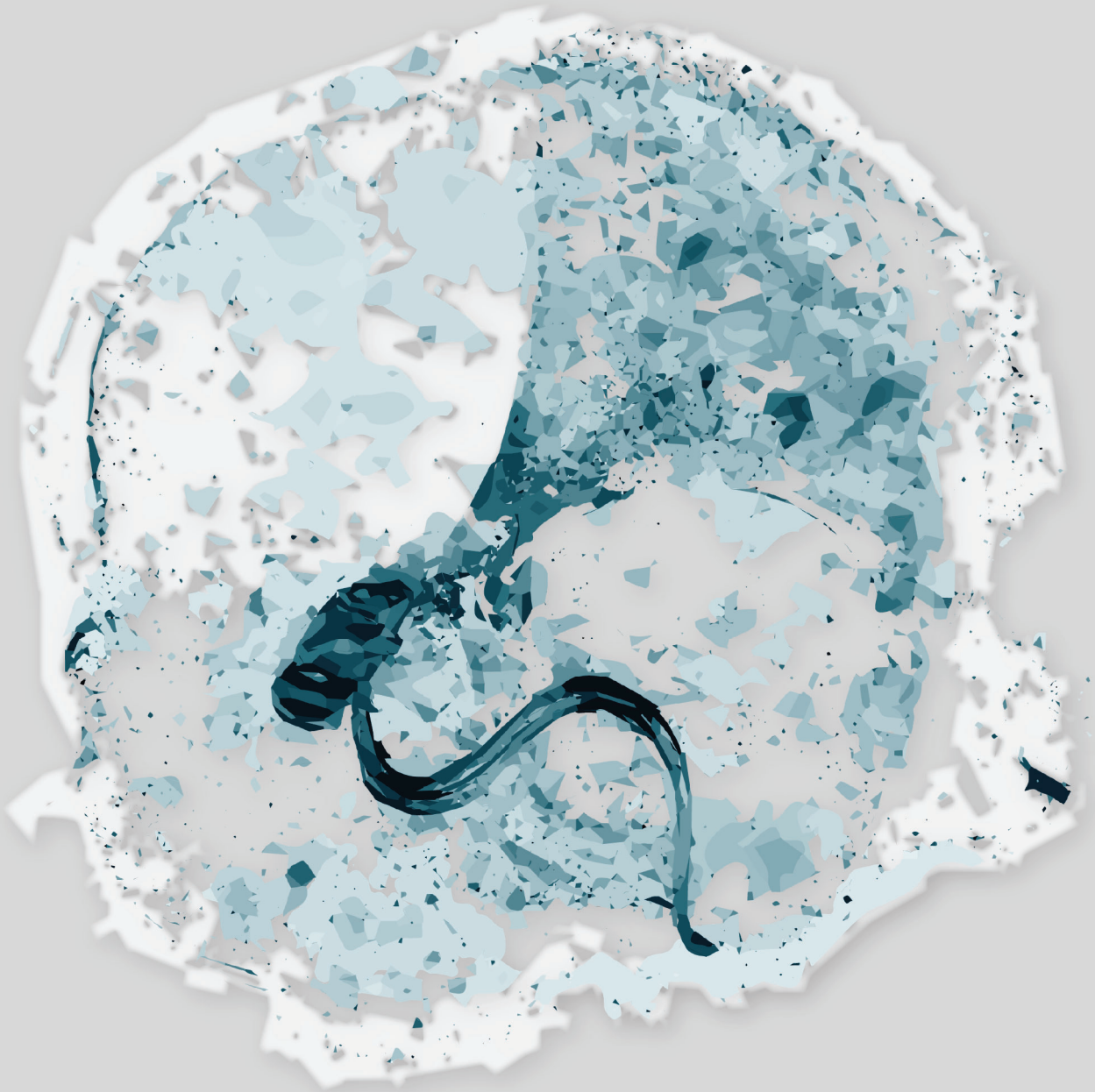
Less, but more: Massive gene losses and expansions in appendicularians stretch the evolutionary limits of the FGF signalling pathway within chordates

Gaspar Sánchez Serna

LESS, BUT MORE:

MASSIVE GENE LOSSES AND EXPANSIONS IN APPENDICULARIANS STRETCH THE EVOLUTIONARY LIMITS OF THE FGF SIGNALLING PATHWAY WITHIN CHORDATES



Gaspar Sánchez Serna

2024

UNIVERSITAT DE BARCELONA

Tesis Doctoral

Universitat de Barcelona

Less, but more: Massive gene losses and expansions in appendicularians stretch the evolutionary limits of the FGF signalling pathway within chordates

Memoria presentada por

Gaspar Sánchez Serna

Para optar al grado de

Doctor por la Universidad de Barcelona

Programa de Doctorado “Genética”

Departamento de Genética, Microbiología y Estadística

Director: Cristian Cañestro García

Tutor: Cristian Cañestro García

Doctorando: Gaspar Sánchez Serna

Barcelona, 28 de julio de 2024

AGRADECIMIENTOS / ACKNOWLEDGMENTS

Bueno… ha llegado el momento de cerrar esta etapa. Quiero empezar confesando que nunca se me ha dado bien eso de separar vida profesional de vida personal, y por eso en estos párrafos me voy a permitir el lujo de escribir de la forma más personal que cabe. Sin haber corrido una maratón de verdad en mi vida, creo que el doctorado ha sido algo parecido. Ha habido momentos buenos y momentos malos, y aunque la balanza no ha parado de oscilar, hoy me alegra saber que de todos he sacado una lección valiosa. Durante todos estos años he tenido la suerte de estar rodeado de personas que recordaré como la cara y la voz de mis vivencias. A todas ellas les estoy profundamente agradecido.

Quiero comenzar agradeciendo a mi director y tutor de tesis, Cristian. Gracias por confiar en mí desde el primer momento, por tu cercanía y por todo tu esfuerzo. He aprendido mucho contigo, y no solo de genética. También quiero dar las gracias a todos los miembros y exmiembros del Oikoteam que han contribuido directamente a este proyecto. A Jordi y a Paula, que abrieron la puerta de los FGF en Oiko; a Laura y a Ana, que dedicaron sus TFGs a avanzar este trabajo; a Nuria, que sin su ayuda aún sudaría frío al oír hablar de ”data”; y a Alfonso, que desde el primer día ha hecho las veces de manual de instrucciones y de navaja suiza del laboratorio, siempre dispuesto a ayudar en lo que hiciese falta. Gracias también a Eva, Constan, Biel, Daniela, y a todas las personas con quienes he coincidido estos años en el lab y que han hecho cada día más ameno entre conversaciones, música, anécdotas y quejas compartidas. Un agradecimiento especial a Sebas, por mantener a las Oikos vivas y coleando todos estos años; y a Ricard, por sus visitas recurrentes al lab con las que, además de instarme a trabajar, siempre me ha sacado unas risas.

Y ahora Marc!! A ti te dedico un párrafo entero, como mínimo. De todos los experimentos que hemos hecho estos años, el de hacernos amigos es de los que han dado mejores resultados (excepto para nuestros pulmones). Ni me hago una idea de cómo habría sido toda esta experiencia sin haberla compartido contigo, tanto dentro como fuera del laboratorio. Muchas gracias de corazón por tu infinita paciencia y por haberme acompañado en cada momento de la mejor manera posible.

I also want to thank Prof. Nicholas Luscombe and the entire team at the Genomics and Regulatory Systems Unit at OIST for welcoming me so warmly and helping me to enhance both my skills and my love for science. Special thanks to Nico, Charles, Mike, and Nick for the insightful discussions, guidance on my work, and your kindness. It’s been a great pleasure to work with you. Osewa ni narimashita!!

Cambiando un poco de ámbito, no puedo mirar atrás a estos casi cinco años sin pensar en las personas que, aunque no han estado directamente involucradas en el trabajo, son con quien he celebrado las alegrías, aliviado las frustraciones, y recargado las pilas para funcionar día tras día. Una maratón es mucho más asequible cuando se tiene a gente animando en el camino y esperando en la meta (supongo). Ahora que estoy llegando al final, quiero resaltar la importancia de esa gente.

Muchas gracias a Sergio, Alba, Salo, Dani y Marc por haber sido un apoyo firme y constante durante todos estos años. Cuando llegué a Barcelona me sentía lejos de casa, y gracias a vosotros aprendí que un hogar puede construirse en cualquier sitio, siempre que los materiales y las manos que lo levantan sean los adecuados. Gracias también a Pao, Helena, Ángel, Eric, Youssef, Ale, Janka... Con vosotros he vivido momentos y etapas que hacen que me lleve los mejores recuerdos de estos años. Barcelona me parece la mejor ciudad del mundo gracias a todos vosotros.

Ahora me dirijo a la Flota family. Sole, Rebeca, Miguel, Irene y Carlos, habéis estado conmigo desde que mis asignaturas favoritas eran “Conocimiento del Medio” y “Alternativa a Religión”. Con vosotras prácticamente aprendí a caminar, así que podría daros las gracias por cada paso que doy. A pesar de la distancia, siempre he sentido vuestro apoyo muy cerca, y cada visita a Murcia o cada campamento Krusty me ha alegrado el mes. Retomando la metáfora del hogar, con vosotros es como si tuviera un chalet en Cabo de Palos.

Si mencionase a todas las personas que me pasan ahora por la mente, esto parecería otro capítulo de la tesis, pero hay algunas más a las que no me quiero dejar. Juanlu, Fran, David, Adri, Ram, Antonio, Pedro… con vosotros empezó todo. No creo que me gustase tanto la biología si no la hubiese aprendido a vuestro lado. Lucía, muchas gracias por estar siempre presente, por tratarme con toda la paciencia del mundo y por seguir creciendo conectados, incluso en la distancia. Raquel, tú fuiste básicamente mi sherpa al empezar el doctorado, y también de las personas con las que más me he reído dentro y fuera de la universidad. Muchas gracias por todos los cafés, cervezas, consejos y charlas infinitas sobre cualquier tema random. Edgar, tú también has sido una figura clave en mi experiencia UB, aunque creo que nos hemos visto más en fiestas que dentro de la universidad. ¡Gracias por todos los buenos momentos! Roberto y Ángela, estos últimos meses os habéis tragado mis peores caras y estados de ánimo, y aun así me lo habéis puesto muy fácil para desconectar cada vez que volvía al piso, os lo agradezco mucho. Muchas gracias también a Misae. Desde que nos conocimos me has motivado con cada conversación y te has reído descaradamente de mis dramas, contagiándome la risa. And also to the other two members of the Gomi family, Oli and Nikochan! Even though we haven’t known each other for long, some of the most unforgettable memmories of my PhD experience have been made by your side. With you I rediscovered the fun side of science and learned to cope with prolonged sleep deprivation. Thank you for always sharing your energy and enthusiasm!

El último párrafo se lo quiero dedicar a mi familia. Gracias por haber creído siempre en mí e impulsarme a tomar este camino que, aunque a veces me cuesta admitirlo, me hace muy feliz. Gracias especialmente a mis padres por apoyarme desde todos los ángulos que existen durante los últimos 5 años y los 23 anteriores. Gracias a vosotros conozco el sentimiento de gratitud en toda su pureza y sin duda es de mis sentimientos favoritos.

En definitiva, muchas gracias a todas las personas que me han acompañado estos años, que hacen que los tragos amargos pasen rápido y que los dulces sean aún más dulces. Esta tesis no habría sido posible sin vosotros.

ABSTRACT

The extensive diversity of forms in the animal kingdom arises from a limited set of evolutionarily conserved genes that coordinate during embryonic development to build complex bodies. Consequently, minor variations in this conserved set can lead to morphological innovations. The impact of gene loss on the evolution of developmental mechanisms has become an important topic in EvoDevo, especially since genomics revealed a high prevalence of gene loss across the tree of life. Over the past decade, the appendicularian *Oikopleura dioica* has emerged as a key model for studying this phenomenon in chordates. Despite a drastic reduction in its genome size and the loss of many developmental genes, *O. dioica* retains a typical chordate body plan, offering a unique opportunity to investigate the organization of gene regulatory networks (GRNs) in a biological system heavily affected by gene loss.

Previous research has shown that *O. dioica* has experienced significant losses in developmental signalling pathways, including the complete dismantling of the retinoic acid (RA) signalling machinery and a major reduction in the number of Wnt families. The extensive interactions between these pathways and the Fibroblast Growth Factor (FGF) signalling pathway in other chordates raise an interest in investigating the evolution of FGF signalling in this species. This study characterizes the FGF signalling pathway in *O. dioica*, revealing an unprecedented remodelling compared to other chordates. The species has lost six of the eight Fgf subfamilies present in the last common chordate ancestor, but the remaining two subfamilies have expanded along with the *FgfR* gene.

Phylogenetic analyses, combined with comparative studies on gene structure and conserved protein motifs, provide robust evidence that the 10 *Fgf* genes identified in *O. dioica* belong exclusively to two subfamilies: Fgf9/16/20 and Fgf11/12/13/14. Analyses of gene structure, putative functional motifs, and developmental expression patterns indicate functional diversification of these paralogs, as well as of the *FgfR* gene, following gene expansion. Additionally, examination of the three main intracellular transduction pathways associated with FgfR activation—MAPK, PLCγ/PKC, and PI3K/AKT—reveals that their main components are preserved and expressed throughout *O. dioica* development. However, structural rearrangements in FGF signal transduction have occurred, including the loss of classical *Ras* and *Spred* genes. Functional studies using the pharmacological inhibitor SU5402 demonstrate that the FGF signalling is crucial for embryonic development in *O. dioica.* Phenotypic analyses via whole-mount in situ hybridization of tissue-specific genes and *omic* approaches reveal that FGF signalling is involved in

gastrulation and cellular lineage differentiation. Finally, by comparing the developmental expression domains of *O. dioica Fgf* genes with those in the ascidian *Ciona robusta*, an evolutionary scenario is proposed where the evolution of FGF signalling in *O. dioica* is related to the transition from an ascidian-like biphasic lifestyle to a fully free-living one. This scenario categorizes expression domain changes into three types: extinction of ancestral domains due to gene loss, functional shuffling among surviving paralogs, and innovation of novel expression domains in new paralogs.

Overall, this doctoral thesis presents a comprehensive study of the FGF signalling pathway in a chordate with unprecedented levels of developmental gene loss. Our findings unveil the evolution of the Fgf family in appendicularians as a paradigmatic example of what we call “less, but more”, where massive gene losses, but also extensive duplications, result in the loss, conservation, and innovation of Fgf expression domains. This research provides new insights into the flexibility and resilience of the FGF signalling pathway in chordates and raises questions about the evolutionary significance of the Fgf9/16/20 and Fgf11/12/13/14 subfamilies. These findings contribute to a broader understanding of gene loss in EvoDevo, highlighting the complex interplay between genetic conservation and innovation.

INDEX

ABBREVIATIONS

AP	Antero-Posterior
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
ATFS	Anterior Tail FGF Source
ATRAS	Anterior Tail Retinoic Acid Source
BAR	Barcelona (O. dioica cryptic species)
Bra	Brachyury
CNS	Central Nervous System
DEG	Differentially Expressed Gene
DGE	Differential Gene Expression
DMSO	Dimethyl sulfoxide
EH	Early hatchling (larvae)
eKO	Evolutionary Knockout
ETB	Early tailbud (embryo)
Fgf	Fibroblast Growth Factor
FgfR	Fibroblast Growth Factor Receptor
GRN	Gene Regulatory Networks
HBS	Heparin Binding Site
HMM	Hidden Markov Models
hpf	hours post-fertilization
Ig	Immunoglobulin
ITB	Incipient tailbud (embryo)
JH	Just hatchling (larvae)
LH	Late hatchling (larvae)
LTB	Late tailbud (embryo)
MH	Mid hatchling (larvae)
mpf	minutes post-fertilization
MTB	Mid tailbud (embryo)
mya	million years ago
NLS	Nuclear Localization Signal
NOR	Norway (O. dioica cryptic species)
OKI	Okinawa ((O. dioica cryptic species))
OSA	Osaka (O. dioica cryptic species)
PTFS	Posterior Tail FGF Source
RA	Retinoic Acid
RBBH	Reciprocal Best Blast Hit
RNA-seq	RNA sequencing
RT	Room Temperature

RTK	Receptor Tyrosine Kinase
SP	Signal Peptide
SSW	Sterilized Sea Water
TK	Tyrosine Kinase
VEGFR	Vascular Endothelial Growth Factor Receptor
WMISH	Whole Mount In Situ Hybridisation

INTRODUCTION

1. ***Oikopleura dioica, an EvoDevo model to study gene loss***
2. ***The Fibroblast Growth Factor signalling pathway as a case study***

1. *Oikopleura dioica*, an EvoDevo model to study gene loss

1.1. EvoDevo

EvoDevo, an abbreviation for "evolutionary and developmental biology," emerges as a scientific discipline at the intersection of evolutionary biology, developmental biology, molecular genetics and genomics. Its framework is based on the notion that the vast morphological diversity observed throughout the tree of life, and especially within the animal kingdom, arises from a limited number of evolutionarily conserved genetic systems. These genetic systems function during embryonic development orchestrating and building the body plan of animals, so minor variations in the system may result in significant morphological innovations. This assumption is based on the alleged "paradox in EvoDevo", which highlights the remarkable fact that a relatively small and conserved developmental genetic toolkit may be used in myriad ways to produce the extensive diversity of life forms. The paradox underscores the importance of gene regulation, expression patterns, and the evolutionary modifications of developmental processes in generating biological diversity (Ferrández-Roldán et al., 2019; Jacob, 1977).

EvoDevo aims to connect the origin and evolution of phylogenetic groups and their distinct morphological traits with concrete alterations in the genetic mechanisms underlying development. Consequently, although the lens of EvoDevo seeks to elucidate evolution in an especially tangible manner, its nature as an integrative science endows it with a remarkable multidisciplinary character, encompassing diverse areas of interest and drawing perspectives from many fields of biology. This PhD thesis focuses on the role of gene loss as an evolutionary force and the extent to which genetic variation can occur without causing major alterations in an organism, using the FGF signalling pathway and the appendicularians tunicate species *Oikopleura dioica* as a case study.

1.2. Conservation and diversity of intercellular signalling mechanisms in metazoan development

EvoDevo focuses on the evolution of developmental mechanisms in organisms that undergo development, which involves transforming from a single cell to a complex multicellular organism through coordinated processes of cell division, differentiation, and morphogenesis. Cellular communication is essential for these developmental processes and, although it most likely emerged with the first living cells, it became increasingly sophisticated in parallel to the evolution of multicellular organisms. The sophistication of cellular communication mechanisms like the complex gene regulatory

networks (GRNs) that guide embryonic development (Babonis & Martindale, 2017; Combarnous & Nguyen, 2020; Gerhart & Kirschner, 1997).

The concept of cellular communication likely dates to the discovery of cells themselves. However, it was not until the early 19th century when Spemann and Mangold provided the first empirical evidence of cellular communication in embryonic development (Bouwmeester, 2001). In 1954, Rita Levi-Montalcini marked the discovery of the first growth factor with the molecular description of the NGF (Nerve-Growth-Factor), and over the following decades the list of secreted proteins, receptors, and mechanisms for cellular communication expanded (Aloe, 2004; Cohen et al., 1954). By the beginning of the genomic era, it was already clear that a limited number of signalling pathways were similarly used across diverse and distant species (Gerhart, 1999). One of the most surprising findings is that only seven classes of highly conserved signalling pathways are responsible for most of animal development, as they have been used repeatedly through the development of organisms and in the evolution of metazoans, activating different subsets of target genes in different developmental contexts. The seven alleged pathways are Transforming-growth-factor-β (TGF-β), canonical Wingless related (Wnt), nuclear receptors (NRs), Notch-Delta, Hedgehog (Hh), receptor tyrosine-kinases (RTK), and Janus kinase/Signal transducer and activator of transcription (JAK/STAT) **(Figure I1)**. They are strikingly diverse in both their complexity and in the biochemical mechanism employed for the transduction of the extracellular signal inside the cells, ranging from direct transcriptional regulation by the NRs to the extensive protein phosphorylation cascades characteristic of RTK pathways. Nonetheless, their primary outcome is the same: activation of specific target genes by signal-regulated transcription factors (Barolo & Posakony, 2002; Pires-daSilva & Sommer, 2003).

The sequencing of early branching metazoans and their unicellular relatives revealed that key components of the seven pathways existed in unicellular eukaryotes, but their full functionality and role in embryonic development evolved within the metazoan lineage, enabling complex body plans (Babonis & Martindale, 2017). During metazoan embryonic development, these seven pathways transduce extracellular signals into transcriptional regulation of target genes and are responsible of coordinating the GRNs that drive the specification of cells from a proliferative progenitor. While their core elements are conserved, the manner in which these pathways are used and regulated have diversified. The patterns of conservation and diversity of mechanisms of cell signalling in eumetazoans reflect the intricate balance between preserving essential cellular functions and allowing for evolutionary innovation. Understanding these processes provides valuable insights into the evolutionary history of life and the molecular basis of development, since their diversification has been crucial for the evolution of complex body plans and specialized functions observed in different animal lineages.

In the clade of vertebrates, the two rounds of whole genome duplication resulted in an expansion of developmental genes that were retained more frequently than non-developmental gene duplicates (Putnam et al., 2008). The expansion of developmental genes including signalling pathways led to the advent of new GRNs and to the rewiring of the existing ones, what likely favoured the appearance of novel features that accounted for the diversification and evolutionary success of vertebrates (Gil-Gálvez et al., 2022; Putnam et al., 2008). However, this also caused an entangling of the GRNs often hindering the study of the genetic mechanisms underlying developmental processes. Thus, the use of closely related animal models with smaller and simpler genomes, such as those from amphioxus or sea squirts that belong to the chordate phylum, has been essential for understanding the bases and the evolution of GRNs governing chordate development.

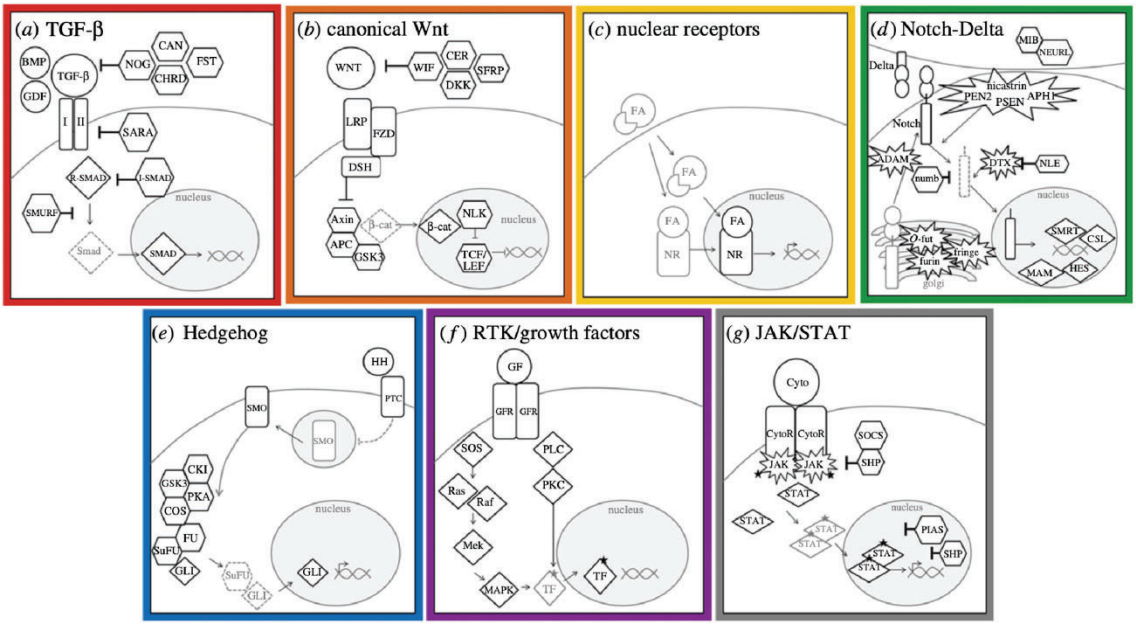


Figure I1. Schematic representation of the major components of the seven highly conserved metazoan signalling pathways. Although the molecular mechanisms employed by each of the seven pathways differ, their ultimate function is the same: to receive extracellular signals and trigger an intracellular response that ultimately results in the transcriptional regulation of transcription factors. Adapted from Babonis & Martindale (2017).

1.3. Tunicates in EvoDevo

Tunicates are, along with vertebrates and cephalochordates, one of the three subphyla that comprises the phylum Chordata. Tunicates are subdivided into three classes: Ascidiacea, also known as sea squirts; Thaliacea, also known as salps; and Appendicularia, also known as larvaceans. All of them are marine filter-feeder animals, but each class exhibits distinctive lifestyles and reproductive strategies. The study of tunicates has been outstanding in the field of evolutionary and developmental biology since the beginning of the last century, and ascidians like *Ciona robusta* (formerly known as *Ciona* intestinalis type A) have been extensively used as animal models for

understanding the GRNs governing chordate development (Cañestro et al., 2003; Stolfi & Christiaen, 2012).

The study of tunicates bridges the gap between invertebrate and vertebrate biology and sheds light on the evolutionary history of chordates for several reasons (Holland, 2016; Lemaire, 2011). First, their phylogenetic position as the sister group of vertebrates offers insights into the last common ancestor of olfactores (the clade comprising tunicates and vertebrates) and the evolutionary origins of vertebrates (Delsuc et al., 2006). Second, most tunicates, particularly in their tadpole-like larval phase, retain the basic chordate body plan and exhibit key morphological and biological features with a great equilibrium of simplicity and complexity, what provides critical insights into the evolutionary pathways that led to more complex organisms within the chordate phylum (Satoh, 2003). Third, their simple and transparent embryos coupled with their short life cycles, the small size of their genomes, and the availability of several techniques to perform genetic manipulation, make them ideal models for experimental and computational approaches (Pennati et al., 2024). Lastly, from a developmental biology perspective, tunicates are invaluable due to their determinate development, where cell fates are fixed early in embryogenesis. This characteristic provides a simplified model to study fundamental processes of cell differentiation and tissue formation (Kumano & Nishida, 2007; Satou, 2020).

The fact that ascidian embryos exhibit a determinative development and invariant cell division patterns has allowed for the precise naming and tracing of their cellular lineages throughout embryogenesis, revealing that most cells have a unique developmental fate by the 112-cell stage **(Figure I2)** (Nishida, 1987; Satou, 2020). The ascidian embryo follows a “mosaic” mode of development, where localized maternal factors are differentially distributed in the first cell divisions and end up setting five distinct developmental territories in the 16-cells embryo. When zygotic expression starts at this stage, each territory displays a different expression profile according to the combination of maternal factors it carried. Subsequently, each blastomere follows a developmental trajectory that depends on its initial expression profile and on its interactions with other blastomeres in the embryo. The output of these GRNs results in fate-specific expression profiles in individual cells of the 112-cell embryo **(Figure I2)**. This simple and highly organized system enables the study of cell interactions and gene expression patterns at a single cell resolution, what provides insights into gene regulation and cell fate determination in early development, highlighting the ascidian embryo as an exceptional model for studying chordate developmental GRNs (Satou, 2020). Despite their simpler and more compact genomes, ascidians share many fundamental developmental processes and gene regulatory mechanisms with vertebrates, providing a clearer and more manageable model to uncover essential principles of gene regulation and cell fate determination (Satou, 2020).

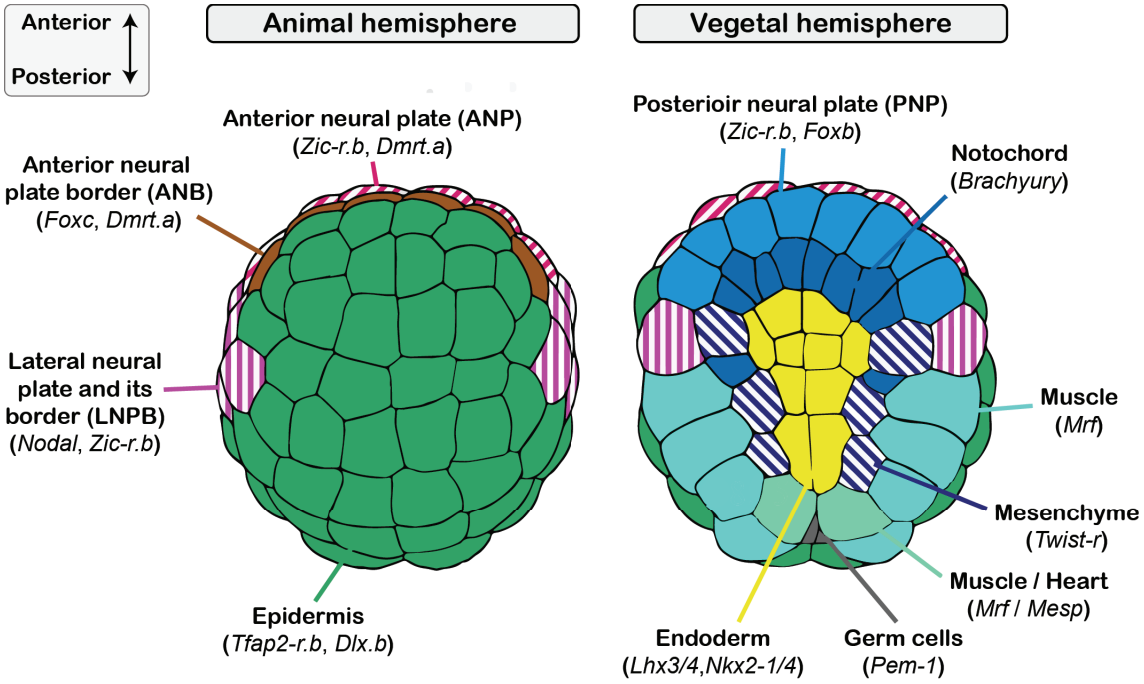


Figure I2. Schematic representation of the ascidian 112-cell embryo. Cells with different developmental fates are marked with different patterns and colours. Representative genes expressed in specific lineages are indicated in parentheses. Adapted from Satou (2020).

Apart from their invaluable role as model organisms for chordate development, and as a proxy to understand vertebrate embryogenesis, tunicates also offer unique and essential insights into animal evolution. This keeps them at the forefront of biological research independently of their evolutionary proximity to vertebrates (Lemaire & Piette, 2015). One of the key aspects that make tunicates especially interesting is their rapid evolutionary rate compared to their closest relatives. The rapid evolution of tunicate genomes, with high rates of gene loss, rearrangements, and divergence in non-coding regulatory elements, offers a unique perspective on how genomic and regulatory network changes can occur without disrupting essential developmental processes. This contrasts with the more conserved genomes of vertebrates and cephalochordates, providing a complementary viewpoint on the stability and plasticity of developmental programs (Holland, 2016; Lemaire, 2011). Tunicates in general are characterized by their "liberal" pattern of genome evolution, which has been revealed mainly by extensive studies on the Ascidiacea class (Somorjai et al., 2018). However, within tunicates, the class of appendicularians has likely undergone the most drastic alteration of genome-organization features accompanied by genome compaction and prevalent gene losses, as revealed by the genomic sequencing of *O. dioica* (Denoeud et al., 2010; Ferrández-Roldán et al., 2019).

1.4. *Oikopleura dioica*, a model organism for the study of gene loss

Oikopleura dioica is the most extensively studied species within the class Appendicularia, commonly referred to as larvaceans due to their retention of a larval-like form throughout their entire lifespan. This endows appendicularians with a fully free-living lifestyle and distinguishes them from ascidians, which undergo a drastic metamorphosis before reaching the adult stage and lose the typical chordate body plan to adopt a sessile lifestyle.

1.4.1. The unique biological and genomic features of *O. dioica*

O. dioica is a semi cosmopolitan planktonic organism found in all oceans except the Arctic and Antarctic. Like other appendicularians, *O. dioica* feeds on phytoplankton obtained through filtration using a mucous house secreted by its specialised trunk epidermis, known as the oikoblast **(Figure I3A)** (Ferrández-Roldán et al., 2019). Under normal conditions, its adult body, which reaches 3-4 millimetres in length, is embedded inside this mucous house and is divided into a trunk and a tail. The trunk contains most organs, while the tail includes the notochord, a dorsal nerve cord and a row of muscle cells on each side **(Figure I3B)**. Notably, the tail of appendicularians exhibits a remarkable 90° counterclockwise rotation relative to the trunk when viewed from the posterior, influencing the direction of the tail beat. In appendicularians, the tail beats in a dorsal-ventral direction, whereas in the rest of chordates the tail usually beats in a left–right direction (Nishida, 2008). The tail is also flanked by lateral fins that facilitate swimming when the animal is outside the house, and the generation of water microcurrents when the animal is inside the house due to the constant beating of the tail. These microcurrents enhance the water flow through the house and the retention of phytoplankton, thus playing a crucial role in the filter-feeding mechanism (Ferrández-Roldán et al., 2019; Troedsson et al., 2009).

Regarding its reproduction and life cycle, *O. dioica* is a gonochoric and semelparous species with external fertilization and high fecundity, with females producing up to 300 eggs in a single spawning event (Ferrández-Roldán et al., 2019). Its life cycle is very short and spans only 5 days at 19°C. Embryonic development is determinate, rapid, and invariant, and cellular lineages have been traced during embryogenesis (Fujii et al., 2008; Stach et al., 2008). Gastrulation begins at the 32-cells stage, and the larva hatches from the chorion at 3 hours and 30 minutes post-fertilization. During the larval phase the animals elongate their body and conclude organogenesis and culminates at 6 hours post-hatching, when the animals undergo a subtle metamorphosis in which the tail angle changes relative to the body **(Figure I3C)**.

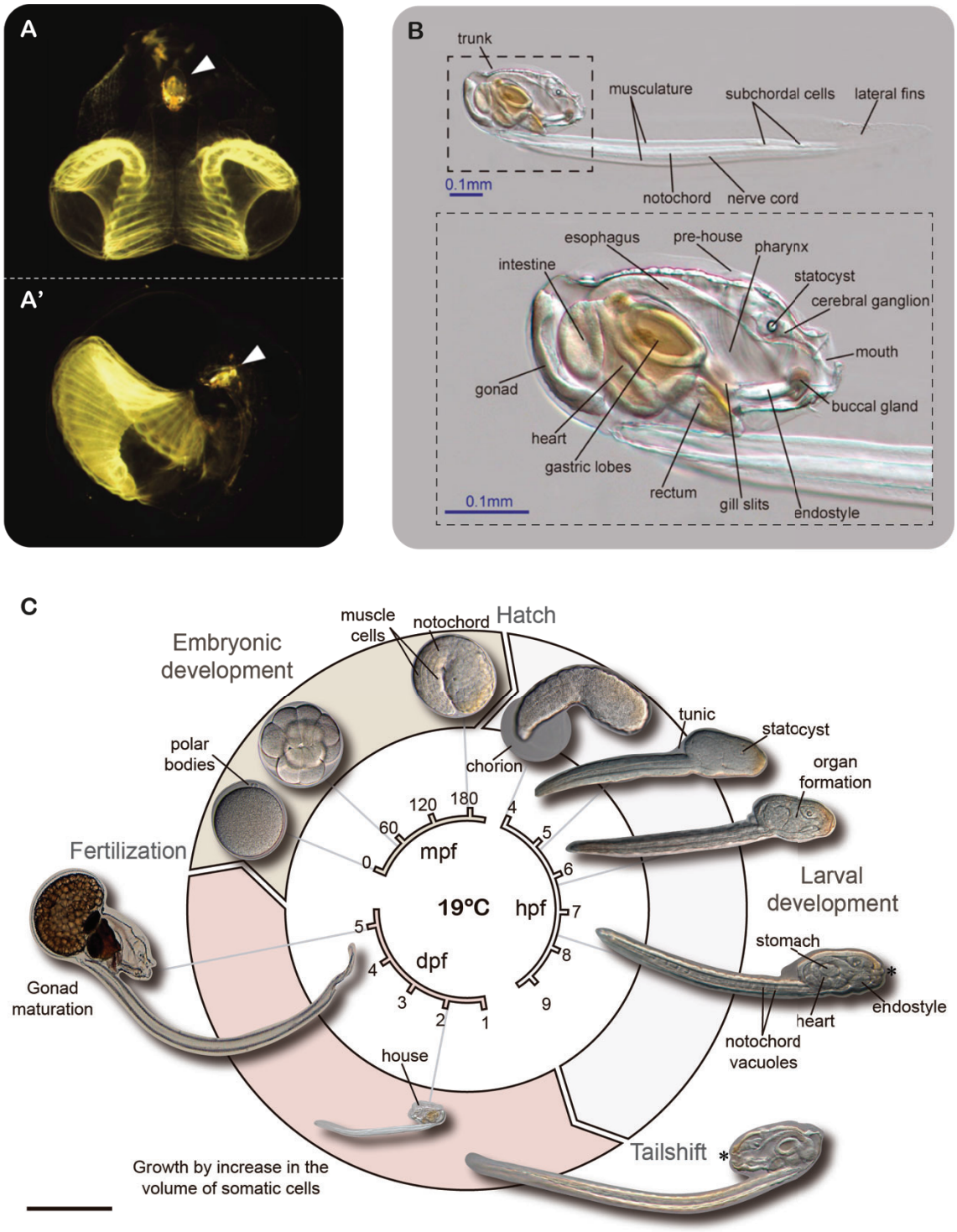


Figure I3. Biology and life cycle of *Oikopleura dioica*. (A) Frontal and (A') lateral views of the *O. dioica* inside the mucous house, made visible by trapped algae. White arrowheads indicate the trunk of the animal inside the house. **(B)** Immature adult specimen of *O. dioica*. Differential interference contrast (DIC) micrograph in right lateral view, with dorsal at the top. The magnified trunk displays the main organs and structures. **(C)** Schematic representation of the O. dioica life cycle at 19°C. Embryonic development begins with the fertilization of the oocyte and concludes when the larva hatches by breaking the chorion. Larval development lasts approximately 6 hours, ending with a 120º reorientation of the tail in a process called tail shift. Note the position change of the mouth (asterisk) relative to the tail in juvenile animals. Over the next 4.5 days, juvenile animals feed, grow, and mature into males or females, which then spawn gametes to complete the cycle. Scale bar represents 100 µm for all stages except for day 2 and mature adults (1 mm). All figures adapted from Ferrández-Roldán et al. (2019).

O. dioica possesses the smallest and most compact genome ever described in a chordate, with a haploid size of 70 Mb and a gene density of 1 gene per 3.9 kb (Denoeud et al., 2010; Ferrández-Roldán et al., 2019; Seo et al., 2001). Furthermore, it is characterized by the scarcity of transposable elements and the minute size of its introns and intergenic spaces. This process of compaction has been accompanied by a high number of gene losses, many of which affect genes considered essential in all other chordates. Examples include the complete DNA-repair machinery of cNHEJ (canonical Non-Homologous-End-Joining) (Denoeud et al., 2010), or all genes necessary for peroxisome biosynthesis and function (Žárský & Tachezy, 2015). This peculiar feature has placed *O. dioica* as an outstanding model organism for studying the evolutionary impact of gene loss. The study of *O. dioica* and its genomic features stands as a gateway to uncover the mechanisms underlying the loss of genes. This will help to understand how gene losses can affect the evolution of developmental mechanisms in chordates and contribute to the diversity among taxa. Furthermore, *O. dioica* offers several practical advantages: it has a straightforward and accessible morphology, a brief generation time and life span, is affordable to culture in the laboratory, and is amenable to experimental manipulation (Albalat & Cañestro, 2016; Ferrández-Roldán et al., 2019; Martí-Solans et al., 2015).

1.4.2. Evolution by gene loss

From a traditional evolutionary perspective, an increase in complexity and genetic variability among organisms has been linked to the expansion of their genetic material, which serves as a substrate for mutations. Gene duplication events have garnered the most attention because they can lead to neofunctionalization or subfunctionalization of the resulting copies through random mutation (Ohno, 1970). Conversely, gene loss, defined as the functional loss of genes regardless of the cause, is usually associated with the absence of selective pressure on a sequence, leading to its degeneration without functional consequences (Albalat & Cañestro, 2016). However, the genomic sequencing of the cnidarian *Nematostella vectensis* in 2007 revealed that the common ancestor of eumetazoans already possessed a complex genome with most of the current gene families (Putnam et al., 2007). This discovery refuted the preconceived notion that an increase in phenotypic complexity was linked to an increase in the number of genes. Instead, it highlighted gene loss as a mechanism to explain the diversification of bilaterians. Tracing the effects of gene loss in the phylogenetic tree of metazoans reveals that gene loss is not only ubiquitous, but also it has differently impacted various evolutionary branches in terms of both the number and nature of the lost genes **(Figure I4)**. The loss of gene families relative to the ancestral eumetazoan has been much greater in protostomes than in deuterostomes (Fernández & Gabaldón, 2020; Guijarro-Clarke et al., 2020). However, despite the weak tendency of deuterostomes to lose genes, the group of tunicates to which *O. dioica* belongs stands as a paradigmatic case of evolution

through gene loss, exhibiting similar rates to those found in ecdysozoans **(Figure I4)** (Albalat & Cañestro, 2016).

