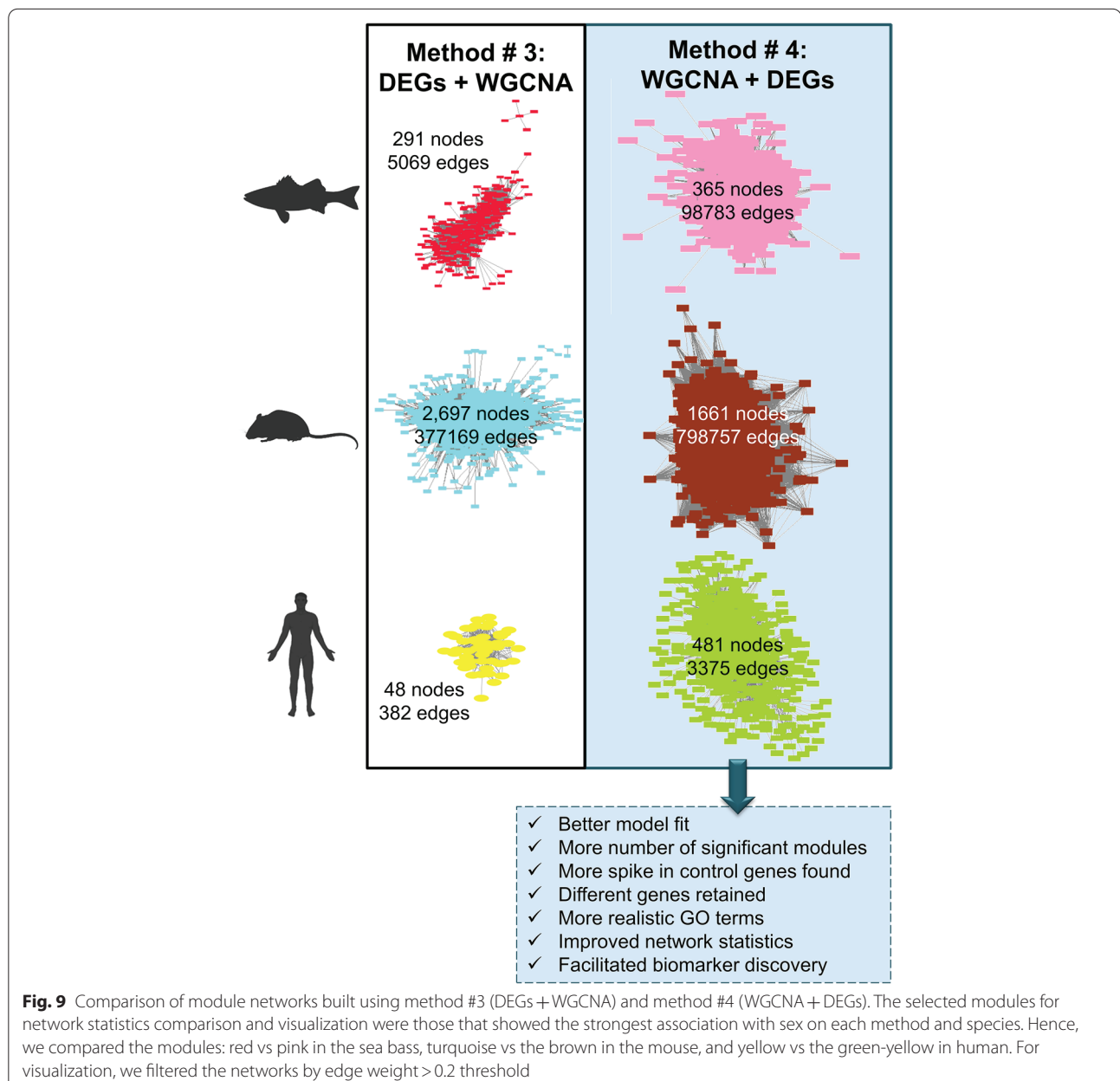
Discussion

Studies based on transcriptomics data can be broadly classified into three different major methods: #1 DEGs, #2 WGCNA, and #3 filtering by DEGs (or other filters such as keeping the 20% with more gene expression variance, or the top 5000 DEGs) before WGCNA [20, 48,

54, 60]. To combine the power from WGCNA with the filter from the traditional and robust statistical analysis without introducing bias, we propose a new pipeline: (4) WGCNA of the entire transcriptome filtered by DEGs only after network construction.

In the present study, we compared the performance of filtering by DEGs before (method #3) or after (method #4) WGCNA in the analysis of one of the most transcriptomically complex organs, the gonads, during the critical period of sex differentiation. And to do so, we used three different vertebrates, a modern fish, the mouse model, and human. The fish has a very different sex-determining system than the mouse and the human that, in contrast, have a very similar sex-determining system. In all species, method #4 (WGCNA + DEGs) clearly and consistently outperformed method #3 (DEGs + WGCNA) (Fig. 9). WGCNA establishes an adjacency matrix of correlation values which is transformed using a power function. The value used to power the matrix is chosen based on parameters that ensure a scale-free topology (in order of importance): signed $R^2 \geq 0.8$, or the β value which maximizes a scale-free independence, high mean connectivity, and the slope ($-\gamma$) of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$ was around -1 . In sea bass, the best scale-free topology model fit using method #3 resulted in index $\beta = 9$ leading to $R^2 = 0.7$, low mean connectivity, and the $-\gamma$ slope of -0.47 . However, using method #4, for the same dataset, a much better scale-free topology model fit ($R^2 = 0.87$) was obtained with a lower power index, $\beta = 5$. Also, higher mean connectivity and $-\gamma$ was slightly closer to -1 . In mouse, using method #3, even a high soft threshold index of $\beta = 14$ lead to a weak model fit $R^2 = 0.55$, low mean connectivity, and the slope of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$ of -0.79 . Using method #4 with the same dataset improved all the parameters during co-expression network construction. The R^2 increased from 0.55 to 0.7, and the mean connectivity increased from hundreds to a few thousands. Furthermore, the $-\gamma$ slope decreased from -0.79 to -0.88 . The improvement of model fit was also observed in human using method #4 ($\beta = 8$, $R^2 = 0.89$, mean connectivity > 200 , and $-\gamma = -1.38$ ($R^2 = 0.87$) when compared to method #3 (method #3: $\beta = 8$, $R^2 = 0.67$, mean connectivity value < 50 , and $-\gamma = -1.02$ ($R^2 = 0.63$). The results from six analyses (using two methods in three species) under different soft threshold values showed that the network model fit is much better when using method #4, objectively.

Another important outcome of our comparison related to the network construction is the number of modules created by each method. Thus, method #3 created 8



modules while method #4 created 29 in sea bass. In the mouse, these figures were 6 and 21, and in human 8 and 36, respectively. These results consistently show that the number of modules associated with the trait of interest is considerably smaller when using method #3, clearly showing that potentially interesting genes and connections are lost already before WGCNA is performed, diminishing its power. These results were consistent with values found in the literature regardless of vertebrate species or question being answered, where method #3 produces fewer modules (5 modules [46]; 12 in [47]; and 6 in

[48], than studies using WGCNA without previous filtering (17 modules in [40]; 23 in [36]; and 14 in [44]).

Several strategies exist to compare gene co-expression networks [53, 71]. We calculated topological parameters for comparison of the network module most associated with the trait of interest using methods #2, #3, and #4. All the network parameters from method #4 outperform method #3 in two or all the three species studied. One of the instances was the characteristic path length in human, which showed a shorter path length in the network built using method #3. This could be because

of the much smaller number of nodes obtained with such a method, which could provide a shorter length despite being a less compacted network, as noticed elsewhere [69]. Some of the other parameters that did not clearly improve using method #4 compared to method #3 resulted from the mouse. Such networks were built with 11 samples (less than the minimum sample size recommended by the developers of WGCNA) [61]. Therefore, results should be considered with precaution. We observed in the literature that out of 20 studies in Table 1 using WGCNA or DEGs + WGCNA, some used $n < 15$ ($n = 13$ [44]; $n = 10$ [43]; $n = 8$ [48]; $n = 14$ [60], and $n = 8$ [52]), although the majority used > 15 samples to construct robust WGCNA ($n = 160$ [36]; $n = 183$ [40]; $n = 25$ [42], $n = 430$ [20], $n = 45$ [58], $n = 36$ [32], $n = 228$ [31], $n = 48$ [27]). Considering the developers' recommendation and our results, we take the opportunity to further emphasize that at least 15 samples in total are required to perform method #2 and extend this recommendation to method #4.

Regarding the results of biological relevance, which was to identify genes related with sex and therefore with reproduction, higher proportions of genes previously known to be involved in sex differentiation were retained for further analysis when screening with method #4. Of note, not only the number of genes captured was larger in both species, but also essential key genes in a given process (here used as key controls), described in previous works, were also better represented with method #4. For example, key genes found in sea bass were as follows: *amh* [72], *gsdf* [73], *ar* [74], and *erb* [75]. Similarly, some of the key genes found in mouse were as follows: *dmrt1* [76], *amh* [77], *ar* [78], *foxl2* [79], *bmp6* [80], and *wnt4* [81]. These genes were retained with method #4 but not with method #3. Thus, our approach results in a more meaningful representation of the biological question under consideration.

Another important insight of this study is that we clearly show that methods #3 and #4 generate different gene lists. As a result, this has an impact on the downstream functional analysis and can affect the resulting GO biological processes, as well as the interpretation of the results (Additional file 5). First, the overrepresented GO biological processes from the gonadal transcriptome of the mouse at 16.5 dpc slightly varied depending on the filtering strategy and the order in using such filters leading to different results. First, we found that GO terms from the category of biological processes related to gonad development functions were more enriched and had a higher significance threshold when using method #4. Second, only method 4 yielded some terms related to the gonadal sex differentiation process. Altogether, these results provide evidence that more accurate gene

lists related to the target phenotypic trait were achieved using method #4 (WGCNA + DEGs) than when using the DEG filter or gene lists from pre-filtered WGCNA as in method #3.

We are aware that the results from gene ontology enrichment analysis should not be considered proof of biological validity in the analysis of high-throughput data considering the multiple sources of bias existing in functional enrichment analysis tools [82]. However, to date, this is one of the main methods used to summarize information from high-throughput experiments and here we showed how the previous filtering of DEGs can affect the results of WGCNA.

In search of early sex differentiation markers in the sea bass, several genes have been identified so far in the literature. Blázquez et al. [83] defined *cyp19a1a* as a suitable molecular marker of ovarian differentiation, which showed the first significant gene expression difference between sexes at 120 dph. Ribas et al. [35] identified the *hsd17b10* gene as an early marker for ovarian differentiation at 110 dpf. At this stage, *cyp19a1a* expression levels were higher in females but not significantly different from males. In the present study, among the key genes involved in early gonadal development, three genes stood out from the combined filter of WGCNA and the subsequent DEG filter at 110 dpf. Among these three genes, *hsd17b10* was found, hence, supporting the results from Ribas et al. [35]. The other gene identified as marker of testis early differentiation was *sox9a*. Noteworthy, this gene could not be identified in previous studies using method #1 [35].

sox9 is a multifunctional transcription factor found in different tissues and plays crucial roles in vertebrate development, including cell proliferation and differentiation [84]. This gene was defined as a "hub" gene of testis differentiation after sex determination in vertebrates. As reviewed elsewhere [84], *sox9* has a conserved role in male gonadal development and a highly conserved protein sequence. In fishes, two orthologs exist due to gene duplication: *sox9a* and *sox9b* [85]. Although the two orthologs are retained, they are related to different functions and tissues depending on the fish species. For example, *sox9a*, is expressed in the ovary of the medaka (*Oryzias latipes*) and the platy fish (*Xiphophorus maculatus*), while in the zebrafish (*Danio rerio*), *sox9a* is expressed only in the testis [85], like the current results found in the sea bass. In this species, *sox9a* was recently found to be involved in the epigenetic regulation of the temperature induced sex ratio during sex differentiation [86].

Based on the parameters obtained in the model fit during network construction, network statistics, and the biological relevance of the results obtained, we highlight

the importance of using method #4 followed by functional enrichment analysis rather than method #3. Even though the computation of method #3 (DEG + WGCNA) is faster (a matter of seconds to a few minutes for the datasets used in this study) and requires less computational power, it does not compensate for the bias introduced to the results. Additionally, we recommend using method #4 instead of method #2 because, although the same network is built and the same groups of genes are formed, method #4 allows to further filter the large gene lists produced by WGCNA without removing the most important (based on gene expression) genes. WGCNA is a powerful method still being used [20, 42–44], with a large proportion of the studies using specifically the inappropriate method #3 [54–60]. More recent alternative methods to WGCNA to identify modules of differentially co-expressed genes have been developed: THD-Module Extractor [22], DiffCoEx [23], and MODA [24]. And, although in the comparison of these methods with WGCNA, the THD-Module Extractor method was claimed more effective in finding modules with higher functional relevance and biological significance than WGCNA, this result could be due to the previous filter of the dataset. Precisely, as stated by the authors, the use of DEGs before WGCNA lead them to the soft threshold of power fit to fail [53]. Therefore, a comparison of THD-Module Extractor method with the method #4 (WGCNA + DEGs) described in this study remains to be done.

Conclusions

In conclusion, WGCNA is a robust and systematic approach commonly used in transcriptomics. With time, several strategies have appeared and became quite common, i.e., method #3, to use the parameters and filters provided by this tool but somehow running against its original philosophy. We evaluated the model fit and the biological relevance of the results obtained using method #3 (DEGs + WGCNA) and our proposed method #4 (WGCNA + DEGs) with the gonadal transcriptome of three different vertebrates, sea bass, mouse, and human, obtained from two high-throughput technologies: microarray and RNA-seq. The results showed that method #4 is more efficient in filtering for smaller gene lists that contain the genes most related to the trait of interest. Thus, we propose to analyze transcriptomic data using WGCNA to build a co-expression network from the entire dataset and, only subsequently, filter by DEGs. Such a strategy combines the powerful method of producing a network with the filter from the traditional and robust statistical analysis without introducing bias. Additionally, we produced new lists and network visualization of genes related to early sex differentiation in the sea bass

with the corresponding functional enrichment analysis and identifying a novel biomarker for testis early differentiation (*sox9a*). Last but not least, the filter by DEGs in method #4 increases the probability of robust biomarker discovery when compared to method #2. We propose that method #4 should be the method to use in future transcriptomic studies regardless of biological system, species, or question being considered.

Methods

After WGCNA development, numerous studies started to use this method in combination with the previous filtering of the dataset (method #3). Combining WGCNA with other filtering methods like DEGs can be a strategy to further filter for the target genes associated with a trait of interest. However, whether this filtering step before WGCNA affects the results and to what extent has never been shown yet. In this study, we aim to compare method #3 with the here proposed new method #4: applying the filtering step after construction of the co-expression network. To determine which one is more efficient, we used both methods to study the gonadal transcriptome of three vertebrates: sea bass, mouse, and human. Hence, a total of six analyses were performed. Then, we compared the model fit, connectivity, and other network statistics to compare both methods for each of the three species. Additionally, we used existing information on key genes for the sexual development and reproduction of sea bass and mouse to measure how many key genes were found with each method. Finally, the enriched genes by GO term enrichment analysis obtained from the gene lists produced by methods #3 and #4 were compared in the mouse. The methods are implemented in R software and all source code has been made publicly available on GitHub as part of the WGCNA_DEGs project at: https://github.com/Nsbaizan/WGCNA_DEGs.

Literature review

Before any comparison of methods #3 and #4 was performed, we wanted to know if the use of method #3 was exceptional or if we could find several studies where this method was used. We used Scopus and Web of Science databases to find published studies using each method. To search a few examples of method #1, we used the following keywords in the topic field: “((DEG) OR (“differentially expressed genes”)) AND ((Microarray) OR (RNA sequencing))”. To find studies using methods #2 or #3, we used the following keywords in the topic field: “((WGCNA) OR (“weighted gene co-expression network analysis”)) AND ((Microarray) OR (RNA sequencing))”. After reading detailed information on the methods section of each paper, we could classify between methods depending on whether WGCNA was used without

(method #2) or after filtering by DEG (method #3). We found that the use of the WGCNA is widespread in transcriptomic data studies and that method #3 is not exceptional but, rather, it is of common use even in recent publications.

Species and processes studied: the sea bass

In sea bass, sex depends on the combination of several pro-male and pro-female autosomal genes plus environmental (temperature) influences [64]. Sex determination is thought to occur between 60 and 100 dpf [87], around 120 dpf the first molecular signs of gonadal sex differentiation appear in the form of differences in the expression of gonadal aromatase, *cyp19a1a* [88], and the first histological differences appear around 150 dpf when fish are around 8 cm long. Sex differentiation proceeds earlier in females than males, and the process is completed by 250 dpf in females and around 350 dpf in males when females are 12.7 ± 5.7 cm (mean $\pm$ SD), and 11.2 ± 0.6 cm of SD in males, and 16 ± 1.3 in females and 14.8 ± 1.1 cm of SD in males, respectively [35]. Gene expression during gonadal sex differentiation in the sea bass has been extensively studied, first by targeted approaches [89–92] and more recently by using a homologous and validated microarray containing 43,803 probes. In the latter case, fish before (110 dpf), during (250 dpf), and after (350 dpf) gonadal sex differentiation were analyzed [35], and DEGs and enriched signaling pathways were identified. Additionally, *hsd17 β 10* was identified as a marker of early ovarian differentiation.

Species and processes studied: the mouse

Sex differentiation in the mouse is well characterized [93] and thus less information will be provided. Briefly, key genes during mouse gonadal sex differentiation include sex-determining region Y (*sry*), *amh*, and *sox9*, while pro-female genes include *wnt4*, R-spondin 1 (*rspo1*), a member of the R-spondin family, and β catenin (*ctnnb1*) [93]. In mouse, sex differentiation starts between embryonic day (E) 11.0–E12.0 and ends at E16.5 [94, 95].

Species and processes studied: the human

Sex differentiation in the human has been studied extensively despite the limited availability of normal fetal human gonads [66]. In brief, the gonadal primordium arises from the coelomic epithelium around the 4th PCW and differentiation reaches the end towards the 17th PCW. Testis differentiation is activated by the expression of the Y-linked transcription factor *sry* during the 6th PCW in supporting cells leading to the expression of *sox9* and *amh* genes. In the absence of *sry*, *rspo1*/*Wnt4*/ β

catenin pathway, *foxl2* activates the transcriptional cascade required for ovarian differentiation.

Datasets

Transcriptomic data during sea bass sex differentiation was previously obtained using a homologous microarray [35] and can be downloaded from the Gene Expression Omnibus (GEO) database with the accession number GSE115841 [96]. For the aim of the present study, we selected data available from fish at two key developmental stages: four females and seven males at the beginning of sex differentiation at 110 dpf with a length of 5.2 ± 0.5 cm (mean $\pm$ SD). At this time, gonads are still morphologically undifferentiated but fish can be sexed measuring *cyp19a1a* expression levels [88]. We also used six females (12.7 ± 5.7 cm) and six males (11.2 ± 0.6 cm) in the middle of sex differentiation period at 250 dpf. Thus, we used 23 fish in total for analysis. The original downloaded file consists of 43,801 probe copies representing 20,978 transcripts with normalized expression values, corrected for batch effect. Microarray intensity values were directly used for the determination of differentially expressed genes and/or network construction.

Transcriptomic data during mouse sex differentiation was obtained using RNA-seq [96] and can be downloaded from the GEO database under the accession number GSE117590 [97]. For the present study, we selected transcriptomic data from twelve samples, three males and three females at two embryonic stages: 12.5 dpc, corresponding to the beginning of gonadal sex differentiation, and 16.5 dpc, at the end of sex differentiation. After trimming the raw reads, the alignment to reference genome (v. GRCm39 GCA_000001635.9) (2.1.0) [98] was performed to obtain a dataset with 55,416 genomic features using featureCounts (v2.0.0) [99]. Counts were processed using edgeR package (v3.34.0) [100]—limma workflow which includes counts pre-processing and exploratory data analysis before obtaining lists of DEGs as described in [101].

Transcriptomic data during human sex differentiation was obtained using RNA-seq [66] and can be found at NCBI GEO under accession number GSE116278 [102]. We selected transcriptomic data from a total of 32 samples which were grouped into three stages as follows: four males and four females at 6 PCW, corresponding to the beginning of gonadal sex differentiation (first stage); 8 males and 8 females at 7 PCW (second stage); as well as two males and two females at 13–14 PCW, and two males and two females at 17 PCW, corresponding to the end of sex differentiation (third stage). The total of 35,194 finely annotated transcripts in [66] were used to test methods #3 and #4.

Transcriptomic data analysis overview

The gonadal transcriptome of the sea bass, mouse, and human during sex differentiation were analyzed with two different methods: #3 (DEGs + WGCNA) and #4 (WGCNA + DEGs). In method #3, we first created a list of the DEGs between males and females at 250 dpf in sea bass and at 16.5 dpc in mouse and then we used this list for network construction. For the human, we used exactly the same list of differentially expressed transcripts between males and females at 6 PCW previously published [66]. In method #4, we first carried out WGCNA and then used the same list of DEGs as before to filter the dataset just after network generation but prior to network visualization. Therefore, we carried out a total of six analyses (two methods $\times$ three species). In the following sections, we describe in detail how we determined the DEGs and how we carried out WGCNA in each case.

Weighted gene co-expression network analysis

We implemented the WGCNA using the R package (v.1.51) according to the authors' recommendations [17, 103] in R statistical software [104, 105]. We first checked for the presence of outliers using the hierarchical clustering of samples with Euclidean distance [68]. In the sea bass dataset, one male at 110 dpf grouped with the male samples at 250 dpf at a height > 15 , distant from the rest of samples and thus removed from further analysis. The rest of the samples were used to generate the Pearson correlation matrices. Among the mouse samples, one was identified as an outlier using the hierarchical clustering of samples. One of the females at 16.5 dpc was at a height > 30 far from the rest of samples and hence, removed from further analysis. Thus, in total we analyzed 22, 11, and 32 samples from sea bass mouse, and human, respectively.

To build unsigned weighted networks, the adjacency matrix was calculated as $a_{mn} = |\text{cor}_{mn}|^\beta$, where a_{mn} is the adjacency between gene m and gene n , cor_{mn} is the Pearson correlation, and β is the soft-power threshold. Unsigned networks allow the connection of genes that are both positively and negatively correlated; the absolute value of the Pearson correlation is used as a co-expression similarity measure [106, 107]. In contrast, a signed network would not include the connections between strong negatively correlated genes [106, 107]. Sexual development in vertebrates is known to be orchestrated by mutually antagonizing male and female pathways [63] in which not only upregulation of a set of genes tilts the balance towards the development of one sex or the other but, importantly, the concomitant mechanism of active downregulation of numerous genes in the opposite sex is also required [95, 108]. Therefore, using the unsigned network allowed drawing the connections between genes or nodes that were either positively or negatively

correlated, since no difference was made between gene inhibition and activation patterns.

To fit the scale-free topology model, we tested several soft threshold powers (ranging from 1 to 20) to which co-expression similarity is raised. To quantify how well a network met the scale-free topology criterion, the model fit was measured as the signed linear regression model fitting index R^2 . After selection of the value leading to the best fit to the scale-free topology model (signed R^2 closest to 1), additional considerations described by the developers were taken into account: high mean connectivity (k), and the slope of the regression line between $\log_{10}(p(k))$ and $\log_{10}(k)$, closest to -1 [17, 103]. Subsequently, the adjacency matrix was transformed into a topological overlap matrix, and gene modules were detected by hierarchical average linkage clustering analysis for the gene dendrogram, setting the parameters as default (minimal gene module size = 30, and the threshold to group similar modules was set to 0.25).

Selection of modules associated with sex

After the modules were defined, the module eigengene (ME) distances were calculated to elucidate potential relationships of modules with two phenotypic traits: sex and age of the samples. For simplicity, we focused on the results related to sex development in the present study, although the study of other traits could be considered following the same workflow. Hence, we chose the significant modules that met the following thresholds regarding sex trait: absolute $R^2 > 0.5$ and $P < 0.001$ in the sea bass and human datasets or $P < 0.01$ for the mouse transcriptome. To assess the correlation strength, we calculated the module significance (MS), the average absolute GS of all the genes involved in the module. The key modules kept for further analysis were those with the highest MS score among all modules produced.

For the genes within the selected modules, we calculated intra-modular membership (k_{IM}) for each gene to determine how well-assigned is a gene within a module and its relationship with the trait of interest as described by Langfelder and Horvath [17]. Additionally, we calculated GS for the sex trait of each gene within the modules. The gene significance is defined as $-\log$ of the p -value of association of the gene with the trait, in our case, sex.

To identify the modules with most interesting genes associated to the trait of interest, we selected the modules with a significant correlation ($|\text{cor}| \geq 0.5$, $P < 0.001$ in sea bass and human; $|\text{cor}| \geq 0.5$, $P < 0.01$ in mouse). To identify the genes most interesting within those modules, we filtered by k_{IM} and GS, where the higher the absolute value of GS of a gene the more biologically significant it is for the trait of interest. Thus, we kept for further analysis all the genes within the selected modules with an

absolute gene significance higher than 0.2 and an absolute intra-modular membership higher than 0.8 [17].

Analysis of differentially expressed genes

The DEGs were determined by fitting a linear model, using the same empirical Bayesian statistics in both datasets. For the sea bass data, we used the Quantile method in the Linear Models for Microarray Analysis (Limma) (v. 3.44.3) R package [109]. We compared sea bass female vs male gonads at 110 and 250 dpf. The same statistical test was applied to the mouse transcriptome dataset using edgeR (v.3.30.3) [100] and GLimma package (v.2.2.0) [110] in R software (v. 3.4.1) [104, 105]. In this case, the comparison of female vs male at 16.5 dpc was used. For all the comparisons, genes with a false discovery rate (FDR) based on the Benjamini–Hochberg method were defined as differentially expressed (adjusted $P < 0.05$). We pre-processed the two transcriptomes using different packages because the dataset from the sea bass was obtained from a microarray experiment (normalized intensity values matrix), and the mouse data were obtained from an RNA sequencing experiment (normalized expression counts matrix). While data pre-processing steps were necessarily different because of the technologies used to obtain the data, the mathematics behind DEG determination were essentially the same.

Key genes with reproduction-related functions

To determine which methodology would unveil more genes relevant for the target process, we generated lists of genes previously known to be involved with gonadal sex differentiation for each of the two species studied. For the sea bass, we selected a total of 28 genes from the literature related to the sea bass or other well-studied fish species like zebrafish [35, 93, 111] (Additional file 6). In addition, we produced a heat map with gene hierarchical analysis of the selected genes in sea bass using gplots (v3.1.1) [112] and ggplot2 (v3.3.5) [113] packages. For the mouse, where gonadal sex differentiation has been thoroughly studied, we selected 50 genes compiled from recent and comprehensive reviews [93, 96, 114] (Additional file 7).

Up to this step, all analyses were performed in sea bass and mouse species to test whether the comparison between methods #3 and #4 was reproducible with transcriptomes of different species and obtained with different technologies (microarray and RNA-seq). From this point onwards, only the results from the sea bass were further explored by network visualization and identification of early gene expression markers of sex differentiation.

Network visualization and statistics

In order to compare the three methods involving network construction, we selected the most associated

module with sex (positively or negatively) to calculate network parameters. The modules were as follows: red (DEGs + WGCNA) vs pink (WGCNA + DEGs) in the sea bass, turquoise (DEGs + WGCNA) vs brown (WGCNA + DEGs) in the mouse, and yellow (DEGs + WGCNA) vs yellow-green (WGCNA + DEGs) in the human. We exported network results to Cytoscape software (version 3.5.1) [115] using the function *exportNetworkToCytoscape* from WGCNA package by including the adjacency matrix of each module as input and defining the adjacency threshold to 0.2. Cytoscape was then used to visualize and to analyze the networks using *NetworkAnalyzer* [69].

Identification of biomarkers of early gonadal sex differentiation

To find genes involved in the early stages of gonadal sex differentiation in the sea bass, we used the gene lists of the four modules associated with sex, filtered by gene significance and intra-modular membership as described above, i.e., absolute gene significance higher than 0.2 and an absolute intra-modular membership higher than 0.8. Additionally, for those genes that met the criteria, we checked which ones were also DEGs between males and females at 110 dpf. Among the selected genes, we checked for the presence of genes previously known to be involved in gonadal sex differentiation in this species.

Functional enrichment analysis of the mouse dataset

We chose to investigate potential differences in GO term analysis caused by different methods using the mouse, a model species with a much more curated and up to date annotation. The co-regulated DEGs associated with sex at 16.5 dpc were enriched for Gene Ontology analysis using the GO Enrichment Analysis bioinformatic PANTHER tool [116], with the list of genes captured by the RNA-seq experiment ($n = 14,088$) serving as the background reference list.

Abbreviations

acox3: Acyl-CoA Oxidase 3, Pristanoyl; *ajuba*: Ajuba LIM Protein; *amh*: Anti-Müllerian Hormone; *aqp10*: Aquaporin 10; *ar*: Androgen Receptor; *bmp2*: Bone Morphogenetic Protein 2; *ccna*: Cyclin A2; *cDNA*: Complementary DNA; *cor*: Correlation; *cox17*: Cytochrome C Oxidase Copper Chaperone COX17; *cox5a*: Cytochrome C Oxidase Subunit 5A; *ctnnb1*: β Catenin; *cyp11a1*: Cytochrome P450 Family 11 Subfamily A Member 1; *cyp11b1*: Cytochrome P450 Family 11 Subfamily B Member 1; *cyp17a1*: Cytochrome P450 Family 17 Subfamily A Member 1; DEGs: Differentially expressed genes; dpc: Days post coitum; dpf: Days post fertilization; DSDs: Disorders of sexual development; E: Embryonic day; *erb2*: Estrogen Receptor beta 2; ESTs: Expressed sequence tags; FDR: False discovery rate; *fgfr2*: Fibroblast Growth Factor Receptor 2; *fshr*: Follicle-Stimulating Hormone Receptor; *fst*: Follicle-statin; *gata6*: GATA Binding Protein 6; GEO: Gene Expression Omnibus; GO: Gene ontology; GS: Gene significance; GSD: Genetic sex determination; *gsdf*: Gonadal soma derived factor; *heph11*: Hephaestlin Like 1; *hsd17b10*: Hydroxysteroid 17-Beta Dehydrogenase 10; *hsd3b*: Hydroxy-Delta-5-Steroid Dehydrogenase, 3 Beta- And Steroid Delta-Isomerase 1; *hsp70*: Heat Shock Protein Family a Member; k: Connectivity;

k_{im} : Intra-modular membership; *krt8*: Keratin 8; *lhr*: Luteinizing Hormone/Choriogonadotropin Receptor; Limma: Linear Models for Microarray Analysis; ME: Module eigengene; MODA: Module differential analysis for weighted gene co-expression network; mRNA: Messenger RNA; MS: Module significance; *pacs13*: Protein Kinase C and Casein Kinase Substrate in Neurons 3; PCW: Post-conceptual week; *pop4*: POP4 Homolog, Ribonuclease P/MRP Subunit; PSD: Polygenic sex determining; *psma6b*: Proteasome 20S Subunit Alpha 6; RNA-seq: RNA sequencing; *rspo1*: R-spondin 1; *ship1*: Schwannomin Interacting Protein 1; SD: Standard deviation; *sox9a*: SRY-Box Transcription Factor 9 a; *sry*: Sex-determining region Y; *star*: Steroidogenic Acute Regulatory Protein; *tuba8*: Tubulin Alpha 8; *usp5*: Ubiquitin-Specific Peptidase 5; *vim*: Vimentin; WGCNA: Weighted gene co-expression network analysis; *wnt4*: Wnt Family Member 4; *zfpd3*: Zinc Finger AN1-Type Containing 3; β : Soft threshold.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-022-01398-w>.

Additional file 1. Table with DEGs in the differentiating gonads of the European sea bass between males and females.

Additional file 2. Table with DEGs in the differentiating gonads of the mouse between males and females.

Additional file 3. Table of co-expressed DEGs in the differentiating gonads of the European sea bass between males and females.

Additional file 4. Table of co-expressed DEGs in the differentiating gonads of the mouse between males and females.

Additional file 5. Significant GO terms enriched from the co-expressed DEGs in the differentiating gonads of the mouse between males and females.

Additional file 6. Key genes involved in the gonadal development of the European sea bass and other teleosts. Data compiled from [93, 111].

Additional file 7. Key genes involved in the gonadal development of the mouse. Data compiled from [93, 114, 117].

Additional file 8: Supplementary figure 1. Sample dendrogram and trait information on sex and age using hierarchical clustering analysis. **Supplementary figure 2.** Flow diagram and number of genes of methods #3 and #4. **Supplementary figure 3.** Determination of soft-threshold power in the WGCNA using the mouse gonadal transcriptome. **Supplementary figure 4.** Determination of soft-threshold power in the WGCNA using the human gonadal transcriptome. **Supplementary figure 5.** Identification of gene modules associated with sex in the mouse. **Supplementary figure 6.** Identification of gene modules associated with sex in the human. **Supplementary figure 7.** Filtering of genes based on network properties in the mouse. **Supplementary figure 8.** Filtering of genes based on network properties in the human. **Supplementary figure 9.** Identification of key genes involved in sex differentiation of mouse. **Supplementary figure 10.** DEGs co-expression network of the magenta module using method #4 in sea bass data. **Supplementary figure 11.** DEGs co-expression network of the pink module using method #4 in sea bass data. **Supplementary figure 12.** DEGs co-expression network of the sky blue module using method #4 in the sea bass data.

Acknowledgements

We thank Dr. Steve Horvath for his expert assistance and Dr. Frédéric Chalmel for his help in obtaining the data from their previous publication in the required format.

Authors' contributions

NS, LR, and FP designed research; NS collected and analyzed the data. NS wrote the first draft of the manuscript. NS and FP wrote the manuscript. All authors revised and approved the final version of the manuscript.

Funding

NS was supported by a Spanish Ministry of Science (SMS) predoctoral scholarship (BES-2017-079744). Research was supported SMS grant no.

PID2019-108888RB-I00 to FP. Also with funding from the Spanish government through the 'Severo Ochoa Centre of Excellence' accreditation (CEX2019-000928-S).

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories. The datasets analyzed are available at the GEO database with the accession numbers GSE115841 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115841>), GSE117590 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117590>), and GSE116278 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116278>).

Declarations

Ethics approval and consent to participate

Not applicable. No human subjects or animals were involved, we exclusively used publicly available datasets.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 22 February 2022 Accepted: 2 September 2022

Published online: 24 September 2022

References

1. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol*. 2017;13(5):e1005457.
2. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*. 1991;252(5013):1651–6.
3. Piétu G, Mariage-Samson R, Fayein NA, Matingou C, Eveno E, Houllatte R, et al. The Genexpress IMAGE Knowledge Base of the Human Brain Transcriptome: a prototype integrated resource for functional and computational genomics. *Genome Res*. 1999;9(2):195–209.
4. Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, et al. Characterization of the yeast transcriptome. *Cell*. 1997;88(2):243–51.
5. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57–63.
6. Parmigiani G, Garrett ES, Irizarry RA, Zeger SL. The analysis of gene expression data: an overview of methods and software. In: Parmigiani G, Garrett ES, Irizarry RA, Zeger SL, editors. *The analysis of gene expression data: methods and software*. New York: Springer; 2003. p. 1–45. Available from: https://doi.org/10.1007/0-387-21679-0_1. Statistics for Biology and Health. cited 2021 Jul 22.
7. Voshall A, Moriyama EN. Next-generation transcriptome assembly: strategies and performance analysis. *Bioinformatics in the era of post genomics and big data*. 2018. p. 15–36.
8. Ayers KL, Davidson NM, Demiyah D, Roeszler KN, Grützner F, Sinclair AH, et al. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol*. 2013;14(3):R26.
9. Clement TM, Anway MD, Uzumcu M, Skinner MK. Regulation of the gonadal transcriptome during sex determination and testis morphogenesis: comparative candidate genes. *Reproduction*. 2007;134(3):455–72.
10. Dergunova LV, Filippenko IB, Stavchansky VV, Denisova AE, Yuzhakov VV, Mozerov SA, et al. Genome-wide transcriptome analysis using RNA-Seq reveals a large number of differentially expressed genes in a transient MCAO rat model. *BMC Genomics*. 2018;19(1):655.
11. Lata C, Sahu PP, Prasad M. Comparative transcriptome analysis of differentially expressed genes in foxtail millet (*Setaria italica* L.) during dehydration stress. *Biochem Biophys Res Commun*. 2010;393(4):720–7.

12. Men X, Ma J, Wu T, Pu J, Wen S, Shen J, et al. Transcriptome profiling identified differentially expressed genes and pathways associated with tamoxifen resistance in human breast cancer. *Oncotarget*. 2017;9(3):4074–89.
13. Ribas L, Robledo D, Gómez-Tato A, Viñas A, Martínez P, Piferrer F. Comprehensive transcriptomic analysis of the process of gonadal sex differentiation in the turbot (*Scophthalmus maximus*). *Mol Cell Endocrinol*. 2016;15(422):132–49.
14. Yu Y, Zhu D, Ma C, Cao H, Wang Y, Xu Y, et al. Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development. *Crop J*. 2016;4(2):92–106.
15. Czerwinski M, Natarajan A, Barske L, Looger LL, Capel B. A timecourse analysis of systemic and gonadal effects of temperature on sexual development of the red-eared slider turtle *Trachemys scripta elegans*. *Dev Biol*. 2016;420(1):166–77.
16. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4(1):Article17.
17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;29(9):559.
18. Duan H, Ge W, Zhang A, Xi Y, Chen Z, Luo D, et al. Transcriptome analyses reveal molecular mechanisms underlying functional recovery after spinal cord injury. *Proc Natl Acad Sci*. 2015;112(43):13360–5.
19. Iancu OD, Colville A, Oberbeck D, Darakjian P, McWeeney SK, Hitzemann R. Coslicing network analysis of mammalian brain RNA-Seq data utilizing WGCNA and Mantel correlations. *Front Genet*. 2015;6:174.
20. Liu B, Zhan Y, Chen X, Hu X, Wu B, Pan S. Weighted gene co-expression network analysis can sort cancer-associated fibroblast-specific markers promoting bladder cancer progression. *J Cell Physiol*. 2021;236(2):1321–31.
21. Zhu L, Ding Y, Chen CY, Wang L, Luo Z, Kim S, et al. MetaDCN: meta-analysis framework for differential co-expression network detection with an application in breast cancer. *Bioinformatics*. 2017;33(8):1121–9.
22. Kakati T, Kashyap H, Bhattacharyya DK. THD-Module Extractor: an application for CEN module extraction and interesting gene identification for Alzheimer's disease. *Sci Rep*. 2016;30(6):38046.
23. Tesson BM, Breitling R, Jansen RC. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics*. 2010;11(1):497.
24. Li D, Brown JB, Orsini L, Pan Z, Hu G, He S. MODA: MODA: Module Differential Analysis for weighted gene co-expression network. Bioconductor version: Release (3.14); 2022. Available from: <https://bioconductor.org/packages/MODA/>. cited 2022 Jan 16
25. Zhao H, Du X, Zhang K, Liu Y, Wang Y, Liu J, et al. Weighted Correlation Network Analysis (WGCNA) of Japanese Flounder (*Paralichthys olivaceus*) embryo transcriptome provides crucial gene sets for understanding haploid syndrome and rescue by diploidization. *J Ocean Univ China*. 2018;17(6):1441–50.
26. Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun*. 2014;5(1):5748.
27. Deist MS, Gallardo RA, Bunn DA, Dekkers JCM, Zhou H, Lamont SJ. Resistant and susceptible chicken lines show distinctive responses to Newcastle disease virus infection in the lung transcriptome. *BMC Genomics*. 2017;28(18):989.
28. Amiri A, Coppola G, Scuderi S, Wu F, Roychowdhury T, Liu F, et al. Transcriptome and epigenome landscape of human cortical development modeled in brain organoids. *Science*. 2018;362(6420):eaat6720.
29. Chen S, Yang D, Lei C, Li Y, Sun X, Chen M, et al. Identification of crucial genes in abdominal aortic aneurysm by WGCNA. *PeerJ*. 2019;8(7):e7873.
30. Wang Q, Liu K, Feng B, Zhang Z, Wang R, Tang L, Li W, Li Q, Piferrer F, Shao C. Transcriptome of gonads from high temperature induced sex reversal during sex determination and differentiation in Chinese tongue sole, *Cynoglossus semilaevis*. *Front Genet*. 2019;22(10):1128.
31. Niemira M, Collin F, Szalkowska A, Bielska A, Chwiałkowska K, Reszec J, et al. Molecular signature of subtypes of non-small-cell lung cancer by large-scale transcriptional profiling: identification of key modules and genes by weighted gene co-expression network analysis (WGCNA). *Cancers*. 2020;12(1):37.
32. Tu J, Huang Z, Wang Y, Wang M, Yin Z, Mei X, et al. Transcriptome analysis of the procession from chronic pancreatitis to pancreatic cancer and metastatic pancreatic cancer. *Sci Rep*. 2021;11(1):3409.
33. Ayers KL, Lambeth LS, Davidson NM, Sinclair AH, Oshlack A, Smith CA. Identification of candidate gonadal sex differentiation genes in the chicken embryo using RNA-seq. *BMC Genomics*. 2015;16(1):704.
34. Piprek RP, Damulewicz M, Kloc M, Kubiak JZ. Transcriptome analysis identifies genes involved in sex determination and development of *Xenopus laevis* gonads. *Differ Res Biol Divers*. 2018;100:46–56.
35. Ribas L, Crespo B, Sánchez-Baizán N, Xavier D, Kuhl H, Rodríguez JM, et al. Characterization of the European sea bass (*Dicentrarchus labrax*) gonadal transcriptome during sexual development. *Mar Biotechnol N Y N*. 2019;21(3):359–73.
36. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, et al. Functional organization of the transcriptome in human brain. *Nat Neurosci*. 2008;11(11):1271–82.
37. Miller JA, Horvath S, Geschwind DH. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc Natl Acad Sci U S A*. 2010;107(28):12698–703.
38. Chen YC, Guo YF, He H, Lin X, Wang XF, Zhou R, et al. Integrative analysis of genomics and transcriptome data to identify potential functional genes of BMDs in females. *J Bone Miner Res Off J Am Soc Bone Miner Res*. 2016;31(5):1041–9.
39. Yin L, Cai Z, Zhu B, Xu C. Identification of key pathways and genes in the dynamic progression of HCC based on WGCNA. *Genes*. 2018;9(2):E92.
40. Cox B, Tsamou M, Vrijens K, Neven KY, Winckelmans E, De Kok TM, Plusquin M, Nawrot TS. A co-expression analysis of the placental transcriptome in association with maternal pre-pregnancy BMI and newborn birth weight. *Front Genet*. 2019;29(10):354.
41. Feltrin AS, Tahira AC, Simões SN, Brentani H, Martins DC. Assessment of complementarity of WGCNA and NERL results for identification of modules associated to schizophrenia spectrum disorders. *PLoS ONE*. 2019;14(1):e0210431.
42. Kim A, Yoon D, Lim Y, Roh HJ, Kim S, Park CI, et al. Co-expression network analysis of spleen transcriptome in rock bream (*Oplegnathus fasciatus*) Naturally Infected with Rock Bream Iridovirus (RBV). *Int J Mol Sci*. 2020;21(5):E1707.
43. Wang G, Yu J, Yang Y, Liu X, Zhao X, Guo X, et al. Whole-transcriptome sequencing uncovers core regulatory modules and gene signatures of human fetal growth restriction. *Clin Transl Med*. 2020;28(9):9.
44. Guillotin D, Taylor AR, Platé M, Mercer PF, Edwards LM, Haggart R, et al. Transcriptome analysis of IPF fibroblastic foci identifies key pathways involved in fibrogenesis. *Thorax*. 2021;76(1):73–82.
45. Deng SP, Zhu L, Huang DS. Mining the bladder cancer-associated genes by an integrated strategy for the construction and analysis of differential co-expression networks. *BMC Genomics*. 2015;16(Suppl 3):S4.
46. Lin X, Li J, Zhao Q, Feng JR, Gao Q, Nie JY. WGCNA reveals key roles of IL8 and MMP-9 in progression of involvement area in colon of patients with ulcerative colitis. *Curr Med Sci*. 2018;38(2):252–8.
47. Zuo Z, Shen JX, Pan Y, Pu J, Li YG, Shao XH, et al. Weighted gene correlation network analysis (WGCNA) detected loss of MAGI2 promotes chronic kidney disease (CKD) by podocyte damage. *Cell Physiol Biochem Int J Exp Cell Physiol Biochem Pharmacol*. 2018;51(1):244–61.
48. Deng T, Liang A, Liang S, Ma X, Lu X, Duan A, et al. Integrative analysis of transcriptome and GWAS data to identify the hub genes associated with milk yield trait in buffalo. *Front Genet*. 2019;10:36.
49. Di Y, Chen D, Yu W, Yan L. Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis. *Hereditas*. 2019;156:7.
50. Ding M, Li F, Wang B, Chi G, Liu H. A comprehensive analysis of WGCNA and serum metabolomics manifests the lung cancer-associated disordered glucose metabolism. *J Cell Biochem*. 2019;120(6):10855–63.
51. Pan S, Zhan Y, Chen X, Wu B, Liu B. Identification of biomarkers for controlling cancer stem cell characteristics in bladder cancer by network analysis of transcriptome data stemness indices. *Front Oncol*. 2019;9:613.
52. Liu S, Ye T, Li Z, Li J, Jamil AM, Zhou Y, Hua G, Liang A, Deng T, Yang L. Identifying hub genes for heat tolerance in water buffalo (*Bubalus bubalis*) using transcriptome data. *Front Genet*. 2019;13(10):209.