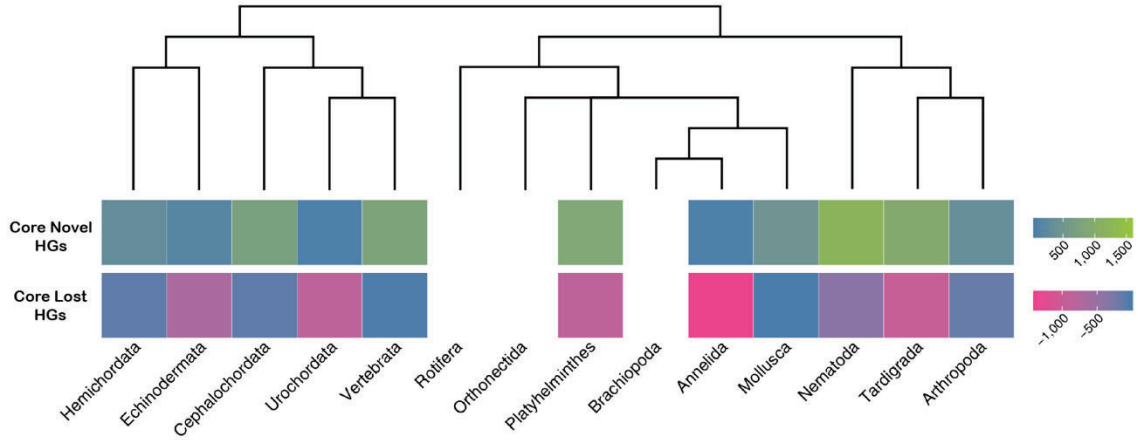


Figure I4. Levels of gene gains and losses at the phylum level. The heat map, normalized by row, illustrates the levels of gene gains and losses across different animal phyla as reported by Guijarro-Clarke et al. (2020). Gene gains are represented in green for the highest numbers and blue for lower numbers, while gene losses are shown in pink for the highest numbers and blue for lower numbers. "Core Novel HGs" refer to homology groups (HGs) that originated in the stem lineage or last common ancestor of the ingroup and are present in 95% of the ingroup taxa. "Core Lost HGs" are HGs that were lost in the stem lineage of a clade before the last common ancestor and are present in 95% of the outgroup taxa. Notably, tunicates (referred to here as Urochordata) stand out within deuterostomes as the clade with the highest number of Core Lost HGs. Adapted from Guijarro-Clarke et al. (2020).

Regarding gene loss, a prerequisite for its occurrence is that the genes are dispensable, meaning their loss does not impede the development or reproduction of the organism. The "knockout paradox" suggests that most genes in an organism are dispensable, what can be explained by the mechanisms of mutational robustness and environment-dependent dispensability (Papp et al., 2011). Thus, the key question is how the loss of these genes can serve as an evolutionary driver. Depending on whether gene losses affect an individual's fitness, they can be considered adaptive or neutral, with their likelihood of fixation being primarily influenced by natural selection or genetic drift, respectively. Numerous examples of fixed gene losses exist, both due to their adaptability, such as the case of the CCR5 leukocyte receptor in regions where AIDS is endemic, or due to their neutrality, as seen in regressive evolution modelled by parasitic and endosymbiotic species (Albalat & Cañestro, 2016). However, while most described examples are understood as cases of microevolution, EvoDevo places the interest of gene losses within the context of the macroevolution of species. When the loss of one or several genes becomes fixed in a population, the evolution and diversification of its descendants will be conditioned by this loss. The consequences are particularly significant when the affected genes have roles in development, as their loss can impact the body plan and affect the GRNs in which it was involved, thereby causing divergence in developmental mechanisms. Although the dispensability of developmental genes is

very limited, their loss in the evolution of species can be explained by three main reasons: (i) loss-of-function mutations are the most frequent and can affect any functional element in various ways, (ii) the dispensability of genes varies over time, influenced by both the environment and the extent to which their function is integrated into development, and (iii) genetic drift allows for the fixation of mutations even if they have a slight negative effect on fitness.

1.4.3. Gene losses affecting developmental functions and signalling pathways in *O. dioica*

From an EvoDevo perspective, the most intriguing gene losses in *O. dioica* involve developmental genes that are highly conserved across the animal kingdom. Notably, among the homeobox genes, which encode homeodomain transcription factors typically involved in developmental processes, *O. dioica* retains only 55 out of the 81 homeobox gene groups thought to be present in the ancestral chordate, representing a loss of more than 30% of these gene groups (Edvardsen et al., 2005). Among the homeobox genes that the *O. dioica* genome lacks are all the central *Hox* genes (*Hox3*, *Hox5*, *Hox6*, *Hox7* and *Hox8*), *Goosecoid*, *Nk3* and *Nk6*, *Vax,* and *Prox,* to name a few (Edvardsen et al., 2005; H.-C. Seo et al., 2004). In addition to homeobox genes, other groups of conserved developmental genes have experienced significant losses in the *O. dioica* lineage. These include the *Sox* genes (Heenan et al., 2016), notochord genes (Kugler et al., 2011), cyclins and cyclin-dependent-kinases (Campsteijn et al., 2012) or miRNAs (K. Wang et al., 2017), as reviewed in (Ferrández-Roldán et al., 2019).

Regarding signalling pathways, the most notable example of loss is the retinoic acid (RA) pathway, that belongs to the class of nuclear receptors (NRs) described in section 1.2. RA, a derivative of vitamin A, acts as a morphogen in determining the anteroposterior and dorsoventral body axes and in chordate organogenesis (Cañestro et al., 2006; Cunningham & Duester, 2015). *O. dioica* has lost all the enzymatic machinery necessary for RA biosynthesis, rendering its development independent of retinoic acid (Cañestro et al., 2006; Martí-Solans et al., 2016). Nodal, which is a member of the class of TGF-β signalling pathways and leads the left-right asymmetry in deuterostomes and lophotrochozoans, has also gone extinct in the genome of *O. dioica* (Onuma et al., 2020). Other well-documented case is the Wnt signalling pathway. In this case, *O. dioica* has not experienced a complete loss of this pathway, but it has undergone an extreme reduction in its components. It has retained only four of the thirteen Wnt subfamilies, becoming the chordate with the minimal Wnt repertoire known to date (Martí-Solans et al., 2021).

1.4.4. The EvoDevo paradox and the concept of evolutionary knockout (eKO)

The extensive loss of developmental genes in *O. dioica* led to the formulation of the “Inverse Paradox in EvoDevo”**,** which describes the phenomenon where organisms

with fundamentally similar morphologies exhibit significant differences in their genetic toolkits **(Figure I5)** (Cañestro et al., 2007). Many of the genes that *O. dioica* has lost are essential components of GRNs that drive the development of specific structures in other chordates. While the loss of these genes has sometimes resulted in the reduction or loss of these anatomical structures, the prevailing trend is that *O. dioica* retains the structures despite the disruptions in the conserved GRNs. The continued development of these structures suggests that the GRNs coordinating their development have undergone significant rewiring in *O. dioica*. This characteristic has established *O. dioica* as a paradigmatic evolutionary knockout (eKO) model. An eKO model refers to species that have naturally lost specific genes, offering a unique opportunity to study the impacts of these losses on their development and evolution. eKOs allow to observe the consequences of gene absence in a natural context, providing unique insights into gene function, genetic redundancy, and the evolutionary mechanisms driving phenotypic diversity. By studying organisms like *O.* dioica, which has lost numerous genes related to key developmental pathways, we can gain a better understanding of how gene loss influences evolutionary adaptations and the development of complex traits (Ferrández-Roldán, 2021).

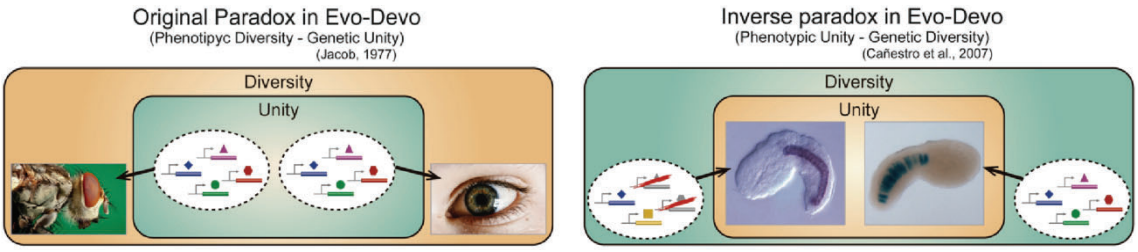


Figure I5. Th inverse paradox in EvoDevo. Contrary to the original paradox, the inverse paradox in EvoDevo suggests that organisms may develop fundamentally similar morphologies (phenotypic unity) despite significant differences in their genetic toolkits (genetic diversity). Adapted from Ferrández-Roldán et al. (2019).

The rewiring of GRNs upon the loss of one of its components usually leaves a trace at genomic level. The loss of a gene does not come as an isolated event, but it can be accompanied by the co-elimination of other genes that are functionally linked to a distinctive pathway (Albalat & Cañestro, 2016). Gene losses can also be accompanied by the duplication of paralogs or other functionally related genes that may increase the robustness of the genetic system and lead to processes of function shuffling, facilitating therefore the losses (Cañestro et al., 2009; McClintock et al., 2001). To understand the impact of the loss of essential genes, such as those governing embryo development, it is necessary to study how events of gene co-elimination and duplication correlate with the loss or survival of ancestral characters still present in sister groups. The study of specific cases is therefore needed to infer in what circumstances gene losses are not detrimental, or how they can be even adaptive.

The dismantling of the entire RA pathway Is a clear example of gene co-elimination in which all the genes involved in the synthesis of RA have been lost in *O. dioica* (Martí-Solans et al., 2016). Given the numerous interactions of the RA pathway with the FGF (Fibroblast-Growth-Factor) signalling pathway in establishing the body plan and the development of certain organs in olfactores, an interest arises in studying the evolution of the FGF signalling in an eKO model for the RA (Diez del Corral et al., 2003; Pasini et al., 2012).

2. The Fibroblast-Growth-Factor (FGF) signalling pathway as a case study

Fibroblast growth factors (Fgfs) and their receptors (FgfRs) form a system of signalling proteins that emerged concomitantly with the origin of Eumetazoans and have been extensively maintained throughout animal evolution (S. Bertrand et al., 2014). These proteins regulate a plethora of important biological processes such as cell proliferation, migration, or differentiation during embryonic development and adult tissue homeostasis (Dorey & Amaya, 2010; Itoh & Ornitz, 2011; Ornitz & Itoh, 2015; Teven et al., 2014). Among the seven highly conserved signalling pathways that are repeatedly used in animal development, the FGF signalling pathway belongs to the class of receptor-tyrosine kinases (RTK) (see section 1.2). The functional mechanism of the FGF signalling, like other RTK-based signalling pathways, involves the binding of an extracellular ligand to a transmembrane receptor (the RTK) on the cell surface. This binding activates the receptor, initiating phosphorylation cascades inside the cells through the receptor's intracellular tyrosine kinase (TK) domain. These cascades trigger one or more intracellular transduction pathways, ultimately leading to the regulation of gene expression (Pires-daSilva & Sommer, 2003).

2.1. Components of the FGF signalling pathway

2.1.1. The Fgf ligands

Fibroblast growth factors (FGFs) are generally small proteins characterized by a conserved FGF core homology domain of 120-130 amino acids arranged in a β-trefoil topology (Plotnikov et al., 2001). The FGF domain contains essential motifs for binding to both heparin or heparan sulphate in the extracellular matrix and to the extracellular portion of the Fgf receptors (FgfR) on the cell surface. Besides the FGF domain, sequences are not conserved among different Fgf subfamilies. Some *Fgf* genes have independently evolved extended N- or C-terminal regions of variable length that are not homologous, and consequently protein alignments among distant Fgf subfamilies are not possible outside the FGF domain (Popovici et al., 2005). These extended regions often include functional motifs such as a signal peptide (SP) or nuclear localization signals (NLS) (Coulier et al., 1997). The canonical SP, important for the secretion of the proteins from the cell, is a short hydrophobic region including the first 15-30 residues of a protein at its N-terminus (Owji et al., 2018). In principle, Fgf proteins that lack a canonical SP remain intracellular, although alternative secretion mechanisms have been described for some Fgfs lacking a canonical SP, what allows them to be released from the cell and perform signalling functions (Kirov et al., 2012; Miyakawa & Imamura, 2003; Revest et al., 2000; Schäfer et al., 2004).

2.1.2. The Fgf receptors

Fgf receptors (FgfRs) are generally large proteins of around 800 amino acids that consist of an extracellular ligand-binding portion with various immunoglobulin (Ig)-like domains, a transmembrane domain, and a split intracellular tyrosine kinase (TK) domain (Itoh & Ornitz, 2004). The prototypical extracellular portion of an FgfR contains a canonical SP, three Ig-like domains, a stretch of acidic residues (acidic box) between Ig-like domains I and II, and a heparin binding domain between Ig-like domains II and III (Itoh & Ornitz, 2004; Ornitz, 2000; Ornitz & Itoh, 2015). Ig-like domains II and III, along with the linker region between them, regulates the ligand binding specificity of the FgfR, while Ig-like domain I and the acidic box are thought to inhibit ligand binding, so alternative splicing of the FgfR transcripts results in different isoforms with different ligand binding properties (Ornitz & Itoh, 2015, 2022; Tulin & Stathopoulos, 2010).

FgfRs are embedded in cellular membranes through their transmembrane helices and tend to form active dimers, although their dimerization is inefficient in the absence of Fgf ligands (Sarabipour & Hristova, 2016; Schlessinger, 2000). The binding of Fgf ligands to the extracellular portion of the FgfRs, along with their interaction with heparan sulphate, allows the formation of a stable dimeric ternary complex, which is essential for the proper activation of the FgfRs (Sarabipour & Hristova, 2016; Schlessinger et al., 2000). The ligand-binding-induced activation is based in an allosteric change that stabilizes the FgfR dimer and brings the intracellular TK domains into the correct proximity and orientation, favouring transphosphorylation and thus activation of the kinases (Chen et al., 2008; Furdui et al., 2006; Sarabipour & Hristova, 2016). The activation of the intracellular TK domain of FgfR, induced by ligand binding to its extracellular portion, enables the receptor to trigger one or several intracellular transduction pathways (Goetz & Mohammadi, 2013).

2.1.3. The RTK transduction pathways

Following ligand binding and activation, the TK domain of the FgfR phosphorylates intracellular substrates. The first targets are the receptors themselves, resulting in numerous autophosphorylated tyrosine residues throughout the cytoplasmic region of the receptor. These phosphotyrosines function as specific recognition sites for the recruitment of downstream factors, leading to the assembly of a signal transduction complex. This makes the activated RTK a hub in a signalling network that transmits information from the exterior to the interior of the cell (De Meyts, 2015; Lemmon & Schlessinger, 2010). The possible outcomes of these intracellular signalling complexes are extremely diverse, and the complete understanding of their functioning remains elusive due to the large number of factors involved, the complexity of their interactions, and the many case-dependent variations in their mechanisms. However, three intracellular signalling pathways are considered to be the major responsive

cascades to FgfR activation: (1) the mitogen-activated protein kinase (MAPK) pathway, (2) the phosphatidylinositol-3-kinase (PI3K)/AKT pathway, and (3) the phospholipase C gamma (PLCγ)/protein kinase C (PKC) pathway (S. Bertrand et al., 2009; Goetz & Mohammadi, 2013).

The activation of the MAPK and PI3K/AKT pathways both start with the recruitment of the Fgf-receptor-substrate-2 (FRS2), which subsequently recruits the GRB2/SHP2/GAB1/SOS complex to the cytoplasmic membrane (S. Bertrand et al., 2009). The recruited SOS then activates the RAS GTPase, initiating the MAPK cascade that results in the activation of the ERK1/2 kinase (alias. MAPK), while the recruited GAB1 leads to PI3K-mediated activation of the AKT kinase (Goetz & Mohammadi, 2013). In the PLCγ/PKC pathway, PLCγ is directly phosphorylated and activated by the TK domain of the FgfR. PLCγ then catalyses the hydrolysis of membrane phosphatidylinositol 4,5-bisphosphate (PIP_2) into inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG). IP_3 activates calcium signalling, while DAG activates the PKC kinase (S. Bertrand et al., 2009). The activation of any of these pathways ultimately results in the translocation of active kinases or transcription factors into the nucleus, where they regulate the transcription of an array of target genes **(Figure I6).**

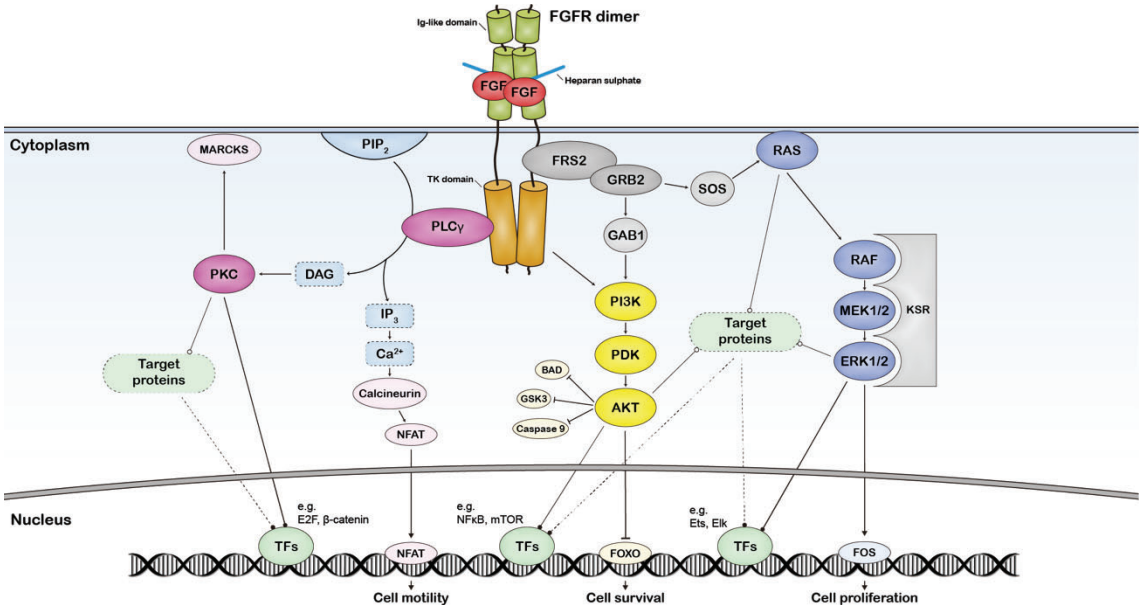


Figure I6. Intracellular transduction pathways associated with FGF signalling. The binding of FGF ligands to the extracellular portion of FGFRs, along with their interaction with heparan sulphate, results in the formation of a dimeric ternary complex, leading to the intracellular activation of the FGFR tyrosine kinase (TK) domains. This activation recruits docking proteins (shown in grey) to the active FGFR TK domain, initiating one or more intracellular transduction pathways. The three main transduction pathways activated by FGFR are: the MAPK pathway (in blue), the PI3K/AKT pathway (in yellow), and the PLCγ/PKC pathway (in magenta). Each of these pathways can lead to a variety of biological outcomes through the regulation of different transcription factors. The most commonly associated cellular outcomes for each pathway are depicted, with the involved proteins shaded in lighter versions of the pathway colours. Alternative outcomes mediated by different factors are shown in green.

The biological outcomes of these transduction pathways vary depending on the cellular context, hindering the association of each pathway with a specific function. Nevertheless, each pathway has been mostly associated with particular types of cellular responses (Goetz & Mohammadi, 2013). Thus, the MAPK pathway is primarily related to cell proliferation, but it can also lead to cell differentiation or cell migration (Böttcher & Niehrs, 2005; Thisse & Thisse, 2005; Turner & Grose, 2010); the PI3K/AKT pathway generally promotes cell survival by inhibiting pro-apoptotic transcription factors such as BAD or FOXO (Brunet et al., 1999; Datta et al., 1997); and the PLCγ/PKC pathway has specific roles in cell migration and cell differentiation, although it can also influence the MAPK and PI3K/AKT pathways **(Figure I6)** (Goetz & Mohammadi, 2013).

Although these general associations are well-established and supported by empirical evidence, several functions have been described for these pathways that do not necessarily align with these general roles. This ambiguity in the biological outcomes of the three pathways is related to the vast variety of transcription factors that are regulated by the effector kinases of each pathway (i.e. ERK1/2, AKT or PKC), either directly or through the activation or repression of cytoplasmic intermediaries (**Figure I6**) (Bai et al., 2009; Kazanietz & Cooke, 2024; Manning & Toker, 2017; Roskoski, 2012).

2.2. Functional mechanisms of the FGF signalling components

According to their functions, Fgf ligands have traditionally been categorized into three groups: (i) canonical Fgfs, which are secreted and bind to heparin and FgfRs, acting as paracrine and autocrine signalling factors by activating FgfRs in cells near the Fgf source; (ii) endocrine Fgfs, which are secreted and bind to the cofactor Klotho instead of heparin, functioning as long-distance signalling molecules in vertebrates; and (iii) intracellular non-secreted Fgfs, also known as Fgf-homologous factors (FHFs), which have intracrine non-signalling functions and serve as cofactors to other proteins (reviewed in Ornitz & Itoh (2015, 2022)).

Most of the described roles of the FGF signalling pathway are attributed to the canonical mechanisms involving the activation of FgfRs on the cell surface. However, there is a growing body of evidence suggesting intracellular functions and interacting partners for non-secreted Fgfs and FgfRs, and several studies have shown biological activity that is independent of FgfR TK activity (Goldfarb, 2001, 2005; Olsnes et al., 2003; Schoorlemmer & Goldfarb, 2002; Wu et al., 2012). Additionally, the presence of nuclear localization signals (NLS) allows some Fgfs and FgfRs to migrate into the nucleus (Antoine et al., 2005; Bryant & Stow, 2005; Du et al., 2015; Popovici et al., 2006; Sheng et al., 2005). Evidence indicates that nuclear FgfRs can directly regulate gene transcription and chromatin conformation through interaction with CREB-binding proteins (Decker et al., 2020; Fang et al., 2005). These nuclear FgfRs are involved in critical functions such as sex determination through transcriptional regulation of the *Sry* gene, or promotion of cell

proliferation through transcriptional activation of *cyclin D1* and *c-Jun* (Du et al., 2015; Reilly & Maher, 2001; Schmahl et al., 2004). Although the intracellular and intranuclear functions of Fgf and FgfR proteins, as well as their putative interaction inside the cells, are far from being fully understood, these findings exemplify the diverse and significant roles of the FGF signalling system (Ornitz & Itoh, 2022; Sluzalska et al., 2021; Stachowiak et al., 2015).

2.3. Origin and evolution of the FGF signalling pathway

2.3.1. Origin of the FGF signalling system

The high sequence variability among Fibroblast Growth Factors (FGFs), combined with the short length of their conserved core, hinders the phylogenetic classification of the *Fgf* genes. The first phylogenetic classification incorporating FGF sequences from several non-vertebrate species was published in 2005 and categorized all extant *Fgf* genes into eight subfamilies (Popovici et al., 2005). Although this classification remains widely accepted, the surge in sequenced genomes over the past two decades and the access to genomic data from a broader range of species has allowed to refine the evolutionary history of the FGF signalling system (Babonis & Martindale, 2017; Bertrand et al., 2011, 2014; Oulion et al., 2012). Current research indicates that bona fide *Fgf* genes most likely originated in the eumetazoan (cnidarian + bilaterians) ancestor from the duplication of an ancestral FGF-like domain, which can be still found outside eumetazoans and in some eumetazoan groups besides vertebrates (S. Bertrand et al., 2014a). Notably, at least two Fgf subfamilies had emerged before the divergence of cnidarians and bilaterians (Oulion et al., 2012). While the precise trajectory leading to the current gene subfamily content in each taxon remains to be elucidated, we know that the evolutionary history of the Fgf family has been marked by multiple events of gene expansion and gene loss affecting each lineage differently (Oulion et al., 2012; Popovici et al., 2005).

Regarding the Fgf receptors, the first bona fide FgfR also appeared in the last common ancestor of eumetazoans along with the Fgf ligands (S. Bertrand et al., 2014). Investigating its origin is challenging due to the extensive range of the Receptor Tyrosine Kinase (RTK) superfamily, which are present in unicellular relatives as distant as choanoflagellates and filastereans (Babonis & Martindale, 2017). However, evidence that the sponge *Oscarella carmela* has a TK domain orthologous to the one in eumetazoan FgfRs, but without Ig-like domains in its extracellular portion; and that the placozoan *Trichoplax adhaerens* possesses an RTK with an FgfR-like domain architecture, but a TK domain orthologous to the one of Cad96Ca; suggests that bona fide FgfR emerged by domain shuffling in the lineage leading to eumetazoans (S. Bertrand et al., 2014). Thus, the emergence of *FgfR* genes is characterized by the origin of the characteristic structure of various extracellular Ig-like domains attached to an

intracellular split TK domain in early metazoans. This was followed by clade-specific duplications and losses that resulted in varying numbers of *FgfR* copies, ranging from 1 to 5, across different animal lineages (Rebscher et al., 2009).

In summary, the FGF signalling system is likely an eumetazoan innovation, and both the Fgf ligands and the Fgf receptors appeared by the co-option and rearrangement of already existing genes and domains, highlighting the modular evolution of cell signalling pathways (Babonis & Martindale, 2017; Bertrand et al., 2014).

2.3.2. Evolution of the FGFs in chordates

When it comes to chordates, the availability of numerous high-quality genomic sequences, along a clearer understanding of the phylum's evolutionary history, has allowed to overcome the limitations of phylogenetic inferences for determining the evolutionary relation of chordate *Fgf* genes (S. Bertrand et al., 2011; Oulion et al., 2012; Satou et al., 2002). The capability for detailed synteny conservation analyses, particularly when comparing vertebrate genomes with those of basal chordates, has significantly refined the evolutionary history of the Fgf family within this clade (Bertrand et al., 2011; Oulion et al., 2012). Recent evolutionary reconstructions suggest the existence of eight chordate Fgf subfamilies: Fgf1/2, Fgf3, Fgf4/5/6, Fgf7/10/22, Fgf8/17/18/24, Fgf9/16/20, Fgf11/12/13/14, and Fgf19/21/23 **(Figure I7)** (Oulion et al., 2012). This classification contrasts with the classical one in that Fgf3 belongs to a different subfamily than Fgf7, Fgf10, and Fgf22, and it raises the total number of Fgf subfamilies to nine (the eight subfamilies present in chordates plus one specific to arthropods and arthropod viruses) (Oulion et al., 2012; Popovici et al., 2005). According to their canonical functions in vertebrates, the Fgf11/12/13/14 subfamily comprises non-secreted intracellular factors, the Fgf19/21/23 subfamily comprises endocrine long-distance signalling FGFs, and the rest of the subfamilies comprise canonical paracrine FGFs **(Figure I7A)**. Notably, even though vertebrate members of the Fgf1/2 and Fgf9/16/20 subfamilies belong to the group of canonical paracrine-acting Fgfs, they do not have a classical signal peptide in their N-terminus and are secreted through alternative mechanisms (Ornitz & Itoh, 2015).

It has been proposed that the basally divergent chordate amphioxus might possess the full catalogue of chordate Fgfs with one single member for each of the eight subfamilies, which contrasts with the burst of the large number of Fgfs within each subfamily in vertebrates due to the various rounds of genome duplication that occurred during the evolution of different lineages (e.g. 19 in sarcopterygians, 22 in mammals, 23 in chicken and 27 in zebrafish) **(Figure I7B)** (Dehal & Boore, 2005; Oulion et al., 2012). In ascidian tunicates, seven *Fgf* genes have been identified, representing at least six of the eight chordate subfamilies. This suggests that Fgf1/2 and Fgf3 were lost during the evolution of the lineage leading to ascidians, while a novel FGF member (Fgf-L) likely

emerged as a duplicate of Fgf7/10/22 **(Figure I7B)** (Dehal & Boore, 2005; Oulion et al., 2012; Popovici et al., 2005; Satou et al., 2002).

2.3.3. Evolution of the FgfRs in chordates

Regarding the Fgf receptor, the presence of a single ortholog of the FgfR gene in cephalochordates and ascidians provides strong evidence that the last common ancestor of chordates also had a single FgfR gene **(Figure I7B)** (S. Bertrand et al., 2011; Kamei et al., 2000b; Satou et al., 2003). In vertebrates, however, the two rounds of whole genome duplication resulted in four FgfR copies. The expansion of the FgfR gene occurred along with the expansion of the *Fgf* genes, and this, coupled with the mechanisms of alternative splicing for the FgfR genes, supposed a boost in the combinatorial options of FgfR and Fgf ligands **(Figure I7B)** (Itoh & Ornitz, 2004; N. Mistry et al., 2003).

The fact that most vertebrates have retained the multiple copies of both the Fgf receptors and Fgf ligands suggests that the expansion of the FGF system, along with that of other cell signalling mechanisms, favoured the diversification and evolutionary success of vertebrates, although it remains to be understood how this gene expansion has participated in the acquisition of vertebrate-specific morphological features (S. Bertrand et al., 2011; Itoh & Ornitz, 2004; Meister et al., 2022). Interestingly, regarding the FgfR architecture, all known vertebrate FgfRs, as well as the cephalochordate and echinoderm orthologs, possess three Ig-like domains in their extracellular regions. This indicates that the ancestral chordate FgfR gene also had this architecture. In contrast, ascidian FgfRs possess only two Ig-like domains, indicating a simplification of the gene in the tunicate lineage leading to ascidians (Rebscher et al., 2009; Satou et al., 2003).

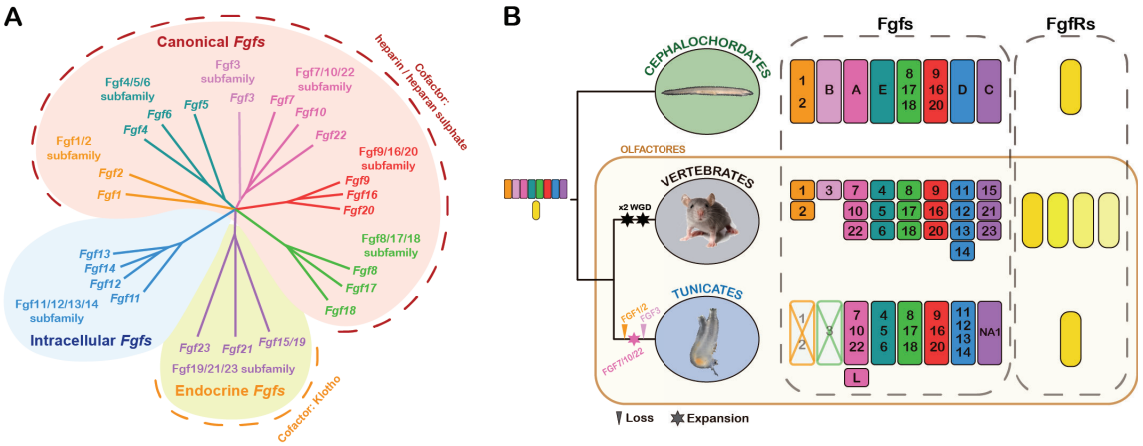


Figure I7. Classification of the Fgfs and evolution of the Fgfs and FgfRs in chordates. (A) Classification of vertebrate Fgf genes into 8 subfamilies. Adapted from Ornitz & Itoh (2015). **(B)** Evolution of the Fgf gene family in chordates. Adapted from Oulion et al. (2012).

2.4. Developmental functions of the FGF signalling pathway

2.4.1. Conserved developmental functions of the FGF signalling pathway in eumetazoans

The FGF signalling pathway acquired important developmental functions shortly after its appearance, as it was revealed by its implication in the gastrulation and the organogenesis of sensory structures in the early branching cnidarian *Nematostella vectensis* (Matus et al., 2007; Rentzsch et al., 2008; Technau, 2020). Among bilaterians, the extrapolation of conserved developmental roles has been problematic due to the lack of functional data in protostomes other than *Drosophila melanogaster* or *Caenorhabditis elegans*. In these organisms, FGF signalling is known to play important roles in the coordination of gastrulation movements, mesoderm patterning, axon guidance, and in more specific morphogenetic events such as tracheal branching in *D. melanogaster* or the migration of sex myoblasts in *C. elegans* (Bülow et al., 2004; de Miguel et al., 2020; García-Alonso et al., 2000; Leptin & Affolter, 2004; Lo et al., 2008; Stathopoulos et al., 2004). The conserved function of the FGF signalling in neural development is also supported by studies in gastropods and platyhelminths (Cebrià et al., 2002; Pollak et al., 2014).

Moreover, recent studies in three lophophorate species emphasised the conserved roles of the FGF signalling in gastrulation and development of neural structures and pointed to an ancestral role in mesoderm induction (Andrikou & Hejnol, 2021). In the case of deuterostomes, numerous studies have identified an extensive variety of developmental functions for the FGF signalling in vertebrates (Dorey & Amaya, 2010; Thisse & Thisse, 2005). Insights from the basal chordate amphioxus, as well as from echinoderms and hemichordates, highlight the ancestral function of the FGF signalling in coordinating gastrulation movements and in the development of mesodermal and neural structures (S. Bertrand et al., 2011; Fan et al., 2018; Röttinger et al., 2008). However, the precise function of the FGF pathway in mesodermal and neural development is still controversial, since it is not fully understood whether it is required for their fate induction or for fate maintenance and patterning in their derivatives, at least in vertebrates (Dorey & Amaya, 2010).

2.4.2. Developmental functions of the FGF signalling in chordates

Within chordates there is much more information about the functions of the FGF signalling pathway thanks to the immense amount of data coming from vertebrates, as well as studies performed in ascidians and cephalochordates. During the evolution of chordates, the expansion of the eight gene subfamilies up to 27 *Fgf* genes in vertebrates due to the two rounds of genome duplication has been linked to the innovation and sophistication of many of the characteristic traits of this clade, including axial patterning,

somitogenesis, limb bud formation, visceral and skeletal development (Thisse & Thisse, 2005), and even the invention of the “new head” in vertebrates (Bertrand et al., 2011). Studies of FGF signalling in ascidian tunicates have contributed to reveal that some of the traits that characterize vertebrates, indeed, were not vertebrate innovations, but were already present in the last common ancestor of olfactores, such as for instance the role of Fgf8/17/18 in the organizer activity and the compartmentalization of the central nervous system, or the role of Fgf8/17/18 and Fgf11/12/13/14 in neural crest and placodal derivatives (Horie et al., 2018; Imai et al., 2009; Kourakis & Smith, 2007; Stolfi et al., 2015; Wagner & Levine, 2012).

FGF signalling plays a crucial role in establishing the main body axes during vertebrate embryogenesis, orchestrating both dorsal-ventral and anterior-posterior patterning. It is instrumental in defining the body’s structural layout by influencing the differentiation and placement of cells along these axes. For the anterior-posterior axis, FGF signalling has a strong posteriorizing effect on developing tissues, including the neural tube, which affects overall body length and orientation (Diez del Corral & Morales, 2017; Hébert, 2011). Similarly, in the context of dorsal-ventral patterning, FGF contributes to the specification of dorsal cell fates while inhibiting ventral fates, directly impacting the development of the spinal cord and associated structures. This axis establishment is vital for the correct anatomical structure and function in the mature organism. It is typically driven by the coordination of several developmental signalling pathways, including the FGF signalling pathway, with signal sources distributed throughout different parts of the embryo (Diez del Corral & Storey, 2004; Dorey & Amaya, 2010; Kumar et al., 2021)

2.4.3. Triangular interactions between the FGF, RA and Wnt signalling pathways

Among the many developmental processes in which the FGF signalling is involved in vertebrates, in some important cases it functions in coordination with the RA and Wnt signalling pathways.

The opposing role of the RA and the FGF signalling in posterior extension and AP axis patterning in vertebrates is a paradigmatic case of antagonistic interaction between signalling pathways (Diez del Corral & Storey, 2004; Dorey & Amaya, 2010; Ribes et al., 2009; Shiotsugu et al., 2004; Zhao & Duester, 2009). Within the caudal stem zone of vertebrate embryos, the expression domains of *Fgf8/Fgf4* and *Raldh2* (the enzyme that synthesizes RA) are mutually exclusive and the ligands from the two pathways form opposing gradients (Diez del Corral & Storey, 2004; Wilson et al., 2009). FGF is expressed in the posterior stem zone and inhibits differentiation, keeping the cells in an undifferentiated proliferative state, while RA provided by the anterior somitic and presomitic mesoderm triggers differentiation. This creates a differentiation front where RA signalling wins over FGF signalling as the body axis extends **(Figure I8)** (Dorey &

Amaya, 2010). The Wnt signalling pathway appears to mediate the interaction between FGF and RA, for instance promoting the expression of FGF in the posterior stem zone or modulating *Raldh2* expression upon the control of FGF in the anterior somitic zone (Aulehla et al., 2003; Olivera-Martinez & Storey, 2007). The triangular interactions of these pathways regulate body extension, neural differentiation, mesoderm segmentation, and rostrocaudal identity through *Hox* gene activation (Diez del Corral et al., 2003; Diez del Corral & Storey, 2004; Olivera-Martinez et al., 2012; Wilson et al., 2009). This opposing interaction of the FGF and RA pathways is not exclusive to the posterior embryo elongation, as similar mechanisms have been described in limb induction and development (Cunningham et al., 2013; Mercader et al., 2000; Mok et al., 2014; Zhao et al., 2009), in neural crests induction and emigration to the trunk (Martínez-Morales et al., 2011; Villanueva et al., 2002), in the anteroposterior patterning of the heart field (Sirbu et al., 2008), or in the development of the otic placode (Maier & Whitfield, 2014). Thus, it seems that the antagonistic interaction between the FGF and RA signalling pathways is a conserved mechanism reiteratively used in vertebrate development.

Beyond vertebrates, a similar antagonistic interaction between FGF and RA with an involvement of the Wnt pathway has been identified in ascidians **(Figure I8)**. In the larval tail of *Ciona robusta*, an opposing gradient of RA coming from the anterior-most muscle cells and FGF coming from the posterior epidermal tail tip was discovered to regulate the AP patterning of the tail epidermis by controlling *Hox* gene expression. This mechanism also plays a role in patterning the peripheral nervous system by controlling the number and distribution of ventral caudal epidermal neurons (Pasini et al., 2012). In contrast, cephalochordates lack this cross-regulation between RA and FGF signaling, suggesting that this mechanism is an evolutionary innovation specific to olfactores **(Figure I8)** (S. Bertrand et al., 2015).

Thus, this triangular interaction of opposing RA and FGF, with Wnt mediating the interaction was present in the last common ancestor of tunicates and vertebrates. Its involvement in the development of several structures suggests that it has been repeatedly recruited to coordinate the developmental GRNs in olfactores, although the evolutionary implications are not fully understood. As detailed in previous sections, the appendicularian tunicate *Oikopleura dioica* is an eKO of the RA signalling (Cañestro et al., 2006; Martí-Solans et al., 2016), and it has also lost most of the ancestral Wnt subfamilies (Martí-Solans et al., 2021), standing as an attractive animal model to study the evolutionary implications of this FGF-RA-Wnt triangular mechanism. However, although several *Fgf* genes have been identified in *O. dioica* (Oulion et al., 2012), there is an important gap of knowledge affecting their evolutionary relation with other chordate *Fgf* genes and their putative developmental functions.

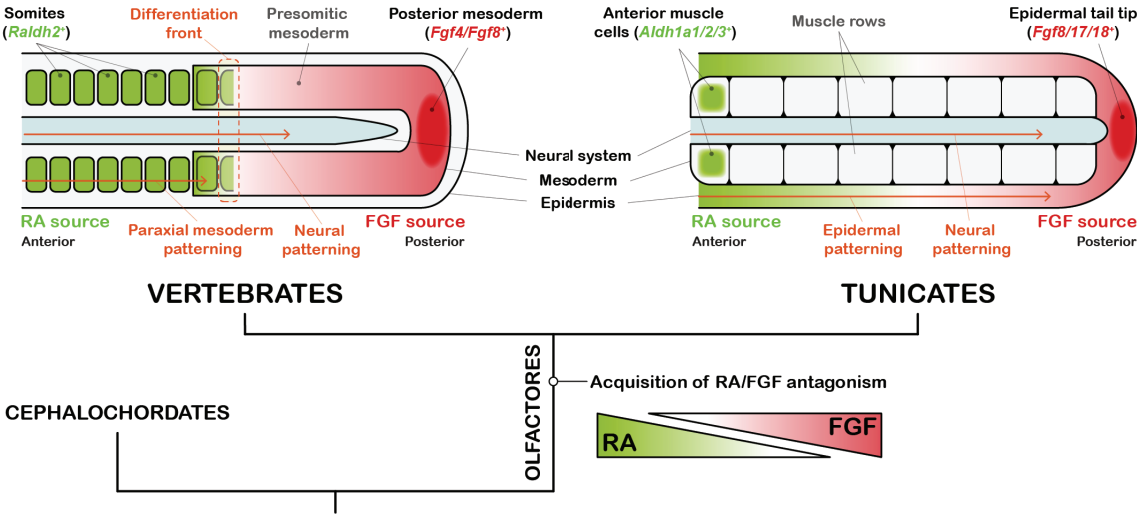


Figure I8. Opposing gradients of FGF and RA signalling in vertebrate and tunicate development. In vertebrates, an FGF source from the posterior mesoderm keeps presomitic cells undifferentiated, while RA from the anterior region promotes their differentiation into somites, creating a "differentiation front" as the body extends. Similarly, in the tunicate *Ciona robusta*, RA and FGF form opposing gradients that control the patterning of the tail epidermis and peripheral nervous system. This conserved mechanism appears to be an innovation within the clade olfactores, as no similar RA/FGF antagonism is found in cephalochordates. The schemes represent a dorsal view of the posterior region of the developing embryo in vertebrates and the larval tail in tunicates.

OBJECTIVES

The **overall objective** of this PhD project was to investigate the impact of gene loss on the evolution of developmental mechanisms within the chordate phylum, using *Oikopleura dioica* as a model organism and the FGF signalling pathway as a case study. As detailed in the introduction, previous research established *O. dioica* as an evolutionary knockout of the RA signalling pathway, and it has also undergone an extensive remodelling of the Wnt signalling pathway, having lost nine of the thirteen Wnt subfamilies that were present in the chordate ancestor. Given the conserved triangular interactions among these two pathways and the FGF signalling pathway in the clade olfactores, the **primary objective** of this PhD project was to exploit the unique evolutionary context in *O. dioica* to explore the evolution of the FGF signalling. This research aims to advance our understanding of the impact of gene loss on the evolution of developmental mechanisms and their putative adaptive implications.

Thus, the specific objectives of this doctoral project are:

- **Objective 1**: To identify and characterize the complete ***Fgf* gene catalogue** in *O. dioica,* and to study the evolutionary relationships with other chordate *Fgf* genes.

- **Objective 2**: To identify and characterize the **Fgf receptors** in *O. dioica* and the main **downstream transduction pathways** typically associated with the FGF signalling in other chordates.

- **Objective 3**: To investigate the **developmental roles of the FGF signalling pathway** in *O. dioica* by inhibiting the receptors and analysing the effects through WMISH and *omic* approaches.

- **Objective 4**: To propose an **evolutionary scenario** describing the impact of gene losses on the evolution of developmental mechanisms in *O. dioica*, with a particular focus on their potential associations to the innovation of a free-living lifestyle in appendicularians.

MATERIALS AND METHODS

1. ***Experimental procedures***
2. ***Informatic and computational procedures***

1. Experimenal procedures

1.1. Animal sampling and culture

Oikopleura dioica specimens were acquired from animal colonies that have been maintained in our laboratory for over five years. The founder individuals were originally obtained from the Mediterranean coast near Barcelona (Catalonia, Spain), using a plankton net at a distance of approximately 200 meters from the shore, and cultured as detailed in Martí-Solans et al. (2015). This project did not raise any ethical concerns since the experimentation conducted on aquatic invertebrate animals does not fall under the regulations pertaining to animal experimentation, as stipulated in *Real Decreto 223 14-3-1998* and *Catalonia Law 5/1995, DOGC2073,5172*. Nonetheless, all experimental procedures adhered to the European Union (EU) guidelines for animal care and were formally approved by the Ethical Animal Experimentation Committee (CEEA-2009) of the University of Barcelona.

1.2. *In vitro* fertilization and animal fixation

In vitro fertilizations were conducted to obtain synchronously developing embryos whenever animals were needed for experimental procedures or for fixation.

For *in vitro* fertilizations, mature females and males were collected from the cultured stock (Martí-Solans et al., 2015), washed in sterilized sea water (SSW), and kept separately in individual bowls until they released eggs or sperm. Eggs or sperm were never kept longer than 2 hours after their release to avoid introducing bias due to a reduced viability. A minimum of three males were used each time to prepare a sperm solution with a concentration of one male's sperm per mL of SSW. The sperm solution was prepared by collecting the concentrated sperm from the males into the corresponding volume of SSW in a 15 mL Falcon tube, shaking vigorously for spermatic activation, and allowing it to rest for at least 10 minutes at 19 ºC before use. The number of females varied depending on the number of eggs required for each experiment.

Released eggs were collected in 4 mL of SSW in a small custom glass bowl or a Ø60 mm petri dish, fertilized with 200-250 µL of sperm solution, and then transferred to a new bowl or petri dish with 4 mL of fresh SSW 5 minutes post fertilization (mpf). The embryos were left to develop at 19ºC and were derived to experimental procedures or fixed at the desired developmental time, according to the time points described in Ferrández-Roldán et al. (2019). The fixation was conducted incubating the animals for 1 hour at room temperature (RT) in a solution containing EGTA 1 mM, $MgSO_4$ 2 mM, MOPS 0.1 M, NaCl 0.5 M and paraformaldehyde 4%. Fixed animals were washed and stored in 70% ethanol at -20ºC until further use.

1.3. Gene cloning

Fragments of the transcripts of interest were amplified by PCR from a mixture of cDNA coming from animals at various developmental stages, or from gDNA in the case of intronless genes. The list of genes cloned in this doctoral project, the primers designed and used for their amplification, and the template used are detailed in **Supplementary Table 1 in Annex 1**.

PCR products were cloned into the pCR™4-TOPO™ vector using the Invitrogen™ TOPO™ TA Cloning™ Kit for sequencing (REF:45-0030) and introduced into One Shot™ TOP10 Chemically Competent *E. coli* (REF:C404003) via heat shock transformation, following the manufacturer instructions. Transformant bacteria were selected based on the resistance to ampicillin (Amp) provided by the vector after incubation on Amp^+ LB Agar plates for over 16 hours at 37ºC. The presence of a recombinant plasmid in the resulting colonies was confirmed by PCR using vector-specific oligonucleotides (M13F and M13R) flanking the insert region. Production and isolation of recombinant plasmids at high concentrations was carried out by liquid culturing of the harbouring strains in Amp^+ LB Broth for over 16 hours at 37ºC and 220 rpm, followed by plasmid extraction using the QIAprep® Spin Miniprep Kit (REF:27104) from 3 mL of the culture. Before their first use, clones were Sanger sequenced to verify the identity and orientation of the insert.

1.4. Synthesis of riboprobes

The synthesis of antisense riboprobes, necessary for *in situ* hybridization experiments, was conducted through *in vitro* transcription of the clones after linearizing the plasmid with restriction enzymes and purifying with the DNA Clean & Concentrator™ Kit from Zymo Research (Cat # D4004). The restriction enzyme and RNA polymerase (T3 or T7) was chosen according to the orientation of the insert so that the riboprobe was complementary to the target mRNA. The transcription reaction employed digoxigenin-labeled uracil (DIG-RNA Labeling Mix, 11277073910 Roche) to enable subsequent probe detection with Anti-digoxigenin-AP, Fab Fragments (11093274910 Roche). After the synthesis reaction, the probes were purified with the RNA Clean & Concentrator™ kit from Zymo Research (Cat # R1015) and stored at -80ºC until use.

1.5. Whole mount in situ hybridization and imaging

Whole mount in situ hybridization (WMISH) assays were performed on animals fixed at the desired developmental stage. Embryos fixed at stages prior to hatching (< 3 hours 30 minutes post-fertilization) were manually dechorionated using glass micro-

needles. This was done by transferring them onto a dry 70% poly-lysine band in a Ø60 mm petri dish filled with 70% ethanol.

WMISH reactions were conducted according to the detailed protocol outlined in **Supplementary Material** in **Annex 1**. The staining reaction from the probes continued until coloration was observed under a magnifying glass, or for a maximum of 14 days, with the staining solution being refreshed at least once daily. Visualization of the stained animals was achieved by mounting the embryos and larvae in 80% glycerol on a custom excavated slide and examining them under a Zeiss Axiophot Transmitted Nomarski DIC Phase Contrast Trinocular Microscope. Images were captured with a Leica DFC300FX camera. Nuclei visualization was conducted after incubating the embryos for 20-30 minutes at 37ºC in 100 µL of a solution of Hoechst 33342 (Thermo Scientific, 62249) 1µM in glycerol

1.6. Pharmacological treatments

SU5402 (Sigma-Aldrich, #SML0443) stock solutions were prepared dissolving the reagent in dimethyl sulfoxide (DMSO) at 10 mg/mL. Wortmannin (Sigma-Aldrich, #W1628) stock solution was prepared at a concentration of 10mg/mL. AZD4547 (BioVision, #9403) stock solutions were prepared dissolving the reagent in DMSO at 80 mg/mL. All stock solutions were stored at -80ºC. Experimental concentrations in treatment solutions were achieved by dissolving the required volume of stock solution in SSW through vigorous vortex agitation. Control solutions were prepared by dissolving an equivalent volume of DMSO in SSW. Stock and treatment solutions were shielded from light exposure at all times.

For performing the treatments, *in vitro* fertilized embryos at the desired developmental time were transferred to 4 mL of treatment or control solutions in Ø60 mm petri dishes. The development of the animals proceeded at 19°C shielded from light exposure. Treated animals at the desired developmental stage were derived to fixation, RNA extraction, nuclei extraction, or developmental success counting. The counting of healthy or malformed animals on each plate was conducted manually with the cell counter plugin of the Fiji package from ImageJ using photographs taken at 4-5 hpf (Schindelin et al., 2012).

1.7. Sample collection for paired RNA-seq and ATAC-seq

We aimed to obtain six replicates of paired RNA-seq and ATAC-seq libraries from SU5402-treated and control embryos originating from the same fertilization batch. To ensure enough genetic material, a minimum of 1200 eggs were collected for each replicate. Synchronously developing *O. dioica* embryos were treated with DMSO 0.2% or

SU5402 50 µM starting from the 2-cell stage. Upon reaching the 64-cell stage, approximately 100 embryos from each condition were frozen in liquid nitrogen and stored at -80ºC for later RNA-seq library preparation. The rest of embryos, approximately 500, accounting for approximately 30.000 cells/nuclei, were immediately used for ATAC-seq library preparation. Due to the less established and generally more challenging procedure of ATAC-seq library preparation compared to RNA-seq library preparation, this process was repeated until obtaining six ATAC-seq libraries with satisfactory electrophoresis profiles when analysed on an Agilent TapeStation System. Frozen samples from the same experiments as the selected six ATAC-seq samples were then processed for RNA extraction and RNA-seq library preparation, and sequenced alongside their ATAC-seq library counterparts.

1.7.1. Nuclei extraction, ATAC-seq library preparation and sequencing

Nuclei were extracted and isolated from treated and control embryos using the reagents and following the protocol detailed in **Supplementary Material** in **Annex 1**. This is an adaptation for fresh *O. dioica* samples of Kaestner Lab ATAC Sequencing Protocol, available at: dx.doi.org/10.17504/protocols.io.bv9mn946. Given the miniature genome of *O. dioica* and the difficulty to obtain enough cells, the volume of Tn5 employed in the transposition reaction was reduced with respect to reference protocols. Moreover, we included an actinase treatment for the 64-cell embryos prior to their resuspension in order to perforate the chorion and facilitate nuclei release, or penetration of the Tn5 through the chorion. Quality control of the libraries was performed through electrophoresis on an Agilent TapeStation 4200 System using the Agilent High Sensitivity D1000 ScreenTape®. ATAC-seq libraries were pooled and sequenced 2x150bp on an Illumina NovaSeq 6000 system.

1.7.2. RNA extraction, RNA-seq library preparation and sequencing

RNA was isolated from samples frozen at -80ºC using TRIzol™ Reagent (Invitrogen, Cat. No. 15596026), and quality control was performed through electrophoresis on an Agilent TapeStation 4200 System using the Agilent High Sensitivity RNA ScreenTape®. RNA-seq library preparation was conducted using the NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina® (NEB #E7760) following the Protocol for use with NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490). PCR enrichment and indexing of the libraries were performed with unique index dual primer pairs from the NEBNext® Multiplex Oligos for Illumina® kit (NEB #E6440), and purified with SPRIselect magnetic beads (Beckman Coulter, #B23319). RNA-seq libraries quality was checked through electrophoresis on a High Sensitivity D1000 ScreenTape®, and re-purified if primer dimers were detected. RNA-seq libraries were pooled and sequenced 2x150bp on an Illumina NovaSeq 6000 system.

2. Informatic and computational procedures

2.1. Genomic searches and identification of genes

Genomic searches were performed with BLASTp or tBLASTn using as queries the protein sequences specified in each case. In most cases, the identification of genes was performed employing the Reciprocal-Best-Blast-Hit (RBBH) method (Ward & Moreno-Hagelsieb, 2014), as detailed in the section *Results and Discussion - 1.1.2.1. Identification of the Fgf genes in the BAR genome.* In the case of the Fgf genes in those species where proteomic or transcriptomic data was available, RBBH searches were complemented with HMM searches, using the consensus FGF domain HMM profile retrieved from the Pfam database (PF00167.21.hmm) (J. Mistry et al., 2021).

O. dioica genomes surveyed were: Bar2_p4 (BAR), OdB3 (NOR), OSKA2016v1.9 (OSA) and OKI2018_I69_1.0 (OKI) (Bliznina et al., 2021; Denoeud et al., 2010; Plessy et al., 2024; K. Wang et al., 2020). Other appendicularian species genomes surveyed were: *Oikopleura albicans* (SCLG01000000), *Oikopleura vanhoeffeni* (SCLH01000000), *Oikopleura longicauda* (SCLD01000000), *Bathochorddaeus sp.* (SCLE01000000), *Mesochordaeus erythrocephalus* (SCLF01000000), and *Fritillaria borealis* (SDII01000000) (Naville et al., 2019). Ascidian species were surveyed on the ANISEED database and included *Ciona robusta, Ciona savignyi*, *Phallusia fumigata, Phallusia mammillata*, *Halocynthia roretzi*, *Halocynthia aurantium*, *Botryllus schlosseri*, *Botryllus leachii*, *Molgula occulta*, *Molgula oculata* and *Molgula occidentalis* (Brozovic et al., 2018). All other species (i.e. vertebrates, cephalochordates, *Drosophila melanogaster* and *Caenorhabditis elegans*) protein sequences were retrieved from either the RefSeq database at NCBI (O'Leary et al., 2016), or the UniProt database (Bateman et al., 2023).

2.2. Sequence alignments and phylogenetic inference

Protein alignments were generated with MUSCLE and MAFFT implemented in AliView v1.28 (Larsson, 2014) and reviewed by hand. Phylogenetic trees were based on Maximum Likelihood (ML) inferences calculated with IQ-Tree (Nguyen et al., 2015), as well as PhyML v3.0 (Guindon et al., 2010). Tree node support was inferred by fast likelihood-based methods including aLRT SH-like and aBayes, and by standard or ultrafast bootstraps (n=100) according to computational capacity.

In the case of the Fgf phylogenetic inference, non-homologous independently extended N- and C-terminus of different subfamilies were aligned in a non-overlapping manner to reduce noise among Fgf subfamilies in which no similarity was detected outside the FGF core homology domain. LG was inferred as the best-fit substitution model according to Bayesian information criterion BIC to Fgf data, with a gamma with 4

categories and a shape alpha of 2.4681 (Kalyaanamoorthy et al., 2017). Phylogenetic trees were inferred both in complete and trimmed protein alignments, the later by removing all extended regions outside the FGF core homology domain, and in both cases producing the same tree topology regarding the FGF subfamily homology of appendicularian genes.

2.3. Protein sequence analyses

The domain architecture and functional motifs of Fgf and other proteins were examined individually with InterProScan, a comprehensive software suite that combines the information stored in several protein databases (including InterPro, PFAM, SMART or PANTHER) to provide *in silico* functional characterization of queried protein sequences (Jones et al., 2014). Hydropathy plots were generated in ProtScale available in Expasy (Gasteiger et al., 2005), using the Kyte-Doolittle hydrophobic scale and an interval of 9 amino acids with a linear weight variation model in a normalized scale, following previous similar analysis reported on Fgf9 (Miyakawa et al., 1999).Sequence identity and similarity for every pair of Fgf or FgfR sequences was obtained from the pairwise sequence alignment with EMBOSS Needle (Madeira et al., 2024). Nuclear Localization Signals (NLS) were predicted with the NLStradammus software using the 4 state HMM static model and a posterior cutoff of 0.5 (Nguyen Ba et al., 2009), or manually identified in the case of classical NLS based on the consensus stated in Lu et al. (2021). Signal peptide (SP) predictions were conducted using the SignalP 6.0 (Teufel et al., 2022) and Phobius (Käll et al., 2004) software.

2.4. Protein structure predictions and manipulation

Three-dimensional structures of *O. dioica* Fgf proteins were predicted *ab initio* with AlphaFold2, and the pLDDT score provided by the same software was used as a measure of quality (Jumper et al., 2021). For each Fgf protein, the top-ranked relaxed model was imported into USCF ChimeraX for its visualization, analysis, and image generation (Pettersen et al., 2021).

In the case of the FgfR proteins, the models presented in this work were predicted *ab initio* with the RoseTTAFold tool at Baker Lab online server (https://robetta.bakerlab.org) (Baek et al., 2021). For each protein, RoseTTAFold produced five models, and all were assessed with the Structure Assessment tool at SWISS-MODEL online server (https://swissmodel.expasy.org) (A. Waterhouse et al., 2018; A. M. Waterhouse et al., 2024). The model displaying the highest QMEANDisCo global score was chosen, and the QMEANDisCo local score was used a measure of quality (Studer et al., 2020).

2.5. RNA-seq and ATAC-seq data processing and analyses

RNA-seq and ATAC-seq data were processed using automated pipelines of the nf-core collection of workflows (Ewels et al., 2020), executed with Nextflow v23.04.4 (Di Tommaso et al., 2017). These pipelines offer a robust, community-curated solution designed to standardize and optimize the analyses of sequencing data. The workflows incorporate multiple quality control, alignment, and quantification steps, ensuring comprehensive and reproducible data processing.

Specifically, the RNA-seq data was processed with the pipeline nf-core/rnaseq v3.12.0 (doi: 10.5281/zenodo.1400710), using STAR as aligner, and Salmon as quantifier. The ATAC-seq data was processed with the pipeline nf-core/atacseq v2.1.2 (doi: 10.5281/zenodo.2634132), using Bowtie2 as aligner and calling narrow peaks with MACS2. The workflows of the two pipelines are detailed in **Supplementary Figure 1**.

2.5.1. Differential gene expression analyses

Several artifacts were noticed on the quantification of reads outcoming from the Salmon software incorporated in the nf-core pipeline. Thus, differential gene expression (DGE) analyses were performed based on an alternative gene count matrix generated with StringTie and employing the DESeq2 package in R (Pertea et al., 2016). For generating the alternative gene count matrix, StringTie was fed with the sorted read alignments (BAM files) outcoming from the nf-core/rnaseq workflow. Volcano plots were generated with the ggplot2 package in R.

2.5.2. GO enrichment analyses

GO terms of all translated coding sequences from *O. dioica* were inferred with eggnog-mapper v2 against the eggnog 5.0 database (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019). Default parameters were employed, except for the e-value cutoff (raised to 0.05), minimum hit bit-score (lowered to 40) and minimum percentage identity (lowered to 20). GO enrichment analyses were performed with the enrichplot package in R.

2.5.3. Condition-exclusive peaks and differential accessibility test

Consensus peaks and their annotations resulted from MACS2, samtools and bedtools integrated in the nf-core/atacseq v2.1.2 workflow. Differential accessibility tests were conducted with DESeq2, based on a peak count matrix generated with IntersectBed using the sorted read alignments (BAM files) and the consensus peaks outcoming from the nf-core/atacseq v2.1.2 workflow (Pallarès-Albanell et al., 2024). Volcano plots were generated with the ggplot2 package in R.

RESULTS AND DISCUSSION

1. ***Evolution of the Fgf gene family in appendicularians***
2. ***Evolution of the Fgf receptors and transduction pathways***
3. ***Testing the function of FGF signalling in Oikopleura dioica by inhibitory treatments and omic approaches***
4. ***An evolutionary scenario for the remodelling of the Fgf signalling in Appendicularians***

1. Evolution of the *Fgf* gene family in appendicularians

1.1. Identification of the *Fgf* gene catalogue in *Oikopleura dioica*

1.1.1. Genome sequencing and assembly of *Oikopleura dioica* cryptic species

The first objective of this doctoral project was to identify the complete catalogue of *Fgf* genes in the genome of *Oikopleura dioica*. At the onset of this project, the only available genome assembly was derived from specimens collected in Bergen, Norway (hereinafter referred to as NOR) (Denoeud et al., 2010, PRJEA42427). According to the annotations available at that time, at least six putative *Fgf* homologous genes were present in the genome of *O. dioica* (Oulion et al., 2012). However, the orthology of these genes remained unclear, and our manual inspection revealed inaccuracies in the annotations that in some cases led to truncated coding sequences that hindered our analysis. Misannotation of genes was a recurrent issue previously encountered in our lab while working with other gene families on the NOR genome assembly. This problem was partially due to the fragmented nature of the assembly (ASM20953v1: 1260 scaffolds; N50 = 395.4 Kb), which caused segmentation of some genes across different scaffolds and increased the probability of concealing crucial genes. Since this undermined the feasibility of conducting reliable comprehensive studies on gene families, a new high-quality genome assembly at the chromosome-scale resolution was generated, using DNA from a single animal from Barcelona.

The genome sequencing and chromosome-scale assembly was conducted in collaboration with the Genomics and Regulatory Systems Unit at the Okinawan Institute of Science and Technology (OIST). This project employed nanopore long-reads and Hi-C data for scaffolding, complemented with Illumina short-reads for base curation. The result was a new reference genomic assembly of *O. dioica* from Barcelona (hereinafter BAR), featuring three chromosomes with a N50 of 13.55 Mb (Plessy et al., 2024). Additionally, we generated RNA-seq and CAGE data from various developmental and adult stages, what greatly contributed to achieving a more robust annotation. In addition to the BAR reference genome, this sequencing project produced two other reference chromosome-scale genomic assemblies of *O. dioica* specimens from different regions, namely Osaka and Okinawa (hereinafter OSA and OKI, respectively) (Bliznina et al., 2021; Plessy et al., 2024). The findings revealed an unprecedented level of genome scrambling when comparing the genomes from the three different regions, what suggested that *O. dioica* populations from different locations constitute cryptic species despite lacking obvious morphological differences **(Figure 1.1)** (Plessy et al., 2024). Moreover, the infertility observed in crosses between animals from Osaka (OSA) and Okinawa (OKI) further supports the notion that different populations are indeed distinct species

(Masunaga et al., 2022). This discovery allowed us to expand our genomic survey about the *Fgf* gene family across the three identified *O. dioica* cryptic species (namely BAR, OSA, and OKI). The precise and manually inspected annotation of the *Fgf* genes performed in this doctoral project was used to evaluate the accuracy of the pipelines for automatic annotation in *O. dioica,* and as a case study to inspect the genomic rearrangements among the three cryptic species. The results from our collaborative project with the Genomics and Regulatory Systems Unit at OIST, including our survey on the *Fgf* genes and other gene families, were recently published in the journal *Genomic Research* under the title "Extreme genome scrambling in marine planktonic *Oikopleura dioica* cryptic species" (Plessy et al., 2024). The full study is provided in **Annex 2.**

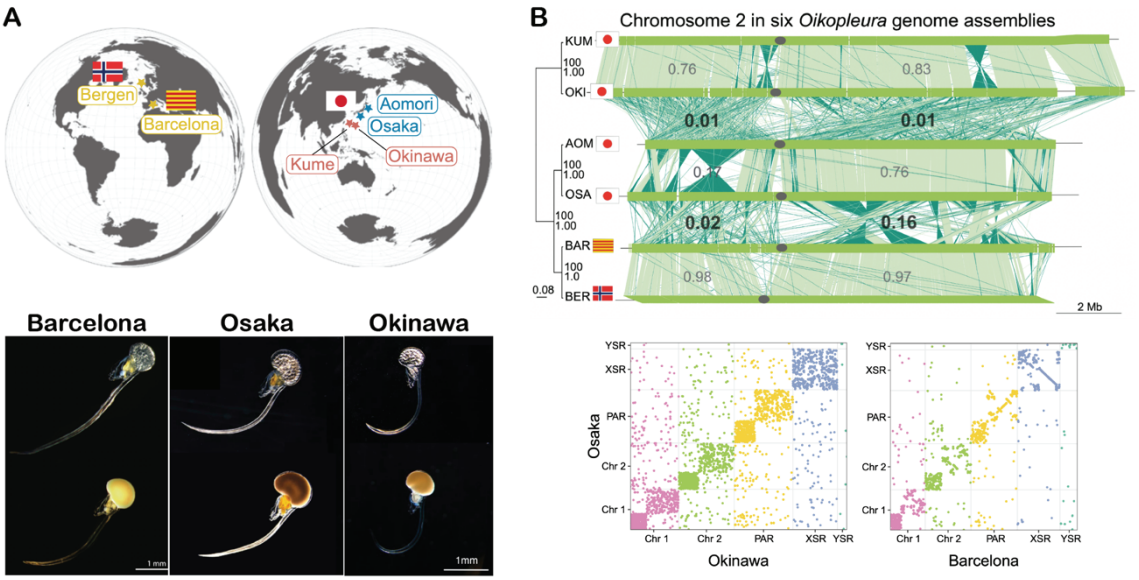


Figure 1.1. Discovery of *O. dioica* cryptic species. (A) Genomic assemblies from *O. dioica* specimens collected in different regions revealed a huge number of chromosomic reorganizations despite not presenting obvious morphological differences. The regions surveyed were the Mediterranean (Barcelona: BAR), the north Atlantic (Bergen: NOR), the southern Japanese Seto Inland Sea (Osaka: OSA), the northern Japanese pacific coast (Aomori: AOM), and the subtropical Ryukyu Islands (Okinawa: OKI, and Kume: KUM). BAR, OSA and OKI genomes were sequenced, polished, assembled to chromosome-scale, and used as reference genomes for each lineage. BER, AOM and KUM genomes were sequenced to contig-level and used for validating the reference genomes. **(B)** The reorganizations have mostly occurred within chromosomal arms. These vast number of genomic reorganizations, together with interbreeding experiments, suggested that different clades represent different cryptic species. Adapted from Plessy et al. 2024.

1.1.2. Identification of the *Fgf* genes in the three *O. dioica* cryptic species

1.1.2.1. Identification of the *Fgf* genes in the BAR genome

With the new genomic assemblies, we identified ten loci encoding for *Fgf* genes in the BAR genome of *O. dioica* **(Table 1)**. This identification was based on sequence similarity searches through two different approaches: (i) probabilistic inferences based on profile Hidden Markov Models (HMM) to identify proteins with an FGF domain, and (ii) BLAST-based searches using reference Fgf protein sequences from other species as queries.

In the first approach we employed the hmmsearch tool from the HMMER software suit to search for proteins matching the FGF domain HMM profile. For the search, we used our transcriptomic data to build a six-frame proteome representing all putative proteins encoded in all transcribed sequences. The HMM search and subsequent validation of the sequences resulted in ten proteins bearing an FGF domain. We mapped these proteins to their coding genes in the BAR genome and annotated the ten putative *Fgf* loci. Given that the hmmsearch scope was limited to genes represented in the transcriptome, we applied a second approach to identify putative *Fgf* genes at the genomic level using BLAST tools. We employed the Reciprocal-Best-Blast-Hit (RBBH) method, starting with all well-known Fgf protein sequences from various chordate and non-chordate species (i.e. *Branchiostoma lanceolatum*, *Homo sapiens*, *Ciona robusta, Drosophila melanogaster* and *Caenorhabditis elegans*). This set included representatives of all known Fgf subfamilies in chordates, as well as the arthropod-specific Bnl subfamily. For each original Fgf query sequence, we performed a tBLASTn search on the BAR genome, selected the five best blast hits, and used them as secondary queries in a reciprocal BLASTp search on the original query-species proteome. Only *O. dioica* hits that returned an Fgf protein as best blast hit on the original species proteome were considered putative *Fgf* loci for further validation. In the case of the ascidian *Ciona robusta,* the closest relative to *O. dioica* in which the *Fgf* genes have been carefully described (Oulion et al., 2012; Satou et al., 2002), only Fgf7/10/22, Fgf9/16/20 and Fgf11/12/13/14 produced significant hits (e-value < 0.05). Notably, all the putative loci identified in *O. dioica* returned either Fgf9/16/20 or Fgf11/12/13/14 as their reciprocal best BLAST hits in *C. robusta* **(Supplementary Table 2)**. Interestingly, the RBBH method allowed us to identify only six *Fgf* loci, compared to the ten loci identified in the HMM approach. Only when these six RBBH-identified Fgf sequences were used as queries in intraspecies tBLASTn searches we were able to recapitulate the ten loci previously discovered through hmmsearch, and no additional putative *Fgf* loci were detected. The discrepancy between the HMM and BLAST approaches suggested a particularly high sequence divergency with respect to the reference query Fgf sequences in those genes that could not be primarily identified with the RBBH method (i.e *Fgf9/16/20b-f*, as explained in following sections).

Overall, our survey of *Fgf* genes in the BAR *O. dioica* genome resulted in ten loci encoding putative *Fgf* genes. Reanalysis of the NOR genome allowed us to conclude that six of the *Fgf* genes from BAR corresponded to the six previously annotated in NOR, and we proved that the four additional genes found in BAR were also present in NOR. However, they had not been predicted by automated annotations or represented in the proteome, and the highly scaffolded nature of the NOR genome assembly had hindered their identification in previous searches. Therefore, we concluded the presence of ten Fgf gene orthologs in *O. dioica* from both Barcelona and Bergen.

Table 1. Fgf gene loci in four *Oikopleura dioica* genomic assemblies. Gene names derive from their classification described in the following subsections. BAR, OSA, and OKI genomic assemblies were conducted at chromosome resolution and represent three distinct cryptic species. The NOR genomic assembly is highly fragmented and likely belongs to the same cryptic species as BAR (Plessy et al., 2024).

Gene	Locus			
	Barcelona (BAR)	Bergen (NOR)	Osaka (OSA)	Okinawa (OKI)
FGF9/16/20a	Chr2:4655677..4656348+	scaffold_102:12088..12757+	Chr2:4057263..4057930+	chr2:1954164..1954855-
FGF9/16/20b	PAR:7533955..7534434-	scaffold_10:5991..6470-	PAR:7751656..7752132-	PAR:8320817..8321311+
FGF9/16/20c	PAR:1413023..1413490+	scaffold_69:77823..78290-	PAR:1133161..1133631	PAR:1213466..1213936+
FGF9/16/20d	Chr2:5643475..5643966-	scaffold_9:881169..881660+	Chr2:8845048..8845539-	chr2:6688155..6688661+
FGF9/16/20e	PAR:11389252..11389686-	scaffold_49:272710..273144+	PAR:12053439..12053873+	PAR:10503013..10503501-
FGF9/16/20f	PAR:3507957..3508448-	scaffold_95:65562..66053+	PAR:2031348..2031806-	PAR:2497535..2498028+
FGF11/12/13/14a	Chr2:9021027..9022429+	scaffold_1:249461..250856	Chr2:6485106..6486499-	chr2:10023082..10024749+
FGF11/12/13/14b	Chr2:9744442..9747660+	scaffold_1:1120013..1122469+	Chr2:9655638..9658003-	chr2:9472380..9474932+
FGF11/12/13/14c	PAR:11284391..11285677+	scaffold_33:397235..398599+	PAR:12145653..12146861-	PAR:11065173..11066400-
FGF11/12/13/14d	PAR:6452375..6453925-	scaffold_25:418164..419636-	PAR:8503048..8504244-	PAR:12086264..12087704+

1.1.2.2. Identification of the *Fgf* genes in OSA and OKI genomes

After identifying the *Fgf* genes in the BAR *O. dioica* genome, we analysed the *Fgf* catalogue in OSA and OKI. Using the hmmsearch approach on transcriptomic data from each cryptic species, combined with intraspecies genomic tBLASTn searches, we identified ten *Fgf* loci in both OSA and OKI. Our RBBH analyses using BAR Fgfs as original queries suggested that the complete *O. dioica Fgf* catalogue comprises ten one-to-one orthologous genes in all cryptic species **(Table 1 and Figure 1.2)**.

The extreme evolutionary rate of *O. dioica*, characterized by a marked tendency to rapidly lose or degenerate non-essential genetic elements, allows for the identification and annotation of functional elements based on sequence conservation among the cryptic species. This, along with the transcriptomic data and gene models prediction for both OSA and OKI, which sometimes differed among themselves and from those in BAR, enabled us to curate or manually annotate the *Fgf* genes in the three cryptic species genomes **(Table 1)**. Phylogenetic analysis of all the *Fgf* genes found in the three species plus the NOR orthologs corroborated the one-to-one orthology and confirmed the absence of secondary lineage-specific gains or losses in any of the cryptic species **(Figure 1.2).** Moreover, it revealed that all *O. dioica* Fgf genes are grouped into two major clades with great node support (99.9/100). This provided the first strong evidence supporting that the ten *O. dioica Fgf* genes could be classified into two subfamilies (red and blue background in Figure 2), as will be assigned in section 1.2.

Regarding the geographical regions from which the four genomic assemblies derived, BAR and NOR always grouped together forming a clade, but contrary to what would be expected according to geographical constraints, OSA and OKI did not form sister clade. Instead, in nine out of the ten cases OSA stood independently as the sister group of the BAR-NOR pair, while OKI stood as the sister group of all the others in a basal position **(Figure 1.2)**. This finding reinforces the predictions made based on whole

genome alignment and phylogenetic inference, in which OKI stood in a basal position relative to OSA and BAR **(Figure 1.2)** (Plessy et al., 2024).

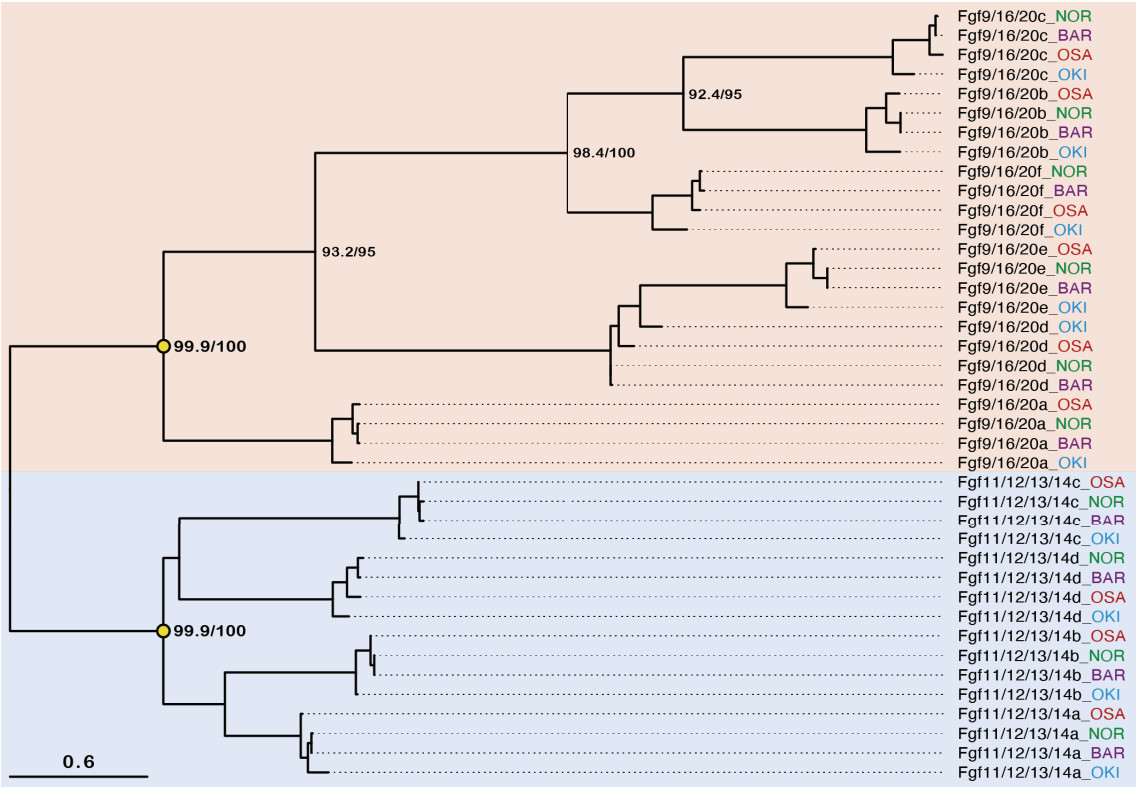


Figure 1.2. Maximum likelihood tree of all *Oikopleura dioica* Fgf genes identified in four populations. The names of the genes derive from their classification described in the following subsections. Node support values are indicated as SH-aLRT support (%) / ultrafast bootstrap support (%).

1.1.3. Comparative genomic context of *O. dioica Fgf* genes among the three cryptic species

To further corroborate the orthology of the identified *Fgf* genes within the *O. dioica* lineage and to better understand their genomic context, we conducted a thorough examination by comparative macrosynteny and microsynteny analyses among the three cryptic species.

On the one hand, analysis of macrosynteny showed that all *Fgf* loci were distributed along two of the three chromosomes that build *O. dioica* genome, namely chromosome 2 (Chr2) and the pseudoautosomic region (PAR) of the sexual chromosome. However, their precise loci have varied substantially among the cryptic species in most cases **(Figure 1.3)**. With *Fgf11/12/13/14b* and *Fgf9/16/20c* showing the most conserved loci in terms of relative position, or *Fgf11/12/13/14d* displaying the most distant loci among the three cryptic species, the general overview pointed to an absence of selective pressure in conserving their genomic loci, although notably each *Fgf* gene has been kept in the same chromosomal arm in the three cryptic species.

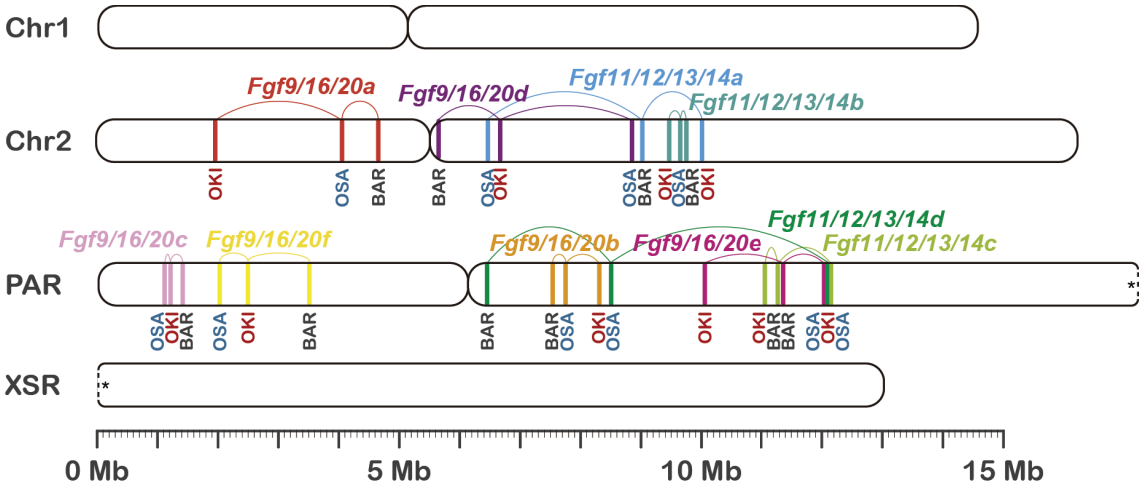


Figure 1.3. Comparative chromosome location of *Fgf* genes in the three *O. dioica* cryptic species. Macrosynteny analysis comparing the position of Fgf genes at chromosome level. Each Fgf gene is labeled with a distinctive color.

On the other hand, analysis of microsynteny revealed a strong conservation among the closest neighbouring genes of most *Fgf* genes in the three cryptic species, providing robust support to their assigned homologies **(Figure 1.4).** Despite the overall conservation at the microsyntenic level, numerous inversions and translocations of neighbouring genes were observed in most *Fgf* loci among the cryptic species. The differential occurrence of these inversions and translocations in each locus allowed us to identify several trends. First, that in agreement with their phylogeny, BAR and OSA *Fgf* genomic contexts are more conserved than any of the two with OKI, as it is case for *Fgf11/12/13/14a* or *Fgf11/12/13/14b*. Second, that genes placed in short chromosome arms (i.e. *Fgf9/16/20a, Fgf9/16/20c* and *Fgf9/16/20f*) show a much more scrambled neighbourhood than the ones placed in long arms. And third, that there is no evident correlation of microsynteny conservation with macrosynteny conservation. This last trend can be easily noticed when comparing *Fgf11/12/13/14a* with *Fgf11/12/13/14b,* as they show an equally high degree of microsynteny conservation even though the genomic loci of the former are much more distant than the ones of the latter; or *Fgf9/16/20c with Fgf11/12/13/14d,* the former showing a very scrambled neighbourhood even though its genomic loci are quite conserved, and the latter showing a quite conserved neighbourhood even though its genomic loci are the most distant **(Figure 1.3 and Figure 1.4)**.

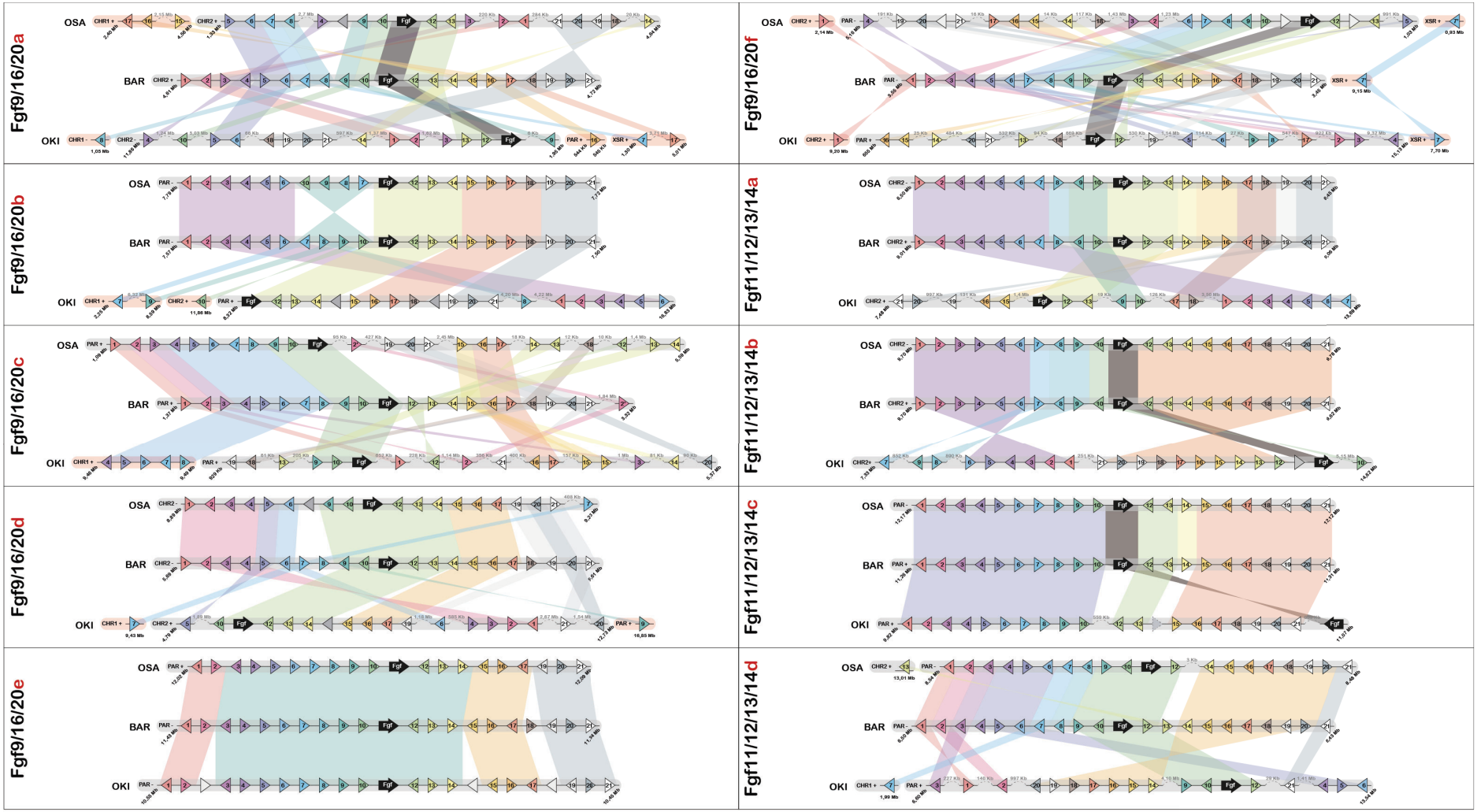


Figure 1.4. Comparative microsynteny conservation in the *Fgf* genes among the three *O. dioica* cryptic species. Comparison of microsynteny conservation between the genomic neighbourhoods of Fgf genes. For each Fgf paralog, the BAR genome was used as the reference in comparisons with OSA and OKI. The block containing the Fgf gene (black arrow) and the ten adjacent genes on each side was mapped in the OSA and OKI genomes.