53. Kakati T, Bhattacharyya DK, Barah P, Kalita JK. Comparison of methods for differential co-expression analysis for disease biomarker prediction. *Comput Biol Med.* 2019;1(113):103380.
54. Bai KH, He SY, Shu LL, Wang WD, Lin SY, Zhang QY, et al. Identification of cancer stem cell characteristics in liver hepatocellular carcinoma by WGCNA analysis of transcriptome stemness index. *Cancer Med.* 2020;9(12):4290–8.
55. Cui W, Ma A. Transcriptome analysis provides insights into the effects of myo-inositol on the turbot *Scophthalmus maximus*. *Fish Shellfish Immunol.* 2020;1(106):691–704.
56. Huang Z, Ma A, Yang S, Liu X, Zhao T, Zhang J, et al. Transcriptome analysis and weighted gene co-expression network reveals potential genes responses to heat stress in turbot *Scophthalmus maximus*. *Comp Biochem Physiol Part D Genomics Proteomics.* 2020;1(33):100632.
57. Zhang J, Wang L, Xu X, Li X, Guan W, Meng T, Xu G. Transcriptome-based network analysis unveils eight immune-related genes as molecular signatures in the immunomodulatory subtype of triple-negative breast cancer. *Front Oncol.* 2020;18(10):1787.
58. Jiang C, Li Z, Wu Z, Liang Y, Jin L, Cao Y, et al. Integrated bioinformatics analysis of hub genes and pathways associated with a compression model of spinal cord injury in rats. *Med Sci Monit Int Med J Exp Clin Res.* 2020;26:e927107-1-e927107-13.
59. Zhu R, Lin W, Tang L, Hu Y. Identification of hub genes associated with adult acute myeloid leukemia progression through weighted gene co-expression network analysis. *Aging.* 2021;13(4):5686–97.
60. Zhao H, Tang X, Wu M, Li Q, Yi X, Liu S, Jiang J, Wang S, Sun X. Transcriptome characterization of short distance transport stress in beef cattle blood. *Front Genet.* 2021;10(12):616388.
61. WGCNA package: Frequently asked questions. Available from: <https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/faq.html>. cited 2022 Jun 1
62. Shao C, Li Q, Chen S, Zhang P, Lian J, Hu Q, et al. Epigenetic modification and inheritance in sexual reversal of fish. *Genome Res.* 2014;24(4):604–15.
63. Capel B. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat Rev Genet.* 2017;18(11):675–89.
64. Vandeputte M, Piferrer F. Genetic and environmental components of sex determination in the European sea bass. 2019. Available from: <https://digital.csic.es/handle/10261/175601>. cited 2021 Jan 4
65. Hughes IA, Houk C, Ahmed SF, Lee PA. Consensus statement on management of intersex disorders. *J Pediatr Urol.* 2006;2(3):148–62.
66. Lecluze E, Rolland AD, Filis P, Evrard B, Leverrier-Penna S, Maamar MB, et al. Dynamics of the transcriptional landscape during human fetal testis and ovary development. *Hum Reprod.* 2020;35(5):1099–119.
67. Lovell-Badge R, Robertson E. XY female mice resulting from a heritable mutation in the primary testis-determining gene. *Tdy Dev Camb Engl.* 1990;109(3):635–46.
68. WGCNA: R package for performing weighted gene co-expression network analysis. Available from: <https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/index.html>. cited 2022 Jun 2
69. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. *Nat Protoc.* 2012;7(4):670–85.
70. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLOS Comput Biol.* 2008;4(8):e1000117.
71. Ovens K, Eames BF, McQuillan I. Comparative analyses of gene co-expression networks: implementations and applications in the study of evolution. *Front Genet.* 2021;12:695399.
72. Rocha A, Zanuy S, Gómez A. Conserved anti-Müllerian hormone: anti-Müllerian hormone type-2 receptor specific interaction and intracellular signaling in teleosts. *Biol Reprod.* 2016;94(6):141.
73. Crespo B, Gómez A, Mazon MJ, Carrillo M, Zanuy S. Isolation and characterization of Ff1 and Gsdf family genes in European sea bass and identification of early gonadal markers of precocious puberty in males. *Gen Comp Endocrinol.* 2013;1(191):155–67.
74. Blázquez M, Piferrer F. Sea bass (*Dicentrarchus labrax*) androgen receptor: cDNA cloning, tissue-specific expression, and mRNA levels during early development and sex differentiation. *Mol Cell Endocrinol.* 2005;237(1–2):37–48.
75. Rodríguez L, Halm S, Mylonas C, Zanuy S. Gene expression of estrogen receptors α and β during early sexual differentiation in the European sea bass (*Dicentrarchus labrax*). *Fish Physiol Biochem.* 2003;28(1):175–175.
76. Raymond CS, Shamu CE, Shen MM, Seifert KJ, Hirsch B, Hodgkin J, et al. Evidence for evolutionary conservation of sex-determining genes. *Nature.* 1998;391(6668):691–5.
77. Staack A, Donjacour AA, Brody J, Cunha GR, Carroll P. Mouse urogenital development: a practical approach. *Differ Res Biol Divers.* 2003;71(7):402–13.
78. Matsushita S, Suzuki K, Murashima A, Kajioka D, Acebedo AR, Miyagawa S, et al. Regulation of masculinization: androgen signalling for external genitalia development. *Nat Rev Urol.* 2018;15(6):358–68.
79. Ottolenghi C, Omari S, Garcia-Ortiz JE, Uda M, Crisponi L, Forabosco A, et al. Foxl2 is required for commitment to ovary differentiation. *Hum Mol Genet.* 2005;14(14):2053–62.
80. Wang X, Su L, Pan X, Yao J, Li Z, Wang X, et al. Effect of BMP-6 on development and maturation of mouse preantral follicles in vitro. *Biotechnol Biotechnol Equip.* 2015;29(2):336–44.
81. Vainio S, Heikkilä M, Kispert A, Chin N, McMahon AP. Female development in mammals is regulated by Wnt-4 signalling. *Nature.* 1999;397(6718):405–9.
82. Timmons JA, Szkop KJ, Gallagher JJ. Multiple sources of bias confound functional enrichment analysis of global -omics data. *Genome Biol.* 2015;16(1):186.
83. Blázquez M, Navarro-Martín L, Piferrer F. Expression profiles of sex differentiation-related genes during ontogenesis in the European sea bass acclimated to two different temperatures. *J Exp Zool B Mol Dev Evol.* 2009;312(7):686–700.
84. Vining B, Ming Z, Bagheri-Fam S, Harley V. Diverse regulation but conserved function: SOX9 in vertebrate sex determination. *Genes.* 2021;12(4):486.
85. Voldoire E, Brunet F, Naville M, Volff JN, Galiana D. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. *PLoS ONE.* 2017;12(7):e0180936.
86. Geffroy B, Besson M, Sánchez-Baizán N, Clota F, Goikotxea A, Sadoul B, et al. Unraveling the genotype by environment interaction in a thermosensitive fish with a polygenic sex determination system. *Proc Natl Acad Sci.* 2021;118(50):e2112660118.
87. Díaz N, Piferrer F. Lasting effects of early exposure to temperature on the gonadal transcriptome at the time of sex differentiation in the European sea bass, a fish with mixed genetic and environmental sex determination. *BMC Genomics.* 2015;4(16):679.
88. Blázquez M, González A, Papadaki M, Mylonas C, Piferrer F. Sex-related changes in estrogen receptors and aromatase gene expression and enzymatic activity during early development and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *Gen Comp Endocrinol.* 2008;158(1):95–101.
89. Mateos J, Mañanos E, Carrillo M, Zanuy S. Regulation of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) gene expression by gonadotropin-releasing hormone (GnRH) and sexual steroids in the Mediterranean Sea bass. *Comp Biochem Physiol B Biochem Mol Biol.* 2002;132(1):75–86.
90. Gornati R, Terova G, Vigetti D, Prati M, Saroglia M, Bernardini G. Effects of population density on seabass (*Dicentrarchus labrax*, L.) gene expression. *Aquaculture.* 2004;230(1):229–39.
91. Socorro S, Martins RS, Deloffre L, Mylonas CC, Canario AVM. A cDNA for European sea bass (*Dicentrarchus labrax*) 11 β -hydroxylase: gene expression during the thermosensitive period and gonadogenesis. *Gen Comp Endocrinol.* 2007;150(1):164–73.
92. Cerdá-Reverter JM, Anglade I, Martínez-Rodríguez G, Mazurais D, Muñoz-Cueto JA, Carrillo M, et al. Characterization of neuropeptide Y expression in the brain of a perciform fish, the sea bass (*Dicentrarchus labrax*). *J Chem Neuroanat.* 2000;19(4):197–210.
93. Nagahama Y, Chakraborty T, Paul-Prasanth B, Ohta K, Nakamura M. Sex determination, gonadal sex differentiation, and plasticity in vertebrate species. *Physiol Rev.* 2021;101(3):1237–308.
94. Menke DB, Koubova J, Page DC. Sexual differentiation of germ cells in XX mouse gonads occurs in an anterior-to-posterior wave. *Dev Biol.* 2003;262(2):303–12.

95. Munger SC, Aylor DL, Syed HA, Magwene PM, Threadgill DW, Capel B. Elucidation of the transcription network governing mammalian sex determination by exploiting strain-specific susceptibility to sex reversal. *Genes Dev.* 2009;23(21):2521–36.
96. Wang J, Tian GG, Zheng Z, Li B, Xing Q, Wu J. Comprehensive transcriptomic analysis of mouse gonadal development involving sexual differentiation, meiosis and gametogenesis. *Biol Proced Online.* 2019;21:20.
97. Wang J, Tian GG, Zheng Z, Li B, Xing Q, Wu J. Comprehensive transcriptomic analysis of mouse gonadal development involving sexual differentiation, meiosis and gametogenesis. GEO Accession viewer. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117590>. cited 2022 Jul 25
98. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–60.
99. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma Oxf Engl.* 2014;30(7):923–30.
100. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40.
101. Law CW, Alhamdoosh M, Su S, Dong X, Tian L, Smyth GK, Ritchie ME. RNA-seq analysis is easy as 1–2–3 with limma, Glimma and edgeR. *F1000Res.* 2016;5:ISCB Comm J-1408.
102. Lecluze E, Rolland AD, Filis P, Evrard B, Leverrier-Penna S, Maamar MB, et al. Dynamics of the transcriptional landscape during human fetal testis and ovary development. GEO Accession viewer. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116278>. cited 2022 Jul 25
103. Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering. *J Stat Softw.* 2012;46(11). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3465711/>. cited 2021 Jan 4
104. R: The R Project for Statistical Computing. Available from: <https://www.r-project.org/>. cited 2022 Jan 16
105. RStudio | Open source & professional software for data science teams - RStudio. Available from: <https://www.rstudio.com/>. cited 2022 Jan 16
106. Mason MJ, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics.* 2009;10(1):327.
107. Dai R, Xia Y, Liu C, Chen C. csuWGCNA: a combination of signed and unsigned WGCNA to capture negative correlations. *bioRxiv.* 2019:288225. <https://doi.org/10.1101/288225>.
108. Munger SC, Natarajan A, Looger LL, Ohler U, Capel B. Fine time course expression analysis identifies cascades of activation and repression and maps a putative regulator of mammalian sex determination. *PLoS Genet.* 2013;9(7):e1003630.
109. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47–e47.
110. Su S, Law CW, Ah-Cann C, Asselin-Labat ML, Blewitt ME, Ritchie ME. Glimma: interactive graphics for gene expression analysis. *Bioinforma Oxf Engl.* 2017;33(13):2050–2.
111. Ribas L, Liew WC, Diaz N, Sreenivasan R, Orbán L, Piferrer F. Heat-induced masculinization in domesticated zebrafish is family-specific and yields a set of different gonadal transcriptomes. *Proc Natl Acad Sci U S A.* 2017;114(6):E941–50.
112. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: various R programming tools for plotting data. 2020. Available from: <https://CRAN.R-project.org/package=gplots>. cited 2022 Jan 16
113. Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. 2021. Available from: <https://CRAN.R-project.org/package=ggplot2>. cited 2022 Jan 16
114. Yildirim E, Aksoy S, Onel T, Yaba A. Gonadal development and sex determination in mouse. *Reprod Biol.* 2020;20(2):115–26.
115. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.
116. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2019;47(Database issue):D419–26.
117. Ribas L, Crespo B, Sánchez-Baizán N, Xavier D, Kuhl H, Rodríguez JM, et al. Characterization of the European sea bass (*Dicentrarchus labrax*) gonadal transcriptome during sexual development. GEO Accession viewer. Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115841>. cited 2022 Jul 25

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

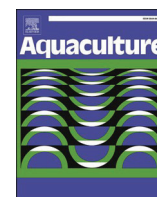
Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions





Development of epigenetic biomarkers with diagnostic and prognostic value to assess the lasting effects of early temperature changes in farmed fish

Alejandro Valdivieso^{a,1}, Núria Sánchez-Baizán^a, Nikos Mitrizakis^b, Nikos Papandroulakis^b, Francesc Piferrer^{a,*}

^a Institut de Ciències del Mar (ICM), Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, Spain

^b Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Center of Marine Research (HCMR), Heraklion, Greece

ARTICLE INFO

Keywords:

DNA methylation-based epigenetic biomarkers
Gene expression
Thermal stress
Key performance indicators
European sea bass aquaculture

ABSTRACT

Epigenetic changes such as DNA methylation modifications at specific loci in the genome in response to environmental cues can appear long before the appearance of transcriptomic or other phenotypic changes with potential consequences for performance. Thus, epigenetic DNA methylation-based biomarkers hold great promise in aquaculture and conservation biology because they can have prognostic value. However, to the best of our knowledge, a clear demonstration of such biomarkers has not been provided yet. Temperature during early stages of development is the main abiotic factor determining larval quality, with effects potentially persisting until adulthood and influencing performance. The objective of this study was to test whether epigenetic biomarkers could be developed as reliable indicators of the quality of the thermal environment during European sea bass (*Dicentrarchus labrax*) early development and their relation to long-lasting phenotypic consequences. We exposed European sea bass embryos and newly hatched larvae to different temperature regimes and sampled fish thereafter at three time points: at larval, juvenile and adult stages. Here, we provide an analytical strategy combining Reduced Representation Bisulfite sequencing (RRBS) and RNA-sequencing (RNA-seq) to develop epigenetic and transcriptomic biomarkers with both prognostic and diagnostic value. Specifically, we report a series of DNA-methylation biomarkers linked to short-term, mid-term, long-term, and very-long term effects of temperature in both somatic and reproductive tissues. Among the different biomarkers identified, one promising example is the methylation status of keratin-associated protein 10-4 (*krtap10-4*), which fulfills the criteria of an epigenetic biomarker with both prognostic and diagnostic value. This study constitutes a first step toward providing the European sea bass farming sector with a set of epigenetic biomarkers that can be used to screen batches of larvae to determine the quality of their thermal environment and to anticipate their performance when juveniles and adults. Our study provides the way for similar developments for other species, farmed or not.

1. Introduction

Larval rearing is one of the most crucial and delicate procedures during modern, intensive fish farming production and, thus, many factors, both biotic and abiotic, can influence its success (Conceicao and Tandler, 2018). Temperature affects many biological functions at different levels of organization (Schulte et al., 2011), being considered also the main abiotic factor determining fish larval quality and survival

(Pepin, 1991). The early stages of development are particularly the most sensitive to temperature (Burggren and Mueller, 2015), while these effects can persist until adulthood influencing growth performance and product quality (Jonsson and Jonsson, 2014).

The European sea bass, *Dicentrarchus labrax*, is, after the Atlantic salmon, *Salmo salar*, and rainbow trout, *Onchorhynchus mykiss*, the third most important species for finfish aquaculture in Europe in terms of volume, and the most important in the Mediterranean region, with over

* Corresponding author at: Institut de Ciències del Mar (ICM), Consejo Superior de Investigaciones Científicas (CSIC), Passeig Marítim, 37-45, 08003 Barcelona, Spain.

E-mail address: piferrer@icm.csic.es (F. Piferrer).

¹ Current address: IHPE, Univ Montpellier, CNRS, IFREMER, Univ Perpignan Via Domitia, Montpellier, France

<https://doi.org/10.1016/j.aquaculture.2022.738918>

Received 28 May 2022; Received in revised form 5 October 2022; Accepted 7 October 2022

Available online 12 October 2022

0044-8486/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

208,000 tons produced in 2019 (FEAP, 2020, <https://feap.info/index.php/data/>). The European sea bass is a eurythermal marine teleost (Pickett and Pawson, 1994). Its thermal tolerance in juveniles acclimated to 20 °C ranges from 4.7 °C to 34.8 °C (Dülger et al., 2012). In the Mediterranean sea, spawning typically takes place during winter. In hatcheries eggs are incubated and hatch at around 13–14 °C, and larval rearing takes place usually between 15 and 21 °C (Morretti, 1999). Temperature has a profound effect on the European sea bass ontogenetic plasticity (Koumoundouros et al., 2001). It is a gonochoristic species with sexual size dimorphism in favor of females and a polygenic system of sex determination (Vandeputte et al., 2007), where genetics and temperature contribute equally to sex ratios (Piferrer et al., 2005). Thus, proper temperature control during their embryonic and first larval stages of development is essential to minimize skeletal deformities (Georgakopoulou et al., 2007), prevent precocious male maturation (Felip et al., 2006), maximize growth, optimize sex ratios (Navarro-Martín et al., 2009), and to increase overall product quality (López-Albors et al., 2008).

Epigenetics concerns the study of phenomena and mechanisms that cause chromosome-bound, heritable changes to gene expression that are not dependent on changes to the DNA sequence (Deans and Maggert, 2015). Epigenetic mechanisms, namely, DNA methylation, histone modifications and non-coding RNAs, regulate gene expression by integrating genomic and environmental information, thus contributing to phenotypic plasticity (Duncan et al., 2014). Epigenetics holds great promise in aquaculture (Gavery and Roberts, 2017; Granada et al., 2018; Piferrer et al., 2015). Epigenetic alterations in response to environmental stimuli, for example, changes in DNA methylation levels in specific loci in the genome that correlate to a certain phenotype, can be used as epigenetic biomarkers. Of note, epigenetic biomarkers not only can have diagnostic but also prognostic value, since sometimes epigenetic changes appear long before gene expression or other phenotypic changes become apparent (Costa-Pinheiro et al., 2015). In the European sea bass, DNA methylation-based biomarkers have already been developed to predict sex (Anastasiadi et al., 2018), assess the early stages of domestication (Anastasiadi and Piferrer, 2019b), determine biological age (Anastasiadi and Piferrer, 2019a), assess the consequences of heat waves (Anastasiadi et al., 2021), and to evaluate the stress response associated with rearing density (Krick et al., 2021). However, and to the best of our knowledge, there are no epigenetic biomarkers in the European sea bass or, for that matter, in any other farmed fish species, that can be used to link the temperature regime during early stages of development with subsequent aquaculture performance. Developing epigenetic biomarkers that link adult phenotypic traits to early environmental conditions is a major goal of epigenetics in aquaculture as well as in conservation biology.

The objective of this study was to assess the effects of temperature changes during embryo and early larval development on the establishment of epigenetic and transcriptomic biomarkers that could correlate with phenotypic traits later in development in both somatic and reproductive tissues. Specifically, we tested whether these biomarkers could be developed as reliable indicators of the quality of the thermal environment during the most critical stages of development of fish, where temperature may have long-lasting phenotypic consequences. This constitutes a novel approach and a first step toward the ultimate goal of providing the aquaculture sector with a set of epigenetic biomarkers with diagnostic and prognostic value to improve management.