1.2. Evolution of the *Fgf* genes in Appendicularians

1.2.1. Phylogenetic analysis and classification of *O. dioica Fgf* genes

To place the *Fgf* genes identified in *O. dioica* within an evolutionary framework, we conducted a phylogenetic analysis of the Fgf family across chordates. This involved retrieving well-annotated Fgf sequences from vertebrates, from the well-studied ascidian *C. robusta,* and from three cephalochordate species available in public databases. To enhance the robustness of our analysis we aimed to include additional tunicate *Fgf* orthologs beyond *C. robusta*, so we searched for *Fgf* sequences in eleven ascidian species available at the ANISEED database (Brozovic et al., 2018). We used the hmmsearch tool to look for proteins carrying an FGF domain on their predicted proteomes, complemented with RBBH searches on their gene models and genomic sequences using the well-annotated *C. robusta Fgf* genes as queries. Furthermore, we expanded our catalogue of appendicularians Fgfs by including sequences from two other species whose genome has been sequenced, namely *Oikopleura albicans* and *Oikopleura vanhoeffeni* (Naville et al., 2019). These species were selected due to the relatively low fragmentation of their genomic assemblies and their phylogenetic proximity to *O. dioica*, which allowed reliable annotation of their putative *Fgf* genes based on sequence conservation. For these appendicularians species where transcriptomic data is not available, we identified *Fgf* genes based on RBBH searches using all *Fgf* sequences from *O. dioica*, *C. robusta*, *H. sapiens*, and *B. lanceolatum* as original queries to ensure a representation of all *Fgf* subfamilies from the three chordate subphyla. The complete list of retrieved sequences is provided in **Supplementary Table 3**.

Phylogenetic inference provided strong evidence with high node support values indicating that all 10 *Fgfs* identified in *O. dioica* belong to only two subfamilies, namely the Fgf11/12/13/14 and Fgf9/16/20 subfamilies **(Figure 1.5)**. Analyses on gene structure and protein sequence motifs further supported this classification, as detailed in the following subsections. Moreover, all Fgfs identified in the other two appendicularians species (i.e. *O. albicans* and *O. vanhoeffeni*) were grouped into the same two subfamilies identified in *O. dioica*. Our survey of 11 ascidian species revealed that most ascidians had seven *Fgf* genes orthologous to the seven Fgfs previously described in *C. robusta*, which are representatives of all chordate subfamilies except Fgf1/2 and Fgf3 (Oulion et al., 2012; Popovici et al., 2005; Satou et al., 2002). The exceptions were *C. savingy*i, in which we could not find an ortholog for *Fgf4/5/6*, and *H. roretzi*, *H. auriantum*, *B. leachii* and *Botryllus spp.*, in which we could not find orthologs for *Fgf-NA1*. These findings, therefore, suggested that the loss of the Fgf1/2 and Fgf3 subfamilies plausibly occurred in the last common ancestor of all tunicates before the split of appendicularians and ascidians. While ascidians have not systematically lost any further Fgf subfamily, appendicularians have lost four additional subfamilies before the radiation of the clade (i.e. Fgf4/5/6, Fgf7/10/22, Fgf8/17/18/24 and Fgf19/21/23 subfamilies) **(Figure 1.5).**

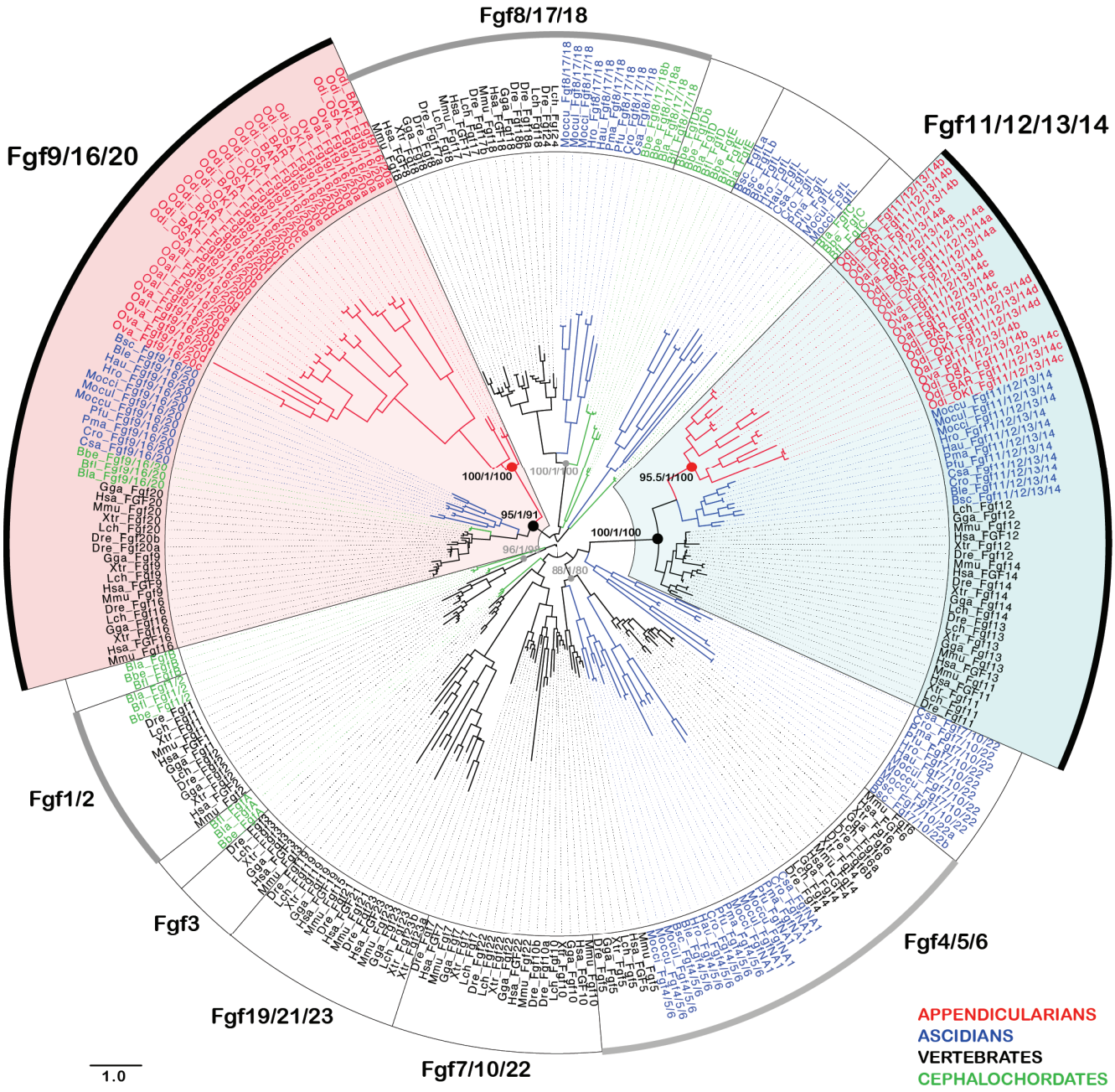


Figure 1.5. Evolutionary tree of the Fgf subfamilies in chordates. ML phylogenetic tree of the Fgf family in chordates reveals that the 10 *Fgf* genes found in *O. dioica*, together with all *Fgf* genes found in other appendicularian species (in red), group in two clusters with high support values (nodes with red solid circles). The tree topology indicates that the two clusters belong to two subfamilies with high support values (nodes with black solid circles). These subfamilies were the Fgf9/16/20 (red background) and the Fgf11/12/13/14 (blue background). The inclusion of Fgfs from ascidians (in blue), vertebrates (in black) and cephalochordates (in green) allowed to infer that appendicularians have lost the Fgf8/17/18, Fgf19/21/23, Fgf7/10/22 and Fgf4/5/6 subfamilies. Well-supported nodes of other Fgf subfamilies (aBayes=1) with members of more than one subphyla are indicated with grey solid circles. Node support values correspond to likelihood-based methods aLRT-SH-like/aBayes/uf-bootstrap. The scale bar indicates amino-acid substitutions. Species abbreviations: Vertebrates (in black): *Danio rerio* (Dre), *Gallus gallus* (Gga), *Homo sapiens* (©), *Latimeria chalumnae* (Lch), *Mus musculus* (Mmu), *Xenopus tropicalis* (Xtr); Ascidian tunicates (in blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Mocci), *Molgula occulta* (Moccu), *Molgula 49ppendi* (Mocul), *Phallusia 49ppendic* (Pfu), *Phallusia mammillata* (Pma); Appendicularian tunicates (in red): *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (in green): *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla).

Our phylogenetic analysis also indicated that the massive loss of Fgf subfamilies during the evolution of appendicularians was accompanied by a burst of duplications of the two surviving ones. This resulted in four *Fgf11/12/13/14a-d* paralogs and six *Fgf9/16/20a-f* paralogs in *O. dioica* **(Table 1 and Figure 1.5)**. The fact that all *Fgf* genes in the other two appendicularians surveyed also appeared as paralogs of these two subfamilies suggests that their expansion predated the radiation of the appendicularians clade. Moreover, each subfamily might have undergone further lineage-specific expansions **(Figure 1.5)**.

1.2.2. Analysis of gene structures in *O. dioica Fgf* genes

To further support the conclusion that all *Fgfs* in *O. dioica* belong to only two subfamilies, namely the Fgf9/16/20 and Fgf11/12/13/14 subfamilies, we analysed and compared the gene structure of all *O. dioica* Fgf genes with the presumptive orthologs of other chordate taxa *(*i. e*. Branchiostoma floridae, Ciona robusta* and *Homo sapiens*; as representatives of cephalochordates, ascidians, and vertebrates, respectively). This analysis provided strong evidence beyond the phylogenetic inference, highlighting the paralogy among *O. dioica* Fgf genes and their orthology with the chordate Fgf9/16/20 and Fgf11/12/13/14 subfamilies. Notably, all Fgf genes in the basal chordate *B. floridae* exhibit two conserved internal introns in the homology core region that encodes the FGF domain (internal core introns 1 and 2: ici1 and ici2). This suggests that the founder gene for each chordate Fgf subfamily also had these two internal core introns **(Figure 1.6)**.

In the case of the Fgf9/16/20 subfamily, all vertebrate genes and the ascidian ortholog have retained the two internal core introns present in the cephalochordate *Fgf9/16/20* ortholog without further modifications to gene structure. In contrast, among *O. dioica Fgf9/16/20* paralogs, only *Fgf9/16/20a* has retained these two ancestral introns, while all other paralogs (*Fgf9/16/20b-f*) exhibit an intron-less gene structure **(Figure 1.6)**. This, coupled with the basal position of *Fgf9/16/20a* in the phylogenetic trees **(Figures 1.2 and 1.5)**, suggests that the Fgf9/16/20 subfamily expansion likely resulted from the integration of a retrotranscribed ancestral *Fgf9/16/20a-like* gene. To test whether the loss of introns could be explained by a retrotransposition event, we examined the 3' region of all *Fgf9/16/20b-f* paralogs for higher adenine content, which could indicate remnants of a recently retrotranscribed PolyA tail. However, none of the intron-less paralogs showed a higher adenine content than the basal *Fgf9/16/20a* **(Supplementary Table 4)**. Given *O. dioica*'s exceptionally rapid evolutionary rate, this outcome was unsurprising and suggests that the putative retrotransposition event was not recent, or that the expansion was due to the total loss of introns by other means followed by multiplication of such intron-less *Fgf9/16/20* paralog. The absence of introns in most *Fgf9/16/20* genes found in other appendicularians species supports the idea that this subfamily expanded ancestrally in this clade.

In the Fgf11/12/13/14 subfamily all vertebrate members have retained the two internal core introns, while in ascidian tunicates, ici1 has been lost, and only ici2 has been preserved. Additionally, we identified two flanking introns external to the core FGF domain in all vertebrate and tunicate *Fgf11/12/13/14* genes (core flanking introns 1 and 2: cfi1 and cfi2) **(Figure 1.6)**. These core flanking introns are absent in all *B. floridae Fgf* genes, what suggests that they are a conserved synapomorphy of this subfamily within the clade olfactores. In *O. dioica*, all presumptive Fgf11/12/13/14 paralogs retain the two olfactores-specific core flanking introns (cfi1 and cfi2), and *Fgf11/12/13/14b* also retains the ancestral internal core intron 2 (ici2), similar to the ascidian ortholog. The exclusive presence of a third internal core intron (ici3) in all *O. dioica Fgf11/12/13/14* genes strongly supports the hypothesis that the four paralogs result from duplications in the appendicularian lineage. Moreover, the presence of two additional and gene-specific internal core introns in *O. dioica Fgf11/12/13/14d*, one in *Fgf11/12/13/14a*, and an OKI-exclusive internal core intron in *Fgf11/12/13/14c* highlights the dynamic evolution of gene structure in this paralogous group within the appendicularians lineage **(Figure 1.6)**.

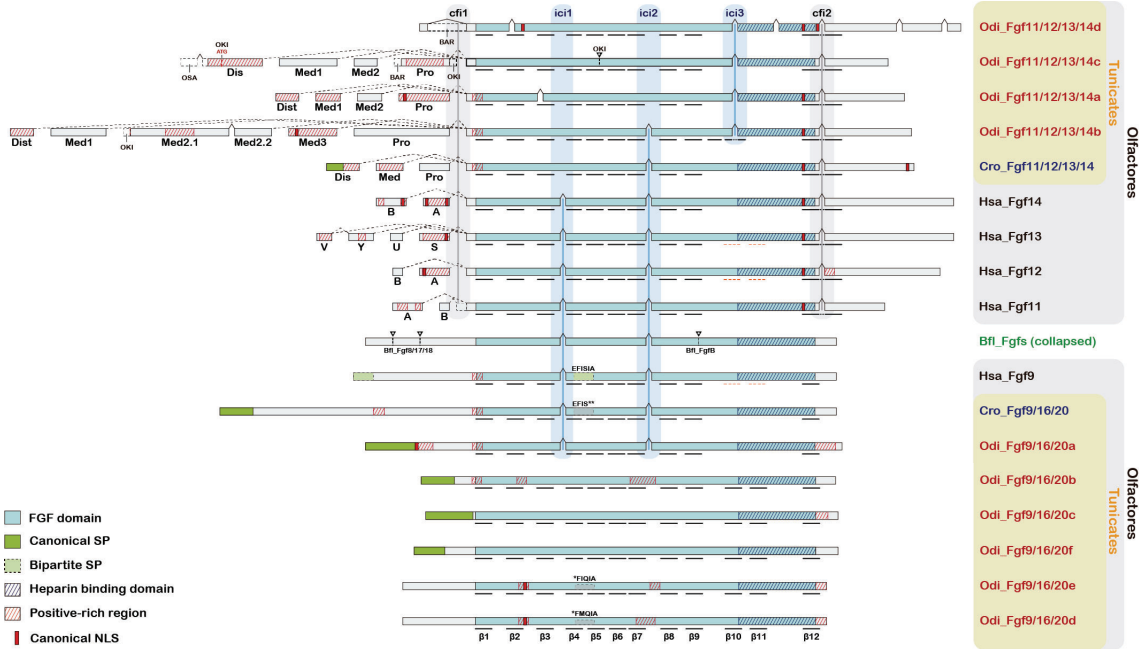


Figure 1.6. Comparative gene structures of *O. dioica* and other chordate *Fgf* genes. Exon-intron organization supports the phylogenetic classification of *O. dioica Fgf9/16/20* and *Fgf11/12/13/14* paralogs. "cfi" denotes conserved core flanking introns, and "ici" denotes conserved internal core introns. Bfl_Fgfs represent the common structure of cephalochordate Fgf genes, featuring two internal core introns (ici) within the FGF domain coding sequence. Gene-specific introns are depicted as arrowheads and dashed lines at their respective locations. Predicted functional motifs are indicated as described in the legend. Black underlines highlight the presence and location of β-sheets as predicted by AlphaFold2. Orange dashed underlines highlight the presence and location of β-sheets that have been empirically determined, even though the AlphaFold2 software does not predict them (Goetz et al., 2009; Olsen et al., 2003; Plotnikov et al., 2001). For comparative purposes, genes and motifs are not drawn to scale. Dashed lines indicate alternative splicing variants of *Fgf11/12/13/14* paralogs (Dis = distal, Med = medial, Pro = proximal), and black dashed lines boxes indicate exon length differences between *O. dioica* cryptic species.

Beyond the FGF domain and its flanking introns, transcriptomic data revealed at least four alternative first exons (distal, proximal, middle 1, and middle 2) in most *Fgf11/12/13/14* paralogs **(Figure 1.6)**. These exons give rise to different isoforms due to alternative splicing and transcription start sites. The presence of alternative isoforms with different N-termini in the Fgf11/12/13/14 subfamily is well-documented in vertebrates (Munoz-Sanjuan et al., 2000; Pablo & Pitt, 2016). Our genome database surveys also identified alternative splice variants of the first exon in the *Fgf11/12/13/14* ortholog of *C. robusta* (GeneID: 445758 in NW_004190431.2) and other ascidian species, suggesting that the presence of alternative first exons is a conserved ancient trait of the Fgf11/12/13/14 subfamily in olfactores. Small differences in alternative splice variants among BAR, OSA and OKI, such as the incorporation of the first intron of *Fgf11/12/13/14d* into the open reading frame in BAR but not in the others, along with the aforementioned cryptic species-specific introns, illustrate the rapid evolution of the *Fgf* genes in appendicularians. These differences provide an example of genetic variation among the cryptic species.

1.2.3. Analysis of protein sequence and conserved motifs

To explore the consequences of the expansion of the Fgf9/16/20 and Fgf11/12/13/14 subfamilies on the diversification of the resulting paralogs, we also analysed their protein sequences and looked for putative functional motifs. Our analysis of sequence similarity among Fgfs revealed that, upon the appendicularians Fgf expansion, one or two members of each subfamily (namely, *Fgf9/16/20a* and *Fgf11/12/13/14a-b*) maintained high similarity with their respective orthologs in other species. In contrast, the other duplicates underwent significant sequence divergence, particularly among the *Fgf9/16/20* paralogs. While protein sequence identity among human Fgf9/16/20 paralogs ranges 62%-69.6% across the entire protein (80.6%-87.8% in the FGF core domain), in *O. dioica* sequence identity among some *Fgf9/16/20* paralogs was as low as 18.7% (21.2% in the FGF core domain), as detailed in **Supplementary Table 5**. To investigate the functional implications of Fgf sequence divergence within each subfamily, we examined their protein domains, conducted protein structure predictions, and looked for potential signal peptides (SP), nuclear localization signals (NLS), and other conserved motifs.

Regarding the preservation of the FGF domain, the high variability in the scores with which HMM profiles were identified in *O. dioica* Fgfs, specially affecting the Fgf9/16/20 paralogs (with e-values ranging from 1^{-35} in *Fgf9/16/20a* to 1^{-7}-1^{-13} in *Fgf9/16/20b-f*), suggested potential variations in the typical β-trefoil fold structure. However, our structural predictions indicated that most *O. dioica Fgf9/16/20* paralogs displayed the twelve β-sheets forming the characteristic β-trefoil fold structure seen in vertebrate Fgf9/16/20 orthologs **(Figure 1.7)**. Additionally, regions enriched in positively charged residues (e.g., arginines and lysines) were found near the end of the FGF domain

in all paralogs, corresponding to the heparan sulphate proteoglycan binding sites (HBS) identified in vertebrates **(Figure 1.7A)** (R. Xu et al., 2012). The presence of a β-trefoil fold and HBS, therefore, suggested that all *O. dioica* Fgf9/16/20 paralogs could potentially interact with Fgf receptors and with heparin to function through the canonical paracrine FGF signalling.

To explore the secretion potential of Fgf9/16/20 paralogs in *O. dioica*, we searched for signal peptides (SP) at their N-termini. All paralogs except *Fgf9/16/20d* and *Fgf9/16/20e* displayed a cleavable SP in their N-termini (likelihood >0.5), suggesting that they can be secreted from the cell through conventional mechanisms **(Figure 1.6)**. This finding is consistent with the description of a similar SP in the ascidian *Fgf9/16/20* ortholog and points to the acquisition of a cleavable SP in *Fgf9/16/20* orthologs as a tunicate synapomorphy (Satou et al., 2002). The presence of a cleavable SP in tunicates explains the degeneration of the EFISIA motif, which in vertebrates and nematodes where Fgf9/16/20 proteins lack a cleavable SP, is required for their extracellular secretion **(Figure 1.6, Figure 1.7A and Supplementary Figure 3)** (Popovici et al., 2004). The lack of a cleavable SP in *Fgf9/16/20d* and *Fgf9/16/20e* can be attributed to a secondary loss in the ancestral *Fgf9/16/20de* gene. Interestingly, this is correlated with a certain conservation or recovery of the EFISIA motifs in *Fgf9/16/20d* and *Fgf9/16/20e* (i.e. TEMQIA and TFIQIA), that produce a hydrophobicity peak similar to the EFISIA motif in other species **(Figure 1.7A and Supplementary Figure 3)** (Miyakawa et al., 1999; Popovici et al., 2004). These findings further support the integration of this paralogous group within the Fgf9/16/20 subfamily, indicate paracrine signalling functions, and suggest potential differences in the secretion mechanisms among paralogs. Further evidence of their classification into the Fgf9/16/20 subfamily is the presence of two cysteine residues in positions that are conserved in all tunicate Fgf9/16/20 orthologs and absent in all other Fgf proteins **(Figure 1.7A)**.

Regarding the Fgf11/12/13/14 subfamily, its members have traditionally been ascribed to intracellular functions, interacting with various proteins as regulators of voltage-gated ion channels (e.g. Na_vs or Ca_vs), as regulators of transcription factors (e.g. islet brain-2 or NEMO), or as players of neuronal cytoskeleton architecture and cell morphology (Pablo & Pitt, 2016). Like all vertebrate Fgf11/12/13/14 proteins, all 53ppendicularians Fgf11/12/13/14 paralogs lack a cleavable SP, providing the first clue of a conserved intracellular function **(Figure 1.6)**. This was reinforced by the conservation in all *O. dioica* Fgf11/12/13/14 paralogs of a Leucine and Arginine pair in positions that have been described as critical for interaction with Na_vs and islet brain-2 **(Figure 1.7A)** (S. K. Olsen et al., 2003; Pablo & Pitt, 2016). These residues are conserved in all vertebrate and ascidian Fgf11/12/13/14 proteins, and absent in all other Fgf subfamilies.

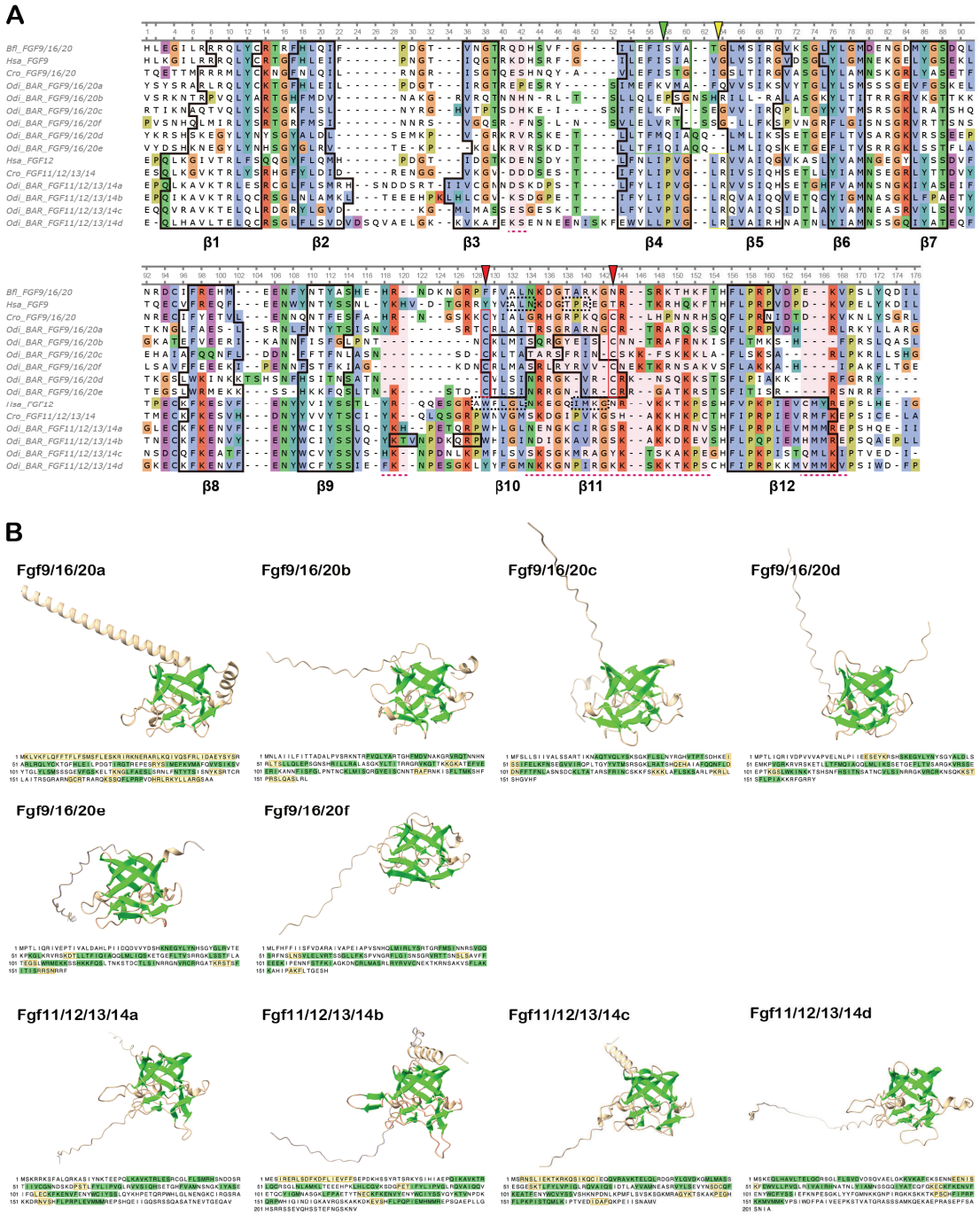


Figure 1.7. (A) Protein alignment of *O. dioica* Fgf orthologs. The alignment includes the FGF domain of each ortholog. Sequence conservation is depicted according to the Clustal X default colouring. Black solid line boxes denote β-sheets as predicted by the AlphaFold2 software, and black dashed line boxes in I_FGF9 and I_FGF12 indicate β-sheets empirically confirmed but not predicted by AlphaFold2 (Goetz et al. 2009; Plotnikov et al. 2001). Red arrowheads and boxes in the *C. robusta* and *O. dioica* Fgf9/16/20 sequences highlight the positions of distinctive and conserved cysteines found in all tunicate Fgf9/16/20 paralogs. Yellow arrowheads and boxes in Fgf11/12/13/14 sequences denote the positions of the Leucine-Arginine pair characteristic of the intracellular Fgf11/12/13/14 orthologs. Magenta shadings and dashed lines indicate the regions involved in binding heparin (Xu et al. 2012). Abbreviations: *Branchiostoma floridae* (Bfl), *Homo sapiens* (I), *Ciona robusta* (Cro), *Oikopleura dioica* (Odi). **(B) Three-dimensional models for *O. dioica* Fgf proteins.** Predicted β-sheets are highlighted in light green in the models as well as in the protein sequences. Predicted α-helices are highlighted in light yellow in the protein sequences. All Fgf11/12/13/14 paralogs models and sequences correspond to the proximal isoform. An alternative colouring of the models according to their predicted local distance difference test (pLDDT) score is provided in **Supplementary Figure 2**.

Upon the growing evidence suggesting that Fgf proteins might also play intranuclear functions (Popovici et al., 2006), we also performed NLS predictions on the protein sequences with NLStradamus and looked for KRVR motifs that have been demonstrated to function as NLS in *O. dioica* (Clarke et al., 2007). Our search revealed that most Fgf11/12/13/14 paralogs consistently had a NLS at the end of the FGF core. Interestingly, we also found NLS in some of the alternative first exons that generated different isoforms diverging at the N-terminus, which suggested that different isoforms not only might have different promoter usage, but also might have different intracellular localizations, as it has been described in vertebrates **(Figure 1.6)** (Pablo & Pitt, 2016). Additionally, we also found NLS in three out of the six Fgf9/16/20 paralogs, including the two that lack a SP and that might have evolved non-secreted functions.

1.2.4. Massive loss of Fgf subfamilies, but a burst of expansion of Fgf paralogs in appendicularians

In summary, our findings unveil *O. dioica* as an unprecedented species among chordates, in which massive gene losses have removed all Fgf subfamilies but two, the Fgf9/16/20 and Fgf11/12/13/14 subfamilies. Moreover, the loss of most Fgf subfamilies has been accompanied by an expansion of the remaining ones, resulting in six Fgf9/16/20 paralogs and four Fgf11/12/13/14 paralogs. Our phylogenetic analyses with a broad set of chordate Fgfs strongly supported the assigned homologies, and analyses on gene structure and conserved sequence motifs provided further support.

Specifically, the inclusion of *O. dioica Fgf9/16/20a-f* paralogs within the chordate Fgf9/16/20 subfamily is supported by several findings:

(i) *Fgf9/16/20a*, which is most similar to other chordate Fgf9/16/20 orthologs, occupies a basal position in phylogenetic trees, with the remaining paralogs appearing as intron-less copies of *Fgf9/16/20a* **(Figures 1.2 and 1.5)**.

(ii) Similar to ascidian *Fgf9/16/20* orthologs, most paralogs exhibit an N-terminal SP for extracellular secretion, in contrast to non-tunicate *Fgf9/16/20* orthologs where the EFISIA motif serves for their secretion. The two paralogs lacking the N-terminal SP exhibit a putative EFISIA motif, which serves as a non-cleavable SP in non-tunicate *Fgf9/16/20* orthologs **(Figures 1.6 and 1.7A, and Supplementary Figure 3)**.

(iii) All paralogs possess two conserved cysteine residues unique to tunicate *Fgf9/16/20* orthologs and absent in other Fgf proteins **(Figure 1.7A)**.

Furthermore, the inclusion of *O. dioica Fgf11/12/13/14a-d* paralogs within the Fgf11/12/13/14 subfamily is supported by the following evidence:

(i) The presence of two FGF-core flanking introns conserved in all *Fgf11/12/13/14* orthologs in olfactores, coupled with the existence of

transcript isoforms that differ in their N-terminus due to several alternative first exons **(Figure 1.6)**.

(ii) All paralogs display a conserved and appendicularian-exclusive internal FGF-core intron **(Figure 1.6)**.

(iii) All paralogs display a Leucine-Arginine pair at conserved positions, crucial for interacting with their intracellular partners in all olfactores **(Figure 1.7A)**.

This particular *Fgf* catalogue is a conserved feature among the three *O. dioica* cryptic species, despite their extreme genomic rearrangements. This finding is robustly supported by our phylogenetic analyses and microsynteny conservation studies. The highly dynamic evolution of Fgf genes in appendicularians becomes evident when comparing the Fgfs of different cryptic species of *O. dioica*. These comparisons reveal divergence in the presence of different isoforms due to alternative splicing, significant sequence divergence outside the FGF domain, the presence of novel introns, and microsyntenic rearrangements. Notably, some *O. dioica Fgf* genes exhibit remarkable sequence divergence. However, the conservation of the FGF domain, with its typical β-trefoil topology and heparin binding sites, supports the hypothesis that they can function as Fgf ligands in other animals. The common presence of secretion motifs in Fgf9/16/20 paralogs suggests that they act extracellularly through the canonical paracrine signalling pathway typical of this subfamily (Itoh & Ornitz, 2011; Ornitz & Itoh, 2015). Conversely, the lack of a signal peptide, the conservation of specific residues for interacting with intracellular proteins, and the presence of nuclear localization signals in Fgf11/12/13/14 paralogs suggest that these proteins have intracellular functions, as has been clearly described for members of this subfamily in other chordates (Pablo & Pitt, 2016; Smallwood et al., 1996).

Overall, our findings from the structural analysis were compatible with paracrine functions for the Fgf9/16/20 subfamily and intracellular functions for the Fgf11/12/13/14 subfamily in appendicularians. The high sequence divergence, variation in the presence of putative SP and NLS, and the formation of different isoforms due to differential splice variants in the N-terminus raised the possibility that multiple functions might have also evolved among the different paralogs duplicated during the expansion of these two surviving families in appendicularians.

1.3. Functional diversification of *O.dioica* Fgf paralogs

1.3.1. Developmental atlas of expression of *O. dioica* Fgf genes

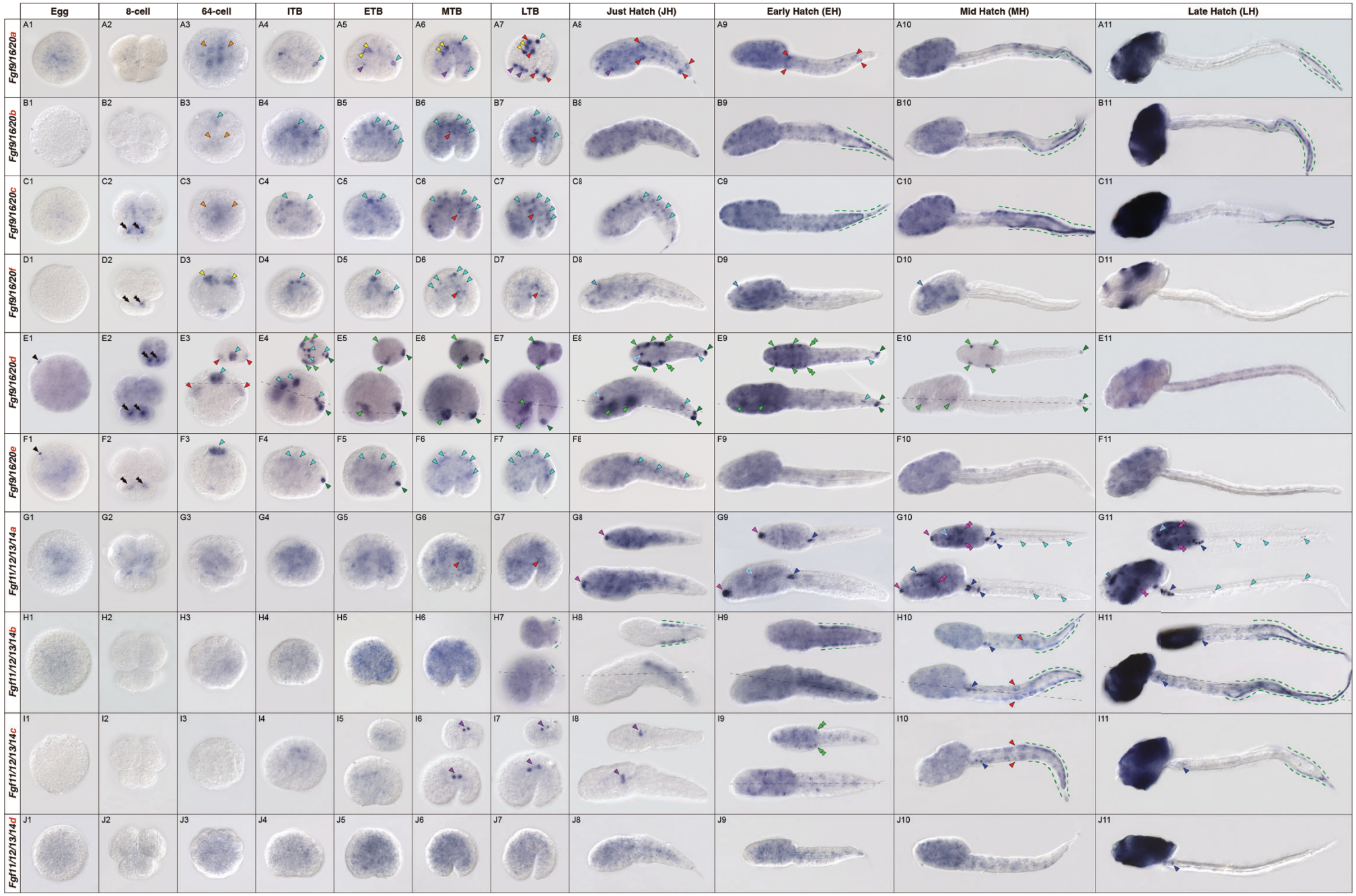
To better understand the functional consequences of the losses and expansions affecting the Fgf subfamilies in *O. dioica*, we conducted an exhaustive expression analysis of all the *Fgf9/16/20* and *Fgf11/12/13/14* paralogs by whole mount in situ hybridization (WMISH) throughout development, from eggs to late-hatchling larval stages **(Figure 1.8**

A-J). In general, the expression signal of most *Fgf* genes was hard to detect, and long periods of staining (up to 14 days) were required to visualize some of the tissue-specific expression domains. Our description mostly focusses on specific *Fgf* expression domains that were repeatedly observed in different embryos or larvae over background signal levels. Nonetheless, we cannot discard that in addition to these specific domains some of the *O. dioica Fgf* paralogs have a generalized basal expression, as it has been described for several *Fgf* genes in other animals (S. Bertrand et al., 2011; Imai et al., 2004).

1.3.1.1. Early expression

In oocytes, some of the *Fgf* genes (i.e. *Fgf9/16/20a,d,e,* and *Fgf11/12/13/14a,b,d*) exhibited a weak staining in the cytoplasm, usually difficult to distinguish over the background noise. This suggested that *Fgf* transcripts are not a major component of the maternal contribution **(Figure 1.8 A-J1)**. However, for *Fgf9/16/20d* and *Fgf9/16/20e*, we observed a small but intense staining spot in the cortical area of many unfertilized eggs (n=11/13 and n=9/11, for *Fgf9/16/20d* and *Fgf9/16/20e*, respectively). The fact that not all eggs exhibited this staining spot suggested that its formation might be transient and difficult to capture, or perhaps not present in all individuals **(Figure 1.8 E1 & F1, black arrowheads)**. In the earliest embryonic stages, from 8-cell to 64-cell, all *Fgf9/16/20* paralogs showed staining signals, what suggests that their expression onset takes place concomitant with the activation of zygotic transcription **(Figure 1.8 A-F2 & A-F3)** (K. Wang et al., 2015).

At the 8-cell stage, many *Fgf9/16/20* paralogs (i.e. *Fgf9/16/20c, Fgf9/16/20d, Fgf9/16/20e* and *Fgf9/16/20f*) showed expression signal restricted to the smaller pair of blastomeres in the vegetal pole **(Figure 1.8 C-F2, black double arrowheads)**. These blastomeres correspond to the A/A4.1 pair in Delsman/Conklin nomenclature, which gives rise to the majority of neural cells, the notochord, and other mesodermal and endodermal derivatives (Nishida, 2008; Stach et al., 2008). At the 64-cell stage, all *Fgf9/16/20* paralogs showed expression signals in the precursor blastomeres of either mesodermal or ectodermal derivatives. Specifically, *Fgf9/16/20a*, *Fgf9/16/20b* and *Fgf9/16/20c* expression was detected in the ingressing endomesodermal blastomeres **(Figure 1.8 A-C3, orange arrowheads)**; *Fgf9/16/20b*, *Fgf9/16/20d* and *Fgf9/16/20e* in the neural plate **(Figure 1.8 B3, E-F3, cyan arrowheads)**; *Fgf9/16/20f* in notochord precursor cells **(Figure 1.8 D3, yellow arrowheads)**; and *Fgf9/16/20d* in muscle precursor blastomeres **(Figure 1.8 E3, red arrowheads)**. Consistent with these observations, most of the specific *Fgf* expression domains observed in later embryogenesis could also be grouped into ectodermal derivatives such as the central nervous system (CNS) and the epidermis, or endomesodermal derivates such as the notochord and muscle cells.