2. Materials and methods

2.1. Experimental facilities

Fish were reared at the Hellenic Center of Marine Research (HCMR) in Crete, Greece. The HCMR aquaculture facilities are certified from the national veterinary authority (code GR94FISH0001) and are licensed for operations of breeding and experimentation with fish issued by Region

of Crete, General Directorate of Agricultural & Veterinary no. 3989/01.03.2017 (approval codes EL91-BIObr-03 and EL91-BIOexp-04). The experimental protocol was approved by the Veterinarian Authority of the Region of Crete with the 255,332/29-11-2017 document. Animal experiments were carried out in accordance with the EU Directive 2010/63/EU.

2.2. Egg incubation and temperature treatments

All eggs used in this study came from the same mass spawn of captive breeders. Approximately 240,000 eggs were distributed into six 500-l tanks at about 60,000 eggs per tank. Eggs were incubated at 14 °C. For the rearing recirculation systems, biological filters were used. Water temperature was adjusted using chillers and heaters, which were automatically controlled with electronic sensors. Hatching and larval rearing were performed by using the “pseudogreen water” methodology (Papandroulakis et al., 2001). With a hatching rate close to 95% the initial number of larvae was approximately 54,000 per tank.

Three groups of fish were subjected in duplicate to 14 °C, 17 °C, and 20 °C (two of the six tanks mentioned above randomly assigned to each group) from the stage of epiboly onset to the end of the yolk-sac larval stage and the beginning of exogenous feeding (Fig. S1A). Given that the protocols for European seabass larval rearing are applied with temperature around 17 °C, the choice of 3 °C difference for the upper and lower limit is considered sufficient for the purpose of the study and potentially applicable at commercial hatcheries.

In all cases, the experimental temperature was set 1 day post fertilization (dpf) and lasted until 8 dpf in the case of 17 °C and 20 °C, while for the 14 °C the duration was until 12 dpf (different durations compensated for the effect different temperatures on growth). In this study, therefore, and following the ARRIVE guidelines (Percie du Sert et al., 2020), there were three experimental groups each carried out in replicate (being the tanks the biological replicates), and the group exposed to 17 °C was the control group. Subsequently, all groups were kept under identical rearing conditions including temperature (17 °C) up to the end of the trial. Acclimation of the eggs from the spawning temperature (14 °C) to the test temperatures, and then to the common temperature for larval rearing (17 °C) was performed at the rate of 0.5 °C and 0.2 °C h⁻¹, respectively. Fish were weighed periodically at the times indicated below.

2.3. Larval and juvenile rearing

Larvae were successively fed on enriched (DHA Protein Selco, INVE) rotifers until they reached 5.5 mm total length (TL), enriched with rotifers and *Artemia* Instar 1 nauplii (AF, INVE) when they were between 6 and 7.5 mm TL, and enriched with *Artemia* nauplii (Easy DHA Selco, INVE) afterwards. Rotifer concentration in the tanks was adjusted three times daily to 6–8 individuals·ml⁻¹. The concentration of *A. nauplii* was adjusted three times daily to 0.5–2 individuals·ml⁻¹. During the rotifer-feeding period, live microalgae *Chlorella* sp. were added three times daily at a concentration of $3 \pm 1 \times 10^5$ cells ml⁻¹. Weaning to inert diets (O-range, INVE) started when fish were approximately 9.5 mm TL. In Table S1 the feeding regime per experimental condition is shown.

Air-blowing skimmers were installed since first feeding until 29 dpf to keep the water surface free from lipids and enhance swim bladder inflation. Water oxygen concentration was 7.68 ± 0.9 , pH 8.37 ± 0.12 , salinity 35–36‰ and photoperiod 12L:12D. After larval rearing (65 dpf) the survival was estimated and fish were transferred for pre-growing in 10 m³ tanks, supplied with seawater from a deep well and operating in an open water system. The survival rate (mean $\pm$ SD) at the end of larval rearing was $19.6 \pm 5.1\%$ for the control group exposed to 17 °C (16.0% and 23.2% for the two replicates); $18.5 \pm 2.4\%$ for the group exposed to 14 °C (16.8% and 20.2% for the two replicates) and $17.2 \pm 3.9\%$ for the group exposed to 20 °C (14.5% and 20.0% for the two replicates). Water oxygen saturation was 7.0 ± 0.9 , pH 7.6 ± 0.3 , salinity 35–36‰, and

photoperiod natural. During pre-growing (65–130 dpf) the larvae were grouped per temperature treatment resulting in one group per experimental temperature. Following the standard husbandry procedures, the juveniles were sorted by size, removing the very small and the very large individuals in order to create a more homogenous size distribution among the three experimental groups and were also selected for swim bladder with a salinity test.

2.4. Coping style test

The test was performed following a standard procedure developed at Marine Biology, Biotechnology and Aquaculture (IMBBC) in the frame of the FP7 project “A new integrative framework for the study of fish welfare based on the concept of allostasis, appraisal and coping styles, COPEWELL, (Grant agreement ID: 265957)”. From each temperature treatment, three subgroups were randomly created when fish had approximately 130 dpf and there was no size difference between the groups. We did not consider sex in this test because the earliest stage in which gonads can be sexed is around 150 dpf histologically (Papadaki et al., 2005). These groups were transferred to cages for on-growing after being subjected to a conditioned place preference test (CPP) to characterize fish coping style. Briefly, the methodology consisted in: (i) restraining the individuals in a 1 m³ containment, placed within the holding tank, for 20 h, (ii) allowing the individuals to pass through three open holes (10 cm diameter), constructed close to the water surface while feed was distributed outside the containment by the use of an automatic feeder, and (iii) recording the number of individuals passed through the holes, during a time course of two hours. At the end of the test, six groups were made according to the temperature treatment and the coping style expressed, mentioned as “*Bold*” for the proactive individuals, i.e. the ones exiting the containment and “*Shy*” for the reactive individuals i.e. the ones remaining in the containment.

2.5. Rearing in sea water

The six original groups (from 2000 to 3000 individuals each) were then transported for further on-growing in net-open cages. At this time fish were approximately 140 dpf and had a mean weight of 1.72 ± 0.40 g. On-growing lasted for nine months and was performed in the pilot scale cage facility of the Institute of Aquaculture (HCMR) located at the North-West of Crete in Souda Bay (35°28'48.3"N 24°06'41.8"E). The mean water depth at the site is about 60 m (ranging from 30 to 90 m) with a continuous water circulation. Water temperature varied from 15 °C to 28 °C, and salinity was constant all year round at about 39 ‰. Water temperature and dissolved oxygen concentration were monitored daily using digital probes (HQ 40d; Hach SA, Loveland, CO, USA).

Rearing was performed in polyethylene cages (rectangular and/or circular) and standard rearing methodologies were applied (Pavlidis and Mylonas, 2011). The cage size at the beginning of the rearing was approximately 35 m³ (3 m high x 3 m wide x 4 m depth) and increased to approximately 50 m³ (3 × 3 × 6 m) when fish reached a mean weight of about 50 g. Rearing conditions were the standard followed for the European seabass with daily feeding using automatic feeders with standard commercial diets based on feeding tables available/adapted in the farm.

2.6. Samplings

The tissues selected for this study were sampled at three time points. After each sampling, tissues were kept in RNA-later solution (ThermoFisher Scientific) and stored at –80 °C until further analysis. Whole larvae at the first feeding stage (8 or 12 dpf depending on the temperature group) were sampled and four pools of ten larvae per temperature group were used for RNA-sequencing. Another four different pools of ten larvae each were made for DNA methylation analysis (see details below). At the past metamorphosis stage (90 dpf), the central part of the body trunk of juveniles, consisting mostly of muscle tissue was taken,

and two different pools with four fish each were used for RNA-seq and DNA methylation analysis, respectively. Adult fish were sampled at 482 dpf, and testes of four individual fish per each temperature treatment were selected for RNA-seq and DNA methylation analysis. The testis was used for two reasons: first, because we had examined the very long-term effects of temperature on this organ in a previous study (Anastasiadi et al., 2021), and thus could compare results; and second because epigenetic information can be transmitted through the sires in the European sea bass (Anastasiadi et al., 2018). To avoid possible size-dependent differences that could confound the results, we selected larvae samples at the first feeding stage of the same size. Similarly, juvenile samples were weighted to ensure that all samples contributed equally to the pool. In addition, during the adult fish sampling at 482 dpf, the total weight was determined in 20 fish (ten males and ten females) randomly taken in each temperature treatment. Phenotypic sex was identified in the same fish by visual inspection, and gonads were weighted. Gonadosomatic index (GSI) was calculated as gonad weight (g) / total fish weight (g) × 100.

2.7. RNA and DNA extraction

Tissues for transcriptomic analysis were removed from the RNA-later® solution, samples were immersed into TRIzol® Reagent (ThermoFisher Scientific), and homogenized manually. RNA was extracted from larvae, juveniles, and adults according to the manufacturer's instructions. RNA quality was evaluated with Agilent RNA 6000 Nano Kit (Agilent), and samples with RNA Integrity Number (RIN) > 8 were used for RNA-seq library preparation. Genomic DNA was extracted using the Blood and Cell Culture DNA Mini kit (13,323, Qiagen) according to the manufacturer's instructions. The quality of the DNA was checked with NanoDrop ND-1000 (Thermo Fisher Scientific), and concentration was measured by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific).

2.8. Sea bass reference genome

The European sea bass reference genome (GCA_000689215.1) (Tine et al., 2014) was used for this study. For methylation analysis, the reference genome was bisulfite converted using Bismark software (v. 0.22.3) applying the `bismark_genome_preparation` function (Krueger and Andrews, 2011). The start and end positions for each gene (gene body region) in the genome were taken from Biomart (Durinck et al., 2005, 2009).

2.9. RNA-sequencing (RNA-seq) and differential gene expression analysis

The RNA libraries of larvae, juveniles, and adults were prepared using the mRNA-Seq preparation kit (Illumina) according to the manufacturer's protocol at the Centre for Genomic Regulation (CRG) service within the Barcelona Biomedical Research Park (PRBB). Briefly, 0.5 µg of RNA were used for poly-A-based mRNA enrichment selection using oligo-dT magnetic beads followed by fragmentation of divalent cations at elevated temperature resulting into fragments of 80–250 base pairs (bp), with the major peak at 130 bp. First-strand cDNA synthesis by random hexamers and reverse transcriptase was followed by the second-strand cDNA synthesis. Double-stranded cDNA was end-repaired, 3'-adenylated, and the 3'-"T" nucleotide with the Illumina adapter was used to ligate the indexed adapters. The ligation product was amplified using 12 PCR cycles. Each library was sequenced using 50 bp paired-end mode on an Illumina HiSeq 2000 instrument following the manufacturer's protocol. The quality of raw reads was checked using FastQC software (v. 0.11.9) (Wingett and Andrews, 2018). Adapters and sequences with a Phred quality score lower than 30 were trimmed with the *Bbduk* software (v. 38.90) (Bushnell, 2014). The HISAT2 software (v. 2.1.0) (Kim et al., 2019) was used first to index the reference genome using `hisat2-index build` command and then to align the trimmed reads with default parameters. Finally, reads were counted with `featureCounts`

software (Liao et al., 2014) (v. 2.0.1) at the gene level. Detailed information of the samples for RNA-seq analysis in each tissue, treatment and control quality statistics are listed in Table S2.

Differential gene expression (DEGs) analysis was conducted using the *edgeR* package (v. 3.28.1) (McCarthy et al., 2012; Robinson et al., 2010). The lowly expressed genes were filtered using the “filterByExpr” function (Chen et al., 2016) among all samples. Retained counts were normalized using the upper-quartile “calcNormFactors” function (Anders and Huber, 2010). Estimates for negative binomial (NB) dispersion parameters were generated using “estimateDisp” function (Chen et al., 2014; Phipson et al., 2016). A quasi-likelihood binomial generalized log-linear model was fitted using “glmQLFit” (Lun and Smyth, 2016; Lund et al., 2012) and the “glmQLTest” function was used to compare the different contrasts within the matrix design (Lund et al., 2012; Chen et al., 2016; Lun and Smyth, 2016; Phipson et al., 2016; Lun and Smyth, 2017). Finally, a false discovery rate correction (FDR or Q-value, < 0.05) was applied in all contrasts (Benjamini and Hochberg, 1995).

2.10. Reduced representation bisulfite sequencing (RRBS)

The RRBS libraries of larvae, juveniles, and adults were prepared using the Premium RRBS kit (Diagenode, C02030032) according to the manufacturer's instructions. Qubit DNA High Sensitivity assays (ThermoFisher Scientific) were used to quantify the final concentrations of the libraries and their quality was checked using the DNA High Sensitivity Assay (Agilent). In total, 36 libraries (i.e., four biological replicates x three stages x three temperatures) were grouped in four pools of nine samples. The four pools of libraries were sequenced on an Illumina HiSeq 2500 instrument using single-end mode and a read length of 50 bp.

Quality control of reads were performed using the FastQC software (v. 0.11.9) (Wingett and Andrews, 2018). Trimmomatic software (v.0.36) (Bolger et al., 2014) was used for trimming the adaptors with parameters: sliding window 4, Phred score > 20, and adaptive quality trimming with largest length of 20 bp. Trimmed reads were aligned to the reference genome using the BSMAP software (v. 2.90) (Xi and Li, 2009) with RRBS mode and filtering for at least 10× coverage. In order to extract methylation data only from CG context (CpG) we used the methratio.py function from BSMAP. The bisulfite conversion efficiency of the libraries was checked following the instructions from the Premium RRBS kit.

The *methylKit* package (v. 1.14.2) (Akalın et al., 2012) was used to calculate differentially methylated cytosines (DMCs). To do so, coverage was normalized across samples for each tissue and only CpGs covered in all samples from the same tissue were maintained for further analysis. To detect DMCs a logistic regression model was applied with the F-test, which identifies methylation differences with the following chosen criteria: methylation differences larger than 15% and Q-value < 0.01 adjusted from the SLIM method (Wang et al., 2011). A detailed information of the samples for the methylation analysis in each tissue and treatment and control quality statistics are listed in Table S3. Finally, the *GenomicRanges* package (v. 1.38) (Lawrence et al., 2013) was used to identify those genes that contained at least one DMC within their gene body region ±1000 bp.

2.11. Gene ontology terms and Kyoto encyclopedia of genes and genomes pathways analysis

The Gene Ontology terms (GO-terms) and the Kyoto Encyclopedia of Genes and Genomes pathways (KEEGs) analyses were carried using the *TopGO* package (v. 2.46.0) (Alexa and Rahnenfuhrer, 2010) adapted for the European sea bass GO-terms database. Only the Biological Process (BP) category was considered relevant for this study. The several GO-terms carried out were simplified using the R code script from Revigo software (Supek et al., 2011) with settings: medium (0.7) and SimRel

semantic similarity method (Schlicker et al., 2006).

2.12. Data analysis

Sex ratio analysis and behavior tests were analyzed within each temperature treatment using Chi-squared test. A two-way ANOVA with Tukey post hoc test was applied to weight and GSI data. For methylomics and transcriptomics, the middle temperature, 17 °C, was considered the reference or control temperature. Therefore, two different comparisons at each one of the three sampling points were carried along in this study: the effects of cold temperature (i.e., 14 °C vs. 17 °C) and the effects of elevated temperature (i.e., 20 °C vs. 17 °C). In all analysis significant differences were considered when $P < 0.05$. Statistical analysis were performed with R software (v. 4.1.2) (R Core Team, 2020) using in Rstudio (v. 2021.09.0) (RStudio Team, 2020). Graphs were created using the *ggplot2* package (v. 3.3.5) (Wickham, 2011).

Samples for transcriptomic and DNA methylation analysis were selected based on the integrity of RNA (> 8 RIN level), the bisulfite conversion efficiency (> 99.0%), and the aggrupation of the samples in a Principal Component Analyses (PCA) to detect the outliers for the same group. Those samples which did not meet these criteria were discarded. In total, 9 and 6 out of a total of 36 and 36 samples were discarded for RRBS and RNA-seq analysis, respectively. The number of samples used for each comparison by each tissue is presented in Table S4.

3. Results

3.1. Growth, gonadal development, sex ratio, and coping style

At 95 dpf, the weight (mean ± SEM) was 1.08 ± 0.26 g for the 17 °C-reared fish; 1.07 ± 0.28 g for the 14 °C-reared fish; and 1.07 ± 0.26 g for the 20 °C-reared fish, without statistically significant differences ($P > 0.05$). At 482 dpf, females tended to have a higher mean weight than males, but no significant sexual dimorphism was observed in any of the three temperatures tested. Although fish exposed to 14 °C were bigger than those exposed at 20 °C, (Two-way ANOVA: $F = 4.37$, $P = 0.017$) there were no statistical differences between fish exposed at 14 °C vs. those exposed at 17 °C and the same applied between fish exposed to 20 °C vs. those exposed to 17 °C (Fig. S1B). Analysis of the GSI showed that females exposed to 20 °C had significantly higher mean values ($P < 0.05$) than females exposed at 14 °C, but there were no differences in the GSI of males (Fig. S1C). Sex ratios progressively increased in favor of males with temperature, without significant differences (Fig. S1D). Regarding copying style, the number of bold fish increased in the group exposed at 20 °C (Fig. S1E).

3.2. Overview of DNA methylation and gene expression changes as a function of temperature and tissue

We carried out transcriptomic and DNA methylation analyses to identify DEGs and DMCs in larvae, juveniles and adults in two different comparisons: 14 °C vs. 17 °C, and 20 °C vs. 17 °C temperatures.

Regarding transcriptomic results for larvae, juveniles, and adults, we obtained 8490, 58 and 28 DEGs between 14 °C vs. 17 °C (Tables S5, S6 and S7, respectively), and 8128, 42 and 53 DEGs between 20 °C vs. 17 °C (Tables S8, S9, and S10, respectively). On the other hand, regarding DNA methylation, the number of DMCs were 27,286 and 22,166 in larvae (Tables S11 and S12), 9884 and 15,886 in juveniles (Tables S13 and S14) and 7254 and 49,610 in adults (Tables S15 and S16) from the 14 °C vs. 17 °C and 20 °C vs. 17 °C comparisons, respectively. Finally, for each tissue and contrast, we identified 7231 genes which had at least one DMC (Table S17).

Next, taking all gene lists together we identified 163 genes in the 14 °C vs. 17 °C comparison and 264 genes in the 20 °C vs. 17 °C comparison with DMCs that were common to larvae, juveniles, and adults (Fig. S2A and B). Furthermore, by comparing these 2 lists we identified 70 genes

in common regardless of the direction of temperature change (Table S18). These are then genes, e.g., amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase (*agl*), the methylation of which is affected in larvae and maintained in both juveniles and adults and represent indicators of long-term epigenetic (but not transcriptomic) effects of temperature on sea bass tissues. This list also includes jumonji domain containing 7 (*jmjd7*) and potassium voltage-gated channel subfamily H member 5 (*kcnh5b*), two genes that are related to environmental signal transduction, and zinc finger FYVE-type containing 9 (*zfyve9a*), which participates in the TGF- β signaling pathway.

Finally, we were interested to know if there were genes that were considered both DEG and also contained at least one DMC in their body region. Results showed that 1350 (Table S19) and 723 (Table S20) genes in larvae, 6 (Table S21) and 7 genes (Table S22) in juveniles, and the gene potassium voltage-gated channel subfamily H member 3 (*kcnh3*), and two genes (ENSDLAG00005004413 and ENSDLA00005021715) in adults were shared in the 14 °C vs. 17 °C and 20 °C vs. 17 °C comparisons, respectively (Fig. S2A and B). These are then genes, e.g., potassium voltage-gated channel subfamily H member 2 and 3 (*kcnh2a* and *kcnh3*), proprotein convertase subtilisin/kexin type 7 (*pcsk7*), protein kinase D4 (*prkd4*), in which temperature affected both their DNA methylation and

expression and can be considered biomarkers of the persistent epigenetic and transcriptomic effects of temperature with potential functional consequences.