Egg
8-cell
64-cell
ITB
ETB
MTB
LTB
Just Hatch (JH)
Early Hatch (EH)
Mid Hatch (MH)
Late Hatch (LH)
Fgf9/16/20a
Fgf9/16/20b
Fgf9/16/20c
Fgf9/16/20f
Fgf9/16/20d
Fgf9/16/20e
Fgf11/12/13/14a
Fgf11/12/13/14b
Fgf11/12/13/14c
Fgf11/12/13/14d

«Figure 1.8. Developmental expression atlas of *O. dioica Fgf* genes. Whole-mount in situ hybridization images of *O. dioica* at various developmental stages: eggs (A-J1), 8-cell embryos (A-J2), 64-cell embryos (A-J3), incipient tailbud (ITB) embryos (A-J4), early tailbud (ETB) embryos (A-J5), mid tailbud (MTB) embryos (A-J6), late tailbud (LTB) embryos (A-J7), just hatchlings (A-J8), early hatchling larvae (A-J9), mid hatchling larvae (A-J10), and late hatchling larvae (A-J11). Central images in each panel are left lateral views oriented anterior toward the left and dorsal toward the top. Upper-right image insets (') are dorsal views of optical cross-sections at the levels indicated by black dashed lines. Black arrowheads label an stained cortical area in unfertilized eggs; black double arrowheads label the A pair blastomeres in 8-cell embryos; orange arrowheads mark ingressing vegetal blastomeres in 64-cell embryos; blue arrowheads label neural derivatives (cyan-blue labels the neural plate in 64-cell and ITB embryos, and the nerve cord in later stages; dark-blue labels the caudal ganglion, and pale-blue labels the anterior brain); light green arrowheads label epidermal domains in the trunk and light green double arrowheads label the primordia of the Langerhans receptors; dark green arrowheads label epidermal domains in the tailbud tip; green dashed lines mark the lateral epithelium of the tail and the fins; magenta arrowheads mark the mouth primordium and magenta double arrowheads mark the pharyngeal slits; yellow arrowheads label notochord cells; red arrowheads label muscle precursor cells and muscle cells in the tail; purple arrowheads label endomesodermal derivatives of unknown identity.

1.3.1.2. Neuroectodermal derivatives

Among neuroectodermal derivatives, the expression signal of *Fgf9/16/20* paralogs detected in the neural plate could be followed in cells of the developing CNS. This neural expression was observed in the developing brain, the caudal ganglion, and the nerve cord up to the mid-hatchling (MH) larvae **(Figure 1.8A-F, different tones of blue arrowheads)**. Notably, *Fgf9/16/20d* displayed a very dynamic neural expression pattern, exhibiting intense staining in the neural plate in 64-cell embryos and in the posterior part of the developing brain in incipient tailbud (ITB) embryos, but disappearing abruptly in later tailbud stages **(Figure 1.8 E3-4, cyan arrowheads)**. It also displayed a clear expression in an undetermined cell within the developing brain in the just-hatched (JH) larvae **(Figure 1.8 E8, light blue arrowhead)**, and in a distinct cell at the posterior tip of the nerve cord in the JH and early-hatchling (EH) larvae **(Figure 1.8 E8-9, cyan arrowheads).** This dynamic expression suggests that some *Fgf9/16/20* paralogs might be precisely regulated in specific subsets of neural cells along the anteroposterior (AP) axis during the formation of the CNS. In these neural derivatives, we also detected expression signals for the *Fgf11/12/13/14* genes. However, in contrast to the *Fgf9/16/20* paralogs that were mostly expressed in embryonic stages, *Fgf11/12/13/14* paralogs were mostly expressed in larval stages. *Fgf11/12/13/14* expression in neural cells was detected in several specific locations along the CNS. These included cells of the brain dorsally located to the sensory vesicle, as well as the ventral region of the ciliary funnel **(Figure 1.8 G10-11, light blue arrowheads)**, specific groups of cells in the caudal ganglion **(Figure 1.8 G9-11, H10-11, I10-11 & J11, dark blue arrowheads)**, and isolated cells in specific positions along the nerve cord **(Figure 1.8 G10-11, cyan arrowheads)**.

1.3.1.3. Epidermal derivatives

Among other ectodermal derivatives, we also observed distinct specific domains of various *Fgf* genes in the epidermis both in the tail and in the trunk **(Figure 1.8, green arrowheads and dashed lines)**. In the tail epidermis, the expression of several *Fgf* genes was detected at different levels of the AP axis, revealing a dynamic expression that

affected two structures: the developing fin and the tip of tail. In the developing fin, we first detected *Fgf11/12/13/14b* expression in a bilateral pair of precursor cells in the tail at the late tailbud (LTB) stage **(Figure 1.8 H7, green dashed lines)**. This expression spread to the anterior third of the tail in the JH larvae, and eventually to the posterior half of the tail in MH and late-hatchling (LH) larvae **(Figure 1.8 H7-11, green dashed lines)**. Notably, in these later MH and LH stages, *Fgf11/12/13/14c*, *Fgf9/16/20a*, *Fgf9/16/20b* and *Fgf9/16/20c* also exhibited expression signals in the developing fin **(Figure 1.8 A10-11, B9-11, C9-11 & I11, green dashed lines)**. In the tip of the tail, a pair of epidermal cells showed strong staining for *Fgf9/16/20d* and *Fgf9/16/20e* as early as in the ITB stage. While *Fgf9/16/20e* expression was downregulated by the mid-tailbud (MTB) stage, *Fgf9/16/20d* expression persisted until the MH larval stage **(Figure 1.8 E4-10 & F4-5, dark green arrowheads)**.

In the trunk epidermis, we detected *Fgf9/16/20d* expression in bilateral groups of epidermal cells at different levels of the AP axis from the ITB to the MH stage **(Figure 1.8 E4-10, light green arrowheads)**. The most posterior group included the area of the primordia of the Langerhans receptors, which also exhibited expression signal for *Fgf11/12/13/14c* **(Figure 1.8 E8-E9 & I9, light green double arrowheads)**. In the most rostral region of the trunk epidermis, *Fgf11/12/13/14a* showed an expression domain in a group of subepidermal cells in the area of the mouth at the JH stage. This expression domain was later extended to the epidermal surface by the MH stage, coinciding with the opening of the mouth **(Figure 1.8 G8-10, magenta arrowheads)**. Interestingly, *Fgf11/12/13/14a* expression was also observed in the pharyngeal slits in the MH and LH larvae **(Figure 1.8 G10-11, magenta double arrowheads)**.

In LH larvae, virtually all *Fgf* genes were strongly expressed in different parts of the house-making organ, namely the oikoblast. Some *Fgfs* showed generalized patterns, while others were restricted or excluded in specific regions. For example, *Fgf9/16/20f* was restricted to cells adjacent to the anterior cells of the field of Fol, the anterior rosette, the field of Martini, the posterior rosette, and its adjacent lateral bands **(Figure 1.8 D11)** (Kishi et al., 2017). In contrast, *Fgf9/16/20a and Fgf9/16/20b* were expressed throughout the entire oikoblast but excluded from the ring of the mouth and the Giant cells **(Figure 1.8 A11 & B11)**. This finding suggests that the FGF signalling may have been recruited for the development of this innovative organ responsible for the formation of the house, as described for many other developmental genes in appendicularians (Mikhaleva et al., 2018).

1.3.1.4. Endomesodermal derivatives

Among endomesodermal derivatives, the notochord exhibited an expression domain for *Fgf9/16/20a* restricted to the first and third cells during the early tailbud (ETB) and LTB stages **(Figure 1.8 A5-7, yellow arrowheads).** In these stages, *Fgf9/16/20a*

expression was also observed in a group of internal cells located bilaterally in the anterior half of the trunk, whose positions were compatible with endomesodermal progenitors of the pharynx, endostyle or buccal glands **(Figure 1.8 A5-8, purple arrowheads)**.

At the LTB stage, a new mesodermal expression domain of *Fgf9/16/20a* appeared restricted to the first and eighth pairs of muscle cells, and it was maintained until the early hatchling stage **(Figure 1.8 A7-9, red arrowheads)**. Other *Fgf* paralogs that showed broad ubiquitous expression often included specific muscle cells with stronger signal than other parts of the embryo. *Fgf9/16/20b*, *Fgf9/16/20c*, *Fgf9/16/20f* and *Fgf11/12/13/14a* were specifically expressed in the MTB and LTB stages in the three anterior left muscle cells of the tail, observed in a ventral position due to the 90º counterclockwise tail rotation **(Figure 1.8 B6-7, C6-7, D6-7 & G6-7, red arrowheads)**. *Fgf11/12/13/14b* and *Fgf11/12/13/14c* were expressed in the posterior pairs of tail muscle cells in the MH larvae (**Figure 1.8 H10 & I10, red arrowheads)**.

From ITB to LTB stages, *Fgf11/12/13/14c* expression was observed in two distinct cells at the right side of the anterior part of the notochord, which later at the EH stage were located anteriorly in a rostral position to the notochord **(Figure 1.8 I6-8, purple arrowheads)**. We could not determine the nature of these endomesodermal cells, but considering their position, it is plausible that they could be related to the development of progenitors of the digestive system or the gonad (L. C. Olsen et al., 2018).

1.3.2. *O. dioica* Fgf expression suggests functional diversification

Overall, the expression atlas of *O. dioica Fgf* genes during embryonic and larval stages revealed a highly dynamic expression throughout development. *Fgf9/16/20* and *Fgf11/12/13/14* paralogs exhibited expression in a variety of tissues, including the CNS, several epidermal domains, and specific notochord and muscle cells, suggesting their involvement in the development of these structures. Interestingly, *Fgf9/16/20* paralogs were broadly expressed from very early in development and during gastrulation, determination of cellular lineages, and early organogenesis, mostly during embryonic stages **(Figure 1.9)**. This suggests a conserved role for the FGF signalling in early developmental processes, as it has been described across very distant taxa (Andrikou & Hejnol, 2021). *Fgf11/12/13/14* paralogs started their expression in later embryonic stages, and throughout larval development they displayed a dynamic expression in neural structures, muscle cells, and epidermal and subepidermal domains **(Figure 1.9)**. Interestingly, we found expression of *Fgf11/12/13/14* paralogs in the primordia of several organs and structures where ciliated sensory cells develop. These included the expression of *Fgf11/12/13/14a* in the mouth primordium, the ciliary funnel in the brain, or the ciliary rings in the pharyngeal slits, as well as the expression of *Fgf11/12/13/14c* in the mechanorreceptors of the Langerhans cells **(Figure 1.9)**. These findings suggested that appendicularian *Fgf11/12/13/14* orthologs might be involved in the development

of placodal derivatives and other structures where epithelial perforation and fusion of tissues occur, as it has been described in other chordates (Bassham et al., 2008; Bassham & Postlethwait, 2005; Kourakis & Smith, 2007).

To conclude, the evolution the Fgf family in *O. dioica* serves as a paradigmatic example of genetic rearrangement of an entire signalling pathway affected by gene loss. Despite the extensive loss of Fgf subfamilies, the expansion and diversification of the remaining Fgf genes appear to have preserved key developmental processes, but also to have innovated some expression domains. This highlights the adaptive evolution of the Fgf gene family in *O. dioica*. The potential evolutionary implications of this drastic rearrangement of the chordate Fgf catalogue in appendicularians will be discussed in detail in section 4. Given the impressive evolutionary remodelling of the Fgf catalogue, we aimed to investigate whether similar events have impacted other components of the FGF signalling pathway in appendicularians. These included the Fgf receptors and the intracellular transduction pathways.

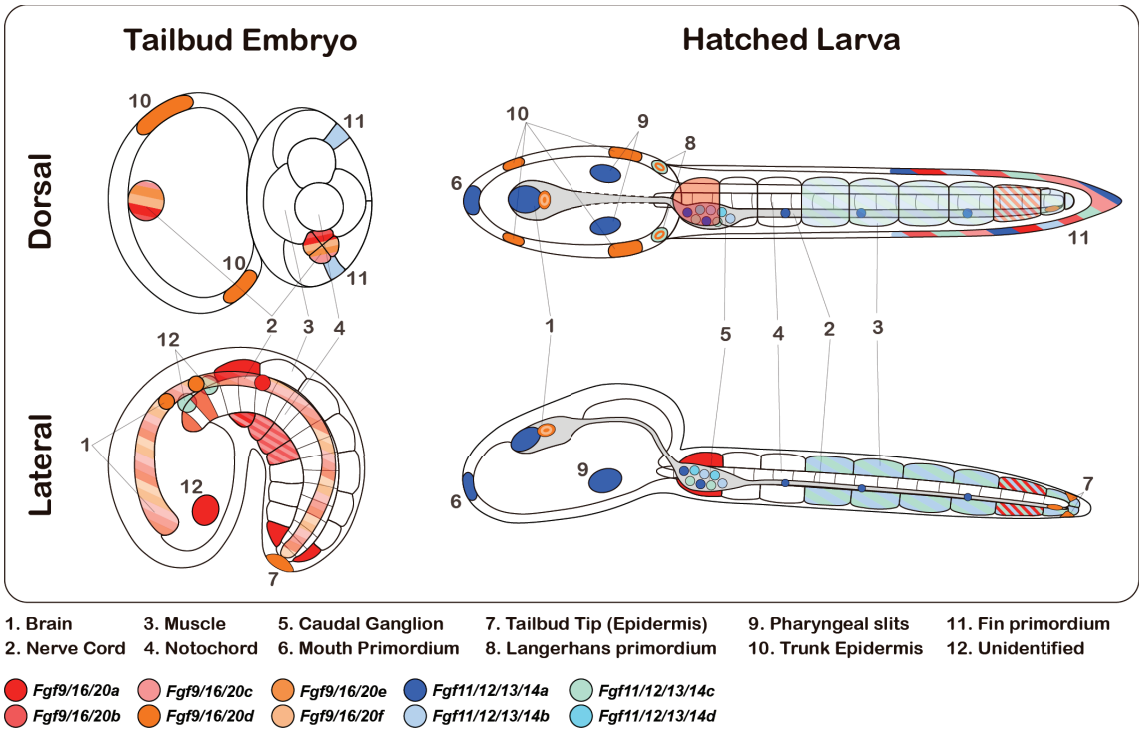


Figure 1.9. Schematic representation of the specific expression domains of *Fgf* genes during the embryonic and larval development of *O. dioica*. On the left, the schematic tailbud embryo highlights specific *Fgf* expression domains observed from the ITB to LTB stages. On the right, the schematic hatched larva highlights specific *Fgf* expression domains observed from the JH to LH larval stages, excluding those in the oikoblast. The CNS in the schematic hatched larvae, including the brain, nerve cord, and caudal ganglion, is depicted in grey for visualization purposes.

2. Evolution of the Fgf receptors and transduction pathways

The massive loss of Fgf subfamilies in appendicularians was accompanied by the expansion and functional diversification of the surviving *Fgf9/16/20* and *Fgf11/12/13/14* paralogs in *O. dioica*. These findings prompted us to investigate the evolution of the Fgf receptor (FgfR) and the downstream components of the FGF signalling pathway involved in transducing extracellular signals into cellular responses.

2.1. The Fgf receptors

2.1.1. Evolution of the Fgf receptor in appendicularians

Our findings showed that in parallel to the expansion and diversification of the surviving Fgf subfamilies, the *FgfR* gene has undergone a similar process of gene expansion and diversification in the appendicularian lineage.

Searches by RBBH allowed us to identify three *FgfR* genes (*FgfRa-c*) in each of the three *O. dioica* cryptic species. Phylogenetic analyses indicated that the three *FgfR* genes hold a relation of one-to-one orthology among the three cryptic species **(Figure 2.1A)**. Considering that ascidians and cephalochordates hold a single *FgfR* gene (D'Aniello et al., 2008; Kamei et al., 2000; Satou et al., 2003), and that the four vertebrate *FgfR* genes resulted from the two rounds of whole genome duplication that occurred early at the base of the vertebrate lineage (Itoh & Ornitz, 2004), the three *O. dioica FgfR* genes are paralogs resulting from an expansion affecting the *FgfR* gene in this species lineage. The fact that we also found several copies of the *FgfR* gene in other appendicularians suggested that the expansion occurred at the base of the appendicularian lineage after their split from the rest of tunicates. However, obtaining *bona fide* FgfR sequences from other appendicularians species was challenging due to the highly fragmented nature of their genomic assemblies and the lack of functional annotations or expression data, coupled with the extended length of *FgfR* gene sequences. In most cases, these genes were fragmented across several scaffolds, making manual reconstruction of the proteins achievable only for the highly conserved tyrosine kinase (TK) domain of the receptor, but not for the extracellular portion typically much more divergent in terms of sequence similarity.

We conducted a phylogenetic analysis including the FgfR TK domain sequences from the three *O. dioica* cryptic species and five other appendicularians species (i.e. *O. albicans*, *O. vanhoeffeni*, *O. longicauda*, *M. erythrocephalus* and *B. stygius*), as well as the FgfR TK domain sequences of various vertebrate, ascidian, and cephalochordate species, and the vascular endothelial growth factor receptor (VEGFR) TK domain

sequences of several chordate species as an outgroup. Results showed that appendicularians FgfRs always grouped together with high node support values **(Figure 2.1A)**. The tree topology suggested that the ancestral appendicularians *FgfR* gene (*FgfRabc*) was duplicated into two paralogs (*FgfRa* and *FgfRbc*), and subsequently one of these paralogs (*FgfRbc*) underwent a second duplication to result in three current paralogs **(Figure 2.1A)**. Overall, the finding of the expansion of the *FgfR* in appendicularians reinforces the notion that the loss of several Fgf subfamilies did not necessarily imply a decline in the complexity of the FGF signalling pathway or a limitation of its potential functions. The expansion of both the remaining Fgf subfamilies and the *FgfR* gene likely provided new opportunities for either subfunctionalization, function shuffling or neofunctionalization of the resulting paralogs.

The rapid evolution affecting the FGF signalling pathway components in *O. dioica* is probably related with the scrambled nature of its genome. Despite the many translocations and other chromosomal rearrangements that characterize the genomes of the three *O. dioica* cryptic species when they are aligned, there is a marked tendency for the orthologous genes and regions to be retained in the same chromosomal arms, as we observed in the case of the *Fgf* genes (see section 1.1.3) (Plessy et al., 2024). In the case of the *FgfR* genes, we observed that *FgfRa* has been kept in the same chromosomal arm in the three cryptic species, as well as *FgfRb* is located in the same arm at least in OSA and OKI genomes, and in an unconnected scaffold in the BAR genome assembly. Nonetheless, *FgfRc* stands as a particular case of chromosomal translocation among the cryptic species, since the orthologs in BAR and OSA genomes are located in the long arm of chromosome PAR, while in the OKI genome the ortholog is found in the short arm of chromosome 2 **(Figure 2.1B)**.

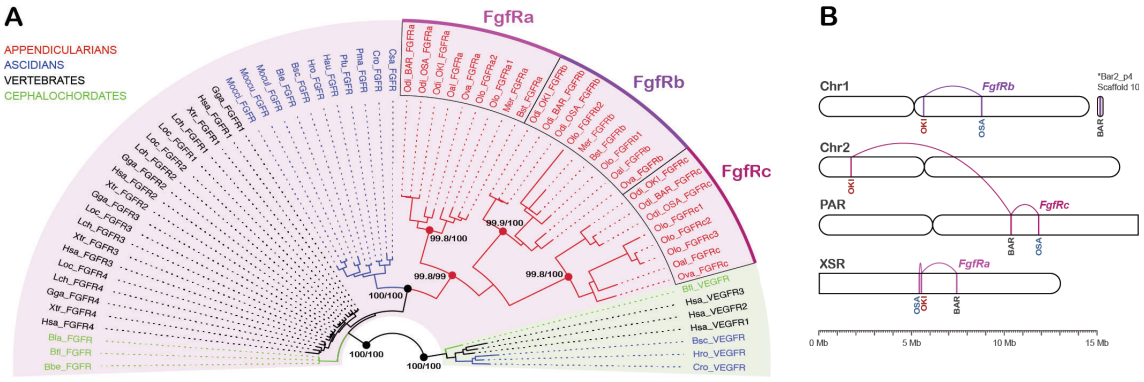


Figure 2.1. Evolutionary relation of appendicularian Fgf receptors. (A) Maximum likelihood phylogenetic tree of chordate FgfR TK domains (shaded in magenta) including the orthologs found in appendicularians. Cephalochordate orthologs are depicted in green, vertebrate orthologs in black, ascidian orthologs in blue, and appendicularian orthologs in red. VEGFR TK domains were added as an outgroup, shaded in green. Support values written on the nodes: SH-aLRT support (%) / ultrafast bootstrap support (%). **(B)** Comparative mapping and chromosome location of *FgfR* genes in the three *O. dioica* cryptic species from Barcelona (BAR), Osaka (OSA) and Okinawa (OKI).

2.1.2. Sequence divergence, but structural conservation in *O. dioica* FgfR paralogs

The three FgfR paralogs in *O. dioica* have evolved to become remarkably divergent when compared to each other and to other chordate FgfR proteins in terms of sequence conservation. To determine whether this sequence divergency corresponded to significant differences in protein structure, we conducted *ab initio in silico* modelling of the three *O. dioica* FgfR paralogs, along with the cephalochordate and ascidian FgfR orthologs for comparative purposes **(Figure 2.2)**. The predictions revealed that despite the remarkable sequence divergence, the three FgfRs displayed similar structures consistent with an FgfR architecture. All contained two immunoglobulin-like (IG-like) domains in their extracellular regions and a split tyrosine kinase (TK) domain in their intracellular regions, connected by a trans-membrane helix **(Figure 2.2)**. This two-IG-like domain architecture differs from the three-IG-like domain architecture found in vertebrate and cephalochordate Fgf receptors (D'Aniello et al., 2008; Ornitz & Itoh, 2015), but it is similar to what has been previously described in *C. robusta* and other ascidians FgfRs (Satou et al., 2003; Shimauchi et al., 2001). This finding indicates that the loss of IG-like domain 1 (IG-L1) was likely ancestral in the tunicate lineage.

Regarding the intracellular portion of the receptors that includes the TK domain, the aminoacidic sequence is well conserved among the three *O. dioica* paralogs, although not to the extent typically found in catalytic domains. The FgfR TK domain exhibits approximately 50% sequence similarity among *O. dioica* BAR paralogs, in contrast to the 80-90% of sequence similarity observed among the TK domains of *H. sapiens* FgfR paralogs **(Supplementary Table 6)**. Additionally, the intron code is completely different among the three *O.dioica* FgfR paralogs **(Figure 2.3A)**. This is most notable considering that the intron code is widely conserved among TK domain-containing protein orthologs, often used to classify a given TK domain-containing protein into a specific protein family **(Figure 2.3B)** (D'Aniello et al., 2008). In contrast to the differences observed among paralogs, the intron code is generally conserved among the orthologs of the three cryptic species, although there are some differences particularly affecting *FgfRc* **(Figure 2.3A)**. The conservation of the TK domain is also noteworthy when assessing sequence similarity among orthologs in different cryptic species. In all cases, BAR FgfR TK domains shared 96-100% sequence similarity with their orthologs in OSA or OKI, with the particular exception of the OKI FgfRc TK domain, where the similarity dropped to 90% **(Supplementary Table 7)**. The lower sequence conservation in the FgfRc TK domain, coupled to the differences in its intron code, may be related to the unique chromosomal change observed in this ortholog among the cryptic species **(Figure 2.1B)**.

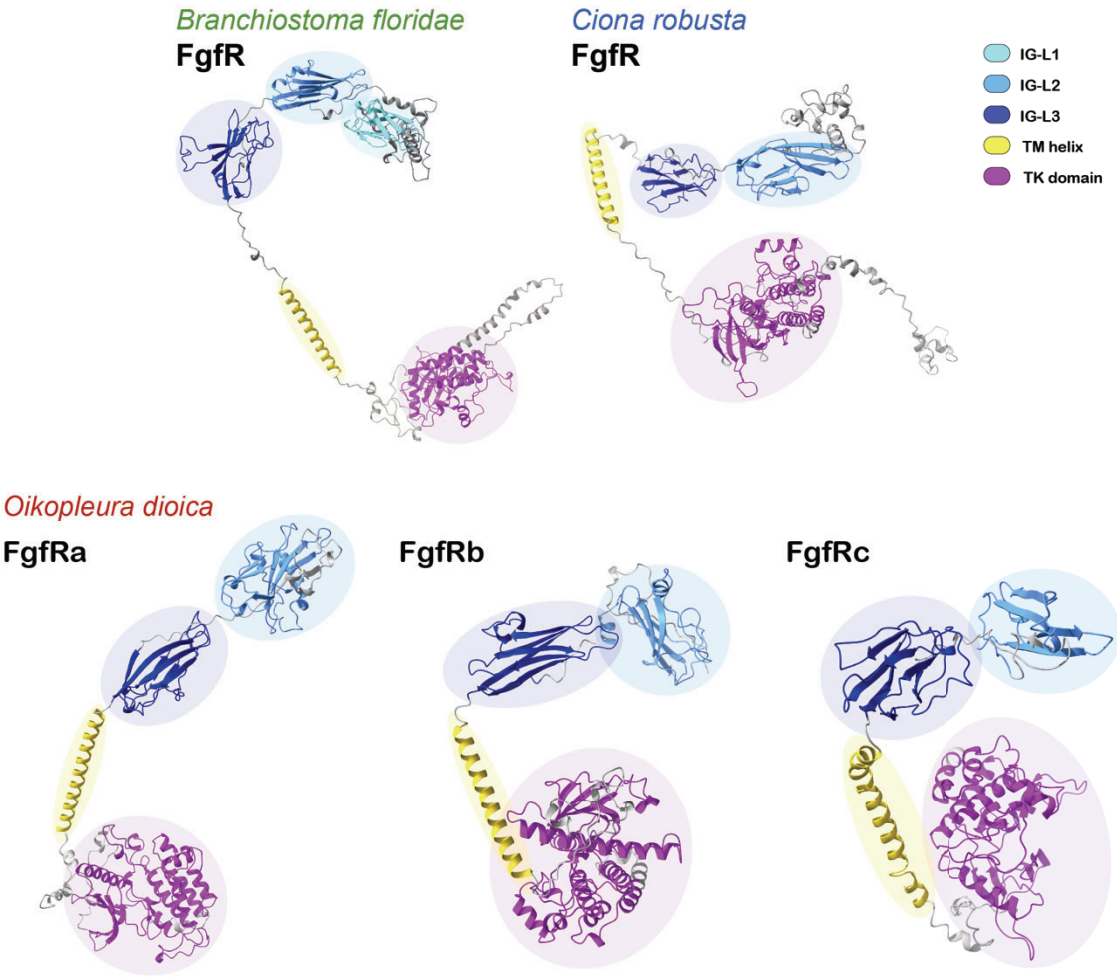


Figure 2.2. Three-dimensional models of the Fgf receptors. *O. dioica* FgfR models are based on the protein sequences of the annotations derived from this project. *B. floridae* FgfR model is based on the protein sequence in XP_035673320.1. *C. robusta* FgfR model is based on the protein sequence in NP_001037820.1. Models were manually coloured to mark the different domains and motifs that build each protein. An alternative colouring according to the local confidence score of the predictions is depicted In **Supplementary Figure 4**.

Regarding the extracellular portion that interacts with Fgf ligands, *O. dioica* FgfRs displayed poor sequence similarity, ranging from 27% to 36% among paralogs. This contrasts with the 64%-74% similarity observed among *H. sapiens* FgfR paralogs **(Supplementary Table 6)**. When we compared the extracellular portion of human or *O. dioica* FgfRs with that of other chordates, we observed that *O. dioica* FgfRs in all cases displayed a lower similarity than human FgfRs, even with their closely related ascidian ortholog **(Supplementary Table 7)**. This indicates an extreme sequence divergency in the extracellular portion of the appendicularians FgfR. Notably, sequence similarity among the FgfR orthologs in the three distinct *O. dioica* cryptic species were significantly lower than for the intracellular TK domain. Similarity values between BAR and OSA orthologs ranged from 90% to 95%, approximately the same range as when comparing the extracellular portions of human and mouse FgfR orthologs. However, similarity values between BAR and OKI orthologs decreased to 79%- 83%, similar to the range observed when comparing human and chick orthologs (70%-90%) **(Supplementary Table 7)**.

Considering that last common ancestor of humans and chickens lived approximately 310 million years ago (mya), and that the last common ancestor of *O. dioica* cryptic species is estimated to have lived 25 mya, the divergence in the extracellular portion of the FgfR among the cryptic species highlights the extreme evolutionary rate of *O. dioica* (Khamsi, 2004; Plessy et al., 2024). The extreme sequence divergency in the extracellular portion of the FgfR paralogs in *O. dioica* may be related with the same degree of divergence observed in the Fgf9/16/20b-f ligands. Further research will be needed to elucidate the specificity of the three FgfRs for binding each Fgf9/16/20 ligand in *O. dioica*. However, it is tempting to speculate that each of the three FgfRs may have become specialized to bind the members of one of the three clades in which O. dioica Fgf9/16/20 ligands can be subdivided (i.e. the basal Fgf9/16/20a, the intron-less SP-carrying Fgf9/16/20bcf, or the intron-less non-SP-carrying Fgf9/16/20de; **Figure 1.2 and 1.6**).

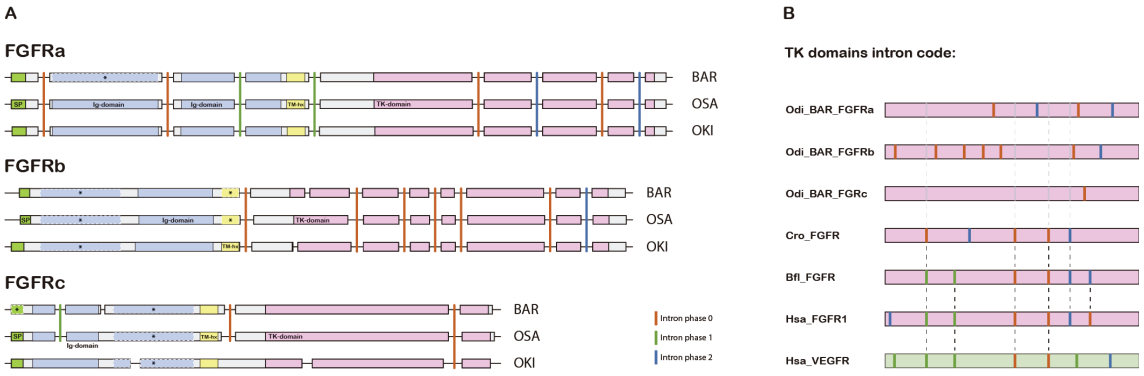


Figure 2.3. *O. dioica FgfR* gene structures. (A) Compared structures of FgfR orthologs from the three *O. dioica* cryptic species. Exons are scaled and depicted as boxes over the non-coding intronic regions (not scaled), depicted as lines. Vertical lines show the relative position of conserved introns, and their phase is indicated with a colour code. Coloured boxes with black solid lines over the grey background illustrate the conserved domains as identified by the InterProScan software. Coloured boxes with white dashed lines marked with an asterisk illustrate domains that were not identified by the InterProScan software, due to a remarked sequence divergency, and whose presence was inferred based on the proteins' 3D structure prediction. **(B)** Intron code of O. dioica FgfR TK domains compared to other chordate orthologs. Solid dashed lines mark the position of conserved introns, and shaded dashed lines mark the position of lost introns in *O. dioica* FgfRs TK domains.

2.1.3. Distinct expression patterns of *O. dioica FgfR* paralogs through development suggest functional diversification

In order to understand the functional consequences of the expansion and diversification of the FgfR genes in *O. dioica*, we performed WMISH assays to detect their expression throughout the different developmental stages **(Figure 2.4)**. Contrary to the generalized low expression of the Fgf genes, two of the FgfR genes (i.e. *FgfRa and FgfRb*) were strongly expressed and the embryos or larvae were completely stained after only 2 hours of coloration. The staining was fast and intense also in the oocytes, revealing therefore that these two genes are part of the maternal component. For the *FgfRc* gene, on the other hand, we did not detect a clear expression signal until the 64-cell stage, and the embryos required longer periods of incubation for the staining to be detected. These observations were consistent with the Gene Expression Matrix data available in the

OikoBase, in which *FgfRa* (GSOIDG00008404001) and *FgfRb* (GSOIDG00009004001) displayed high levels of expression, especially in the earliest stages of development, while *FgfRc* (GSOIDG00010261001) expression was not detected at all **(Supplementary Table 8)** (Danks et al., 2013).

The expression signal of *FgfRa* and *FgfRb* was ubiquitous all over the embryo in virtually all stages, although some obvious tissue specific domains could be distinguished transiently over a generalized staining. *FgfRa* expression signal was detected homogenously all over the embryo until the early hatch stage, when a specific and more intense staining could be distinguished in the muscle cells of the tail **(Figure 2.4 A9-10, red dashed lines)**. This specific staining was maintained until the late hatch stage, the time at which the tail muscles become functional for the larvae to start swimming. Similarly, *FgfRb* expression signal was ubiquitous and homogeneous in the embryos until the late hatch stage, but over the generalized signal a specific and more intense expression domain was evident in the developing notochord from the incipient tailbud to the early hatch stage **(Figure 2.4 B4-9, yellow dashed lines)**. This time frame coincides with some of the main processes of organogenesis of the notochord, including the proliferation, convergency and extension, fusion, and vacuolization of the notochord cells (Søviknes & Glover, 2008). In contrast to the two aforementioned receptor genes, *FgfRc* showed a restricted pattern of expression in all developmental stages examined. Its expression first became clear at the 64-cell stage, when the staining could be detected in some of the inner cells that give rise to endomesodermal derivatives **(Figure 2.4 C3, orange arrowheads)**. In tailbud stages and in the just-hatched larvae, the *FgfRc* gene was specifically expressed in the anterior most left muscle cells in the tail **(Figure 2.4 C4-C8, red arrowheads)**, coincident with the expression detected for *Fgf9/16/20b*, *Fgf9/16/20c* and *Fgf9/16/20f* (see section 1.3); and in the anterior region of the notochord **(Figure 2.4 C7-C8, yellow arrowheads)**. *FgfRc* expression was also detected in the developing neural system, including the developing brain from the MTB embryo to the MH larvae, when the expression signal became generalized in the trunk **(Figure 2.4 C6-9, light blue arrowheads)**, and the caudal ganglion and neuronal bodies along the nerve cord in the EH and MH larvae **(Figure 2.4 C9-C10, dark blue and cyan arrowheads)**.

In LH larvae, when most organs and structures were completing their development to become functional, the three *FgfR* genes showed novel specific expression domains. *FgfRa* and *FgfRb* expression signal was detected in the gonad primordium **(Figure 2.4 A11 and B11, purple arrowheads)**; *FgfRa* expression was also detected in the walls of the pharynx **(Figure 2.4 A11, magenta arrowhead)**; *FgfRb* expression in the pharyngeal slits **(Figure 2.4 B11, magenta double arrowheads);** and *FgfRc* expression was detected in the ventral organ **(Figure 2.4 C11, green arrowhead)**. Moreover, at this LH stage all three *FgfR* genes showed a clear signal of expression in some fields the oikoplastic epithelium **(Figure 2.4 A11, B11 and C11, black arrowheads)**, as did most of the *Fgf* genes.

All in all, our results showed that the three *FgfR* paralogs in *O. dioica* were expressed during the embryonic and larval development, and the fact that they were expressed with different intensities and in different structures suggested that they acquired different developmental functions after their expansion. While *FgfRa* and *FgfRb* expression signal was ubiquitous and their putative specific functions seemed to be mostly related to certain mesodermal structures (i.e. the tail muscles and the notochord), and endemesodermal derivatives (i.e. pharynx, ventral organ and gonad primordium), *FgfRc* expression seemed to have become more specifically related to ectodermal derivates (i.e. nervous system and the oikoplastic epithelium).

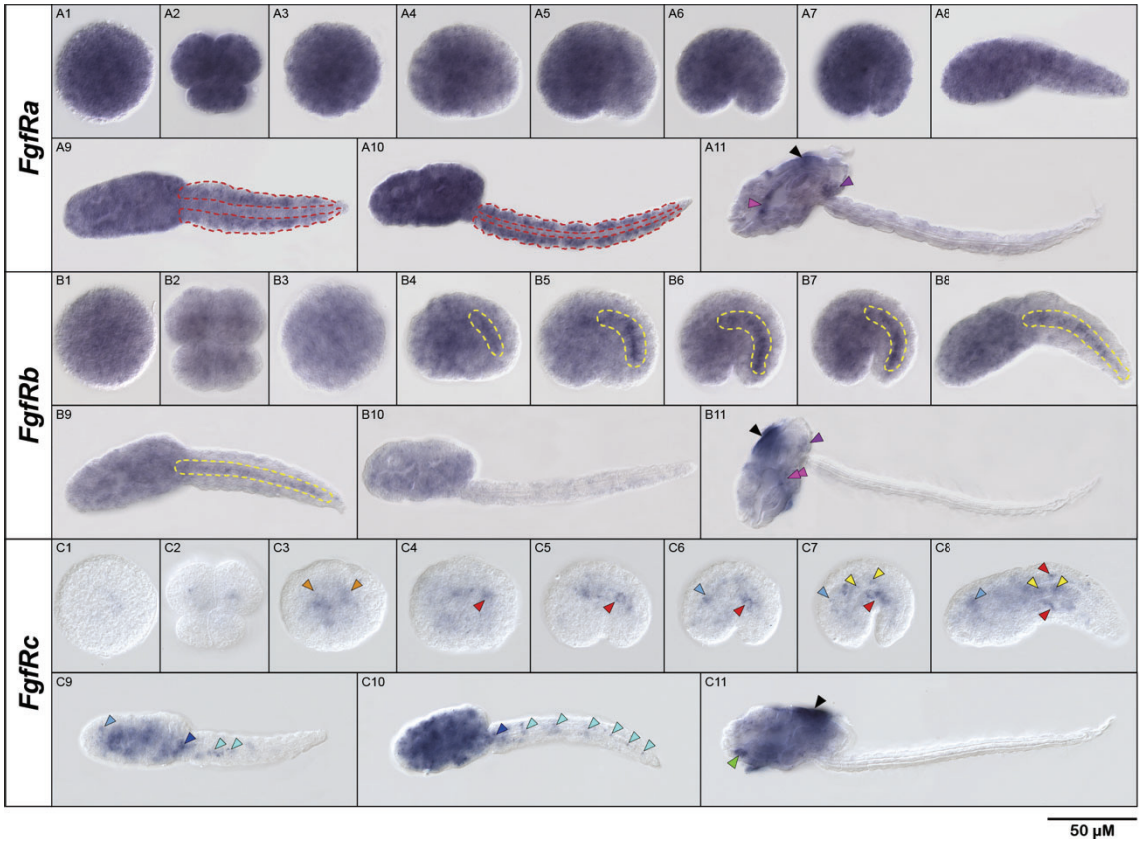


Figure 2.4. *FgfR* gene expression throughout *O. dioica* development. ABC1 are eggs, ABC2 are 8-cell embryos, ABC3 32-cell embryos, ABC4,5,6 and 7 are incipient, early, mid, and late tailbud embryos, respectively; ABC8,9,10 and 11 are just-, early-, mid- and late-hatched larvae, respectively. Images of tailbud embryos and larvae are left lateral views, with the anterior to the left and the dorsal to the top. Red dashed lines mark the muscles of the tail, yellow dashed lines mark the notochord, orange arrowheads mark endomesodermal precursor blastomeres, red arrowheads mark muscle cells, light blue arrowheads mark the developing brain, dark blue arrowheads mark the developing caudal ganglion, cyan arrowheads mark neuronal bodies in the nerve chord, yellow arrowheads mark notochord cells, magenta arrowheads mark the walls of the pharynx, magenta double arrowheads mark the pharyngeal slits, green arrowheads mark the ventral organ, purple arrowheads mark the gonad, and black arrowheads mark the oikoplastic epithelium.

2.2. Transduction pathways

2.2.1. Identification of the components of the three main transduction pathways

Upon activation of the FgfR by an Fgf ligand, a diverse array of proteins is recruited to the territory of the receptors TK domains and subsequently phosphorylated to propagate the signal through one or multiple transduction cascades. In our study in *O. dioica*, we focused on analysing the three primary transduction pathways associated to the FgfR: the MAPK pathway, the PLCγ/PKC pathway, and the PI3K/AKT pathway (Goetz & Mohammadi, 2013). To assess the status of these transduction pathways in *O. dioica*, we performed a careful survey on the BAR genome looking for orthologs of some of the main genes involved in their functioning, using human and ascidian genes as queries through RBBH.

Our survey revealed that most of the main components of the three abovementioned pathways are conserved in the genome of *O. dioica* **(Supplementary Table 9)**. Regarding the PI3K/AKT pathway, we found orthologs for *PI3KCA, PDK* and *AKT*; regarding the PLCγ/PKC pathway we found orthologs for *PLCγ* and for the four groups of vertebrate *PKC* (Newton, 2010); and regarding the MAPK pathway we found orthologs for *Raf*, *MEK1/2*, *and ERK1/2*. Interestingly, we noted an exception in the MAPK signalling pathway, since none of the classical Ras proteins (i.e. H-Ras, K-Ras and N-Ras) were found in the genome of any of the three *O. dioica* cryptic species **(Figure 2.5 and Supplementary Table 9)**. The absence of classical *Ras* genes in *O. dioica* parallels previous observations in ascidians, where these genes are also absent (Keduka et al., 2009). However, in ascidians an alternative Ras family member, namely *M-Ras*, has been identified as assuming the role of classical Ras proteins in signal transduction following FgfR activation, at least during neural development (Keduka et al., 2009). To investigate the potential for a similar mechanism in *O. dioica*, we expanded our genomic searches to include the ascidian *M-Ras* gene and other closely related Ras family members (i.e., *R-Ras*, *Rap1*, *Rap2*, *Ral*, and *Rheb*). Although no clear orthologs for *M-Ras* or *R-Ras* were identified in *O. dioica*, orthologs for *Rheb*, *Ral*, and *Rap* were found, along with other *Ras-like* genes whose orthology to human or ascidian counterparts could not be determined. Considering the magnitude and complexity of the Ras family and the resemblance of their molecular mechanisms (Wennerberg et al., 2005), coupled with the precedent of classical Ras proteins being functionally substituted by alternative Ras members in ascidians, it is probable that some of the many Ras-like proteins present in *O. dioica* mediate the transduction of this signalling pathway. Therefore, the absence of classical *Ras* genes does not necessarily indicate a disruption of the MAPK transduction cascade, a hypothesis further supported by the conservation of the three downstream kinases of the pathway (i.e. *RAF*, *MEK1/2* and *ERK1/2*).

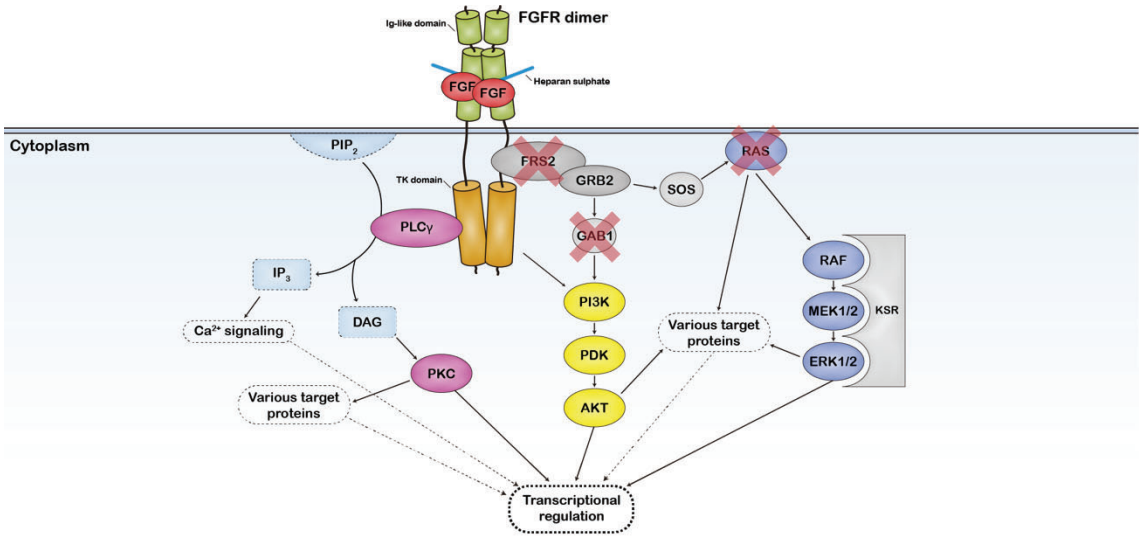


Figure 2.5. Intracellular transduction pathways activated by FgfR in *O. dioica*. Proteins included in our genomic survey of *O. dioica* are depicted with solid lines. Proteins marked with a red cross indicate those that were not found and are presumed to be lost at the genomic level in *O. dioica*.

In addition to the absence of classical *Ras* genes, we found that *O. dioica* genomes also lack genes encoding signal-transducing adaptor proteins. These proteins typically act as accessories to the main proteins in transduction cascades, facilitating protein-protein interactions to form larger signalling complexes (Luo & Hahn, 2015). For instance, we observed the absence of *Fibroblast-growth-factor-receptor-substrate-2*α (*FRS2*α) and *Grb2-associated binder-1* (*Gab1*), the latter of which is also absent in ascidian *C. robusta*. Altogether, our genome survey revealed that the main components of the three classical transduction pathways upon FgfR activation are present in *O. dioica*, what points to a conservation in the potential outcomes following activation of the receptors. Interestingly, the absence of some of the adaptor proteins, as well as the absence of classical *Ras* genes, suggests that the arrangement of these initial transduction complexes and the mechanisms by which the activation of the FgfR is transmitted to downstream kinases differ between *O. dioica* and vertebrates.

2.2.2. Developmental expression pattern of key components in *O. dioica* transduction pathways

To investigate potential relations between the Fgf/FgfR and the downstream classical transduction pathways, we analysed the expression pattern of some of their components throughout *O. dioica* development. The genes chosen for this approach were *MEK1/2* and *ERK1/2* as representatives of the MAPK pathway, *PLC*γ as a representative of the PLCγ/PKC pathway, and *PIK3CA*, *PDK* and *AKT* as representatives of the PI3K/AKT pathway.

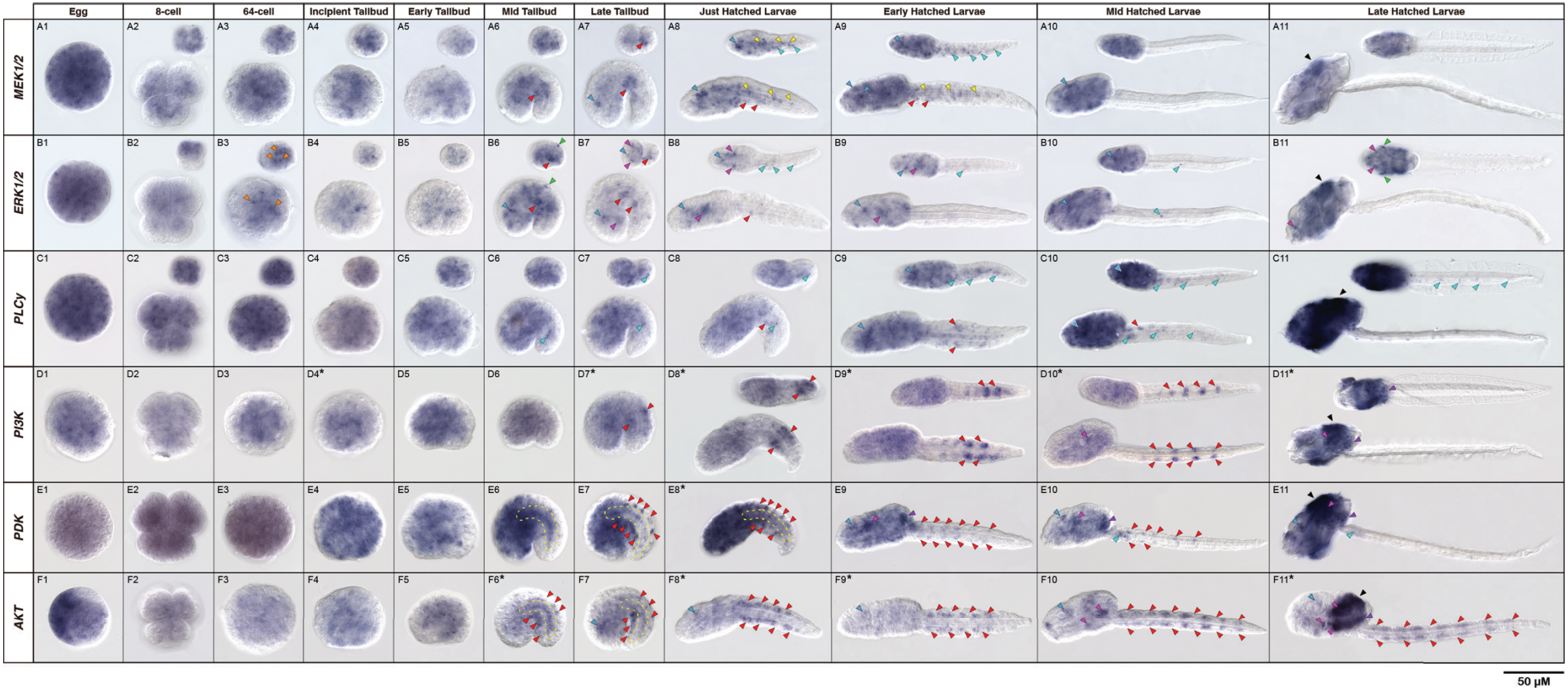


Figure 2.6. Developmental expression patterns of *O. dioica* genes involved in intracellular transduction pathways. (A-F) Expression panels of the genes selected as representatives of the three main pathways. Orange arrowheads mark endomesodermal precursor blastomeres, red arrowheads mark muscle precursors and muscle cells, light blue arrowheads mark the developing brain, cyan arrowheads mark neuronal bodies in the caudal ganglion and the nerve chord, yellow arrowheads mark notochord cells and yellow dashed lines mark the whole notochord, magenta arrowheads mark undetermined domains in the trunk, green arrowheads mark a epithelial domains, and purple arrowheads mark the gonad primordium. All panels except those marked with an asterisk are left lateral views of the embryo or larvae with the anterior to the left and the dorsal to the top. Pannels marked with an asterisk are right views of the embryo or larvae, mirrored for aesthetic purposes. All insets are dorsal views of the same embryo pictured in the panel.

2.2.2.1. The MAPK pathway

In the case of the MAPK pathway, *MEK1/2* and *ERK1/2* expression signal was strongly detected in the earliest stages from the egg to the tailbud embryo, what suggested that they were both part of the maternal mRNA contribution. In these stages, the staining appeared fast all over the embryo, indicating a ubiquitous distribution of the mRNA. Moreover, in the case of *ERK1/2,* a more intense staining could be detected in the vegetal blastomeres at the 32-cell stage, what points to the specific zygotic expression of the gene in these endomesodermal precursor cells when they start ingressing the gastrula **(Figure 2.6 B3)**. In later stages, from the tailbud onwards, the embryos required longer periods of incubation to get stained, what pointed to a milder expression of the genes, and the staining signal appeared to get restricted to the trunk and the anterior region of the tail. Although the staining still showed a generalized distribution in these regions with no particular structures showing an obvious restricted expression, specific expression domains could be recognized as more intense or more rapidly stained regions over the generalized background. In the case of *MEK1/2*, specific expression domains could be detected in the developing brain **(Figure 2.6 A7-9, light blue arrowheads)**, in the anterior ventral muscle cells of the tail **(Figure 2.6 A6-9, red arrowheads)**, in the notochord **(Figure 2.6 A8-9, yellow arrowheads)**, and in neural bodies along the nerve chord **(Figure 2.6 A8-9, cyan arrowheads)**. In the case of *ERK1/2,* the expression signal resembled that of *MEK1/2* in the developing brain, the anterior ventral muscle cells of the tail and in neuronal bodies along the nerve cord **(Figure 2.6 B6-10, blue, red and cyan arrowheads)**. Moreover, specific expression could also be observed in a pair of bilateral domains in the trunk at the late tailbud stage, compatible with the area where pharyngeal slits will develop, and were maintained until the early hatchling **(Figure 2.6 B7-9, magenta arrowheads)**. Interestingly, *ERK1/2* also showed a very specific expression in a unilateral pair of epidermal cells in the tail in late tailbud embryos **(Figure 2.6 B6, green arrowhead)**.

2.2.2.2. The PLCγ/PKC pathway

As for the PLCγ/PKC pathway, *PLCγ* expression signal was detected ubiquitously all over the embryo form the egg to the mid tailbud stage **(Figure 2.6 C1-6)**. In these early stages, the embryos were homogeneously and rapidly stained, and we could not detect any obvious signal of specific expression. From the late tailbud to the early hatched stages, the expression signal appeared to become more restricted to the trunk and the anterior region of the tail, and some structures started to show a more intense staining, including the developing brain, the developing nerve chord, and some muscle cells in the tail **(Figure 2.6 C7-9, blue, cyan and red arrowheads)**. In the mid- and late-hatched stages the trunk kept showing a general homogeneous staining with only the brain outstanding over the background. Contrarily, in the tail where the generalized signal of

expression did not appear anymore, we detected some neuronal bodies showing specific expression of *PLCγ* along the nerve cord **(Figure 2.6 C10-11, cyan arrowheads)**.

2.2.2.3. The PI3K/AKT pathway

Regarding the PI3K/AKT pathway, the WMISH results also revealed that the three genes surveyed (i.e. *PI3KCA*, *PDK* and *AKT*) were expressed during *O. dioica* embryonic development and showed a ubiquitous and maternal distribution until the tailbud stages, when a more restricted and tissue-specific expression pattern was revealed. In these late stages, the PI3K gene showed a generalized expression signal in the trunk, although a more intense staining could be distinguished in the developing gut and in the gonad primordium of the late-hatched larvae **(Figure 2.6 D10-11, magenta and purple arrowheads)**. The PDK gene showed a similar widespread expression pattern in the trunk, and from the early-hatched stage its expression signal became notably more intense in the developing brain, the gut and the gonad primordium **(Figure 2.6 E9-11, light blue, magenta and purple arrowheads)**. Likewise, *AKT* was generally expressed in the trunk, although it displayed specific expression domains more clearly than the two other genes of the pathway. A specific expression domain could be detected in the developing brain from the late tailbud embryo to the late hatched larvae **(Figure 2.6 F7-11, light blue arrowheads)**. Also, from the early hatched stage onwards, *AKT* staining in the trunk became progressively more restricted to the posterior endoderm and finally, in the mid and late hatched larvae, the trunk showed an intense expression in the stomach lobes, the posterior pharynx floor and the migrating buccal glands, as well as in the gonad primordium **(Figure 2.6 F10-11 magenta and purple arrowheads)**.

Besides these expression patterns in the trunk, the *PI3K*, *PDK* and *AKT* genes appeared to be specifically expressed in the tail's muscle cells in all hatching stages; but while *AKT* and *PDK* seemed to be expressed indistinctly in all the muscle cells, *PI3K* was not expressed in the most anterior pairs. Nuclear staining with Hoechst was conducted in the early- and mid-hatched larvae outcoming from the *PI3K* and *AKT* WMISH assays to determine with precision the identities of the muscle cells expressing *PI3K*, in contrast to *AKT* and *PDK* generalized expression **(Figure 2.7)**. This confirmed that the expression of *AKT* was detectable with a similar staining intensity within the first 8 muscle cell pairs of the tail across all developmental stages post-hatch. Conversely, the *PI3K* gene showed a dynamic expression pattern with different expression intensities in the cells along the tail. In the early hatched stage, *PI3K* showed a particularly strong expression signal in the 5th and 6th muscle cells, a milder intensity in the 3rd, 4th, 7th and 8th pairs, and no expression at all in the 1st and 2nd pairs **(Figure 2.7)**. In the late hatched larvae, while the more intense staining in the 5th and 6th muscle pairs seemed to be maintained, the staining in the 3rd, 4th and 7th paired was notably weaker, and none of the larvae examined appeared to have staining in the 1st, 2nd, or 8th pair **(Figure 2.7)**. This restricted and differential expression pattern of *PI3K* throughout the tail muscles, together with

the fact that the PI3K kinase is the most upstream actor in this transduction cascade, suggested that the PI3K/AKT pathway could be playing a specific role in the tail's musculature axial identity and development.

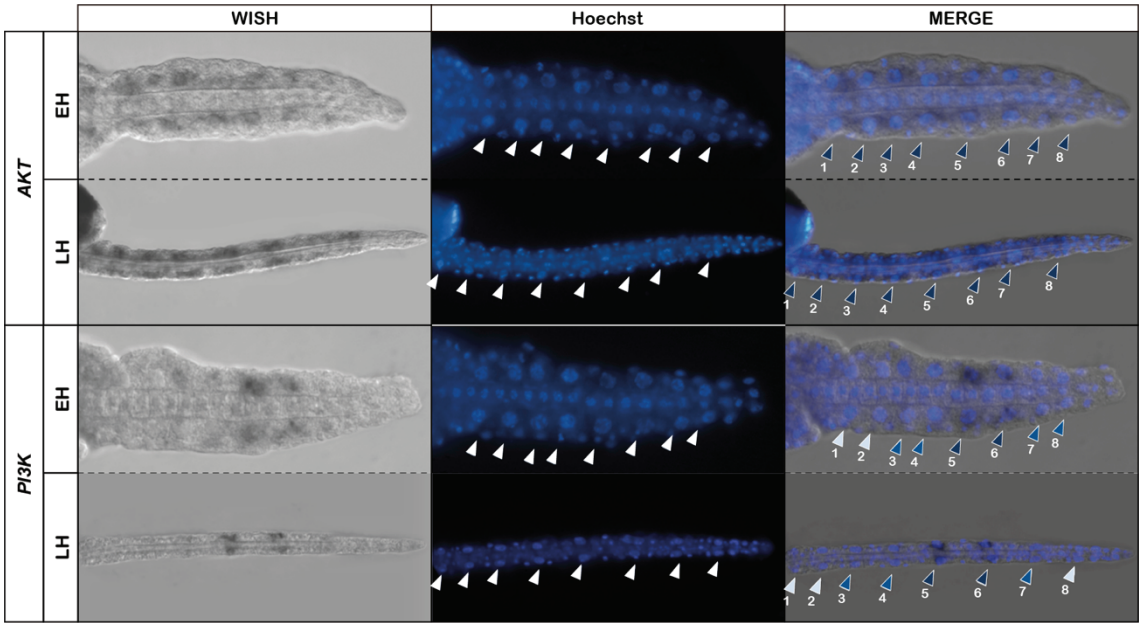


Figure 2.7. Detailed expression of *PI3KCA* and *AKT* genes in the developing muscles of the tail of early- and mid-hatchled larvae. All pictures are left lateral views of the tail of larvae stained with the WISH dye or with Hoechst, and an overlap of the two images to artificially generate a merged staining. Numbered arrowheads mark the position of muscle cell pairs along the tail, and they are coloured with different intensities of blue according to the different intensities of staining observed in the WISH results.

In summary, the expression analysis of the MAPK, PLCγ/PKC, and PI3K/AKT pathways has provided significant insights into their potential developmental roles in *O. dioica*. A common feature among all the genes surveyed is their contribution as part of the maternal mRNA, showing ubiquitous expression in the earlier stages. This is followed by a transition to more localized expression patterns in the trunk and tail during the late tailbud and larval stages. The shift from uniform to more localized expression suggests an evolving functional specificity of these transduction pathways during development. This transition is logical, as virtually all early embryonic cells are actively involved in morphogenetic processes, but as development progresses cells become more restricted to their final fates, and their location and morphology approach their final states. This reduces the diversity of cellular responses needed by most cells or favours one particular response over the others, what can be reflected in the expression of the components of certain transduction pathways. In later stages, all transduction pathways were found to be specifically expressed in the brain and other parts of the nervous system.

In terms of specific expression domains that could be related to the ones found for the *Fgf* or the *FgfR* genes, we found several cases that are worth mentioning. For example, *MEK1/2* and *ERK1/2* exhibited specific expression domains in the anterior left muscle cells of the tail in the MTB and LTB embryo **(Figure 2.6 A6-8 & B6-8, red**

arrowheads), similar to the expression found for *Fgf9/16/20b*, *Fgf9/16/20c*, *Fgf9/16/20f* and *FgfRc.* Also, *ERK1/2* exhibited specific expression from the LTB to the EH stages in a region in the trunk compatible with the pharyngeal slit primordia **(Figure 2.6 B7-9, magenta arrowheads)**, where *Fgf11/12/13/14a* and *FgfRb* are expressed in later stages. *PI3K*, *PDK* and *AKT* exhibited specific expression in the gonad primordium in the LH larvae **(Figure 2.6 D11, E11 & F11, purple arrowheads)**, similar to the expression found for *FgfRa* and *FgfRb*. Moreover, members of the three intracellular transduction pathways exhibited expression in the developing CNS, including the brain, the caudal ganglion, and the nerve cord, similar to most *Fgf* genes and *FgfRc*. The identification of these overlapping expression domains can be considered the starting point for future investigation to study putative functions of the FGF signalling in specific developmental processes in *O. dioica*. Notably, the PI3K/AKT pathway exhibited the most distinct and dynamic expression changes, with differential expression of *PI3K* in tail muscles, what could suggest a role in axial identity.

2.2.3. PI3K/AKT pathway involvement in tail development

Considering the expression pattern of PI3K/AKT in specific subsets of muscle cells of the tail, we decided to explore the potential involvement of this transduction pathway in tail muscle development. To do so, we conducted experiments using wortmannin, a specific inhibitor of the PI3K kinase (Wipf & Halter, 2005). To determine the optimal conditions, we assessed various concentrations of the inhibitor (from 0.1 to 10 µM) to treat the embryos from different developmental stages: before gastrulation (2-cell stage), during gastrulation (32-cell stage), and prior to hatchling and tail elongation (MTB stage) **(Figure 2.8A)**. Analysis of the resulting phenotypes revealed varying degrees of severity depending on both the concentration and timing of the treatment. Generally, treated animals when examined at the larval stage exhibited a spectrum of affected morphologies, ranging from normal, healthy-looking larvae to those with minor developmental irregularities, aberrant hatchlings, and embryos arrested before hatching. In summary, the severity of embryonic abnormalities induced by wortmannin correlated with concentration and treatment timing, with higher concentrations and earlier treatments yielding higher proportions of severely affected embryos **(Figure 2.8A)**. Notably, treatments at earlier stages or with higher concentrations increased the incidence of arrested or severely affected embryos, where proper differentiation between the tail and trunk failed to occur or where the tail appeared excessively affected, hindering the identification of individual cells. Given that these phenotypes likely stem from disrupted PI3K/AKT pathway function in processes beyond tail muscle development, probably during gastrulation and cellular lineage determination and proliferation, we focused on treatments at a concentration of 5 µM from the mid-tailbud stage. These resulted in the majority of embryos successfully hatching but displaying evident tail malformations. Specifically, larvae exhibited a rounder trunk and a smaller, slightly curved tail, with reduced tail mobility compared to controls. Upon closer

examination, treated larvae lacked a distinct separation between the tail and trunk, suggesting involvement of the PI3K/AKT pathway in this process.

Embryos treated at a 5 µM concentration from the MTB stage were fixed and utilized to examine abnormalities in *Troponin T* (*TnnT*) expression. The *TnnT* gene family served as a muscle cell marker and the analysis of expression of different troponin paralogs allowed us to track muscle cell identity and patterning. Previous research from our lab showed that *Troponin 3* (*TnnT3*) expression was confined to posterior muscle cells derived from the b6.5 blastomere (4th to 8th muscle pairs); *Troponin 4* (*TnnT4*) expression was restricted to cardiac muscle cells and the three anterior tail muscle pairs derived from the B6.3 blastomere (1st to 3rd muscle pairs); and *Troponin 5* (*TnnT5*) was expressed in the 3rd to 8th muscle pairs, its anterior boundaries coinciding with those of *Hox9B* expression (unpublished data). To assess the effects of PI3K inhibition in the determination of muscle cell identity, we performed a WMISH assay of control and treated embryos with riboprobes for the three aforementioned troponins as well as for muscular actin, as a general marker of muscle differentiation. Control embryos displayed the expected *Troponin T* expression pattern and *Actin* expression in all tail and heart muscle cells. Treated embryos exhibited similar expression patterns for all the genes, with no significant difference in staining intensity for actin or troponins compared to controls **(Figure 2.8B)**. This suggested that the inhibition of the PI3K/AKT pathway did not affect *Actin* or *TnnT* expression, nor did it disrupt muscle cell differentiation or the maintenance of muscle cell populations, at least when the pathway is inhibited from the mid-tailbud stage.

However, the WMISH results showed that at least the first two muscle cell pairs failed to fully separate from the trunk **(Figure 2.8B)**. A closer examination of the treated larvae outcoming from the WMISH assays for *TnnT3* and *TnnT4* confirmed that the first two muscle cells of the tail completely remained inside the trunk, and the third muscle cell stood in the middle of an ambiguous boundary between the trunk and the tail **(Figure 2.8C)**. These observations pointed to a critical role for the PI3K/AKT signalling pathway in the morphogenesis of the trunk and the first two muscle pairs and its disruption leading to an impaired tail-trunk separation. This is of special interest considering that the genes involved in the PI3K/AKT transduction pathway did not display a particularly strong expression signal in the anterior-most muscle cells in the WMISH assays, while *MEK1/2* and *ERK1/2* were indeed found to be strongly expressed in these cells **(Figure 2.6)**. Thus, the effect of the inhibition of PI3K on the proper separation of the anterior most muscle cells from the trunk could reflect an interplay between these two pathways.

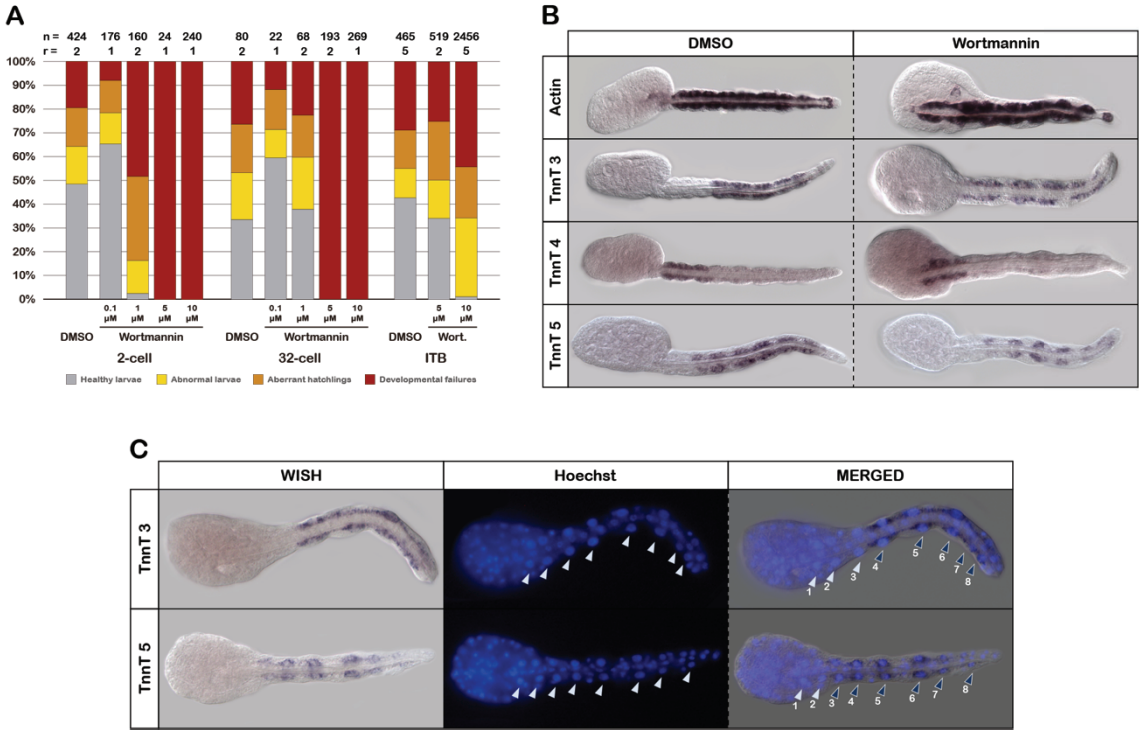


Figure 2.8. Effects of the inhibition of PI3K in *O. dioica* tail morphogenesis. (A) Wortmannin effects on *O. dioica* embryonic development at 6 hpf. The plot collects the proportions of altered phenotypes in the animals treated with different wortmannin concentrations or DMSO (control), at the 2-cells, 32-cells or mid-tailbud (MTB) stages. Number of analysed embryos (n) and number of replicas (r) are indicated above each treatment bar. **(B)** WMISH results for *Actin*, *Tnnt3*, *Tnnt4* and *Tnnt5* genes in 6 hpf larvae treated from the MTB stage at a 5 µM wortmannin (right) or DMSO (left). DMSO panels display lateral views of the larvae with the anterior to the left and the dorsal to the top. Wortmannin panels display lateral views of the larvae with the anterior to the left, but the left/right orientation could not be determined due to the severe malformations **(C)** Hoechst staining of the WMISH results of the *Tnnt3* and *Tnnt5* genes in embryos treated from the MTB stage at a 5 µM wortmannin. Muscle cell pairs are marked with numerated arrowheads, coloured blue if staining was observed in the WMISH results.

2.3. Attenuators of the FGF signalling: Sprouty and SPRED

2.3.1. Gene losses affecting the attenuators of the FGF signalling in appendicularians

The fact that the main components of the three classical transduction pathways displayed such a generalized expression pattern, especially in the earlier stages of development, encouraged us to seek for more downstream genes that could serve as markers for the activation of the Fgf receptors.

Two of such potential genes were *Sprouty* and *SPRED* (*Sprouty-related Proteins with an EVH1 Domain*), which encode well known non-catalytic attenuators of RTK signalling in vertebrates (H. J. Kim & Bar-Sagi, 2004; Neben et al., 2019). Despite these two related proteins have been generally associated to the negative feedback-regulation of RAS-MAPK activation, their molecular mechanisms and other potential functions are not fully understood, and Sprouty proteins have been found to also regulate the PLCγ response and other transduction pathways apart from the MAPK cascade (Akbulut et al., 2010; Ayada et al., 2009; Chow et al., 2009; Nutt et al., 2001). Interestingly, there is a line

of evidence that points to these two factors as the main determinants in triggering either the MAPK or the PLCγ/PKC pathway upon the activation of the FgfR in certain circumstances. This alleged bimodal switch-like mechanism puts forward Sprouty as an inhibitor of the PLCγ/PKC pathway, and SPRED as an inhibitor of the MAPK pathway and would therefore explain how the cells can exhibit different responses (e.g. fate induction or migration) to the Fgf signalling in complex morphogenetic environments, such us the gastrulating *Xenopus tropicalis* embryo (Sivak et al., 2005).

Our searches through tblastn in the BAR genome of *O. dioica* allowed us to identify easily a putative ortholog for *Sprouty*, but we could not find any clear ortholog for *SPRED,* what pointed to a possible loss of the gene. Sprouty and SPRED proteins are closely evolutionarily related as they share an exclusive cysteine-rich domain, the SPRY domain, close to their C-terminus. While Sprouty proteins do not carry any other conserved domain, SPRED proteins also bear an EVH1 domain close to their N-terminus (Kawazoe & Taniguchi, 2019). This EVH1 domain is closely related to the one found in Ena/VASP proteins, a conserved family of actin regulatory proteins (Krause et al., 2003). Thus, SPRED proteins look like a chimera of an Ena/VASP and a Sprouty protein.

To test the possible loss of *SPRED*, we searched for all the orthologs for *Sprouty*, *SPRED*, and *Ena/VASP* genes in appendicularians, ascidians, vertebrates and cephalochordates, and performed a phylogenetic reconstruction with all the sequences retrieved. Our results confirmed the lack of *SPRED* in the cryptic species of *O. dioica* and showed that it was a shared feature with all the other appendicularians species surveyed, as none of the identified sequences containing either a SPRY or a EVH1 domain were grouped with SPRED proteins **(Figure 2.9A)**. Given the fact that this gene is present in all other chordates, our results indicated a loss of the *SPRED* gene in the appendicularians lineage after their split from the rest of tunicates and prior to their radiation **(Figure 2.9B)**.

The loss in of one of the two proposed components of the hypothesized switch-like mechanism abovementioned automatically discarded its viability, at least in the appendicularians clade. Instead, the loss of *SPRED* could be interpreted as part of a coelimination process that included the classical *Ras* and other *Ras*-related genes (see section 2.2.1).

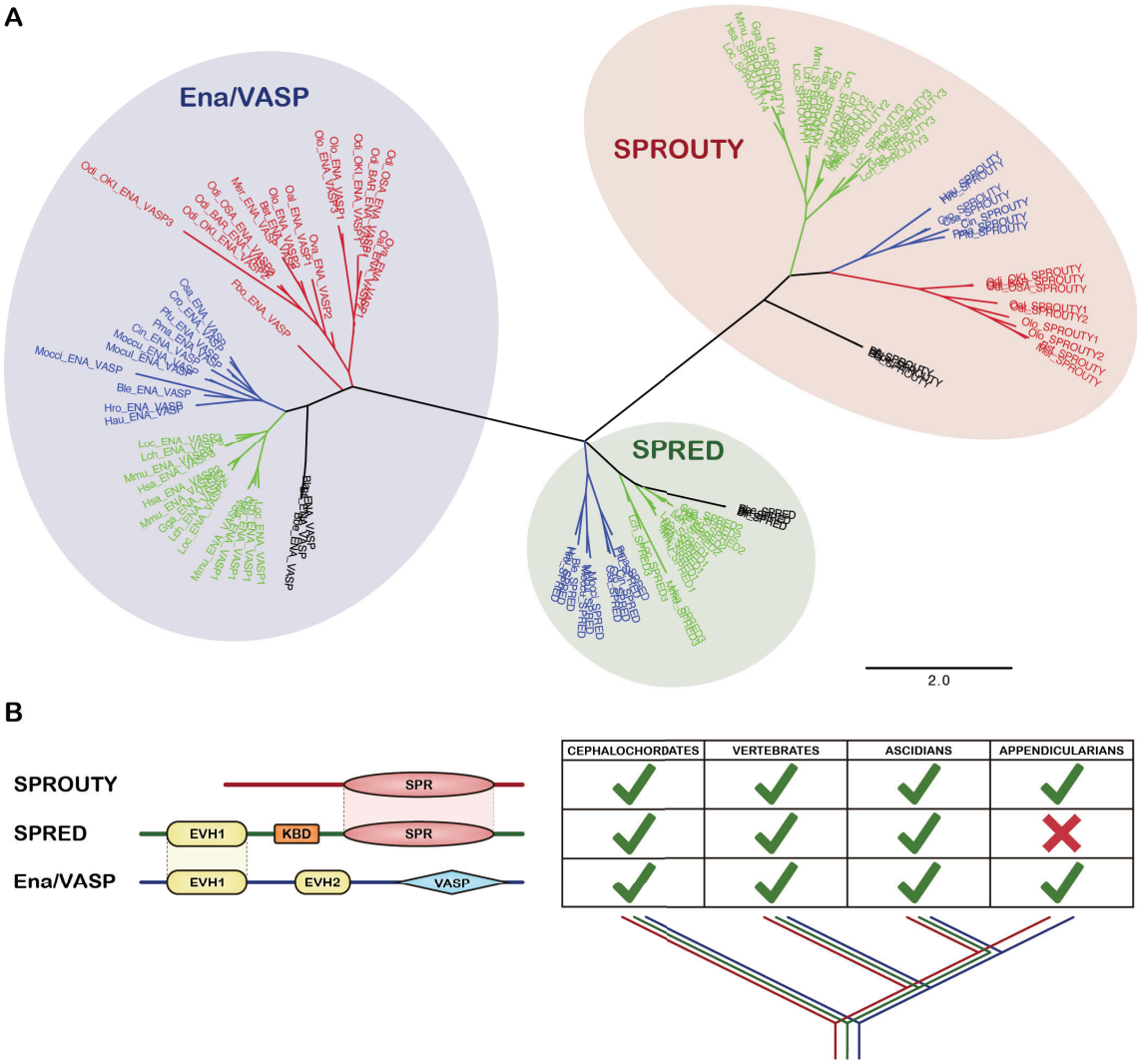


Figure 2.9. Evolution of Sprouty and SPRED proteins in appendicularians. (A) Maximum likelihood phylogenetic inference of Ena/VASP, Sprouty and SPRED proteins from cephalochordates (black), vertebrates (green), ascidians (blue) and appendicularians (red). **(B)** Graphic representation of Sprouty, SPRED, and Ena/VASP domain architectures and their presence in the chordate clades surveyed. The loss of SPRED proteins only affects the appendicularian lineage.

2.3.2. *Sprouty* expression during *O. dioca* development

To learn more about the possible functions of the *Sprouty* gene, we analysed its expression throughout the different developmental stages in *O. dioica* **(Figure 2.10A)**. In the earlier stages, from the egg to the 32-cell stage embryo, we did not find any signal of expression of the gene, what indicates that it is not part of the maternal mRNA contribution. At the 32-cell stage, when gastrulation begins, the expression signal starts to be detected in the nuclei of some blastomeres. In order to determine more confidently the identity of the blastomeres expressing *Sprouty* in these early stages, we performed a one-colour double WMISH assay using *Brachyury* as a marker of the notochord precursors, whose positions are precisely determined in the 32-cell and 64-cell embryos **(Figure 2.10B)**. The expression signal revealed that *Sprouty* started to be expressed in three territories in the 32-cell embryo (according to Delsman's/Conklin's

nomenclature): the B22/B6.1 pair of vegetal blastomeres, that will give rise to endomesodermal derivatives including the whole endodermal strand; the A11/A6.4 and A12/A6.3 pairs within the neural plate, that will develop into most of the nervous system; and the a11/a6.8 pair of animal blastomeres that will give rise to part of the epidermis. At the 64-cell stage, once the embryos were undergoing gastrulation, *Sprouty* expression signal persisted in the whole neural plate, in the nuclei of endomesodermal precursor cells derived from the B22/B6.1 pair, and in the animal blastomeres derived from the a11/a6.8 pair. Moreover, two new expression domains appeared in the 64-cell embryo, one in the b7.9 and b7.10 pairs of blastomeres, that at this time were already restricted to the muscle fate, and the other in a posterior pair of animal blastomeres that will develop into epidermis **(Figure 2.10 A3-4 and Figure 2.10B)**.

In later stages, the expression signal of *Sprouty* was detected mainly in structures derived from the blastomeres that were stained in the 32-cell and 64-cell embryo. These were neural structures, including the posterior region of the developing neural cord **(Figure 2.10 A5, cyan arrowheads)** and the brain during most of its development **(Figure 2.10 A5-10, blue arrowheads)**; muscle cells of the tail, especially the three anterior most pairs in the mid and late tailbud stages **(Figure 2.10 A6-7, red arrowheads)**; undetermined endomesodermal structures within the trunk **(Figure 2.10 A5-10, magenta arrowheads)**; and different regions of the trunk epidermis, including the oikoblast in the late hatched larvae **(Figure 2.10 A5-11, green arrowheads)**. All these domains overlap, at least partially, with expression domains of *Fgf* and *FgfR* genes, what suggests that Sprouty might indeed be involved in the cellular response to the activation of the FgfR. Moreover, *Sprouty* expression signal could also be observed in the anterior most region of the notochord in the late tailbud and just hatched larvae stages **(Figure 2.10 A7-8, yellow arrowheads)**, similarly to the specific staining observed for *Fgf9/16/20a* WMISH.

Among the epidermal regions stained, a specific rostral domain in the mid and late tailbud embryos was revealed as of special interest **(Figure 2.10 A6-7, double green arrowheads)**. This epidermal region is known to develop into the oral placode, whose homology to vertebrate placodes was confirmed by the expression of placodal markers such as *Eya*, *Pitx* and *Six3/6* (Bassham & Postlethwait, 2005). The clear expression of *Sprouty* in some cells within this region suggests that its function is compatible with that described in vertebrates as an attenuator of the RTK pathway, and could therefore be essential for the proper patterning and development of placodes **(Figure 2.10C)**.

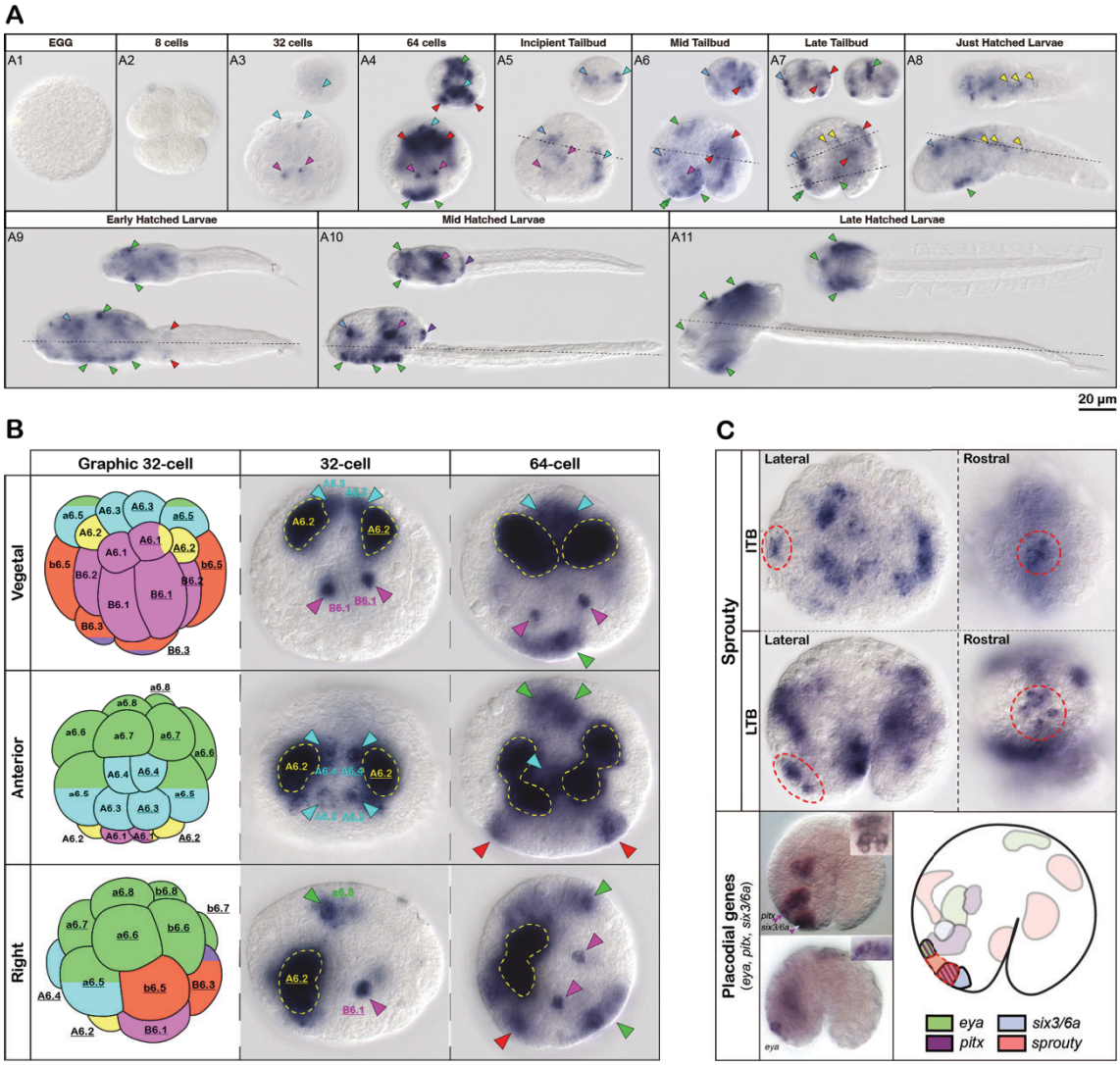


Figure 2.10. Analyses of *Sprouty* gene expression. (A) *Sprouty* gene expression throughout *O. dioica* development. All panels are left lateral views of the animals with the anterior to the left and the dorsal to the top. All insets are dorsal views of the animals with the anterior to the left at the focal plane indicated with dashed lines in the main panel. The scale bar does not account for the insets. Cyan arrowheads mark the neural plate and the developing nerve chord, magenta arrowheads mark endomesodermal precursors and their derivatives, red arrowheads mark muscle precursors and muscle cells, light blue arrowheads mark the developing brain, green arrowheads mark epidermal domains, green double arrowheads mark the oral placode, yellow arrowheads mark notochord cells, and purple arrowheads mark the gonad primordium. **(B)** Left column is a graphic representation of the 32-cells embryo displaying blastomeres with the same colour code as the arrowheads in (A) and numbered according to Conklin's nomenclature. Mid and right columns are 32-cell and 64-cell embryos, respectively, outcoming from a one-colour double WMISH for *Sprouty* and *Brachyury* genes. Yellow dashed lines mark the notochord precursor cells expressing *Brachyury* and coloured arrowheads mark *Sprouty* expression domains. Each row displays a different orientation of the embryos, as indicated in the lettering. **(C)** Upper panels are detailed images of *Sprouty* gene expression in the region of the oral placode, surrounded by a red dashed line. In the bottom, images of WMISH for placodal genes (i.e. *eya*, *pitx* and *six3/6a*) were extracted from *Bassham & Postlethwait, 2005*. Bottom right panel is a graphic representation of a late tailbud embryo in lateral left view with expression domains for *eya*, *pitx*, *six3/6a* and *Sprouty* collapsed.