3.3. Genes whose expression is consistently up- or downregulated by temperature

In larvae, we wanted to know how many genes were consistently upregulated or downregulated by the temperature effects. There were 4839 DEGs that were common in the list of the 8490 DEGs between 14 °C vs. 17 °C and in the list of 8128 DEGs between 17 °C vs. 20 °C comparisons. From these 4839 genes, we filtered those that their transcripts levels were consistently upregulated (i.e., 153 genes) and downregulated (i.e., 58 genes) with the effects of temperature (Table S23). The raw transcript values from these 153 and 58 genes were normalized and plotted (Fig. 1A and B). These 211 genes (Table S23) included signal transducing adaptor family member 2 (*stap2b*), mannose receptor C-type 1 (*mrc1a*), and DAB adaptor protein 2 (*dab2*). Then, we searched for the biological role of these 153 and 58 list of genes separately and results showed 44 and 33 associated GO-terms, respectively (Fig. 1C and D). The most relevant GO-terms related with genes upregulated by

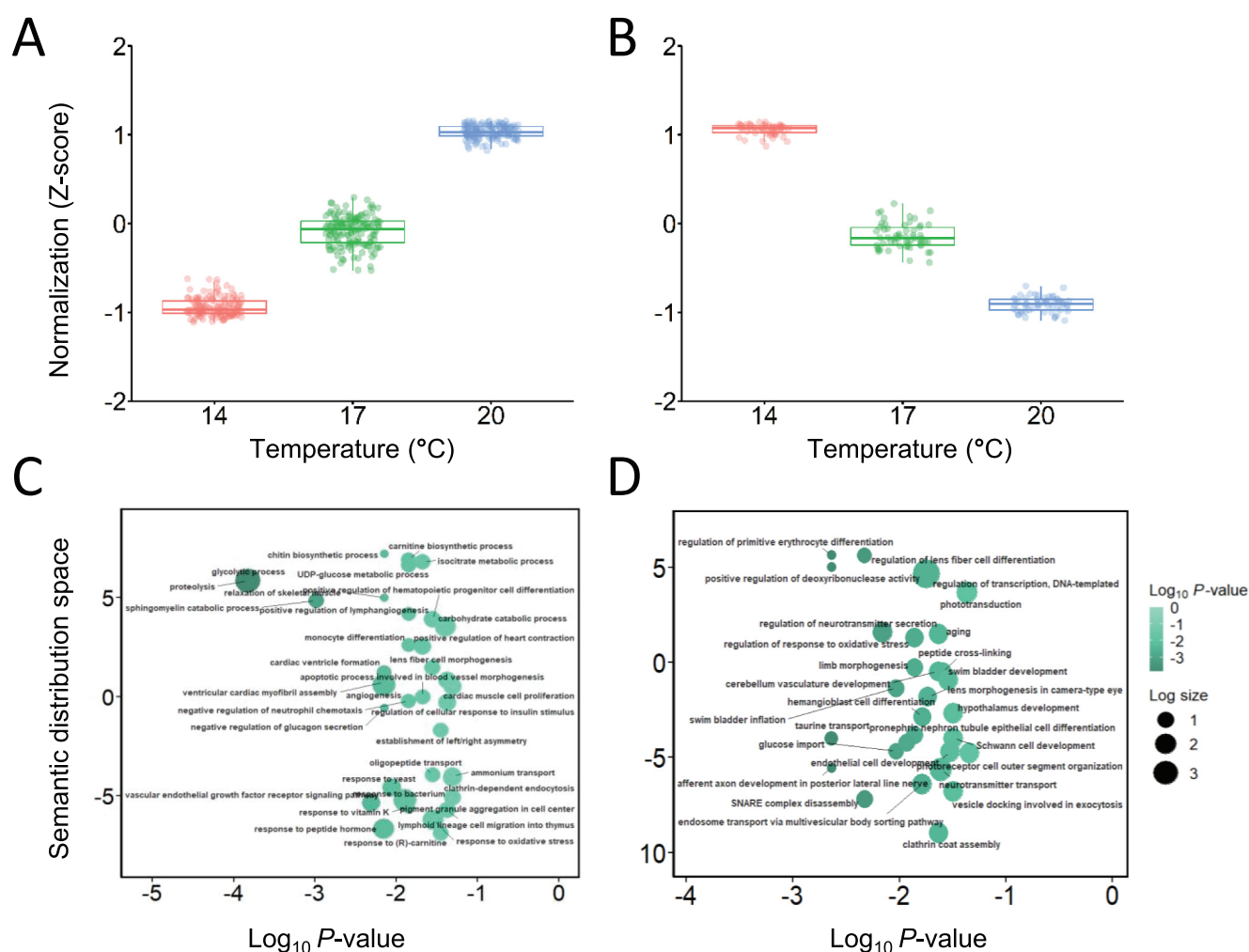


Fig. 1. (A-B) Box plots of the normalized (Z-score) gene expression of common differentially expressed genes (DEGs) between 14 °C vs. 17 °C ($n = 153$) and 20 °C vs. 17 °C ($n = 58$) that were consistently (A) upregulated (Kruskal-Wallis: chi-squared = 407.11; P -value < 0.01) or (B) downregulated (Kruskal-Wallis: chi-squared = 153.78; P -value < 0.01) along with the effects of temperature. (C-D) Semantic similarity scatter plots summarizing the enriched Gene Ontology (GO) terms of the Biological Process category (BP) from the (C) 153 DEGs into 44 GO-terms and from the (D) 58 common DEGs into 33 GOs for the genes that consistently increase or decrease expression with temperature, respectively. The scatterplots show GO-terms as circles. Significance was shown with the green gradient color palette (log₁₀ P -value), and the sizes of the plotted circles were scaled by the frequency of GO-terms they represent. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

temperature were: response to oxidative stress (GO:0006979), negative regulation of glucagon secretion (GO:0070093) and response to peptide hormone (GO:0043434). On the other hand, the most relevant GO-terms related with genes downregulated by temperature were: glucose import (GO:0046323), aging (GO:0007568) and photo-transduction (GO:0007602).

3.4. DEGs in larvae that contain at least one DMC regardless of temperature change

Next, we searched in larvae those genes that in addition of being considered DEGs also contained at least one DMC and appeared in both temperature comparisons (i.e., common genes between the 1350 genes at 14 °C vs. 17 °C and the 723 genes at 20 °C vs. 17 °C). We found 210 genes fulfilling this criterion (Table S24). To understand their

relationship, these genes were filtered and their raw transcript counts data were normalized per individual sample and represented in a heatmap (Fig. 2A). The heatmap showed that samples clearly grouped according to temperature treatment with a close relationship between the 14 °C and 17 °C samples compared to samples from 20 °C treatment. Thus, of the 210 genes identified, 95 were clearly upregulated at 20 °C while they were mostly downregulated at 14 °C and 17 °C treatments. In contrast, 115 genes had just the opposite pattern, i.e., downregulated at 20 °C and upregulated at 14 °C and 17 °C (Fig. 2A). These 95 and 115 genes include: major intrinsic protein of lens fiber (*mipb*), ATPase sarcoplasmic/endoplasmic reticulum Ca²⁺ transporting 1 (*atp2a1*), glutathione S-transferase omega 2 (*gst2*), and calcium binding protein 2 (*cabp2a*).

We also searched for the biological role of these 210 genes to better understand the effects of temperature. Results showed that these genes

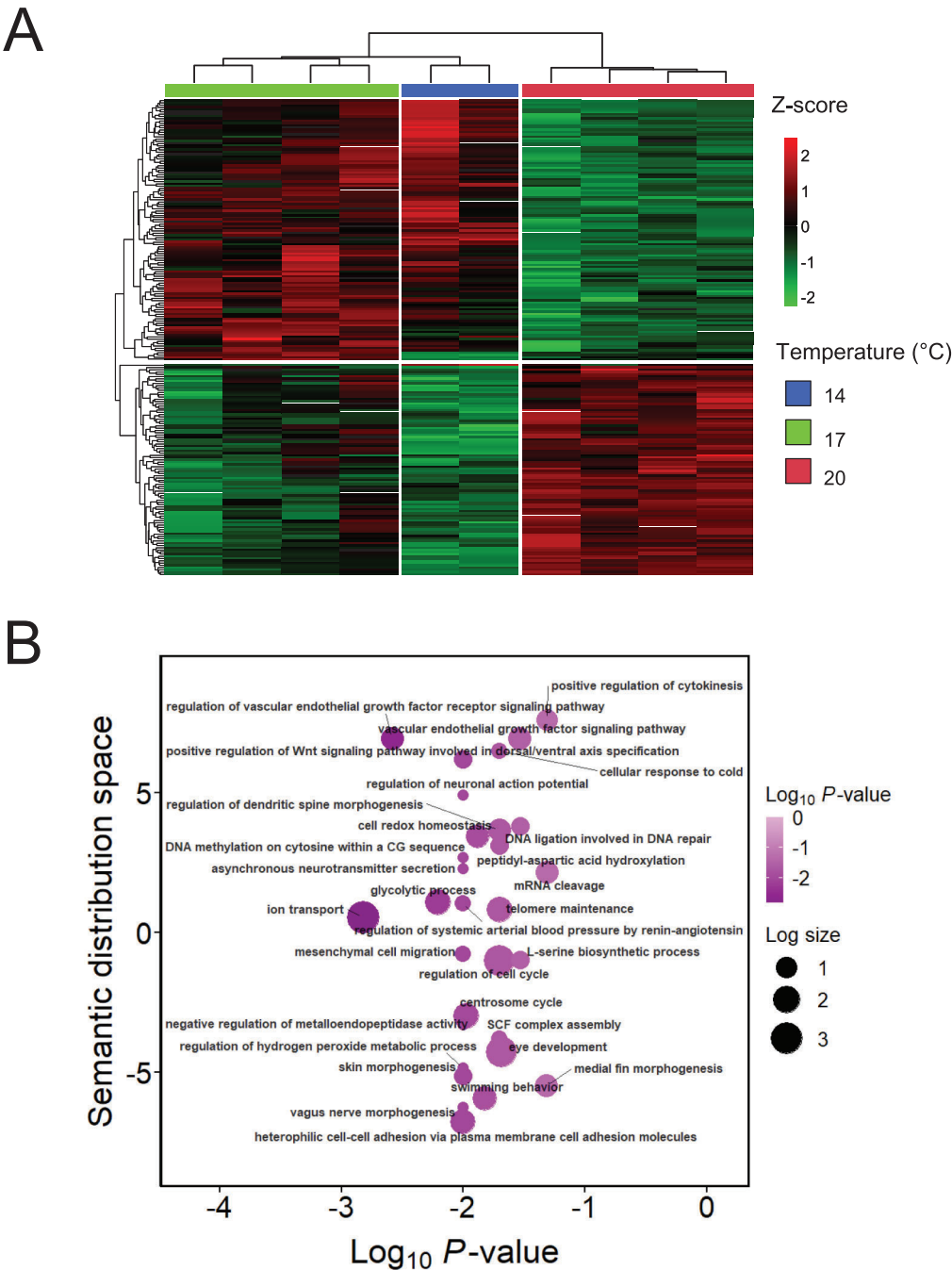


Fig. 2. (A) Heatmap of expression levels in larvae tissue from 210 common differentially expressed genes (DEGs) which at least contain one differentially methylated cytosine (DMC) between 14 °C vs. 17 °C ($n = 1350$) and 20 °C vs. 17 °C ($n = 723$). The color gradient represents the level of expression scaled by gene (red: high expression and green: low expression). The dendrograms indicate the similarity between genes and between samples. The bands on the top indicate the temperature treatment of each sample (blue: 14 °C, green: 17 °C, and red: 20 °C). (B) Semantic similarity scatter plot summarized the enriched GO-terms of the BP category from the 210 common DEGs in both temperature treatments with at least one DMC. The scatterplot showed GO-terms as circles and the significance was shown with the purple gradient color palette ($\log_{10} P$ -value). The sizes of the plotted circles were scaled by the frequency of GO-terms they represent. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

were participating in 30 GO-terms, of which the most relevant were: DNA methylation on cytosine within a CG sequence (GO:0010424), positive regulation of Wnt signaling pathway involved in dorsal/ventral axis specification (GO:2000055) and cellular response to cold (GO:0070417) (Fig. 2B).

3.5. Relationship between DNA methylation-gene expression in larvae

The 95 and 115 genes identified above in larvae are genes that not only their expression is affected by temperature but also their methylation levels are affected. However, whether those two changes are linked is not known. Thus, next we were interested to find out those genes whose expression was affected by the temperature treatment as well as their DNA methylation levels. To do so, we started with the list of 1350 and 723 genes that were DEGs and contained at least one DMC in the 14 °C vs. 17 °C and the 20 °C vs. 17 °C comparisons, respectively (Fig. S2A and B). For each selected gene, we calculated the average number of transcripts with logarithm transformation and also their

mean methylation value by averaging the methylation values of all their DMCs among the samples belonging to the same temperature treatment. Next, for each temperature comparison we identified those genes whose expression was either significantly positively or negatively correlated with DNA methylation levels. Results showed that there were 577 (Table S25) and 344 (Table S26) genes with negative and positive correlation, respectively, in the 14 °C vs. 17 °C comparison (Fig. 3A and C), while there were 102 (Table S27) and 250 genes (Table S28) with negative and positive correlation in the 20 °C vs. 17 °C comparison (Fig. 3B and D). There were no common genes between the 577 and 102 gene lists, whereas only ENSDLAG00005009958 was common in the 344 and 250 gene list that were more methylated and upregulated by temperature.

3.6. Persistent temperature effects on the transcriptome of larvae, juveniles, and adults

Once epigenomic and transcriptomic changes resulting from temperature effects had been well characterized in larvae, we were

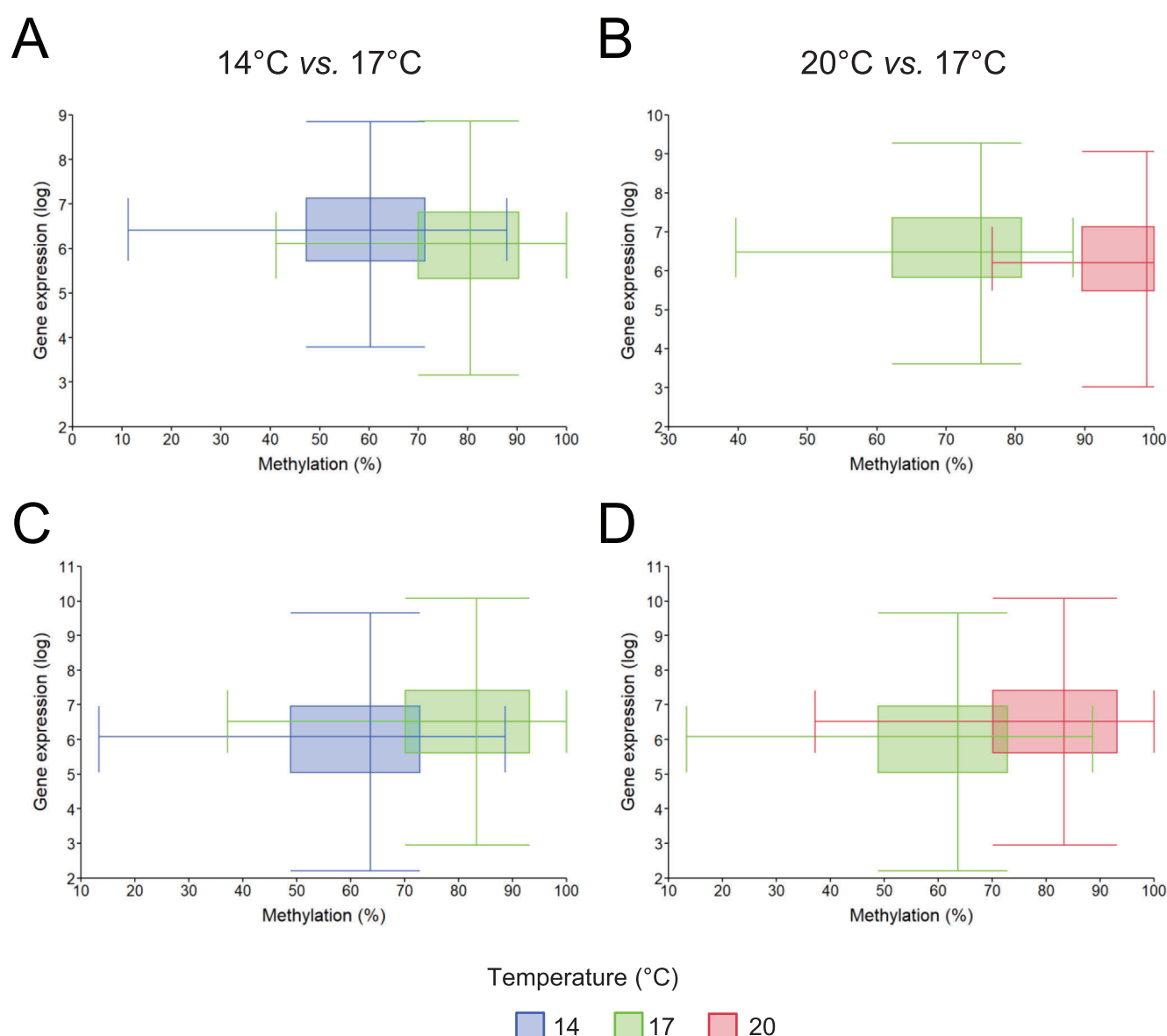


Fig. 3. Boxplots of the common differentially expressed genes (DEGs) between 14 °C vs. 17 °C and 20 °C vs. 17 °C which at least have one differentially methylated cytosine (DMC) in the gene body region (gene ± 1000 base pairs, bp), and that on top showed DNA methylation negatively correlated with gene expression in (A) 14 °C vs. 17 °C, and in the (B) 20 °C vs. 17 °C comparison. The genes that met the same criteria with a positive correlation between DNA methylation and gene expression were plotted in (C) for the 14 °C vs. 17 °C comparison, and in (D) for the genes from the 20 °C vs. 17 °C comparison.

interested to know how persistent those changes were in other tissues. Specifically, we wanted to know which of the DEGs found in larvae were still affected in juveniles and in reproductive (testis) tissues long after the temperature treatments had ceased during larvae development. For this, we compared the DEGs lists from larvae, juveniles, and adults for each temperature comparison. Results showed that not a single gene was shared among three tissues in both comparisons. However, in the 14 °C vs. 17 °C comparison, we identified 29 and six genes between larvae-juveniles and larvae-adults, respectively (Fig. S2A and B). From these 29 DEGs that were common in larvae and juveniles, 16 genes, including chromobox 4 (*cbx4*), showed the same direction of change, either consistently upregulated or consistently downregulated (Fig. S3A). The same applied to three out of the six DEGs found in common between larvae and adults (Fig. S3B). In the 20 °C vs. 17 °C comparison, 21 DEGs in larvae-juveniles and 14 DEGs in larvae-adults were also identified. From these two gene lists, only five and six genes showed the same gene expression tendency, respectively (Fig. S3C and D).

We searched the GO-terms to better understand their biological function due to low temperature (Fig. 4). From the 29 and six common DEGs genes in larvae-juveniles and larvae-adults from 14 °C vs. 17 °C comparison, we found ten and three GO-terms, respectively (Fig. 4A and B). The most relevant GO-terms from larvae-juveniles were: response to hypoxia (GO:0001666), cell redox homeostasis (GO:0045454) and positive regulation of CREB transcription factor activity (GO:0032793), while from larvae-adults were: integrin-mediated signaling pathway (GO:0007229), cell-matrix adhesion (GO:0007160) and small GTPase mediated signal transduction (GO:0007264). From common genes in larvae-juveniles and larvae-adults in the 20 °C vs. 17 °C comparison, we identified 21 and 14 genes, respectively (Fig. S2C and D). We also searched the GO-terms to better understand their biological function due to the effects of elevated temperature (Fig. 4). We found twelve and three GO-terms from larvae-juveniles and larvae-adults, respectively (Fig. 4C and D). The most relevant GO-terms from larvae-juveniles were: cellular response to oxidative stress (GO:0034599), muscle cell cellular homeostasis (GO:0046716), and cardiac muscle tissue morphogenesis (GO:0055008), while from larvae-adults was: oocyte maturation (GO:0001556).

Additionally, we investigated in which pathways these set of GO-terms were participating. For juveniles-larvae we found metabolic and MAPK signaling pathways in 14 °C vs. 17 °C and 20 °C vs. 17 °C, respectively (Fig. 4E). In larvae-adults, for both comparisons we found 6 KEGG pathways of which the most relevant were: focal adhesion in 14 °C vs. 17 °C, and oocyte meiosis, P53 signaling and progesterone-mediated oocyte maturation pathways in 20 °C vs. 17 °C. Together, these results suggested that the biological functions and pathways affected were temperature-dependent and tissue specific.

3.7. Persistent temperature effects on the methylome in larvae, juveniles, and adults

We wanted to know the biological role of the genes found in larvae, juveniles, and adults that had DMCs within their body region due to temperature. For this analysis we selected the 163 and 264 list of genes found previously (Fig. S2A and B).

From the 163 genes in the 14 °C vs. 17 °C comparison, 25 GO-terms were identified (Fig. 5A) and the most relevant were: regulation of histone deacetylation (GO:0031063), cell adhesion (GO:0007155), smooth muscle tissue development (GO:0048745), and negative regulation of Wnt signaling pathway involved in dorsal/ventral axis specification (GO:2000054). On the other hand, from the 264 genes found in 20 °C vs. 17 °C, a total of 45 GO-terms were identified (Fig. 5B) and the most relevant were: neural crest cell migration (GO:0001755), negative regulation of transforming growth factor beta receptor signaling pathway (GO:0030512), and histone H3-K9 methylation (GO:0051567). Six and one KEGG pathways were present in 14 °C vs. 17 °C and 20 °C vs. 17 °C, respectively, for instance included TGF-beta signaling pathway

and ECM-receptor interaction pathways (Fig. 5C). These results pointed at relevant pathways irreversibly affected by temperature changes and, in particular, to the contribution of histone modifications in maintaining certain environmentally-induced epigenetic changes.

3.8. Persistent temperature effects on transcriptome and methylome in larvae and juveniles

Given that we did analyses of the persistent DEGs between tissues and the persistent genes with affected DMCs between tissues, we were interested in those GO-terms and KEGG pathways. We only selected the 1350 and 723 genes from larvae and the six and seven genes from juveniles and we discarded the adults results due to very low number of genes found either in 14 °C vs. 17 °C or 20 °C vs. 17 °C comparisons (Fig. S2A and B).

In the 14 °C vs. 17 °C comparison, and from larvae and juveniles, a total of 36 and three GO-terms were identified from the 1350 and six genes, respectively (Fig. 6A and B). The most relevant GO-terms for larvae were: developmental growth (GO:0048589), somatic muscle development (GO:0007525), calcium ion transmembrane transport (GO:0070588), and regulation of multicellular organism growth (GO:004001), and for juveniles were: digestion (GO:0007586), sarcomere organization (GO:0045214) and embryonic hemopoiesis (GO:0035162). In the 20 °C vs. 17 °C comparison, and from larvae and juveniles, a total of 52 and three GO-terms were identified from the 723 and seven genes, respectively (Fig. 6C and D). The GO-terms for larvae included positive regulation of canonical Wnt signaling pathway (GO:0090263), methylation on cytosine within a CG sequence (GO:0010424), histone H3-K9 demethylation (GO:0033169) and neural crest cell differentiation (GO:0014033), and for juveniles tissue were: hemoglobin biosynthetic process (GO:0042541), iron ion transport (GO:0006826) and response to bacterium (GO:0009617).

We only found KEGG pathways in larvae but did not in juveniles due to the low number of genes represented in the two comparisons. Thus, in larvae, nine and eleven KEGGs pathways were identified in the 14 °C vs. 17 °C and 20 °C vs. 17 °C comparisons, respectively (Fig. 6E). KEGG pathways included Wnt signaling and MAPK signaling pathways in the 14 °C vs. 17 °C comparison and intestinal immune network for IgA production and biosynthesis of amino acids in the 20 °C vs. 17 °C comparison. Thus, in particular, Wnt signaling and histone H3K9 demethylation seemed to emerge as key targets of temperature effects.