3. Testing the function of FGF signalling in *Oikopleura dioica* by inhibitory treatments and *omic* approaches

3.1. Pharmacological inhibition of the FGF signalling pathway in the embryonic development of *Oikopleura dioica*

In order to address the roles of the FGF signalling in the embryonic development of *O. dioica*, pharmacological treatments were conducted on developing embryos with SU5402, a well-known inhibitor of the FgfR tyrosine kinase activity that causes a systemic inhibition of the FGF signalling (Gudernova et al., 2016; Mohammadi et al., 1997), and with DMSO as control. To assess the effect of SU5402, we performed the inhibitory treatments on clutches of embryos at different concentrations (i.e. 50 μM and 100 μM) and different time windows starting at different developmental stages (i.e. 2-cell stage 25 mpf, 32-cell stage at 80 mpf, and incipient tailbud stage 140 mpf) **(Figure 3.1A)**. The embryos were examined at 4h30'-5h post fertilization, when most of the control animals had already hatched and elongated their body. We classified the resulting morphologies of treated and control embryos into four categories according to the severity of their malformations: (i) healthy larvae or larvae with minor malformations, (ii) abnormal larvae with obvious malformations, (iii) aberrant hatchlings with a drastic failure in tail and/or trunk elongation, and (iv) severe developmental failures in which the trunk and the tail could not be distinguished **(Figure 3.1A)**. The obtained phenotype proportions were compared among treatments and controls.

We found that the severity of the treatments was clearly related to the concentration of the inhibitor and to the developmental time at which the treatment started. Treatments started at the 2-cell stage with the 50 μM concentration caused a severe developmental failure in most embryos, resulting in a mass of cells in which the trunk and the tail could not be recognized **(Figure 3.1B)**. The fact that the cells kept dividing and that in most cases organized structures could be distinguished in the embryos revealed that the treatments did not cause a total arrest of development. When the treatments at 50 μM concentration were started at the 32-cell stage, the severity and abundance of the malformations drastically decreased, and almost half of the embryos could develop into healthy looking larvae **(Figure 3.1B)**. In contrast, treatments at 100 μM concentration from the 32-cell stage caused a severe developmental failure in most individuals. Lastly, when the treatments were started at the ITB stage, the 50 μM concentration did not have any apparent effect, and the 100 μM made the vast majority of embryo to develop into aberrant hatchlings or larvae with obvious malformations **(Figure 3.1B)**.

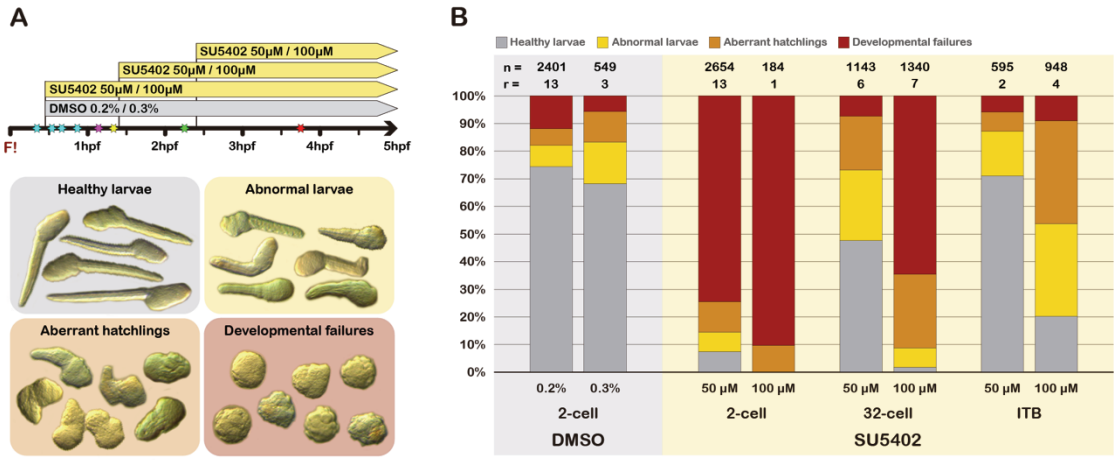


Figure 3.1. Impact of FgfR inhibition on the embryonic development of *O. dioica*. (A) Schematic representation of SU5402 treatments and their resulting morphologies. Coloured stars on the timeline denote key developmental events: cyan indicates the first four synchronized cell divisions, magenta indicates the onset of gastrulation, yellow indicates the beginning of neurulation, green marks the start of the tailbud stage, and red marks the hatching. **(B)** The average percentage of *O. dioica* individuals classified by morphology after SU5402 treatment under various conditions. The total number of analyzed embryos (n) and the number of replicates (r) are indicated above each treatment.

3.1.1. Time-window dependent effects of the inhibition reveal developmental functions of the FGF signalling

Our results demonstrated that the inhibition of FGF signalling has an obvious impact on *O. dioica* embryonic development, with sensitivity to SU5402 treatments varying based on the initiation time of inhibition. Early inhibition from the 2-cell stage, which affects initial cell divisions and the entire gastrulation process, resulted in severe developmental failure in most embryos at a 50 µM concentration of SU5402. Conversely, when the treatments started at the 32-cell stage, after the onset of gastrulation and determination of most cellular lineages, the impact of the inhibition was notably less severe. At this stage, a 50 µM concentration of SU5402 allowed most embryos to progress through tailbud stages normally, with noticeable effects only manifesting after the hatchling. The differential sensitivity likely arises from a critical role of the FGF signalling during gastrulation and cellular lineage induction, periods marked by high cellular communication and where the FGF signalling pathway plays conserved roles across eumetazoans (Matus et al., 2007). The frequent occurrence of the most severely affected phenotype when performing the inhibitory treatments at 100 µM from the 32-cell stage may be related to faster diffusion of the inhibitor into the embryos, what could disrupt the still ongoing gastrulation in delayed-developing embryos, or to the disturbance of later less sensitive morphogenetic events in which the FGF signalling is involved. Supporting this, treatments initiated at the ITB stage showed no evident effects at 50 µM, but 100 µM treatments led to the development of abnormal hatchlings or larvae, indicating importance of the FGF signalling in later developmental stages as well.

The higher frequency of less severely affected Individuals In the treatments initiated from the 32-cell and ITB stages allowed us to observe abnormal morphologies that became evident only when the animals developed into advanced larval stages, pointing to putative functions of the FGF signalling in late developmental events. These events include the proper separation of the tail and the trunk and the elongation of the tail, as indicated by the abundance of aberrant hatchlings and larvae with shortened, rounded tails and slightly malformed trunks. Given that all *O. dioica FgfR* paralogs are expressed in the tail and trunk up to this stage, attributing these malformations to a specific receptor is challenging. However, it is plausible to associate the shortened tail with the expression of *FgfRb* in the developing notochord and the aberrant separation of the trunk and the tail with the expression of *FgfRc* in the anterior muscle cells of the tail, located in the ventral region where the trunk and the tail are defined (see Figure 2.4 in section 2.1). Regarding the Fgf ligands, the involvement of the FGF signalling pathway in these processes could be related to the expression of *Fgf9/16/20b*, *Fgf9/16/20c* and *Fgf9/16/20f* in the same region where *FgfRc* is expressed; to the expression of *Fgf9/16/20a* in the anterior-most and posterior-most muscle cells of the tail; or to the expression of *Fgf9/16/20d* and *Fgf9/16/20e* at the tip of the tailbud during the tailbud and hatchling stages (see Figure 1.8 in section 1.3).

3.1.2. Inhibition of the FgfRs with SU5402 causes a downregulation in *Fgf9/16/20a* expression

To determine whether the inhibition of the FgfR affected the expression of Fgf ligands, we conducted a WMISH assay of *Fgf9/16/20a* on control embryos and embryos treated with SU5402 50 μM from the 100-cell stage at 100 mpf to the late tailbud (LTB) stage. Our results indicated a significant reduction in *Fgf9/16/20a* expression in treated embryos, with some expression domains becoming undetectable through WMISH **(Figure 3.2)**. In control embryos, intense staining was observed within approximately 24 hours revealing all previously identified expression domains (i.e., two bilateral endomesodermal domains, the 1st and 8th pairs of muscle cells in the tail, the 1st and 3rd cells of the notochord, and a dorsal domain compatible with the caudal ganglion). In contrast, treated embryos exhibited only slight staining in limited areas after 24 hours, and even after 48 hours, staining was not observed in all expected domains. Notably, the 1st and 3rd notochord cells were the most robustly stained in treated embryos, often being the only stained domains **(Figure 3.2)**.

These findings suggest that FgfR inhibition negatively impacts the expression of Fgf ligands, at least for *Fgf9/16/20a*. This may be due to the disruption of a positive feedback loop wherein FGF signalling activation promotes the expression of certain FGF ligands. The relative resistance of certain expression domains, such as the notochord cells, to FgfR inhibition might indicate either a distinct regulatory mechanism governing *Fgf*

expression or a higher baseline expression level of *Fgf9/16/20a* under normal conditions.

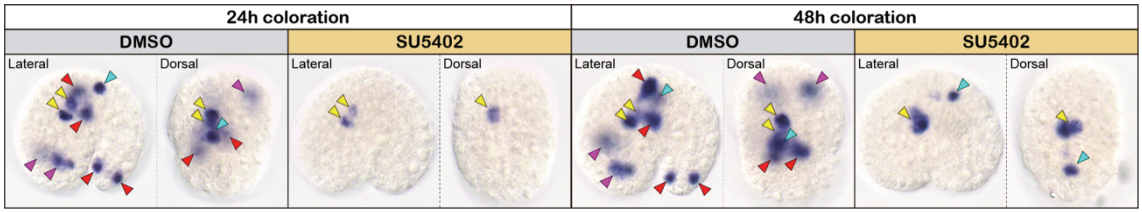


Figure 3.2. Effect of SU5402 on the expression of *FGF9/16/20a*. WMISH results for *Fgf9/16/20a* in control-DMSO and treated-SU5402 embryos after 24h or 48h of coloration. Magenta arrowheads mark endomesodermal domains, yellow arrowheads mark notochord cells, red arrowheads mark muscle cells of the tail, and cyan arrowheads mark a caudal ganglion precursor cell.

3.2. Effect of FGF inhibition on the development of cellular lineages

The differentiation of cellular lineages is a crucial aspect of embryonic development, in which the FGF signalling is known to play significant and generally conserved roles in all chordates (Bertrand et al., 2011; Dorey & Amaya, 2010; Satou, 2020). FGF signalling is involved in early embryogenesis in virtually all eumetazoans where it has been studied, with a distinct role in the induction and patterning of the endomesoderm and endomesodermal derivatives, influencing cell fate decisions and morphogenesis (Andrikou & Hejnol, 2021). Results from the previous section highlighted a strong impact of the systemic inhibition of FGF signalling on the overall embryonic development of *O. dioica*, especially when the inhibitory treatments were applied before cellular lineage determination and gastrulation. Given that the FGF signalling has undergone a drastic reorganization of its components in *O. dioica* compared to other chordates, we aimed to examine if such reorganization has influenced the conserved functions of the FGF signalling in chordate early embryogenesis. Thus, we investigated the effects of SU5402 on the determination of cellular lineages, with a focus on those derived from the endomesoderm. By analysing the expression of marker genes associated with these lineages in embryos treated with SU5402, we sought to elucidate how FGF signalling influences their induction and differentiation.

3.2.1. Effect of SU5402 on the induction of mesodermal derivatives

To assess the effect of FGF inhibition on the determination of mesodermal lineages, we focused on the notochord, using *Brachyury* as a marker gene for notochord differentiation. Under normal conditions, *Brachyury* starts to be expressed at the 32-cell stage in two paired precursor cells once their fate is restricted to the notochord (Bassham & Postlethwait, 2000). By treating embryos with SU5402 at 50 µM concentration from different developmental points, we found that *Brachyury* expression at the 64-cell stage was almost abolished when the treatment was initiated from the

fertilized egg or the 2-cell stages. However, this effect dropped when the treatment began at the 4-cell stage, with most embryos expressing *Brachyury* normally in the two pairs of notochord precursor cells **(Figure 3.3)**. Nuclear staining with Hoechst in the most affected embryos treated from the fertilized egg or the 2-cell stage confirmed that most embryos had more than 32 nuclei/cells, indicating that the suppression of *Brachyury* expression was not due to a developmental delay or arrest prior to the onset of *Brachyury*. A similar effect was observed when we examined the expression of *Actin* at the ITB stage in embryos treated from the 2-cell or the 4-cell stages. Treatments from the 2-cell stage caused a major inhibition of *Actin* expression, with most embryos showing no expression signal. Conversely, treatment from the 4-cell stage resulted in most embryos exhibiting *Actin* expression domains, although their distribution revealed aberrant morphologies when compared to the control **(Figure 3.3)**.

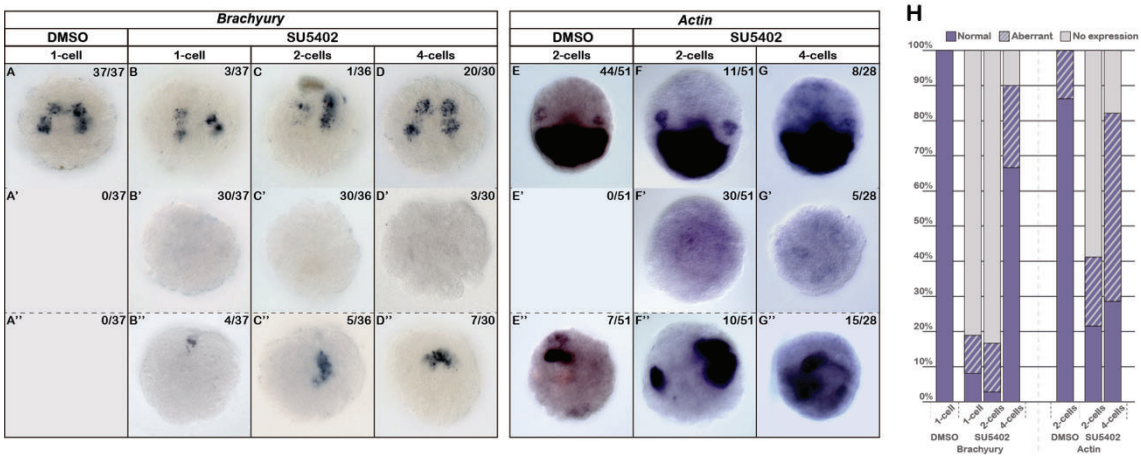


Figure 3.3. Effect of SU5402 in the induction of mesodermal derivatives. Treatments with SU5402 were performed at 50 µM concentration and DMSO controls at 0.2% concentration. **(A-D)** *Brachyury* expression in control embryos (A) and embryos treated with SU5402 (B-D) from different developmental stages. **(E-F)** *Actin* expression in control embryos I and embryos treated with SU5402 (F-G) from different developmental stages**.** Upper panels (A-G) show images of the embryos with normal morphologies and expression of the marker genes in the notochord/muscle precursor cells. Middle panels (A'-G') show images of the embryos with no expression signal detected. Lower panels (A''-G'') show images of embryos with aberrant morphologies and/or aberrant domains of expression for the marker genes. **(H)** Bar graph showing the percentage of normal, aberrant and expression-less embryos obtained in each of the treatments.

Overall, our findings indicate that FGF signalling plays a crucial role in the early determination of the notochord and muscle lineages. Since these are the two lineages derived from the mesoderm in *O. dioica,* our results suggest that FGF inhibition suppresses the development of mesodermal derivatives. It is important to consider that there is a delay between applying the treatment and the inhibitor fully diffusing through the chorion and cell membranes to inhibit the FgfR. Given the short intervals that separate the first cell divisions in *O. dioica*, less than 10 minutes each, it is likely that FgfR inhibition occurs one or even two rounds of cell divisions after applying the treatment. For treatments starting at the 2-cell stage (25 mpf), inhibition could be occurring at the 8-cell stage (40 mpf). Our results also align with recent studies in ascidians, where FGF

signalling is involved in the activation of zygotic transcription at the 16-cell stage (Treen et al., 2023). In *O. dioica*, it is likely that the inductive signals mediated by FGF in early embryogenesis commence at the 8-cell stage, coinciding with the onset of expression of some of the *Fgf9/16/20* paralogs, as described in section 1.3. This would be consistent with the observed trend of "one-cleavage" earlier in *O. dioica* compared to ascidians (Ferrández-Roldán, 2021; Fujii et al., 2008; Stach et al., 2008). Whether the FGF signalling is involved in the initiation of zygotic expression in *O. dioica* as it is in ascidians will require further investigation.

3.2.2. Effect of SU5402 on the development of other derivatives

To evaluate the impact of inhibiting the FGF signalling on the determination and morphogenesis of cellular lineages beyond mesodermal derivatives, we conducted inhibitory treatments with SU5402 at 50 µM concentration from the 2-cell stage and analysed the expression of various marker genes on embryos at the ITB stage. The selected marker genes were: *Brachyury* and *ActnM1* as indicators of the notochord and muscle lineages and as positive controls; *Zic1* as a marker of the nervous system; *Tis11b* as an endodermal marker; *SoxBb* as a marker for the posterior epidermis; and *Tis11a* as a marker for multiple lineages, including the anterior epidermis, nervous system, and the tailbud (Torres-Águila et al., 2018).

Our results demonstrated a downregulation of all surveyed gene markers in treated embryos, with an average of 52.7% of the embryos, and always over 20%, exhibiting no expression at all for the genes surveyed **(Figure 3.4A-F)**. The muscle lineage was the most severely affected, with only 4 out of 27 embryos showing *ActnM1* expression. The notochord lineage, labelled by *Brachyury* expression, and the posterior epidermis, labelled by *SoxBb* expression, followed the muscle lineage in the severity of the inhibition. *Tis11a* was the least affected marker gene in terms of positive expression in treated embryos, although their morphology was always aberrant. Given that *Tis11a* is expressed in several cellular lineages, it remains unclear whether the treatment had a specific inhibitory effect on any of them over the others.

Notably, the proportion of embryos expressing *Brachyury* exceeded the expectations based on results in the previous section **(Figure 3.3)**, although this expression was aberrant in all cases **(Figure 3.4A)**. This may be explained due to a more advanced developmental stage of the observed embryos, allowing the spurious expression of certain lineage specific genes. To address this, we repeated the inhibitory treatments with a shorter developmental period and assessed the expression of marker genes whose normal expression begins early, specifically *Brachyury*, *Zic1*, and *Tis11b*, and examined their expression at the 100-cell stage. In this case, the treatment strongly inhibited the expression of all three surveyed genes, specially of *Brachyury* **(Figure 3.4 G-I)**.

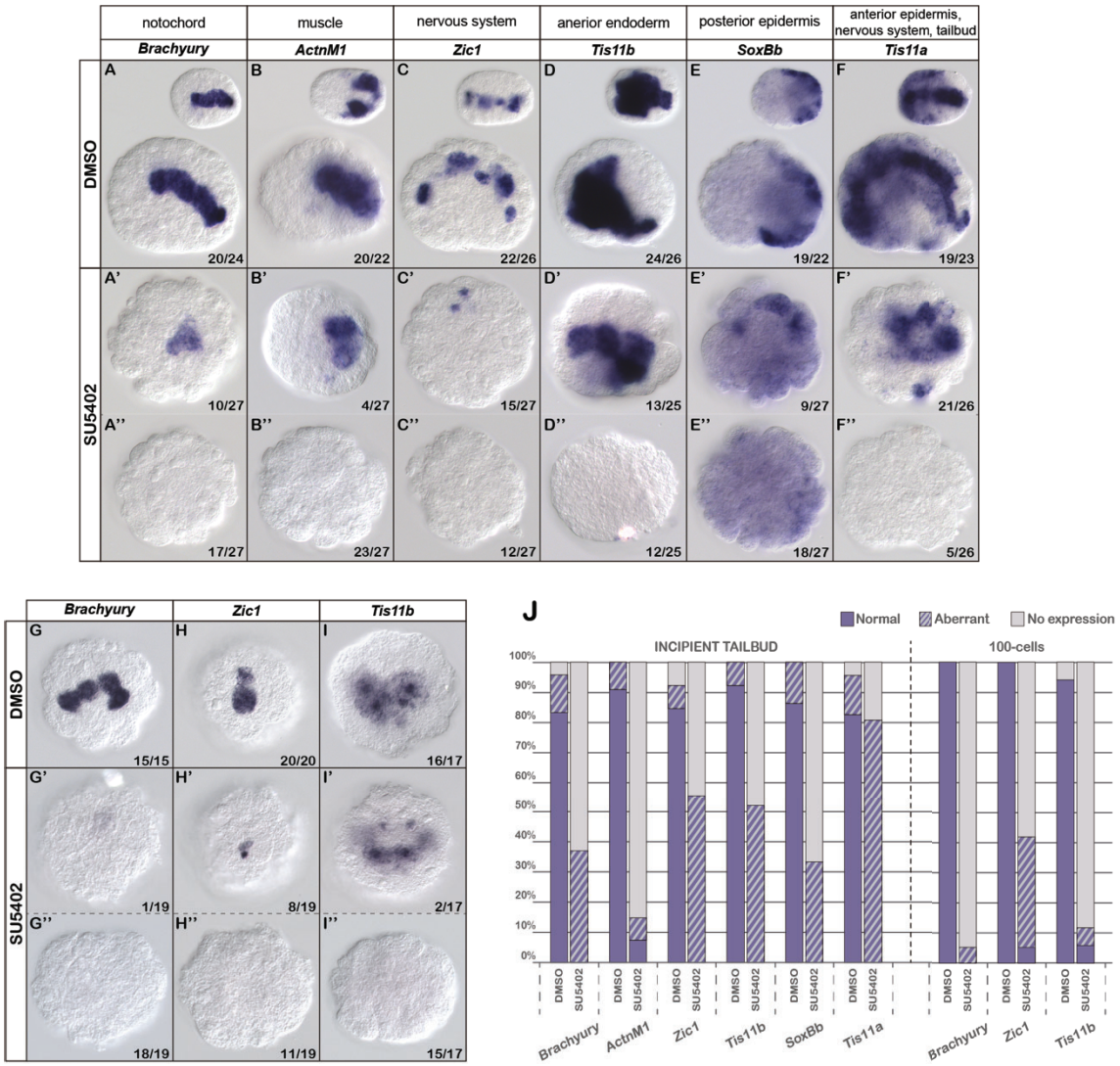


Figure 3.4. Effect of SU5402 in the induction of various cellular lineages. Treatments with SU5402 were performed at 50 µM concentration and DMSO controls at 0.2% concentration. **(A-F)** WISH results of various tissue specific genes in control and treated embryos at the ITB stage. Upper panels (A-F) show control embryos with normal expression patterns; main images are left lateral views of the embryos with the anterior to the left and the dorsal to the top and insets are dorsal views of the same embryos with the anterior to the left. Middle panels (A'-F') show treated embryos with aberrant expression patterns. Lower panels (A''-F'') show treated embryos with no expression detected. **(G-I)** WISH results of *Brachyury*, *Zic1* and *Tis11b* in control and treated embryos at the 100-200-cells stage. Upper panels (G-I) show control embryos with normal expression patterns; middle panels (G'-I') show treated embryos with aberrant expression patterns; and lower panels (G''-I'') show treated embryos with no expression detected. **(J)** Bar graph showing the percentage of normal, aberrant and expression-less embryos obtained for each of the marker genes in the different tested conditions.

Overall, our findings indicate that inhibiting the FGF signalling pathway from the 2-cell stage results in a pronounced suppression of the initial expression of all surveyed marker genes, likely affecting the induction of the corresponding cell fates. However, when treated embryos are allowed to develop for an extended period, the expression of most marker genes eventually initiates in many embryos, regardless of FGF signalling inhibition. This suggests a transition in the GRN involved in the expression of this marker genes during the normal development of *O. dioica* from an FGF-dependent state to an FGF-independent state. A possible scenario would be that, during the gastrula stages

when most cell fates and lineages are being established, FGF signalling is crucial for the derepression of these genes at the precise moment of their induction in specific lineages. As development progresses and cellular lineages are determined, the repressive mechanisms controlling the temporal regulation of these genes may relax allowing a continuous expression. Notably, the spurious expression of tissue-specific marker genes was observed mostly in surface cells of treated embryos, whereas in the controls the tissues expressing those markers were internal **(Figure 3.4A'-C')**. This supports the idea that FGF inhibition disrupts gastrulation movements in *O. dioica*.

3.2.3. Effect of SU5402 on later development of the notochord

In ascidians, FGF signalling is also involved in the morphogenesis of the notochord beyond its initial induction, specifically in fate maintenance and in the convergence and extension of notochord cells (Shi et al., 2009; Yasuo & Hudson, 2007). To investigate if this function is conserved in *O. dioica*, we performed inhibitory treatments with SU5402 at 50 μM concentration starting at the 16-cell and 100-cell stages. Regarding the treatments from the 16-cell stage, just before the onset of *Brachyury* expression, the results confirmed that FgfR inhibition does not significantly impact the induction of notochord precursor cells when the treatment is initiated later than the 2-cell stage. Additionally, it did not affect the early proliferation of the notochord cellular lineage, at least during its first cell division **(Figure 3.5A)**. We also examined whether FGF signalling plays a role in the convergence and extension of the notochord by treating embryos from the 100-cell stage, once the two pairs of notochord precursor cells have coalesced at the dorsal midline of the embryo, but before they start proliferating. These embryos were analysed at the LTB stage, when notochord morphogenesis was nearly complete. Results indicated that a significant portion of the treated embryos exhibited a shorter and wider notochord compared to control embryos **(Figure 3.5B)**. However, the similarity of this morphology to the wild-type appearance of the notochord at earlier developmental stages, along with other signs of delayed development in the overall shape of the embryo (the tailbud), hindered the determination of a causal relationship between the inhibition of the FgfR and the convergence and extension of the notochord. Notably, none of the treated embryos exhibited *Brachyury* expression in a cell adjacent to the distal tip of the notochord, in contrast to control embryos, which consistently showed *Brachyury* expression in this cell **(Figure 3.5B, yellow arrowhead)**.

In ascidians, *Fgf9/16/20* is essential for the induction of the notochord lineage, but *Fgf8/17/18* is also required for fate maintenance of the notochord precursor cells (Yasuo & Hudson, 2007). Moreover, *Fgf7/10/22* (referred to as *Fgf3* in previous studies) is required for notochord convergence and extension in *C. robusta* (Shi et al., 2009). In *O. dioica* it is tempting to speculate that notochord development might have become independent of FGF signalling beyond its initial induction, what could be related with the

loss of Fgf8/17/18 and Fgf7/10/22 subfamilies and the retention and expansion of the Fgf9/16/20 subfamily. Overall, our findings suggest that inhibiting FGF signalling does not affect the early proliferation of notochord precursor cells and its role in the convergence and extension remains unclear. The absence of a *Brachyury*-expressing cell in all treated embryos indicates a potential involvement of the FGF pathway in late notochord morphogenesis or in the development of structures derived from the notochord. Nevertheless, further research is necessary to elucidate the specific function of FGF signalling in *O. dioica* notochord development beyond lineage induction.

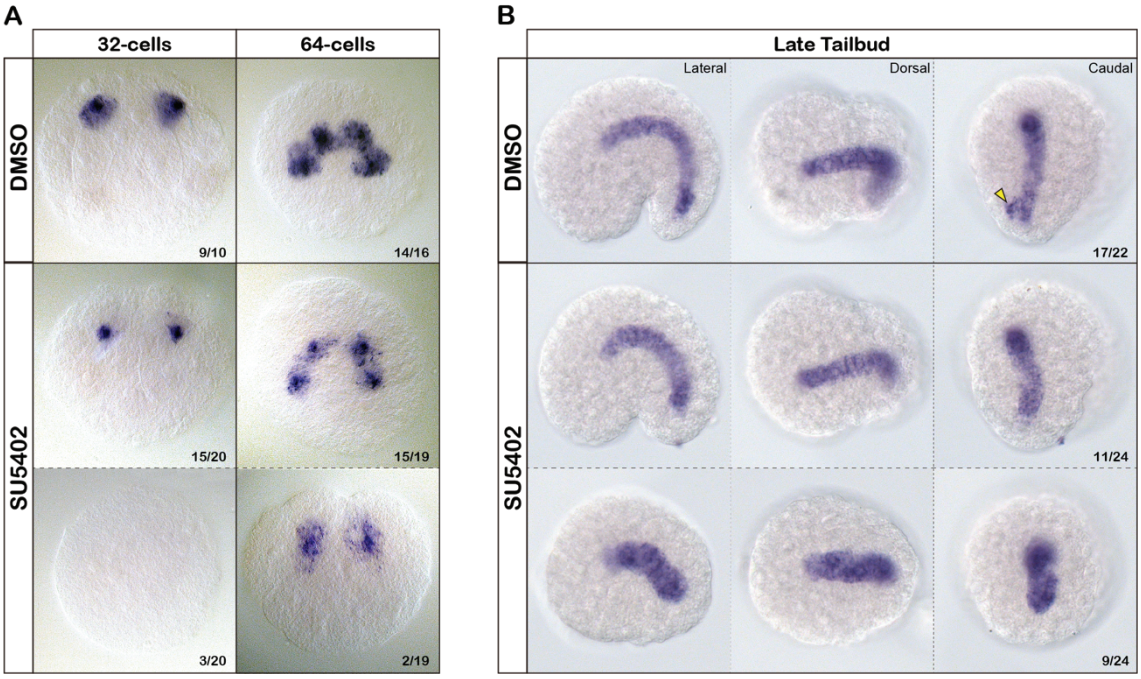


Figure 3.5. Effect of SU5402 on the morphogenesis of the notochord. (A) *Brachyury* expression at the 32-cell and 64-cell stages in embryos treated from the 16-cell stage. Upper and middle panels account for embryos with a pair of notochord precursor cells. Bottom left panel depicts embryos with no expression of *Brachyury* detected, and bottom right panel embryos where the notochord precursors cells have not divided. Embryos not shown in the pictures exhibited aberrant expression of *Brachyury*. **(B)** *Brachyury* expression at the LTB stage in embryos treated from the 100-cell stage. Upper and middle panels account for embryos with a well-developed notochord, and bottom panels account for embryos with a shorter and wider notochord. Embryos not shown in the pictures showed either aberrant expression or no expression of *Brachyury*.

3.2.4. Effect of FGF inhibition on cardiac development

Within the clade olfactores, composed of tunicates and vertebrates, the heart lineage originates from the cardiopharyngeal mesoderm, which in vertebrates gives rise to both the cardiac and pharyngeal muscles (Tolkin & Christiaen, 2012). In the ascidian *C. robusta*, the cardiopharyngeal mesoderm is represented by a group of mesenchymal cells known as trunk ventral cells (TVC). The TVCs are split from the precursor cells of the anterior tail muscles and migrate to the midventral line of the embryo, where they undergo a series of asymmetric divisions. These asymmetric divisions give rise to both heart precursors and atrial siphon muscle (ASM) precursors, the ascidian homolog of the

vertebrate pharyngeal muscles (Davidson, 2007). FGF signalling plays crucial roles at least at two points in ascidian heart development: first, in inducing the TVC fate in the siblings of the anterior tail muscle cells, and second, in the segregation of the cardiac lineage and ASM lineages by inducing the ASM fate after asymmetrical divisions (Davidson et al., 2006; Razy-Krajka et al., 2018; Tolkin & Christiaen, 2012). In our study, we assessed the putative role of FGF signalling in heart development in *O. dioica* by treating embryos with SU5402 and AZD4547 from the 2-cell and the 32-cell stages, two different developmental points prior to heart induction in *O. dioica* (Ferrández-Roldán et al., 2021). We then analysed the expression of *ActnM1*, a general muscle marker highly expressed in all muscle cells from the onset of their differentiation (Almazán et al., 2019), and *NK4,* a specific heart marker whose expression is only appreciable at the ITB stage (Ferrández-Roldán et al., 2021).

Embryos treated with FgfR inhibitors from the 2-cell stage exhibited disrupted gastrulation and abnormal phenotypes when examined at the ITB stage using *ActnM1* WMISH. Specifically, mesodermal derivatives displayed either abnormal domains **(Figure 3.6B'-C')** or were entirely absent **(Figure 3.6B"-C")**. Moreover, embryos that developed relatively normal ITB morphologies still showed the presence of cardiac precursor cells **(Figure 3.6A-C, red arrowheads)**. When the inhibitory treatment was initiated at the 32-cell stage, most treated embryos exhibited *Nk4* expression in cardiac precursor cells at the ITB stage **(Figure 3.6D-F, red arrowheads)**, even when showing noticeable notochord abnormalities **(Figure 3.6E, yellow dashed lines)**. Only in embryos with severe abnormalities or developmental arrest were cardiac precursors indistinguishable from other Nk4 expression domains **(Figure 3.6E'-F')**. WMISH of *ActnM1* in DMSO-control embryos **(Figure3.6G)** and embryos treated with FgfR inhibitors from the 32-cell stage to the early hatchling stage revealed that most treated embryos had abnormal tails, showing impaired elongation and rotation **(Figure 3.6H'-I')**. Additionally, while cardiac precursors in control embryos converged near the midline into a single cardiac field, many embryos with tail malformations had cardiac precursors that remained bilaterally separated on the right and left sides of the trunk **(Figure 3.6H'-I', red arrowheads)**. These findings suggested that FGF signalling pathway plays a role in tail elongation and rotation, as well as in late cardiac morphogenesis, but not in cardiac induction or differentiation.

Our results indicated that the formation of cardiac precursors and the onset of *Nk4* expression were not altered in embryos where the FGF signalling pathway had been inhibited. This indicates that the determination and differentiation of cardiac precursor cells and the onset of the cardiogenic kernel have become independent of the FGF signalling pathway during the evolution of appendicularians. These findings contributed to unveiling the deconstruction of the cardiopharyngeal GRN in appendicularians, supporting an evolutionary scenario in which ancestral tunicates had a sessile ascidian-like adult lifestyle. The full study was published in the journal *Nature* under the title

"Cardiopharyngeal deconstruction and ancestral tunicate sessility", provided in **Annex 3** (Ferrández-Roldán et al., 2021).

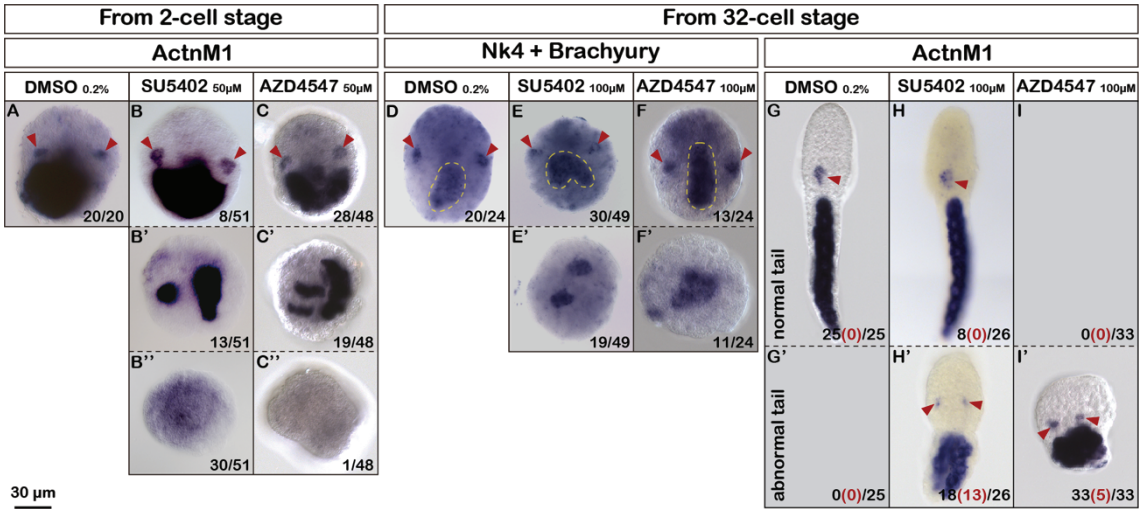


Figure 3.6. FGF inhibition during hear development in *O. dioica*. (A-C) WMISH of *ActnM1* in control embryos (A) and embryos treated with FGFR inhibitors (B-C) from the 2-cells stage until the ITB stage. Upper panels (A-C) show embryos with normal morphologies where the cardiac precursors could be recognised. Middle panels (B'-C') show embryos with aberrant morphologies and abnormal *ActnM1* expression domains due to failures in gastrulation, in which cardiac precursors could not be distinguished. Bottom panels show embryos with no *ActnM1* expression (B"-C"). **(D-F)** Double one-colour WMISH of *Brachyury* and *NK4* in control embryos (D) and embryos treated with FgfR inhibitors (E-F) from the 32-cells stage until the ITB stage. Upper panels (D-F) show embryos with normal morphologies where the cardiac precursors and the notochord were recognised. Bottom panels (E'-F') show embryos with abnormal morphologies in which cardiac precursors and the notochord could not be distinguished. **(G-I)** WMISH of *ActnM1* in control embryos (G) and embryos treated with FGFR inhibitors (H-I) from the 32-cells stage until the early-hatchling stage. Upper panels (G-I) show embryos with normal morphologies where the tail is elongated, and cardiac precursors had converged near the midline into a single cardiac field. Bottom panels (G'-I') show embryos with abnormal tails in which the elongation and rotation had been affected. In some of these embryos (red numbers in brackets) the cardiac precursors had failed to converge in the midline and were still bilaterally separated in the right and left sides of the trunk. Red arrowheads mark the cardiac precursor cells and yellow dashed lines mark the notochords. All images are dorsal views of the embryos with the anterior to the top, except for those embryos with aberrant morphologies that could not be properly oriented.

3.3. *Omic* approaches to investigate the impact of FGF inhibition on early embryogenesis in *O. dioica*

As shown in previous sections, inhibiting the FGF signalling pathway in *O. dioica* severely affected embryonic development, with varying impacts depending on the stage at which the inhibitory treatment started. The most notable effects were observed when inhibition started early enough before gastrulation, resulting in completely aberrant morphologies in most embryos **(Figure 3.1)**. Furthermore, inhibition of the expression of several tissue specific marker genes suggested a failure in the determination of cellular lineages **(Figure 3.4)**. To elucidate the molecular mechanisms underpinning the effects of FGF signalling inhibition during early embryogenesis in *O. dioica*, we conducted RNA sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) on DMSO-control embryos and embryos treated with SU5402 50 µM, from the 2-cell stage to the 64-cell stage. The RNA-seq data was used to perform

differential gene expression (DGE) analyses and identify genes that were up- or down-regulated due to FGF signalling inhibition. The ATAC-seq data was used to assess changes in chromatin accessibility resulting from FGF signalling inhibition, and to determine whether these changes correlated with alterations in the expression of genes located in those genomic regions. The chosen developmental point for sampling RNA and DNA from treated embryos was the 64-cell stage, since this is the time when virtually all cellular lineages have been determined and the embryos are undergoing gastrulation and neurulation in normal conditions.

This resulted in the sequencing of 6 paired RNA-seq and ATAC-seq replicates, which were further processed with the corresponding nf-core pipelines (Ewels et al., 2020) (see Material and Methods). By integrating data from these complementary approaches, we aimed to take the first steps in uncovering the GRNs and epigenetic modifications influenced by FGF signalling.

3.3.1. DGE analyses show downregulation of developmental genes

Sequencing of the six RNA-seq libraries yielded approximately 20 million reads per sample. Following data processing and principal component analyses (PCA), two of the six replicates were excluded due to their lower QC scores and outlier behaviour in the PCA plots **(Supplementary Figure 5)**.

DGE analyses between control and treated embryo revealed a total number of 140 differentially expressed genes (DEGs) with a $\log_2$FoldChange < -1 or > 1, and a p-adjusted value ≤ 0.05. Out of these 140 DEGs, 90 genes were downregulated and 50 were upregulated upon inhibition of the FGF signalling pathway **(Figure 3.7A)**. A Gene Ontology (GO) enrichment analysis resulted in 380 GO terms significatively enriched (p-adjusted value ≤ 0.01) among the DEGs in relation to the whole translated transcriptome of *O. dioica*. When we examined the top 50 enriched GO terms, we found that the vast majority (45/50) corresponded to biological processes related to chordate development, including the morphogenesis of numerous organs and structures, cell fate commitment, or gastrulation. The other 5 enriched GO terms corresponded to molecular functions related to the DNA binding of transcription factors and the regulation of transcription **(Figure 3.7B)**.