3.9. Relationship with metastable epialleles and very-long term effects

Next, we wanted to know what of the genes identified in the present study coincide with the genes identified by Anastasiadi et al. (2021) that were differentially methylated still three years after exposing the European sea bass (adulthood) to a simulated heat wave. We did this to further determine the usefulness of the genes identified in the present study as putative biomarkers of the effects of temperature during larvae development that can have very long-term effects and thus potential lifelong consequences, i.e., well beyond the duration of this study.

First, we started with a quite stringent criteria and thus looked which ones of the genes we identified being DEGs and containing at least one DMC coincided with the nine genes containing metastable epialleles, i. e., loci the methylation of which changes in response to temperature and regardless of tissue being considered, in that case brain, liver, muscle and testis (Anastasiadi et al., 2021), of which the latter two coincide with the tissues investigated in the present study. We found two of these genes, *spata51l* and *agl*, among the list of the 1350 DEGs that also had at least one DMC in larvae but not in juveniles or adults in the 14 °C vs. 17 °C comparison. None of the 9 metastable epialleles found by Anastasiadi et al. (2021) were found in larvae, juveniles or adults of the 17 °C vs. 20 °C comparison.

Given the low number of genes identified with the above stringent strategy, we relaxed the criteria and selected the 527 genes that showed

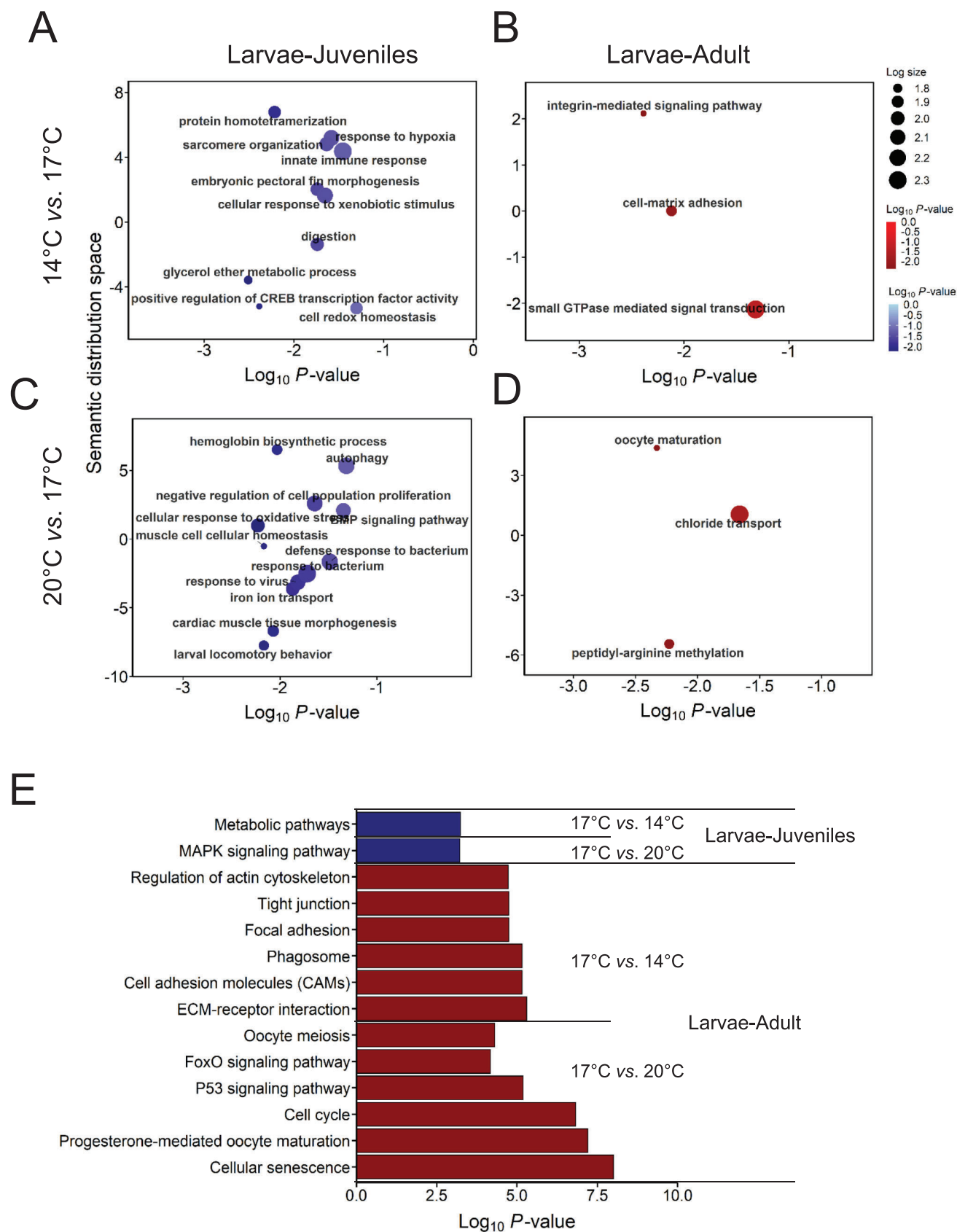


Fig. 4. Scatterplots of the enriched Gene Ontology (GO-terms) of the Biological Process (BP) category from the (A) Common differentially expressed genes (DEGs) in larvae and juveniles and (B) Common DEGs in larvae and gonad between 14 °C vs. 17 °C, (C) common DEGs in larvae and juveniles between 14 °C vs. 17 °C, and (D) Common DEGs in larvae and gonads between 20 °C vs. 17 °C. (E) The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment of the DEGs persistently affected by temperature in larvae and juveniles or in larvae and gonads in each temperature comparison.

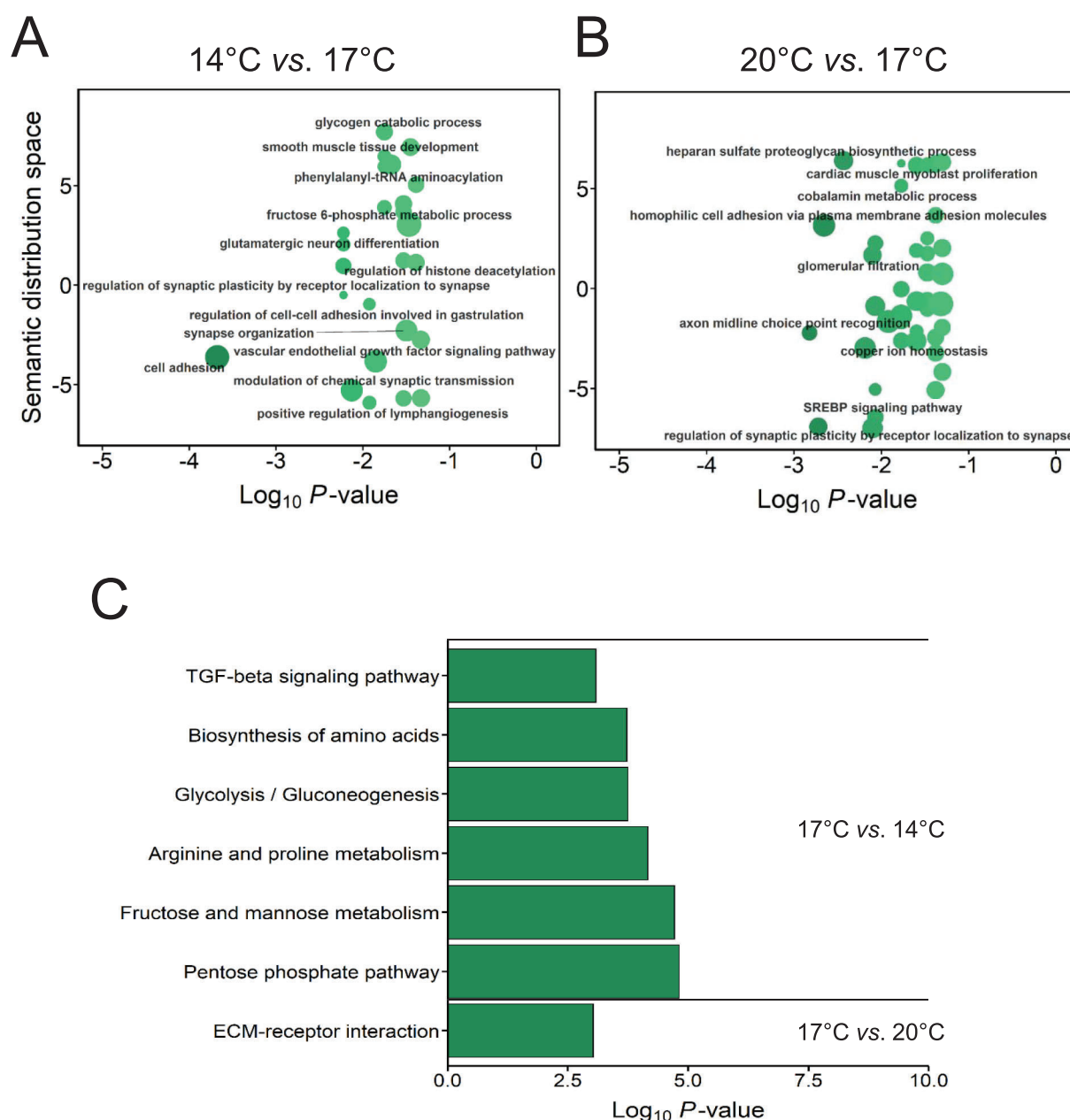


Fig. 5. Scatterplots of the enriched Gene Ontology (GO)-terms of the Biological Process (BP) category from the common genes which at least have one differentially methylated cytosine (DMC) in the gene body (gene ±1000 base pairs, bp) in larvae, gonads, and juveniles in (A) 20 °C vs. 17 °C comparison and in (B) 14 °C vs. 17 °C comparison. (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of the genes which at least have one DMC in the gene body (gene ±1000 bp) in larvae, juveniles and gonads in each temperature comparison.

DMR in muscle from (Anastasiadi et al., 2021) and we compared to our 2199 (14 °C vs. 17 °C; Fig. S2A) and 2990 (20 °C vs. 17 °C; Fig. S2B) list of genes with at least one DMC in juveniles. We identified 64 genes in the 14 °C vs. 17 °C (Table S29) and 95 genes in the 20 °C vs. 17 °C (Table S30). Next, we wanted to know the numbers of genes present in juveniles regardless of temperature change comparing these 64 and 95 lists of genes. Results showed a total of 43 genes in common between these two lists (Table 1).

As we did in the previous section, we did the same analysis but in this case comparing with the list of genes with DMR in the list of 825 (14 °C vs. 17 °C) and 921 (20 °C vs. 17 °C) genes with at least one DMC in testis (Anastasiadi et al., 2021). We identified 16 (Table S31) and 20 genes (Table S32) in 14 °C vs. 17 °C and 20 °C vs. 17 °C lists, respectively. Next, we wanted to know the numbers of genes present in adults

regardless of temperature change comparing these 16 and 20 lists of genes. We found seven genes in common between these two lists. These genes were: *agl*, heat shock protein family B (small) member 11 (*hspb11*), neuroligin 1 (*nlgn1*), serine/threonine kinase receptor associated protein (*strap*), ADP ribosylation factor guanine nucleotide exchange factor 2 (*arfgef2*), and regulatory factor X5 (*rxf5*).

3.10. Epigenetic biomarkers with both diagnostic and prognostic value

Finally, one of the interesting characteristics of epigenetic biomarkers is that they can have both diagnostic and prognostic value, i.e., epigenetic changes can anticipate transcriptomic changes. In order to explore whether we could find genes fulfilling these characteristics, we first searched genes with at least one DMC in larvae (sampled at 5–8 dpf)

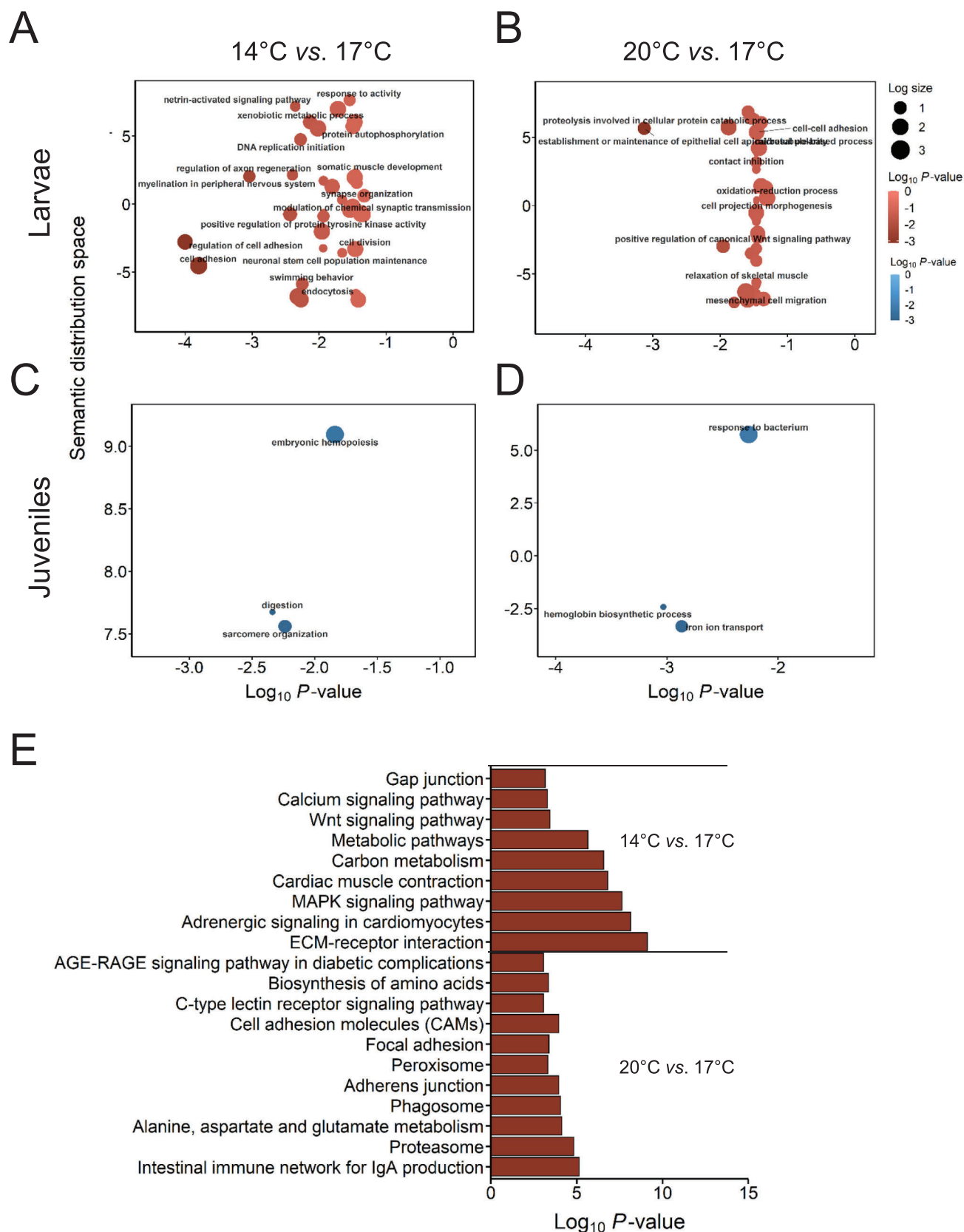


Fig. 6. Scatterplots of the enriched GO-terms of the BP category from the DEGs between (A) 14 °C vs. 17 °C comparison or (B) 20 °C vs. 17 °C comparison, which at least have one differentially methylated cytosine (DMC) in the gene body region (gene ± 1000 base pairs, bp) in larvae tissue. GO-terms scatterplots of the DEGs between (C) 14 °C vs. 17 °C or (D) 20 °C vs. 17 °C comparison, which at least have one DMC in the gene body (gene ± 1000 bp) in juveniles. (E) KEGG enrichment analysis of the DEGs between the two temperature comparisons and which at least contain one DMC in the larvae tissue.

Table 1

List of the 43 genes common between the 64 and 95 genes with at least one DMC in 14 °C vs. 17 °C and in 20 °C vs. 17 °C comparisons, respectively, that also contained differentially methylated regions (DMR) in juveniles tissue in the European sea bass after exposure to elevated temperature (Anastasiadi et al., 2021).

Gene ID	Gene name	Gene description
ENSDLAG00005009668	<i>abcc9</i>	ATP-binding cassette, sub-family C (CFTR/MRP), member 9
ENSDLAG00005019673	<i>agl</i>	amylase-1, 6-glucosidase, 4-alpha-glucanotransferase
ENSDLAG00005002947	<i>astn1</i>	astrotactin 1
ENSDLAG00005011333	<i>atf7ip</i>	activating transcription factor 7 interacting protein
ENSDLAG00005007690	<i>bcas3</i>	BCAS3 microtubule associated cell migration factor
ENSDLAG00005017692	<i>cacna2d4</i>	calcium voltage-gated channel auxiliary subunit alpha2delta 4
ENSDLAG00005007417	<i>cadm1a</i>	cell adhesion molecule 1a
ENSDLAG00005015894	<i>csm2</i>	CUB and Sushi multiple domains 2
ENSDLAG00005008332	<i>cxxc4</i>	CXXC finger 4
ENSDLAG00005016884	<i>dhfr13a.1</i>	dehydrogenase/reductase (SDR family) member 13a, tandem duplicate 1
ENSDLAG00005019105	<i>dusp16</i>	dual specificity phosphatase 16
ENSDLAG00005023582	<i>fgfr3</i>	fibroblast growth factor receptor 3
ENSDLAG00005013628	<i>fstl3</i>	follicle-stimulating-like 3 (secreted glycoprotein)
ENSDLAG00005021658	<i>gpc3</i>	glypican 3
ENSDLAG00005002969	<i>hivp1</i>	HIVP zinc finger 1
ENSDLAG00005002265	<i>hoxc4a</i>	homeobox C4a
ENSDLAG00005004570	<i>hoxd4a</i>	homeobox D4a
ENSDLAG00005017996	<i>inhbaa</i>	inhibin subunit beta Aa
ENSDLAG00005009376	<i>kcnip2</i>	potassium voltage-gated channel interacting protein 2
ENSDLAG00005019431	<i>ksr2</i>	kinase suppressor of ras 2
ENSDLAG00005025443	<i>mcm5</i>	minichromosome maintenance complex component 5
ENSDLAG00005003802	<i>mios</i>	missing oocyte, meiosis regulator, homolog (Drosophila)
ENSDLAG00005011341	<i>mtss1</i>	MTSS I-BAR domain containing 1
ENSDLAG00005001244	<i>nckap5l</i>	NCK-associated protein 5-like
ENSDLAG00005014329	<i>ndufs4</i>	NADH:ubiquinone oxidoreductase subunit S4
ENSDLAG00005006976	<i>nlgn1</i>	neuroligin 1
ENSDLAG00005005359	<i>nup88</i>	nucleoporin 88
ENSDLAG00005023392	<i>ptk7</i>	protein tyrosine kinase 7 (inactive)
ENSDLAG00005003997	<i>rad54l2</i>	RAD54 like 2
ENSDLAG00005010100	<i>rally</i>	RALY heterogeneous nuclear ribonucleoprotein
ENSDLAG00005009049	<i>rbfox1</i>	RNA binding fox-1 homolog 1
ENSDLAG00005022557	<i>rmf25</i>	ring finger protein 25
ENSDLAG00005000657	<i>runx3</i>	RUNX family transcription factor 3
ENSDLAG00005018542	<i>samd4a</i>	sterile alpha motif domain containing 4A
ENSDLAG00005000254	<i>scap</i>	SREBF chaperone
ENSDLAG00005010637	<i>shf</i>	Src homology 2 domain containing F
ENSDLAG00005010165	<i>spata5</i>	spermatogenesis associated 5
ENSDLAG00005018273	<i>strn</i>	striatin, calmodulin binding protein
ENSDLAG00005019987	<i>syde2</i>	synapse defective Rho GTPase homolog 2
ENSDLAG00005009862	<i>tmcc1</i>	transmembrane and coiled-coil domain family 1
ENSDLAG00005009409	<i>ttc7a</i>	tetratricopeptide repeat domain 7A
ENSDLAG00005026035	<i>wnt9a</i>	wingless-type MMTV integration site family, member 9A
ENSDLAG00005013567	<i>xrn2</i>	5'-3' exoribonuclease 2

and juveniles (sampled at 90 dpf) and differentially expressed in juveniles but not in larvae. In the 14 °C vs. 17 °C comparison we found one such gene, *prkd4*, while in the 20 °C vs. 17 °C we found two genes, PDZ Domain Containing 7 (*pdzd7*) and an unannotated gene, ENSDLA00005003985. As for the adults, we found one gene, *kcnh3* in the 14 °C vs. 17 °C comparison and two unannotated genes, ENSDLA00005021715 and ENSDLA00005004413 in the 20 °C vs. 17 °C comparison. However, when we looked at the data at the single nucleotide level we found that we had data for the same exact DMC for only one unannotated gene, ENSDLA00005004413. Regarding the DMC

within this gene, the same CpG locus that was hypomethylated in larvae reared at 20 °C was also hypomethylated in the adults of fish reared at 20 °C at 482 dpf. Therefore, this gene can be considered an epigenetic marker with both prognostic and diagnostic value (Fig. 7). To investigate further the functionality of the biomarker, we searched for similar protein sequences across all genomes available in NCBI database (Database Resources of the National Center for Biotechnology Information, 2016) using blastp (Altschul, 1997) with the protein sequence of ENSDLA00005004413. As a result, the keratin associated protein 10–4 (*krtp10-4*) from the brown trout, *Salmo trutta*, gave the highest sequence similarity (Percentage identity = 43%,) with high statistical significance ($P = 3E^{-25}$). This gene is associated with pathways of keratinization and developmental biology. Although the percentage of identity was low, the alignment was significant, allowing us to suggest the hypothesis that the identified biomarker is likely a coding gene that could be related to keratin production.

4. Discussion

In this study we provide a set of epigenetic and transcriptomic biomarkers associated with the quality of the thermal regime experienced by the European sea bass during early development, when temperature changes can have long-lasting phenotypic consequences for the performance of this species in aquaculture settings. Although different temperature regimes have been tested for longer periods (Blázquez et al., 1998; Person-Le Ruyet et al., 2004), to our knowledge, this study is one of the few that concentrates on the effects of temperature during the early development and, in particular, until exogenous feeding of larvae is initiated. We are aware that methylome and transcriptome is tissue-specific and that differences are not only stage-dependent but also, and importantly, tissue-dependent. Taking into account this, we wanted to know what changes persist rather where they persist because the goal of this study was to search for biomarkers. If the same change induced by temperature observed in whole larvae is also present in the muscle of treated juveniles (somatic tissue) and/or in testis of treated adults (germline tissue) but not in the muscle or testis of the control fish this means that the effect is due to temperature and not stage and therefore has informative value. Thus, as far as this lasting effect is robust it does not matter the contribution of tissue.

Before discussing the results proper, however, three considerations on the reliability of the data obtained seem appropriate. First, the temperature treatment was carried out during the thermosensitive period (Piferrer et al., 2005; Vandeputte and Piferrer, 2019) but, as regards to effects on sex ratio, there were no differences among the three tested temperatures, although higher temperature resulted in a slightly higher number of males. The lack of significant differences are likely due to the short duration of the temperature treatment, less than ten days, when treatments of thirty days or longer are necessary to elicit permanent sex ratio responses (Navarro-Martín et al., 2009), and are regarded as positive in the context of the present study because we wanted to assess the effects of temperature during the most critical phase of development, and not to elicit significant sex ratio changes. Hence, the deliberate short temperature treatment period used. Second, at 482 dpf, fish reared at 14 °C until 12 dpf were bigger than those reared at 20 °C until 8 dpf, a phenomenon that probably indicates better performance of the cold-exposed fish, a result that should be further studied without excluding the possibility of compensatory growth, as reported previously in the European sea bass (Díaz and Piferrer, 2015). However, the lack of differences in growth between fish reared at 14 °C, 17 °C and 20 °C at 90 dpf, nor between adult fish reared at 14 °C vs. 17 °C or between 20 °C vs. 17 °C (i.e., the two comparisons made in this study), coupled with the lack of differences in GSI values among the males of the different temperature regimes, provide confidence that the observed differences in DNA methylation or in gene expression were due only to temperature and not to differences in somatic or gonadal growth. Third, and regarding the analyses performed, transcriptomic data was of good

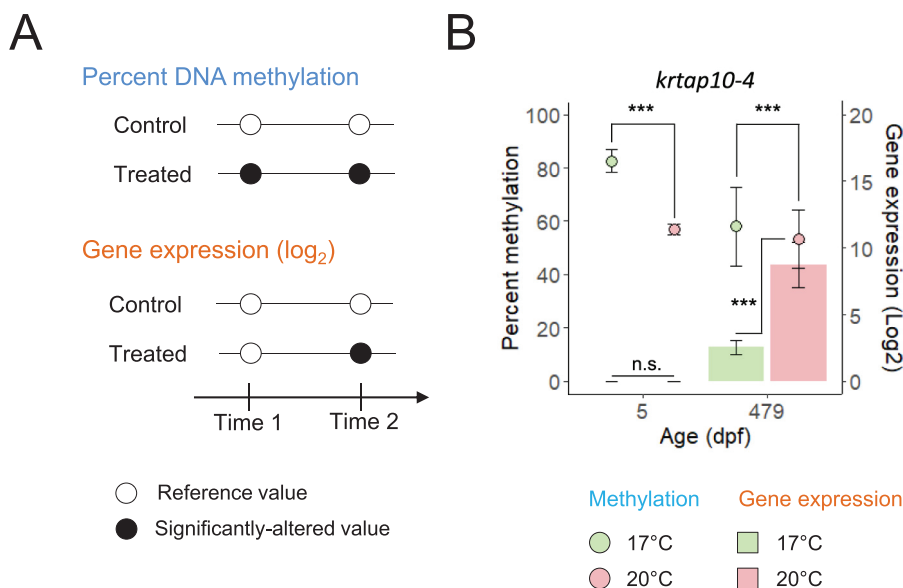


Fig. 7. Epigenetic marker with both prognostic and diagnostic value. (A) Conceptual diagram of an epigenetic marker. The integration of gene expression data and DNA methylation data in two time points under two conditions can reveal genes, regions or loci with prognostic and diagnostic value. (B) The dot plot with the primary y-axis (left) shows the DNA methylation data of the same DMC within ENS-DLAG00005004413 gene body region ± 1000 base pairs, bp in larvae at 5 and adults at 482 days post fertilization (dpf). The bar plot with the secondary y-axis (right) shows the gene expression data (log₂) for this gene in larvae at 5 and adults at 482 dpf. The color code indicates the temperature treatment: 17 °C in green, and 20 °C in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

quality since only samples with RIN > 8, average alignment rate > 88% and > 30 million assigned features were used. The same applies to the epigenomic data, based on an average > 70% alignment rate and > 99% bisulfite conversion quality indicators. Taken together, this indicates that epigenetic and transcriptomic results do indeed reflect temperature effects. It should also be noted that in this study we concentrated on DNA methylation-based biomarkers for their simplicity of measurement (Levenson, 2010) and because, once known the nature of the genes involved, no longer a whole-genome or genome-wide determination method is required. Instead, the measure of their DNA methylation can be achieved by a simple PCR-based method (Hernández et al., 2013), as done, for example, in the development of some epigenetic clocks, another type of DNA methylation-based biomarker (Aliferi et al., 2018; Naue et al., 2017). Other type of epigenetic markers do exist, like those based on histone modifications such as H3K4me3, H3K9me3, H3K79me, and H3K79me3, which have been used, for example, to assess proper oocyte development in cyprinid fishes (Zhou et al., 2019); and those based on the quantification of miRNAs, e.g., the differentially expressed miR-122, novel-miR6, miR-193a-3p, and miR-27a-5p, in the head kidney, the expression levels of which were linked to mechanisms of inflammation and injury during cadmium exposure in the common carp, *Cyprinus carpio* (Q. Liu et al., 2020). Concentrating in these other type of biomarkers was beyond the scope of study but efforts should be aimed at their identification in future studies. Finally, it is worth mentioning that we not only used epigenetic changes but also investigated gene expression changes, meaning that we reported epigenetic changes that also resulted in phenotypic (transcriptomic) changes, thus reinforcing the usefulness of the epigenetic biomarkers presented below. However, we are aware that beyond the proof-of-concept that this study represents, additional phenotypic endpoints would be desirable.

The relationship between markers and their usefulness to behavior is still poorly understood not only in the European sea bass but in fish in general. Regarding the coping style from our results, we found that in the group of fish treated at 20 °C there were more fish that were *Bold* than *Shy*. Therefore, the development of biomarkers that allow selecting fish reared at 14 °C and 17 °C (both groups fish performed similar results in the CPP test), are at the same time increasing the probability of selecting for *Shy* behavior, a trait preferred in an easily stressed species like the European sea bass. The study of the methylome and the transcriptome of the two different behavior types could hold interesting results for aquaculture purposes. These could be further explored in future studies since recent evidence showed differences in

transcriptomic analyses between *Bold* and *Shy* copying styles for this species (Sadoul et al., 2022). However, a behavior assays study using the olive flounder (*Paralichthys olivaceus*) only found 144 DEGs out of the entire transcriptome when comparing *Bold* vs. *Shy* fish (Zou et al., 2021). For these reasons, we prioritized the development of biomarkers for thermal stress and not for the copying style.

Among all the genes identified that could potentially be used as biomarkers of the thermal quality of the early development, there are some that repeatedly showed up (Table 2) and thus we will focus on these genes. We classified the effects as short-term if they were present only in larvae, right at the end of the temperature treatment, mid-term effects if they were present both in larvae and juveniles at 90 dpf, long-term if they were present both in larvae and adults at ~500 dpf, and very-long term if they were affected in larvae, 90 or ~ 500 dpf. In addition, we studies whether the affected genes coincide with the genes the methylation of which was found to be affected in adults three years after the end of the exposure to elevated temperature as is shown in a previous study with the European sea bass (Anastasiadi et al., 2021).

Among the 70 genes that were differentially methylated regardless of the direction of the temperature change and that were affected in larvae, juveniles and adults, it is worth mentioning *jmjd7*, *kcnh5b*, and *zfyve9a*. The former encodes a highly conserved enzyme with a JmjC domain that recognizes and cleaves monomethylated and dimethylated arginine residues of histones H2, H3 and H4. Other members of this family include: (i) *jmjd3* (also known as lysine demethylase 6B, *kdm6b*), which has been implicated in environmental signal transduction in connection to temperature-dependent sex determination in reptiles, including the Australian bearded dragon, *Pogona vitticeps*, (Deveson et al., 2017), the red-eared slider turtle, *Trachemys scripta* (Ge et al., 2018), and involved in sex-related processes in the orange-spotted grouper fish, *Epinephelus coioides* (Zhang et al., 2021); and (ii) jumonji and AT-rich interaction domain containing 2, *jarid2a*, upregulated in response to elevated temperature in the European sea bass (Díaz and Piferrer, 2015) but down-regulated when fish were also exposed to estrogen (Díaz and Piferrer, 2017). On the other hand, *kcnh5b* encodes a member of voltage-gated potassium channel, also involved in environmental signal transduction in different situations (Yang and Zheng, 2014). However, whether these genes are involved in capturing temperature effects on sex ratios is still not known. Lastly, *zfyve9a* encodes a double zinc finger motif-containing protein that interacts directly with SMAD2, and recruits SMAD2 to the TGFβ receptor (Tsukazaki et al., 1998). The *zfyve9a* gene regulates the proliferation of hepatic cells during zebrafish embryogenesis (Liu et al.,

Table 2

Summary of the biomarkers identified in this study related to the effects of early temperature according to their informative properties. Abbreviations: T = transcriptomic; E = epigenetic; B = both; L = larvae; J = juvenile; A = adults.

Properties	Type	Tissue	Lower temperature	Both	Elevated temperature
Short-term effects (L)	E	L		<i>agl, kcnh2a, depdc5, prkd2, gemin7</i> ; (Table S 17) *	
	T	L		<i>cabp2a, eloa1, serpina7, mrc1a, gpx1b, sncga</i> ; (Table S23) *	
	B	L		<i>hivep1, kcnq5a, pcdh1b, hmgcr, dusp16</i> ; (Table S24) *	
Mid-term effects (L & J)	E	L & J		<i>agl, dlga1a, kcnh2a, scap, rsu1</i> ; (Table S17) *	
	T	L & J		<i>ENSDLAG00005005752, ENSDLAG00005023917</i> (Table S5, S6 and S8, S9) *	
	B	L & J	<i>capn9, ENSDLAG00005003493, ENSDLAG00005005752</i> ; (Table S19, S21)		<i>tfa</i> (Table S20, S22)
Long-term effects (L & A or L & J & A)	E	L & A		<i>agl, pvalb7, brinp2, astn1, nup98</i> ; (Table S 17) *	
	E	L & J & A		<i>agl, pvalb7, brinp2, astn1, xdh</i> ; (Table S17) *	
	T	L & A		<i>ENSDLAG00005000425</i> (Table S5, S7 and S8, S10) *	
Very long-term effects (L or J or A & Anastasiadi et al., 2021)	E	J		<i>agl, spata5, fgfr3, cacna2d4</i> ; (see more examples in Table 1) *	
	E	A		<i>agl, hspb11, nlgn1, strap, arfgef2, rfx5</i> * (Table S31, S32) *	
	B	J	<i>spata5l1, agl, prkd4</i>		N/A
Putative prognostic and diagnostic	B	L & J			<i>pdzd7, ENSDLAG00005003985, ENSDLAG00005021715, krtap10-4</i>
	B	L & A	<i>kcnh3</i>		<i>krtap10-4</i>
Strong prognostic and diagnostic	B	L & A	N/A		

* Genes affected by both lower (14 °C) and higher (20 °C) temperatures with respect to the control temperature (17 °C).

2013).

The 211 genes identified in which their expression either consistently increased or decreased from 14 °C to 17 °C and also from 17 °C to 21 °C included *stap2b*, *mrc1a* and *dab2*. These genes constitute, therefore, robust short-term only transcriptomic biomarkers of the effects of temperature during European sea bass early development. Their expression levels could be used to assess the on-the-spot quality of the egg incubation thermal regime and larvae until first feeding. Among the most relevant GO-terms related with the above mentioned genes upregulated by elevated temperature was the response to oxidative stress. It is well established that elevated temperature accelerates developmental rate as does mitochondrial respiration, leading to reactive oxygen species (ROS) accumulation and oxidative stress (Sun et al., 2015). Changes in ROS status and oxidative stress have been implicated as key elements in environmental signal transduction in poikilothermic vertebrates (Castelli et al., 2020). Identifying the mechanisms that regulate these molecular oxidative stress responses can help to develop new approaches to monitor the appearance of undesirable conditions, especially in an aquaculture context (Lazado et al., 2020).

There were 210 genes that in addition of being DEGs in larvae they also contained DMCs. Of these, 95 genes were upregulated at 20 °C and downregulated at 14 °C and 17 °C treatments, while the remaining 115 genes had the opposite pattern. Two considerations are worth being made here. First, the grouping in the heatmap of the 14 °C and 17 °C treatments, on one hand vs. the 20 °C treatment, on the other, suggests that in terms of DNA methylation and gene expression the temperature of 20 °C is perceived as altered when compared to the “natural” temperature range of 14–17 °C, appropriate for larval development in the European sea bass (Vazquez and Munoz-Cueto, 2014). In fact, this observation matches and supports the observations that temperatures > 17 °C alters sex ratios in this species while temperatures < 17 °C do not (Navarro-Martín et al., 2009). Second, these 210 genes, among which we find several genes related to calcium signaling such as *atp2a1*, *gsto2*

and *cabp2a*, involved also in environmental signal transduction (Castelli et al., 2020), constituted robust short-term both epigenomic and transcriptomic biomarkers of the effects of temperature during the European sea bass early development. As above, they can be used to assess the quality of the egg incubation thermal regime and larvae until first feeding. Furthermore, the most relevant GO-terms linked to these 210 genes were DNA methylation, regulation of Wnt signaling pathway and cellular response to cold. All of these GO-terms results are consistent with the fact that DNA methylation, Wnt signaling and cellular response to cold has been previously implicated in other studies examining the effects of temperature on animals (Al-Fageeh and Smales, 2006; Risha et al., 2021; Yao et al., 2022).

Regarding the 577 and 344 genes with negative and positive correlation, respectively, in the 14 °C vs. 17 °C comparison and the 102 and 250 genes with negative and positive correlation, respectively, in the 20 °C vs. 17 °C comparison it is interesting to note that gene expression changed on average <10-fold (less than one log value). In contrast, DNA methylation changes were around 20–25%. These differences are of sufficient magnitude suitable to use them as biomarkers of effects of temperature during early development. There were no common genes between the 577 and 102 gene lists, meaning there were no genes the expression of which is epigenetically and consistently downregulated in response to temperature. As for the 344 and 250 genes more methylated and upregulated by temperature, only one gene, ENSDLAG00005009958, was common, implying that its expression is epigenetically regulated in a consistent manner in response to temperature changes and thus constitute a powerful biomarkers of the short-term effects of early temperature changes in larvae.

We only found two genes, *spata5l1* and *agl*, that coincided with the metastable epialleles identified previously in the European sea bass (Anastasiadi et al., 2021). This low number is not surprising given the stringent criteria we used and the requirements needed to be considered a metastable epiallele. Using instead genes with DMCs in juveniles and/

or adults that coincided with genes differentially methylated in the same tissues in the above mentioned study returned other gene lists (43 shown in Table 1) that constituted robust epigenetic biomarkers for the very long-term effects of temperature alterations during early development on European sea bass juveniles and adults. Some of them are independent of the direction of temperature change and hence can be used as a reliable method to determine whether animals had the optimal temperature during larval rearing. *Agl* and *nlg1* genes are present in both lists. *Agl* encodes the glycogen debrancher, an enzyme involved in glycogen degradation and *agl* mutations are linked with glycogen storage disease (Sentner et al., 2012) found differentially expressed in fast- vs. slow-growing rainbow trout, *Onchorynchus mykiss* (Danzmann et al., 2016) but no effects due to temperature have been reported.

One of the most important insight of this study is the report, to our knowledge for the first time, of epigenetic biomarkers with prognostic and diagnostic value. Of the five genes that in principle met the criteria for having these properties the affected DMCs were not consistently on the same CpGs position and thus can be considered as potential epigenetic markers with both prognostic and diagnostic value. The gene ENSDLA00005004413, for which we had gene expression data and DNA methylation data on exactly the same CpG position in two time points in larvae and adults, was not differentially expressed either at 14 °C nor at 20 °C when compared to 17 °C at 5 dpf but became it was upregulated in the adults in fish reared at 20 °C at 482 dpf. Regarding the DMC within this gene, the same CpG that was hypomethylated in larvae reared at 20 °C was also hypomethylated in the adults of fish reared at 20 °C at 482 dpf. Thus, hypomethylation of this gene in larvae exposed to elevated temperature precedes higher expression in adults ~500 dpf. Therefore, this gene can be considered an epigenetic biomarker with both prognostic and diagnostic value and hence their full identification is warranted. However, further analysis on the same gene and all the other biomarkers listed in the present study should be performed in a larger number of individuals from different families and from an independent experiments to confirm their usefulness as biomarkers. Given the protein sequence alignment of ENSDLA00005004413 with *kpa10-4* gene in the brown trout, we suggest that the identified biomarker could be related to keratin production. This could be of relevance for the fitness of the sea bass since they produce keratin in the skin mucus (Cordero et al., 2015). The skin mucus is a crucial first barrier of protection from both biotic and abiotic factors because it protects the skin epidermis, where fish in general have few or no keratinized cells (Brinchmann, 2016). Thus, since methylation is maintained, measuring the methylation of this gene in fish samples from a given batch at any time would indicate whether that the fish of that batch had been exposed to an abnormally high temperature during the critical stages of early development.

In this study, we focused on the most studied environmental factor, i. e., temperature, which has a huge impact on stress in many vertebrate and invertebrate marine farmed species. However another environmental factor, with detrimental effects on European sea bass development is rearing density (Sammouth et al., 2009), albeit without effects on sex ratio (Saillant et al., 2003) as temperature does. In an experiment with 6 month-old European sea bass exposed to short intervals of acute density stress, 57 DMCs in 51 distinct genes measured in blood samples were identified (Krick et al., 2021). Interestingly, among these 51 genes, two genes were also present in our list of genes with DMCs in three tissues: roundabout homolog 2-like (*robo3*) in 20 °C vs. 17 °C and zinc finger matrin-type 4b (*smat4*) in 14 °C vs. 17 °C. These two genes seem to be involved in behavior linked with stress conditions (Rey et al., 2013).

In conclusion, we report, to the best of our knowledge, the first identification of epigenetic biomarkers with both diagnostic and prognostic value to anticipate the long-lasting effects of thermal alterations during early development in fish. One can envisage the determination of DNA methylation levels (e.g., by a PCR-based approach) of CpGs within some of the genes identified in this study as a rapid and economic means to monitor the quality of the early environment, in our case the thermal

environment, to screen samples of the European sea bass larvae as a diagnostic and prognostic tool for present and future performance. Our study shows the way for similar developments for other farmed species as well as for applications to conservation biology.

Data availability

BS-Seq and RNA-Seq reads are available at the NCBI-SRA under the BioProject accession number: PRJNA888946.

CRediT authorship contribution statement

Alejandro Valdivieso: Methodology, Investigation, Formal analysis, Writing – original draft. **Núria Sánchez-Baizán:** Methodology, Investigation, Formal analysis, Writing – original draft. **Nikos Mitri-zakis:** Methodology, Investigation, Formal analysis. **Nikos Papan-droulakis:** Conceptualization, Resources, Funding acquisition, Supervision, Methodology, Investigation, Formal analysis, Writing – original draft. **Francesc Piferrer:** Conceptualization, Resources, Funding acquisition, Supervision, Writing – original draft.

Declaration of Competing Interest

The authors declare they have no competing interests.

Data availability

Data enclosed in the Supplementary Information

Acknowledgments

We thank Dr. Katherina Moutou for providing the adults samples. This study was supported by the European Union H2020 project Performfish, reference number 727610 and by Spanish Ministry of Science and Innovation (PID2019-108888RB-I00) grant to FP. With funding from the Spanish government through the ‘Severo Ochoa Centre of Excellence’ accreditation (CEX2019-000928-S). AV was supported by a Performfish contract and NS was supported by a Spanish Ministry of Science and Innovation predoctoral scholarship (BES-2017-079744).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aquaculture.2022.738918>.

References

- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., Mason, C.E., 2012. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13 (10), R87. <https://doi.org/10.1186/gb-2012-13-10-r87>.
- Alexa, A., Rahnenfuhrer, J., 2010. TopGO: enrichment analysis for gene ontology. *R Package Version 2* (0), 2010.
- Al-Fageeh, M.B., Smales, C.M., 2006. Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *Biochem. J.* 397 (2), 247–259. <https://doi.org/10.1042/bj20060166>.
- Aliferi, A., Ballard, D., Gallidabino, M.D., Thurtell, H., Barron, L., Syndercombe Court, D., 2018. DNA methylation-based age prediction using massively parallel sequencing data and multiple machine learning models. *Forensic Sci. Int. Genet.* 37, 215–226. <https://doi.org/10.1016/j.fsigen.2018.09.003>.
- Altschul, S., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25 (17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
- Anastasiadi, D., Piferrer, F., 2019a. Data from: a clockwork fish. Age-prediction using DNA methylation-based biomarkers in the European seabass (p. 2287594534 bytes) [data set]. Dryad. <https://doi.org/10.5061/dryad.m0cfxpnz4>.
- Anastasiadi, D., Piferrer, F., 2019b. Epimutations in developmental genes underlie the onset of domestication in farmed European sea bass. *Mol. Biol. Evol.* 36 (10), 2252–2264. <https://doi.org/10.1093/molbev/msz153>.
- Anastasiadi, D., Vandeputte, M., Sánchez-Baizán, N., Allal, F., Piferrer, F., 2018. Dynamic epimarks in sex-related genes predict gonad phenotype in the European sea bass, a

- fish with mixed genetic and environmental sex determination. *Epigenetics* 13 (9), 988–1011. <https://doi.org/10.1080/15592294.2018.1529504>.
- Anastasiadi, D., Shao, C., Chen, S., Piferrer, F., 2021. Footprints of global change in marine life: inferring past environment based on DNA methylation and gene expression marks. *Mol. Ecol.* 30 (3), 747–760. <https://doi.org/10.1111/mec.15764>.
- Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11 (10), R106. <https://doi.org/10.1186/gb-2010-11-10-r106>.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57 (1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Blázquez, M., Zanuy, S., Carillo, M., Piferrer, F., 1998. Effects of rearing temperature on sex differentiation in the European sea bass (</>*Dicentrarchus labrax* L.</>). *J. Exp. Zool.* 281 (3), 207–216. [https://doi.org/10.1002/\(sici\)1097-010x\(19980615\)281:3<207::aid-jez6>3.0.co;2-r](https://doi.org/10.1002/(sici)1097-010x(19980615)281:3<207::aid-jez6>3.0.co;2-r).
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Brinckmann, M.F., 2016. Immune relevant molecules identified in the skin mucus of fish using -omics technologies. *Mol. Biosyst.* 12 (7), 2056–2063. <https://doi.org/10.1039/c5mb00890e>.
- Burggren, W.W., Mueller, C.A., 2015. Developmental critical windows and sensitive periods as three-dimensional constructs in time and space. *Physiol. Biochem. Zool.* 88 (2), 91–102. <https://doi.org/10.1086/679906>.
- Bushnell, B., 2014. *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).
- Castelli, M.A., Whiteley, S.L., Georges, A., Holleley, C.E., 2020. Cellular calcium and redox regulation: the mediator of vertebrate environmental sex determination? *Biol. Rev.* 95 (3), 680–695. <https://doi.org/10.1111/brv.12582>.
- Chen, Y., Lun, A.T., Smyth, G.K., 2014. Differential expression analysis of complex RNA-seq experiments using edgeR. In: *Statistical Analysis of Next Generation Sequencing Data*. Springer, pp. 51–74.
- Chen, Y., Lun, A.T., Smyth, G.K., 2016. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research* 5, 1438. <https://doi.org/10.12688/f1000research.8987.2>.
- Conceicao, L., Tandler, A., 2018. *Success Factors for Fish Larval Production*. John Wiley & Sons.
- Cordero, H., Brinckmann, M.F., Cuesta, A., Meseguer, J., Esteban, M.A., 2015. Skin mucus proteome map of European sea bass (*Dicentrarchus labrax*). *PROTEOMICS* 15 (23–24), 4007–4020. <https://doi.org/10.1002/pmic.201500120>.
- Costa-Pinheiro, P., Montezuma, D., Henrique, R., Jerónimo, C., 2015. Diagnostic and prognostic epigenetic biomarkers in cancer. *Epigenomics* 7 (6), 1003–1015. <https://doi.org/10.2217/epi.15.56>.
- Danzmann, R.G., Kocmarek, A.L., Norman, J.D., Rexroad, C.E., Palti, Y., 2016. Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics* 17 (1), 60. <https://doi.org/10.1186/s12864-016-2363-5>.
- Database Resources of the National Center for Biotechnology Information, 2016. *Nucleic Acids Res.* 44 (D1), D7–D19. <https://doi.org/10.1093/nar/gkv1290>.
- Deans, C., Maggert, K.A., 2015. What do you mean, “epigenetic”? *Genetics* 199 (4), 887–896. <https://doi.org/10.1534/genetics.114.173492>.
- Deveson, I.W., Holleley, C.E., Blackburn, J., Marshall Graves, J.A., Mattick, J.S., Waters, P.D., Georges, A., 2017. Differential intron retention in *Jumonji* chromatin modifier genes is implicated in reptile temperature-dependent sex determination. *Sci. Adv.* 3 (6), e1700731. <https://doi.org/10.1126/sciadv.1700731>.
- Díaz, N., Piferrer, F., 2015. Lasting effects of early exposure to temperature on the gonadal transcriptome at the time of sex differentiation in the European sea bass, a fish with mixed genetic and environmental sex determination. *BMC Genomics* 16 (1), 679. <https://doi.org/10.1186/s12864-015-1862-0>.
- Díaz, N., Piferrer, F., 2017. Estrogen exposure overrides the masculinizing effect of elevated temperature by a downregulation of the key genes implicated in sexual differentiation in a fish with mixed genetic and environmental sex determination. *BMC Genomics* 18 (1), 973. <https://doi.org/10.1186/s12864-017-4345-7>.
- Dülger, N., Kumlu, M., Türkmen, S., Ölcü, A., Tufan Eroldoğan, O., Asuman Yılmaz, H., Oçal, N., 2012. Thermal tolerance of European sea bass (*Dicentrarchus labrax*) juveniles acclimated to three temperature levels. *J. Therm. Biol.* 37 (1), 79–82. <https://doi.org/10.1016/j.jtherbio.2011.11.003>.
- Duncan, E.J., Gluckman, P.D., Dearden, P.K., 2014. Epigenetics, plasticity, and evolution: how do we link epigenetic change to phenotype? *J. Exp. Zool. B Mol. Dev. Evol.* 322 (4), 208–220. <https://doi.org/10.1002/jez.b.22571>.
- Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., Huber, W., 2005. BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 21 (16), 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525>.
- Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4 (8), 1184–1191. <https://doi.org/10.1038/nprot.2009.97>.
- Federation of European Aquaculture Producers, 2020. European Aquaculture Production Report 2014–2020. Available from: <https://feap.info/index.php/data/>. (Accessed 10 May 2022).
- Felip, A., Zanuy, S., Carrillo, M., 2006. Comparative analysis of growth performance and sperm motility between precocious and non-precocious males in the European sea bass (*Dicentrarchus labrax*, L.). *Aquaculture* 256 (1–4), 570–578. <https://doi.org/10.1016/j.aquaculture.2006.02.014>.
- Gavery, M.R., Roberts, S.B., 2017. Epigenetic considerations in aquaculture. *PeerJ* 5, e4147. <https://doi.org/10.7717/peerj.4147>.
- Ge, C., Ye, J., Weber, C., Sun, W., Zhang, H., Zhou, Y., Cai, C., Qian, G., Capel, B., 2018. The histone demethylase KDM6B regulates temperature-dependent sex determination in a turtle species. *Science* 360 (6389), 645–648. <https://doi.org/10.1126/science.aap8328>.
- Georgakopoulou, E., Sfakianakis, D.G., Kouttoui, S., Divanach, P., Kentouri, M., Koumoundouros, G., 2007. The influence of temperature during early life on phenotypic expression at later ontogenetic stages in sea bass. *J. Fish Biol.* 70 (1), 278–291. <https://doi.org/10.1111/j.1095-8649.2007.01305.x>.
- Granada, L., Lemos, M.F.L., Cabral, H.N., Bossier, P., Novais, S.C., 2018. Epigenetics in aquaculture – the last frontier. *Rev. Aquac.* 10 (4), 994–1013. <https://doi.org/10.1111/raq.12219>.
- Hernández, H.G., Tse, M.Y., Pang, S.C., Arboleda, H., Forero, D.A., 2013. Optimizing methodologies for PCR-based DNA methylation analysis. *BioTechniques* 55 (4), 181–197. <https://doi.org/10.2144/000114087>.
- Jonsson, B., Jonsson, N., 2014. Early environment influences later performance in fishes: effects of early experiences. *J. Fish Biol.* 85 (2), 151–188. <https://doi.org/10.1111/jfb.12432>.
- Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37 (8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
- Koumoundouros, G., Divanach, P., Anezaki, L., Kentouri, M., 2001. Temperature-induced ontogenetic plasticity in sea bass (*Dicentrarchus labrax*). *Mar. Biol.* 139 (5), 817–830. <https://doi.org/10.1007/s002270100635>.
- Krick, M.V., Desmarais, E., Samaras, A., Guéret, E., Dimitroglou, A., Pavlidis, M., Tsigonopoulos, C., Guinand, B., 2021. Family-effects in the epigenomic response of red blood cells to a challenge test in the European sea bass (*Dicentrarchus labrax*, L.). *BMC Genomics* 22 (1), 111. <https://doi.org/10.1186/s12864-021-07420-9>.
- Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 27 (11), 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.
- Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., Carey, V.J., 2013. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9 (8), e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>.
- Lazado, C.C., Voldvik, V., Breiland, M.W., Osório, J., Hansen, M.H.S., Krasnov, A., 2020. Oxidative chemical stressors alter the physiological state of the nasal olfactory mucosa of Atlantic salmon. *Antioxidants* 9 (11), 1144. <https://doi.org/10.3390/antiox9111144>.
- Levenson, V.V., 2010. DNA methylation as a universal biomarker. *Expert. Rev. Mol. Diagn.* 10 (4), 481–488. <https://doi.org/10.1586/erm.10.17>.
- Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30 (7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
- Liu, N., Li, Z., Pei, D., Shu, X., 2013. Zfyve9a regulates the proliferation of hepatic cells during zebrafish embryogenesis. *Int. J. Dev. Biol.* 57 (9–10), 773–778. <https://doi.org/10.1387/jidb.130065xs>.
- Liu, Q., Yang, J., Gong, Y., Cai, J., Zheng, Y., Zhang, Y., Yu, D., Zhang, Z., 2020. MicroRNA profiling identifies biomarkers in head kidneys of common carp exposed to cadmium. *Chemosphere* 247, 125901. <https://doi.org/10.1016/j.chemosphere.2020.125901>.
- López-Albors, O., Abdel, I., Perriago, Ma.J., Ayala, Ma.D., Alcázar, A.G., Graciá, C.M., Nathanailides, C., Vázquez, J. Ma, 2008. Temperature influence on the white muscle growth dynamics of the sea bass *Dicentrarchus labrax*, L. Flesh quality implications at commercial size. *Aquaculture* 277 (1–2), 39–51. <https://doi.org/10.1016/j.aquaculture.2008.02.015>.
- Lun, A.T., Smyth, G.K., 2016. Csq: a bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* 44 (5), e45.
- Lun, A.T., Smyth, G.K., 2017. No counts, no variance: allowing for loss of degrees of freedom when assessing biological variability from RNA-seq data. *Stat. Appl. Genet. Mol. Biol.* 16 (2), 83–93.
- Lund, S.P., Nettleton, D., McCarthy, D.J., Smyth, G.K., 2012. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Stat. Appl. Genet. Mol. Biol.* 11 (5).
- McCarthy, D.J., Chen, Y., Smyth, G.K., 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 40 (10), 4288–4297.
- Morretti, A., 1999. *Manual on Hatchery Production of Seabass and Gilthead Seabream*, 1. Food & Agriculture Org.
- Naue, J., Hoefsloot, H.C.J., Mook, O.R.F., Rijlaarsdam-Hoekstra, L., van der Zwalm, M.C. H., Henneman, P., Kloosterman, A.D., Verschure, P.J., 2017. Chronological age prediction based on DNA methylation: massive parallel sequencing and random forest regression. *Forensic Sci. Int. Genet.* 31, 19–28. <https://doi.org/10.1016/j.fsigen.2017.07.015>.
- Navarro-Martín, L., Blázquez, M., Viñas, J., Joly, S., Piferrer, F., 2009. Balancing the effects of rearing at low temperature during early development on sex ratios, growth and maturation in the European sea bass (*Dicentrarchus labrax*). *Aquaculture* 296 (3–4), 347–358. <https://doi.org/10.1016/j.aquaculture.2009.07.022>.
- Papadaki, M., Piferrer, F., Zanuy, S., Maingot, E., Divanach, P., Mylonas, C.C., 2005. Growth, sex differentiation and gonad and plasma levels of sex steroids in male- and female-dominant populations of *Dicentrarchus labrax* obtained through repeated size grading. *J. Fish Biol.* 66 (4), 938–956. <https://doi.org/10.1111/j.0022-1112.2005.00639.x>.
- Papandroulakis, N., Divanach, P., Anastasiadis, P., Kentouri, M., 2001. The pseudo-green water technique for intensive rearing of sea bream (*Sparus aurata*) larvae. *Aquac. Int.* 9 (3), 205–216.

- Pavlidis, M.A., Mylonas, C.C., 2011. *Sparidae: Biology and Aquaculture of Gilthead Sea Bream and Other Species*. John Wiley & Sons.
- Pepin, P., 1991. Effect of temperature and size on development, mortality, and survival rates of the pelagic early life history stages of marine fish. *Can. J. Fish. Aquat. Sci.* 48 (3), 503–518. <https://doi.org/10.1139/f91-065>.
- Percie du Sert, N., Hurst, V., Ahluwalia, A., Alam, S., Avey, M.T., Baker, M., Browne, W. J., Clark, A., Cuthill, I.C., Dirnagl, U., Emerson, M., Garner, P., Holgate, S.T., Howells, D.W., Karp, N.A., Lalic, S.E., Lidster, K., MacCallum, C.J., Macleod, M., Würbel, H., 2020. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. *PLoS Biol.* 18 (7), e3000410 <https://doi.org/10.1371/journal.pbio.3000410>.
- Person-Le Ruyet, J., Mahé, K., Le Bayon, N., Le Delliou, H., 2004. Effects of temperature on growth and metabolism in a Mediterranean population of European sea bass, *Dicentrarchus labrax*. *Aquaculture* 237 (1–4), 269–280. <https://doi.org/10.1016/j.aquaculture.2004.04.021>.
- Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S., Smyth, G.K., 2016. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* 10 (2), 946–963. <https://doi.org/10.1214/16-aos920>.
- Pickett, G.D., Pawson, M.G., 1994. *Sea Bass: Biology*, 12. Springer Science & Business Media.
- Piferrer, F., Blázquez, M., Navarro, L., González, A., 2005. Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (*Dicentrarchus labrax* L.). *Gen. Comp. Endocrinol.* 142 (1–2), 102–110. <https://doi.org/10.1016/j.ygcen.2005.02.011>.
- Piferrer, F., Anastasiadi, D., Díaz, N., Ribas, L., 2015, June 21. Integrating Epigenetics into Aquaculture Research [International Symposium on Genetics in Aquaculture XII : Abstract Book: 34 (2015)].
- R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: <https://www.R-project.org/>.
- Rey, S., Boltana, S., Vargas, R., Roher, N., MacKenzie, S., 2013. Combining animal personalities with transcriptomics resolves individual variation within a wild-type zebrafish population and identifies underpinning molecular differences in brain function. *Mol. Ecol.* 22 (24), 6100–6115. <https://doi.org/10.1111/mec.12556>.
- Risha, M.A., Ali, A., Siengdee, P., Trakooljul, N., Haack, F., Dannenberg, D., Wimmers, K., Ponsuksili, S., 2021. Wnt signaling related transcripts and their relationship to energy metabolism in C2C12 myoblasts under temperature stress. *PeerJ* 9, e11625. <https://doi.org/10.7717/peerj.11625>.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140.
- RStudio Team, 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. Available from: <http://www.rstudio.com/>.
- Sadoul, B., Alfonso, S., Gool, C., Pratlong, M., Rialle, S., Geffroy, B., Bégout, M.-L., 2022. Transcriptomic profiles of consistent risk-taking behaviour across time and contexts in European sea bass. *Proc. R. Soc. B Biol. Sci.* 289 (1975), 20220399. <https://doi.org/10.1098/rspb.2022.0399>.
- Saillant, E., Fostier, A., Haffray, P., Menu, B., Laureau, S., Thimonier, J., Chatain, B., 2003. Effects of rearing density, size grading and parental factors on sex ratios of the sea bass (*Dicentrarchus labrax* L.) in intensive aquaculture. *Aquaculture* 221 (1–4), 183–206. [https://doi.org/10.1016/S0044-8486\(02\)00539-2](https://doi.org/10.1016/S0044-8486(02)00539-2).
- Sammouth, S., d'Orbcastel, E.R., Gasset, E., Lemarié, G., Breuil, G., Marino, G., Coeurdacier, J.-L., Fivelstad, S., Blancheton, J.-P., 2009. The effect of density on sea bass (*Dicentrarchus labrax*) performance in a tank-based recirculating system. *Aquac. Eng.* 40 (2), 72–78. <https://doi.org/10.1016/j.aquaeng.2008.11.004>.
- Schlicker, A., Domingues, F.S., Rahnenführer, J., Lengauer, T., 2006. A new measure for functional similarity of gene products based on gene ontology. *BMC Bioinform.* 7 (1), 302. <https://doi.org/10.1186/1471-2105-7-302>.
- Schulte, P.M., Healy, T.M., Fangue, N.A., 2011. Thermal performance curves, phenotypic plasticity, and the time scales of temperature exposure. *Integr. Comp. Biol.* 51 (5), 691–702. <https://doi.org/10.1093/icb/ict097>.
- Sentner, C.P., Vos, Y.J., Niezen-Koning, K.N., Mol, B., Smit, G.P.A., 2012. Mutation analysis in glycogen storage disease Type III patients in the Netherlands: novel genotype-phenotype relationships and five novel mutations in the AGL gene. In: Brown, G., Morava, E., Peters, V., Gibson, K.M., Zschocke, J. (Eds.), *JIMD Reports—Case and Research Reports*, 2012/4, 7. Springer, Berlin Heidelberg, pp. 19–26. https://doi.org/10.1007/8904_2012_134.
- Sun, B.-J., Li, T., Gao, J., Ma, L., Du, W.-G., 2015. High incubation temperatures enhance mitochondrial energy metabolism in reptile embryos. *Sci. Rep.* 5 (1), 8861. <https://doi.org/10.1038/srep08861>.
- Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6 (7), e21800. <https://doi.org/10.1371/journal.pone.0021800>.
- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierre, N., Volckaert, F.A.M., Bargelloni, L., Power, D.M., Bonhomme, F., Reinhardt, R., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nat. Commun.* 5 (1), 5770. <https://doi.org/10.1038/ncomms6770>.
- Tsukazaki, T., Chiang, T.A., Davison, A.F., Attisano, L., Wrana, J.L., 1998. SARA, a FYVE domain protein that recruits Smad2 to the TGF β receptor. *Cell* 95 (6), 779–791. [https://doi.org/10.1016/S0092-8674\(00\)81701-8](https://doi.org/10.1016/S0092-8674(00)81701-8).
- Vandeputte, M., Piferrer, F., 2019. Genetic and environmental components of sex determination in the European sea bass. In: Wang, H.-P., Piferrer, F., Chen, S.-L., Shen, Z.-G. (Eds.), *Sex Control in Aquaculture*. John Wiley & Sons, Ltd., pp. 305–325. <https://doi.org/10.1002/9781119127291.ch14>.
- Vandeputte, M., Dupont-Nivet, M., Chavanne, H., Chatain, B., 2007. A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. *Genetics* 176 (2), 1049–1057. <https://doi.org/10.1534/genetics.107.072140>.
- Vazquez, F.J.S., Munoz-Cueto, J.A. (Eds.), 2014. *Biology of European Sea Bass*, 0 ed. CRC Press. <https://doi.org/10.1201/b16043>.
- Wang, H.-Q., Tuominen, L.K., Tsai, C.-J., 2011. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. *Bioinformatics* 27 (2), 225–231. <https://doi.org/10.1093/bioinformatics/btq650>.
- Wickham, H., 2011. Wiley interdisciplinary reviews. *Comput. Stat.* 3, 180–185. <https://doi.org/10.1002/wics.147>.
- Wingett, S.W., Andrews, S., 2018. FastQ screen: a tool for multi-genome mapping and quality control. *F1000Research* 7, 1338. <https://doi.org/10.12688/f1000research.15931.2>.
- Xi, Y., Li, W., 2009. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinform.* 10 (1), 232. <https://doi.org/10.1186/1471-2105-10-232>.
- Yang, F., Zheng, J., 2014. High temperature sensitivity is intrinsic to voltage-gated potassium channels. *ELife* 3, e03255. <https://doi.org/10.7554/eLife.03255>.
- Yao, Z.L., Zhao, Y., Chen, H.J., Wang, H., Ji, X.S., 2022. Integrated analysis of DNA methylation and RNA transcriptome during high-temperature-induced masculinization in sex-undifferentiated Nile tilapia gonad. *Aquaculture* 552, 738043. <https://doi.org/10.1016/j.aquaculture.2022.738043>.
- Zhang, J., Yu, Z., Zhang, W., Shi, H., Tao, M., Chen, J., Liu, Y., Zhao, M., Li, S., Zhang, Y., Lin, H., 2021. Jumonji family histone demethylases in *Epinephelus coioides* gonad. *Reprod. Breed.* 1 (1), 1–10. <https://doi.org/10.1016/j.repbre.2021.02.002>.
- Zhou, R., Shang, R., Gong, D., Xu, X., Liu, S., 2019. Characterization of H3 methylation in regulating oocyte development in cyprinid fish. *Sci. China Life Sci.* 62 (6), 829–837. <https://doi.org/10.1007/s11427-018-9346-6>.
- Zou, H., Shi, M., Li, R., Zhang, X., Lu, W., 2021. Comparative transcriptome and methylome analysis of the hindbrain in olive flounder (*Paralichthys olivaceus*) considering individual behavior-type and oxygen metabolism. *Comp. Biochem. Physiol. D: Genom. Proteom.* 38, 100799 <https://doi.org/10.1016/j.cbd.2021.100799>.