We found that most of the DEGs that matched these GO terms and thus are putatively involved in developmental processes were downregulated. Among them, we found *Brachyury* (referred to as *T* in **Figure 3.7A**) and *Zic1*, validating the downregulation that we noticed with the WMISH assays on treated embryos. Other genes that resulted among the downregulated DEGs and whose closest match were well-known transcription factors included *Ets1*, *Tbx2*, *Gata1*, *FoxF1*, *Pax5*, *Hes4*, *Otx1* or *Eya1* (provisionally named after their closest match in eggNOG-mapper) (Cantalapiedra et al.,

2021). Since these transcription factors are known to be involved in the determination of cellular lineages and cell differentiation in other chordates, the fact that they were downregulated upon FgfR inhibition in 64-cell *O. dioica* embryos supports the hypothesis that the FGF signalling pathway also plays an important role in such processes in appendicularians.

Apart from transcription factors, other set of downregulated genes putatively involved in developmental processes comprised several components of the Wnt signalling pathway. Three genes identified as Wnt ligands, and one receptor (fzd3) were found among the 55 most significantly downregulated genes **(Figure 3.7A)**. Considering that the Wnt catalogue in *O. dioica* consists of only 7 *Wnt* genes and 5 receptors, it can be stated that the inhibition of the FGF signalling pathway caused a major downregulation of the Wnt signalling pathway (Martí-Solans et al., 2021).

As for the upregulated genes, GO enrichment analyses only on the subset of upregulated DEGs did not result in any significantly enriched GO term. Even if we lowered the p-adjusted cut-off value of significantly upregulated genes to 0.1, only two GO terms were significantly enriched and corresponded to the biological processes of "regulation of skeletal muscle cell differentiation" and "regulation of skeletal muscle tissue development", which are also related to chordate development. The fact that no GO terms related with apoptosis or detoxification were significatively enriched among the DEGs provided further support to the fact that the alterations observed in SU5402 treated embryos are due to a specific inhibition of the FgfR and not because of a general toxicity of the treatments.

Interestingly, a significant proportion of the DEGs (53 out of 140) did not have hits against the EggNOG 5.0 database with an e-value cut-off of 0.05 or were not found to have seed orthologs at all. Even when the eggNOG-mapper cut-off values were lowered to the minimal, only a small group (7/53) of these orphan genes found a hit on the EggNOG 5.0 database. This suggests that these genes are either extremely divergent compared to their orthologs in other species, or maybe novel genes only found in appendicularians. When this same set of proteins was analysed using a broader database such as the InterPro consortium, only 25 out of 53 found matches in any of the databases that comprise the consortium, including Pfam, SMART, SUPERFAMILY, or PANTHER, and the remaining majority remained as orphan genes. The emergence of novel genes in appendicularians has been previously documented and linked to appendicularians innovations, such as the *oikosin* genes involved in house building (Hosp et al., 2012). Although the identity and nature of our orphan DEGs require more thorough investigation, it is tempting to speculate that the expansion and diversification of the FGF signalling components in appendicularians is related to the innovation of regulatory mechanisms for the expression of novel genes with currently unknown functions.

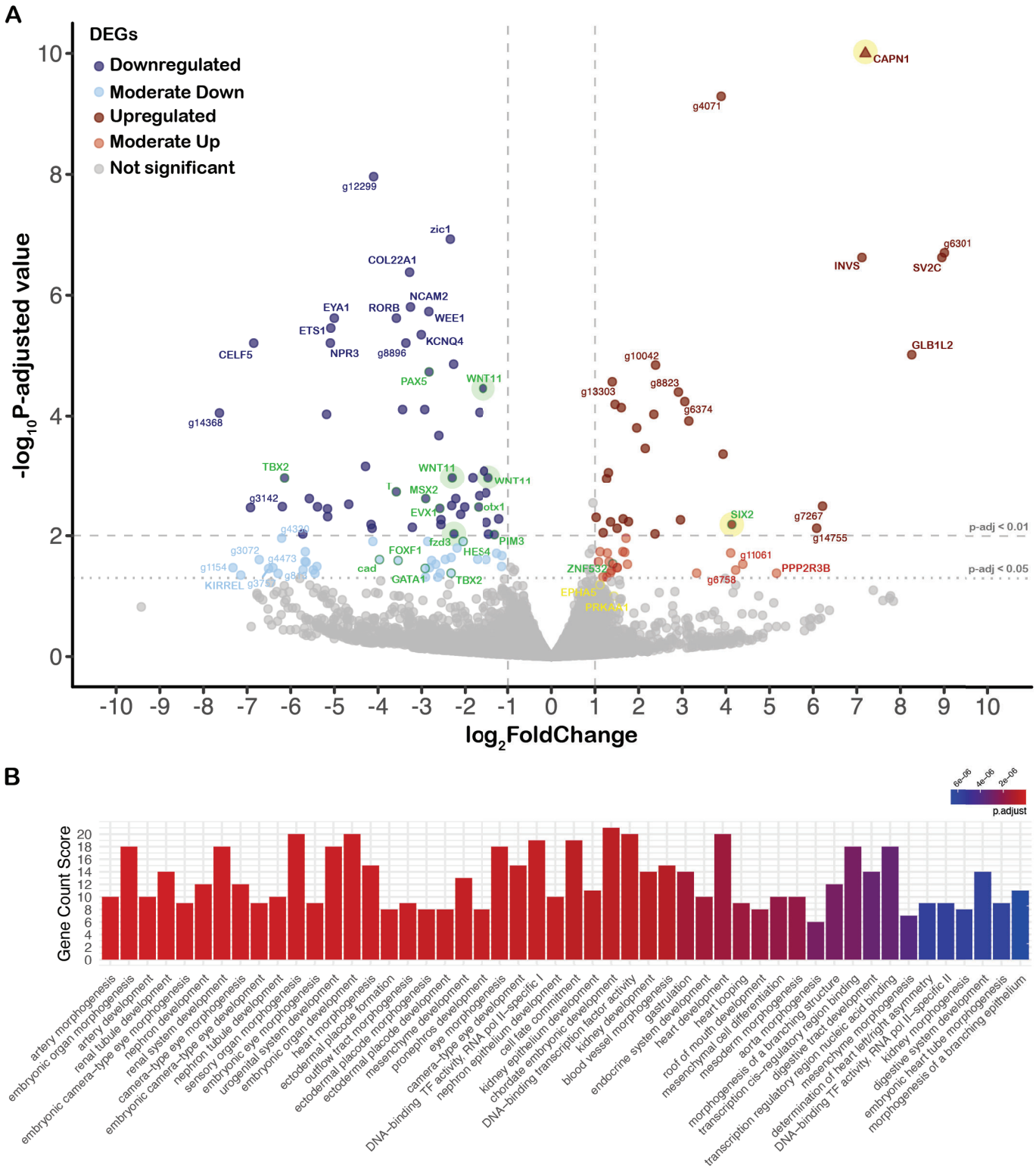


Figure 3.7. Differential Gene Expression analysis. (A) Volcano plot illustrating the differentially expressed genes upon FgfR inhibition with SU5402. The labels correspond to the preferred gene names according to EggNOG mapper results. Genes labeled with their ID (e.g. g12299) did not find an ortholog in the EggNOG 5.0 database. Blue and red labels as indicated in the legend denote the top 10 down- or up- regulated genes based on their p-adjusted value and log2FoldChange. Green labels highlight a custom selection of developmental genes. Green shadings highlight the components of the Wnt signalling pathway. Yellow labels and shadings indicate genes involved in the regulation of skeletal muscle development, which is the only upregulated GO term even with a relaxed p-adjusted value cutoff. **(B) Top 50 enriched GO terms** among the differentially expressed genes. GO terms were ranked according to their p-adjusted value.

3.3.2. ATAC-seq suggests divergence in chromatin accessibility and gene expression

Sequencing of the six ATAC-seq libraries revealed significant differences among replicates, which could be subdivided into two subsets. Libraries from replicates 1 to 3 were overamplified, resulting in an excessive enrichment of shorter fragments. These libraries were also overrepresented in the sequenced pool, causing a substantial difference in sequencing depth compared to the other subset of libraries from replicates 4 to 6 (**Supplementary Figure 6**).

This disparity in sequencing depth affected peak calling, with replicates 1 to 3 yielding significantly more peaks than replicates 4 to 6 **(Figure 3.8A)**. Despite these differences, the annotation of peaks and their association with genomic features yielded similar relative results **(Figure 3.8A)**, as did the distribution of filtered reads with respect to annotated genes **(Figure 3.8B)**. Merging all replicates for each condition resulted in approximately 55 million filtered and mapped reads and around 27,000 consensus peaks per condition. Notably, a proportion of these peaks were exclusively found in either the DMSO or the SU5402 samples, suggesting differences in chromatin accessibility due to the inhibition of the FGF signalling **(Figure 3.8C)**.

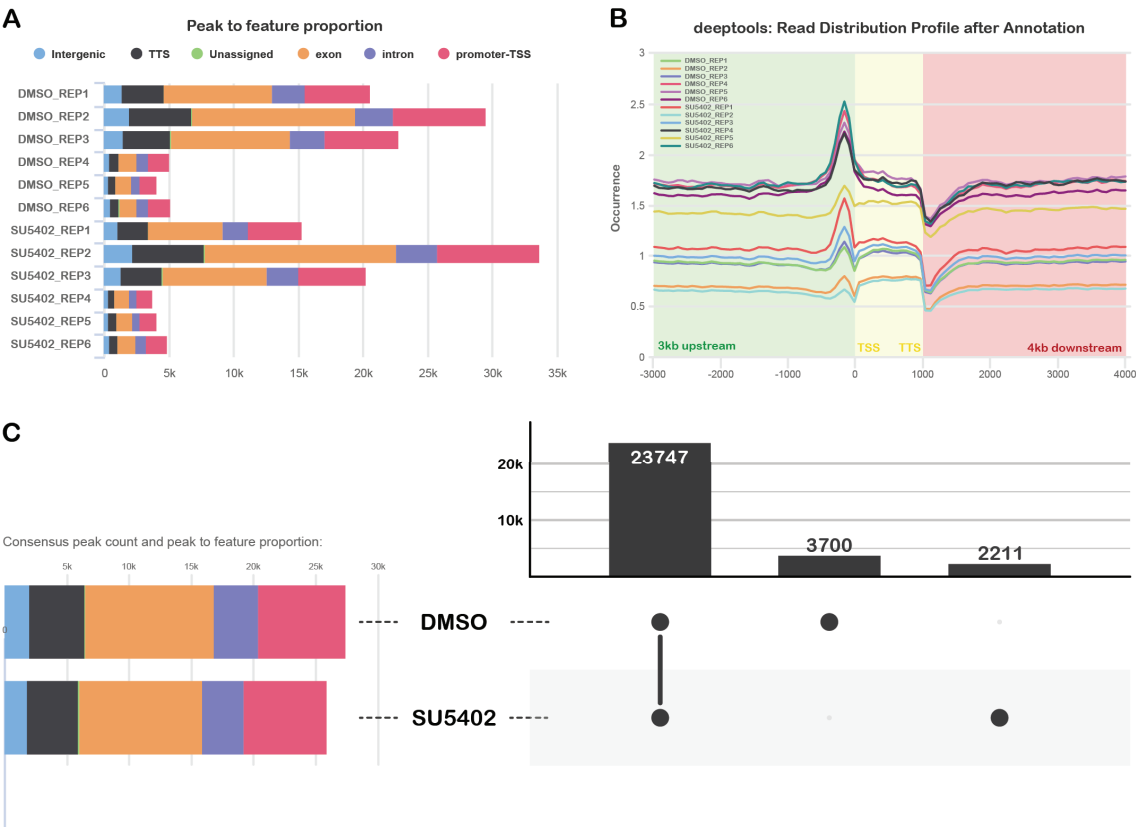


Figure 3.8. (A) Peak counts and their annotations per sample. **(B) Read distribution profile** after peak calling and annotation, showing the accumulated view of filtered reads distribution relative to the nearest annotated gene. All annotated genes have been normalized to the same size. **(C) Merged replicates**. The left graph shows peak counts and their annotation per condition after merging all replicates. The right graph compares the number of peaks common to both DMSO-control and SU5402-treated samples with the number of peaks exclusive to either DMSO-control or SU5402-treated conditions.

Given the significant discrepancy between the number of condition-exclusive peaks and the number of DEGs, establishing an association between these two features was challenging. In fact, only 30% of the DEGs (42 out of 140) were the closest genomic feature to one or various condition-exclusive peaks. Moreover, many of these 42 DEGs were associated with both DMSO-exclusive and SU5402-exclusive peaks. This discrepancy in the number of DEGs and condition-exclusive peak was not surpassed when we conducted the same analysis with only the subset of non-overamplified replicates 4 to 6. In this case, our analyses revealed 7159 common peaks, 2726 DMSO-exclusive peaks and 2419 SU5402-exclusive peaks, and only 32% of the DEGs (45 out of 140) were the closest genomic feature to one or various of these condition-exclusive peaks. Consequently, we could not correlate the differential expression of genes upon SU5402 inhibition with the presence or absence of accessible chromatin domains.

3.3.3. Differentially accessible peaks

Analyses of differential peak accessibility using all replicates did not yield significant results. This lack of significant findings was probably due to the bias introduced by the overamplification and overrepresentation of replicates 1 to 3. Therefore, we conducted the differential accessibility test using only the subset of replicates 4 to 6. This analysis identified two significant peaks with a p-adjusted value below 0.05 **(Figure 3.9A)**. Specifically, peak3068 was associated with the promoter of g5714, a *Gap-junction-beta2-like* gene (connexin), and peak2850 was associated with the promoter of g5474, a *Chondroitin sulfate-synthase-like* gene. However, none of these genes were found among our list of DEGs, nor did they show a tendency of differential expression upon FGF inhibition. When the p-adjusted value cut-off was lowered to 0.1, nine additional peaks resulted as differentially accessible **(Figure 3.9A)**. Among these, peak36 had one of the DEGs (g136) as the closest genomic feature. g136 is one of the aforementioned *O. dioica* orphan genes, and the also downregulated peak36 was associated to its promoter region.

To explore further associations between differentially accessible peaks and DEGs beyond their close proximity to transcription start sites, we mapped both the DEGs and the 11 differentially accessible peaks to the genome (**Figure 3.9B**). This revealed that, even if the differentially accessible peaks are not directly linked to the promoter of a DEG, they are often in the vicinity of DEGs or clusters of them. For example, peak3068 is near three differentially downregulated genes, and peak840 is situated within a cluster of six DEGs (**Figure 3.9B**). Most of these peaks are located in exonic or promoter regions of annotated genes, and thus their differential accessibility could be due to a transcriptional downregulation of these genes even if they didn't appear in our DGE analyses. However, some others are placed at intergenic or intronic regions (i.e. peak840, peak3068, peak3903 and peak5899), and thus could constitute cis-regulatory elements whose activity is affected by the inhibition of the FGF signalling. Further investigation is

necessary to understand whether these peaks are related to the transcriptional regulation of the nearby DEGs.

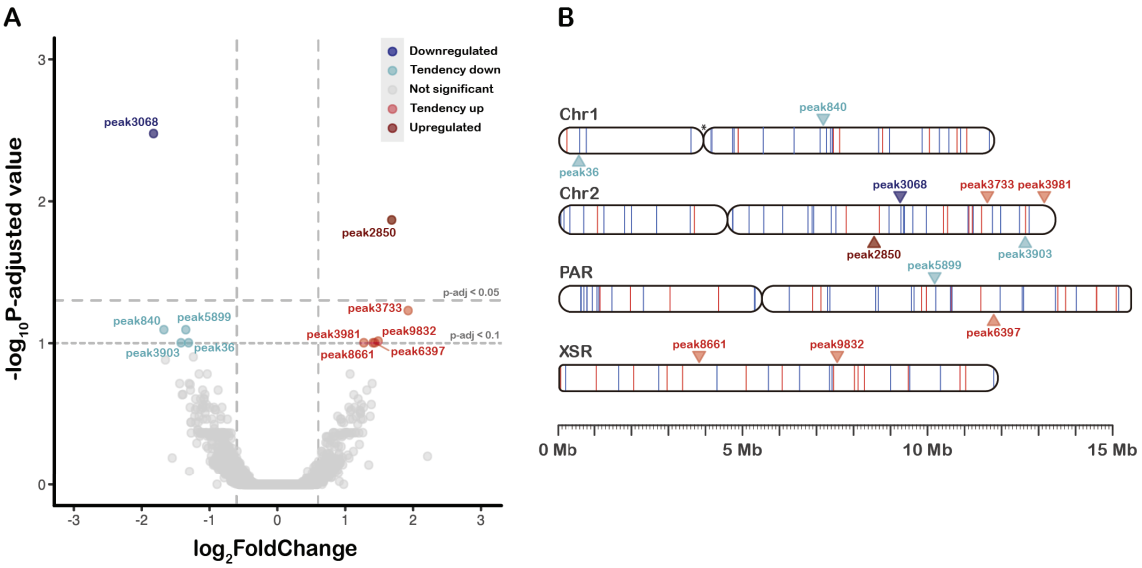


Figure 3.9. Differentially accessible peaks. (A) Volcano plot. This panel displays the differentially accessible peaks identified using replicates 4, 5, and 6. Peaks are represented based on their statistical significance and log2-fold change. Significant peaks with a p-adjusted value below 0.05 are highlighted with dark blue (downregulated) or dark red (upregulated), while those identified with a p-adjusted value below 0.1 are indicated with lighter versions of the same colours. **(B) Genome mapping.** This panel maps the differentially expressed genes (DEGs) and the eleven differentially accessible peaks to the genome of *O. dioica*. Blue lines within the chromosomes represent downregulated genes, and red lines represent upregulated genes. Peaks are represented as arrowheads, colour-coded to match the significance levels shown in panel (A). This mapping highlights the spatial relationship between DEGs and differentially accessible peaks, providing insights into potential regulatory interactions.

3.3.4. Conclusions and future perspectives

These *omic* approaches were conducted towards the end of this doctoral project at the Genomics and Regulatory Systems Unit at OIST, with the primary aim of paving the way for future research. The experiments employed animals from the Osaka *O. dioica* cryptic species, along with the corresponding genome and annotations (Plessy et al., 2024; K. Wang et al., 2020). Besides the valuable insights into the early developmental function of the FGF signalling pathway in *O. dioica*, this data will be useful for comparing the genetic response to FGF signalling inhibition across different cryptic species, and to investigate whether the massive genomic scrambling has led to divergence in developmental mechanisms.

The RNA-seq data and DGE analyses indicated a predominant downregulation of developmental genes due to FgfR inhibition, including multiple transcription factors. The role of these transcription factors in the development of specific cellular lineages or structures, and their dependency on FGF signalling, requires further validation and investigation. Nonetheless, our results highlight several promising areas for future

research. For instance, the transcription factor *otx1* was found among our DEGs. Otx transcription factors are crucial in the complex patterning of the brain in vertebrates (Huang et al., 2018; Simeone, 1998), and its expression in the developing brain of *O. dioica* suggests conserved functions among appendicularians and vertebrates (Cañestro et al., 2005). The presence of otx1 among the downregulated DEGs, along with our findings that some *Fgf9/16/20* paralogs are expressed in the neural plate and that FgfR inhibition impairs the proper CNS differentiation, hints at a conserved function of the FGF signalling in brain patterning in appendicularians.

Another noteworthy finding is the downregulation of three *Wnt11* genes, which arise from an appendicularian-specific expansion of the Wnt11 family (Martí-Solans et al., 2021). Previous research indicated that some of these *Wnt11* paralogs are expressed during embryonic development in *O. dioica* from Barcelona, while others are not, aligning with the gene expression data available at the OikoBase (Danks et al., 2013; Martí-Solans et al., 2021). Intriguingly, two of our downregulated *Wnt11* genes in *O. dioica* from Osaka (Wnt11c_OSA1 and Wnt11c_OSA2) are the orthologs of two of the genes whose expression has not been detected in the BAR/NOR cryptic species (namely GSOIDG00013856001 and GSOIDG00009921001) (Danks et al., 2013; Martí-Solans et al., 2021). This finding points to these two *Wnt11* genes as candidates of differentially regulated developmental genes among the distinct *O. dioica* cryptic species, providing initial directions for future studies about the evolvability of developmental GRNs in appendicularians.

Regarding the ATAC-seq data, the lack of more associations between differentially accessible peaks and DEGs could be attributed to several factors. Firstly, limitations in our ATAC-seq samples, including overamplification of replicates 1 to 3 and low sequencing depth of the other replicates, resulted in no samples meeting the parameters for reliable ATAC-seq analyses. Secondly, the techniques' resolution and the biological features they assess, as mRNA counts can vary more much more than the accessibility of DNA regions producing those mRNAs. This is particularly relevant when genes are expressed in only a few cells in an embryo and the samples consist of bulk embryos, as it is our case. For example, in a 64-cell *O. dioica* embryo, *Brachyury* gene expression in only two pairs of cells yields significant mRNA increase, but DNA accessibility measurements might not reflect this due to the small number of cells with open promoters. Thirdly, the state-of-the-art of the techniques plays a role., RNA-seq is well-established and optimized for various conditions, including *O. dioica*, whereas ATAC-seq is relatively recent and, to our knowledge, this is the first time that it was used to study differential conditions in *O. dioica*. Given the limitations of our samples, these results should be considered just as a preliminary indication of how the ATAC-seq approach can be utilized to study the impact of inhibiting a specific developmental pathway on the transcriptional regulation of *O. dioica*.

4. An evolutionary scenario for the remodelling of the FGF signalling in Appendicularians

4.1. The minimal Fgf subfamily catalogue in chordates

The evolution of the FGF signalling in *O. dioica* has revealed an unprecedented remodelling in the catalogue of *Fgf* and *FgfR* genes when compared to the rest of chordates **(Figure 4.1)**. This remodelling has been marked not only by the massive loss of most Fgf subfamilies, but also by the expansion of the remaining ones. Moreover, the appendicularian expansion of the Fgf9/16/20 and Fgf11/12/13/14 subfamilies has been paralleled by a similar expansion of the *FgfR* gene. Notably, the *FgfR* gene is present as a single copy in cephalochordates and ascidians, and its expansion in vertebrates, along with the expansion of *Fgf* genes, has been associated to an increase in developmental complexity during the radiation of the phylum (Itoh & Ornitz, 2004). The appendicualarian expansion of *Fgf* and *FgfR* genes, together with the diversification of the resulting paralogs, depicts a complex process of gene gains and losses. This implied the conservation or loss of ancestral functions, but also the gain of new functions that may have been key for the evolutionary innovations in the appendicularian lineage.

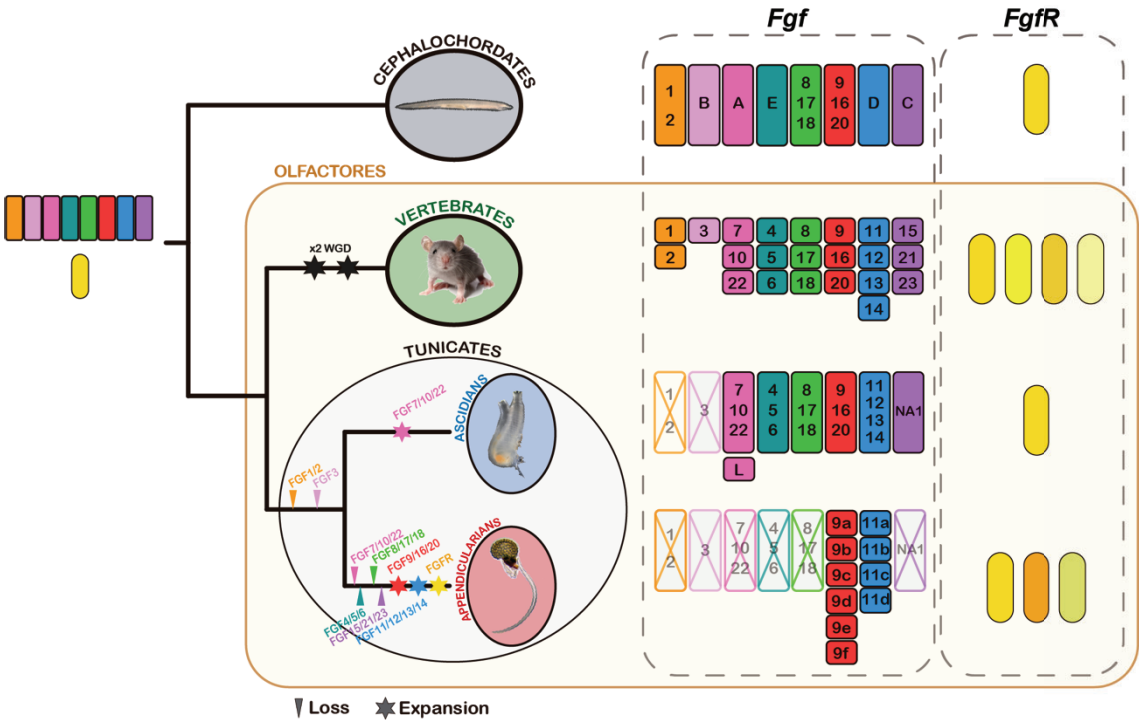


Figure 4.1. Evolution of Fgf subfamilies and the FgfR gene in the chordate phylum. The sequence of losses and expansions in the appendicularian branch is not intended to represent the chronological order. Based on Oulion et al. (2012).

Whether the losses preceded the expansions or vice versa remains enigmatic, but in any case, our findings reveal that *O. dioica* is an evolutionary knockout for the chordate Fgf subfamilies Fgf1/2, Fgf3, Fgf4/5/6, Fgf7/10/22, Fgf8/17/18 and Fgf19/21/23. The fact that the surviving subfamilies were the Fgf9/16/20 and the Fgf11/12/13/14, one classically associated to paracrine signalling functions and the other to intracellular functions, points to evolutionary restrictions in the evolvability of a minimal Fgf catalogue. What has been the selective constraints that led to their survival, or what has been the selective advantage that led to the retention of their duplicates, likely allowing the loss of the other subfamilies, are some of the key questions that the results of this doctoral project elucidate.

4.1.1. The Fgf11/12/13/14 subfamily

Our results shed light on innovative aspects of the Fgf11/12/13/14 subfamily related to the evolution of olfactores. Previous research on ascidians identified seven Fgf genes spanning six different subfamilies, including the Fgf9/16/20 and Fgf11/12/13/14 subfamilies. However, WMISH analyses have not yet elucidated the developmental expression of the ascidian *Fgf11/12/13/14* ortholog. Our results reveal for the first time the expression atlas of an *Fgf11/12/13/14* ortholog in a tunicate, suggesting significant involvement in the development of specific structures and tissues. The expression of *O. dioica Fgf11/12/13/14* paralogs in neural structures through embryonic and larval development, including specific regions of the brain, subsets of cells in the caudal ganglion, and isolated neuronal bodies along the nerve cord, points to a central role in the formation of neuroectodermal derivatives, as previously described in vertebrates (Ornitz & Itoh, 2015; Zhang et al., 2012). Recent single cell RNA-seq studies on the ascidian *C. robusta* suggest the presence of *Fgf11/12/13/14* transcripts in specific neurons of the larval tail, including the bipolar tail neurons, which share properties with the neural-crest-derived dorsal root ganglia in vertebrates (Horie et al., 2018; Stolfi et al., 2015). We also identified alternative splice variants that differ in the first exon in most *O. dioica Fgf11/12/13/14* paralogs, a feature shared with vertebrate *Fgf11/12/13/14* paralogs crucial for their neural function (Laezza et al., 2009; Pablo & Pitt, 2016). The discovery of alternative isoforms differing in the first exon in both appendicularians and ascidians suggests a conserved and significant role of alternative splicing in these genes' functional mechanism. Recent phylogenetic reconstructions based on gene content and synteny conservation indicated that *FgfD* is the chephalochordate ortholog of Fgf11/12/13/14 (S. Bertrand et al., 2011; Oulion et al., 2012). However, our analysis of gene structure and conserved protein motifs did not identify any of the conserved features between vertebrates and tunicates in the cephalochordate *Fgf* genes, indicating that they are conserved synapomorphic traits of the Fgf11/12/13/14 subfamily in olfactores. Future studies of these genes in appendicularians and other tunicates might be relevant to better understand the role of the Fgf11/12/13/14 subfamily in the origin

and evolution of olfactores, with a special focus in the evolution of the nervous system and other developmentally related structures such as the neural crests.

In vertebrates, mutations in some of the Fgf11/12/13/14 paralogs are linked to human neural diseases (Ornitz & Itoh, 2015; C. Wang et al., 2011; Zhang et al., 2012). These mutations have been typically associated to the best-known function of vertebrate *Fgf11/12/13/14* paralogs as modulators of neuronal excitability through their interaction with voltage-gated ion channels in an FgfR-independent manner (Pablo & Pitt, 2016). Over recent years, the list of intracellular partners interacting with vertebrate *Fgf11/12/13/14* paralogs has expanded to include IB2 (MAPK8IP2, Mitogen-activated protein kinase 8-interacting protein 2), β-tubulin, NEMO (NF-κB essential modulator), hypoxia-inducible factor-1a (HIF-1a), and casein kinase 2 (CK2) (Ornitz & Itoh, 2022). However, the functional implications of these novel interacting partners remain mostly unknown. Additionally, recent studies have demonstrated that some vertebrate *Fgf11/12/13/14* paralogs can activate FgfRs, contrarily to classical belief, suggesting that they may perform concealed FgfR-dependent signalling functions (Lin et al., 2019; Sochacka et al., 2020). Future studies on tunicates may be of great interest for a better understanding of the functions of the Fgf11/12/13/14 subfamily members and to develop new animal models to study the molecular bases of related human neurological disorders

4.1.2. The Fgf9/16/20 subfamily

Regarding the Fgf9/16/20 subfamily, the expression of *O. dioica* paralogs points to deeply conserved developmental roles. In early development, four of the *O. dioica Fgf9/16/20* paralogs of (i.e. *Fgf9/16/20c*, *d*, *e* and *f*) are expressed in the smallest pair of vegetal blastomeres in the 8-cell embryo. This expression is equivalent to that of the ascidian co-ortholog *Fgf9/16/20* in the A4.1 derived vegetal blastomeres of *C. robusta* (V. Bertrand et al., 2003; Hudson et al., 2016; Imai et al., 2002a; Satou, 2020). This expression has been related to an ancestral bilaterian function of the FGF signalling in initiating mesodermal and endodermal GRNs (Technau & Scholz, 2003). Our (DGE analyses, showing a predominance of downregulated developmental genes upon inhibition of FgfRs, further support the hypothesis that the FGF signalling pathway is crucial for the initiation of GRNs driving early embryogenesis in *O. dioica*. In ascidians, other Fgf orthologs are also involved in early development, such as *Fgf8/17/18* in mesodermal fate maintenance (Yasuo & Hudson, 2007), or *Fgf7/10/22* in the convergent extension of the notochord (Shi et al., 2009). However, in these processes, *Fgf9/16/20* usually determines the cellular fate (Satou, 2020), and orthologs from other subfamilies act secondarily for fate maintenance or for the coordination of later morphogenetic processes. It is tempting to speculate that the induction of cellular lineages imposed a stronger selective constraint than the secondary processes that cells undergo once they are fated. Moreover, the early expression of F*gf9/16/20* paralogs at the eight-cell stage

in *O. dioica* also aligns with recent findings in ascidians showing that *Fgf9/16/20* acts as a timer for zygotic genome activation, whose responsiveness sharply starts between the 8-cell and 16-cell stages (Treen et al., 2023). The involvement of Fgf9/16/20 genes in the initiation of early GRNs or even zygotic transcription could have imposed a strong selective constraint for its loss in appendicularians.

Another example of conservation of ancestral functions was shown by *Fgf9/16/20d-e* in the developing CNS from the 64-cell to tailbud stages. These expression domains are comparable to the expression of the ascidian *Fgf9/16/20* co-ortholog in the CNS at the level dorsal to the anterior tip of the notochord (Imai et al., 2002b; Miyazaki et al., 2007). The conservation of this expression domain suggests preserved *Fgf9/16/20* ancestral functions in the AP patterning of the CNS among tunicates, what might have also imposed a strong selective constraint to retain this Fgf subfamily.

In summary, the conservation of ancestral functions likely imposed selective constraints for the retention of the Fgf9/16/20 subfamily in appendicularians. Moreover, its expansion and diversification may have been favoured by an adaptive trend to recruit FGF signaling in various developmental processes.

4.2. Fgf evolution reflects evolutionary innovations in appendicularian tunicates

Appendicularians have been classically considered the sister clade of ascidians, but whether the ancestral tunicate had a free-living lifestyle like appendicularians or a biphasic lifestyle like ascidians remained enigmatic (Stach, 2007; Stach et al., 2008; Stach & Turbeville, 2002). Our work on the evolution of the cardiopharyngeal GRN supports the hypothesis that appendicularians indeed descend from an ascidian-like sessile ancestor (**Annex 3**) (Ferrández-Roldán et al., 2021).

This interpretation implies that appendicularians had to transition from a biphasic lifestyle with a motile larva and a sessile adult to a completely free-living lifestyle. By comparing our results on *Fgf* gene expression in *O. dioica* with the current knowledge in ascidians, we can draw some hypothesis of how the evolution of the FGF signalling pathway may reflect the transition from an ascidian lifestyle to a completely free-living lifestyle. Our comparative expression analysis between *O. dioica* and ascidians reveal examples of different evolutionary patterns that can be classified into three categories: extinction of ancestral expression domains and functions linked to gene losses, function shuffling among surviving paralogs upon the loss of genes, and innovation of novel expression domains in novel paralogs **(Figure 4.2A)**.

4.2.1. Extinction of ancestral expression domains and functions linked to gene losses

In ascidians, *Fgf7/10/22* (formerly referred to as *Fgf3*) is strongly expressed throughout the neural tube during tadpole stages **(Figure 4.2B)** (Imai et al., 2004). Its signalling function is crucial for the convergent extension of the notochord located just above the neural tube (Shi et al., 2009). Moreover, ascidian *Fgf7/10/22* knockouts fail to reabsorb the tail during metamorphosis, suggesting that *Fgf7/10/22* serves as an inductive cue for this process (Treen et al., 2014). In *O. dioica*, the *Fgf7/10/22* subfamily has been lost, implying that the convergent extension of the notochord may have become independent of FGF signalling from the neural tube. Moreover, in this context it is tempting to speculate that the loss of *Fgf7/10/22* in appendicularians could be also related with the loss of a drastic metamorphosis and the lack of absorption of the tail **(Figure 4.2A)**.

Additionally, ascidian metamorphosis involves mesenchymal tissue, which consists of mesodermal cells that remain in a pluripotent state until metamorphosis and then differentiate into adult tissues and structures. The ascidian *Fgf8/17/18* ortholog plays a role in the early differentiation of mesenchymal cells, with its expression maintained throughout embryonic development **(Figure 4.2B)** (Imai et al., 2004; Satou, 2020). Therefore, the loss of *Fgf8/17/18* along with *Fgf7/10/22* in appendicularians could be related to the loss of mesenchymal tissue and the metamorphic process **(Figure 4.2A)**.

4.2.2. Function shuffling among surviving paralogs upon the loss of genes

In this second category, our study identifies instances of function shuffling, a process in which paralogs interchange functions, typically occurring after gene duplications (McClintock et al., 2001) and particularly associated to gene losses (Cañestro et al., 2009). We identified two potential cases of function shuffling between *Fgf* orthologs lost in appendicularians and the resulting paralogs of the Fgf9/16/20 subfamily expansion, specifically *Fgf9/16/20d* in *O. dioica.*

The first case involves *Fgf8/17/18*, which in ascidian larvae exhibits an epidermal expression domain at the tip of the tail and likely acts as a secreted posterior tail FGF source (PTFS; **Figure 4.2B**) (K. Kim et al., 2020; Pasini et al., 2012). In *O. dioica* the Fgf8/17/18 subfamily has been lost, but *Fgf9/16/20d* displays a similar epidermal expression domain at the posterior tip of the tail. This suggests that *Fgf9/16/20d* could function as an equivalent posterior tail FGF source **(Figure 4.2B)**.

The second case involves both *Fgf8/17/18* and *Fgf7/10/22*, which in ascidians are expressed during the development of the atrial siphon, a structure related to the evolution of otic placode homologs (Kourakis & Smith, 2007). In *O. dioica*, the *Fgf8/17/18* and Fgf7/10/22 subfamilies have been lost, but *Fgf9/16/20d*, as well as

Fgf11/12/13/14c, show an equivalent expression domain in the Langerhans receptor primordium, which has been proposed as a homologous placodal structure in appendicularians **(Figure 4.2B)** (Bassham & Postlethwait, 2005).

These instances of function shuffling suggest the existence of a selective pressure to conserve the function of lost Fgf subfamilies in appendicularians and highlight the "promiscuity" of the FGF signalling (Beenken et al., 2012; Green et al., 1996).

4.2.3. Innovation of novel expression domains in novel paralogs

In the third category, among expression domains innovated by paralogs resulting from the appendicularian Fgf expansion, we find at least three cases in *O. dioica* (**Figure 4.2A**).

The first case focuses on *Fgf11/12/13/14a*, which is specifically expressed in the stomodeum of *O. dioica* in early hatchling stages, when the mouth is opening (**Figure 4.2B**). In contrast, the ascidian ortholog of *Fgf11/12/13/14* shows no detectable expression during development (Satou et al., 2002; Treen et al., 2014). Notably, in ascidians the stomodeum and mouth opening derive from the anterior neuropore (Veeman et al., 2010), whereas in *O. dioica* these structures develop in the most rostral part of the trunk, directly connected to the pharynx. The expression of *Pax2/5/8a* in the stomodeum primordium of *O. dioica* may be associated with cellular functions such as perforation, adhesion, and fusion of epithelial openings, including the mouth, as suggested by Bassham et al. (2008). The shared expression of placodal markers like *Pitx* suggests a deep genetic homology among mouths and adenohypophysis-like organs, such as the ciliary funnel in tunicates and the Hatschek's pit in cephalochordates (Bassham & Postlethwait, 2005). However, it is possible that the mouths of ascidians and appendicularians have independent evolutionary origins, recruiting a common cassette of placodal genes, as has been suggested for other placodal-derived structures (Bassham & Postlethwait, 2005). Fgfs are involved in the late development of the mouth and gill slits across diverse taxa, suggesting that the FGF signalling might have been repeatedly recruited during the evolution of perforated structures (S. Bertrand et al., 2011a; Crump et al., 2004; Fan et al., 2018; Rees et al., 2024; Röttinger et al., 2008). In *O. dioica*, the high expression of *Fgf11/12/13/14a* in the stomodeum, pharyngeal slits, and the rostro-ventral part of the brain related to the ciliary funnel coincides with the expression of *Pax2/5/8a* (Bassham et al., 2008). This suggests that *O. dioica* might have uniquely recruited *Fgf11/12/13/14a* for the development of ciliary and perforated structures. To our knowledge, this would be the first case in which an intracrine Fgf-ligand has been associated with mouth development.

The second case is related to the expression of several Fgfs in mesodermal derivatives, such as the notochord or muscle cells, particularly in the anterior region of the tail during tailbud and hatchling stages. In *O. dioica, Fgf9/16/20a* is strongly

expressed in the most posterior pair of muscle cells in the tail, as well as in the most anterior pair and in the first and third notochord cells. Additionally, *Fgf9/16/20b-d* expression is observed in the three most anterior and ventral muscle cells at tailbud stages. This indicates that the anterior mesoderm of the tail could function as an anterior tail FGF source (ATFS) of secreted Fgf9/16/20 ligands in appendicularians **(Figure 4.2B)**. This pattern contrasts sharply with that of ascidians, where the *Fgf9/16/20* source is restricted to the most posterior part of the tail (PTFS), often associated with tail elongation and posterior cell identity, similar to vertebrates (Diez del Corral & Storey, 2004; Imai et al., 2004; Olivera-Martinez et al., 2012; Pasini et al., 2012). In ascidians, the most anterior muscle cells of the tail express *Aldh1a*, serving as a source of RA (anterior tail retinoic acid source, ATRAS) **(Figure 4.2B)** (Nagatomo & Fujiwara, 2003). Given the conserved antagonistic interaction between FGF and RA signalling in ascidians and vertebrates, where down-regulation of RA signalling often leads to an increase of Fgf expression (Paschaki et al., 2013; Pasini et al., 2012), it is plausible that the evolutionary innovation of the ATFS in appendicularians is related to their loss of the RA signalling (**Figure 4.2**) (Cañestro & Postlethwait, 2007; Martí-Solans et al., 2016). The loss of RA signalling in appendicularians might have reduced selective constrains, allowing some *Fgf9/16/20* genes to gain novel expression domains in the anterior part of the tail. This drastic difference may represent a major shift in developmental signalling sources between appendicularians and ascidians, potentially driving the divergent evolution of developmental processes associated to the distinct body plans and lifestyles characteristic of these two groups of tunicates.

The third case of novel Fgf expression domains in *O. dioica* is related to the patterning of the epidermis in late hatchling stages. In the tail, the lateral wings that will develop into the fins express three *Fgf9/16/20* and two *Fgf11/12/13/14* paralogs. In the trunk, the oikoblast, the epidermal organ responsible for building the house, has recruited the expression of nearly all *Fgf9/16/20* and *Fgf11/12/13/14* paralogs. Most of these paralogs are expressed broadly, though some appear to be restricted or excluded from certain regions pattern this complex organ. Notably, no comparable epidermal Fgf expression domains have been observed in the tail or trunk of the ascidian larva (Imai et al., 2004; Satou et al., 2002). This suggests that the recruitment of Fgf for epidermal patterning may be an appendicularian innovation. This innovation is likely linked to the evolution of the house-making organ, the oikoblast, and the tail movements that characterize the fully free-living lifestyle of appendicularians.

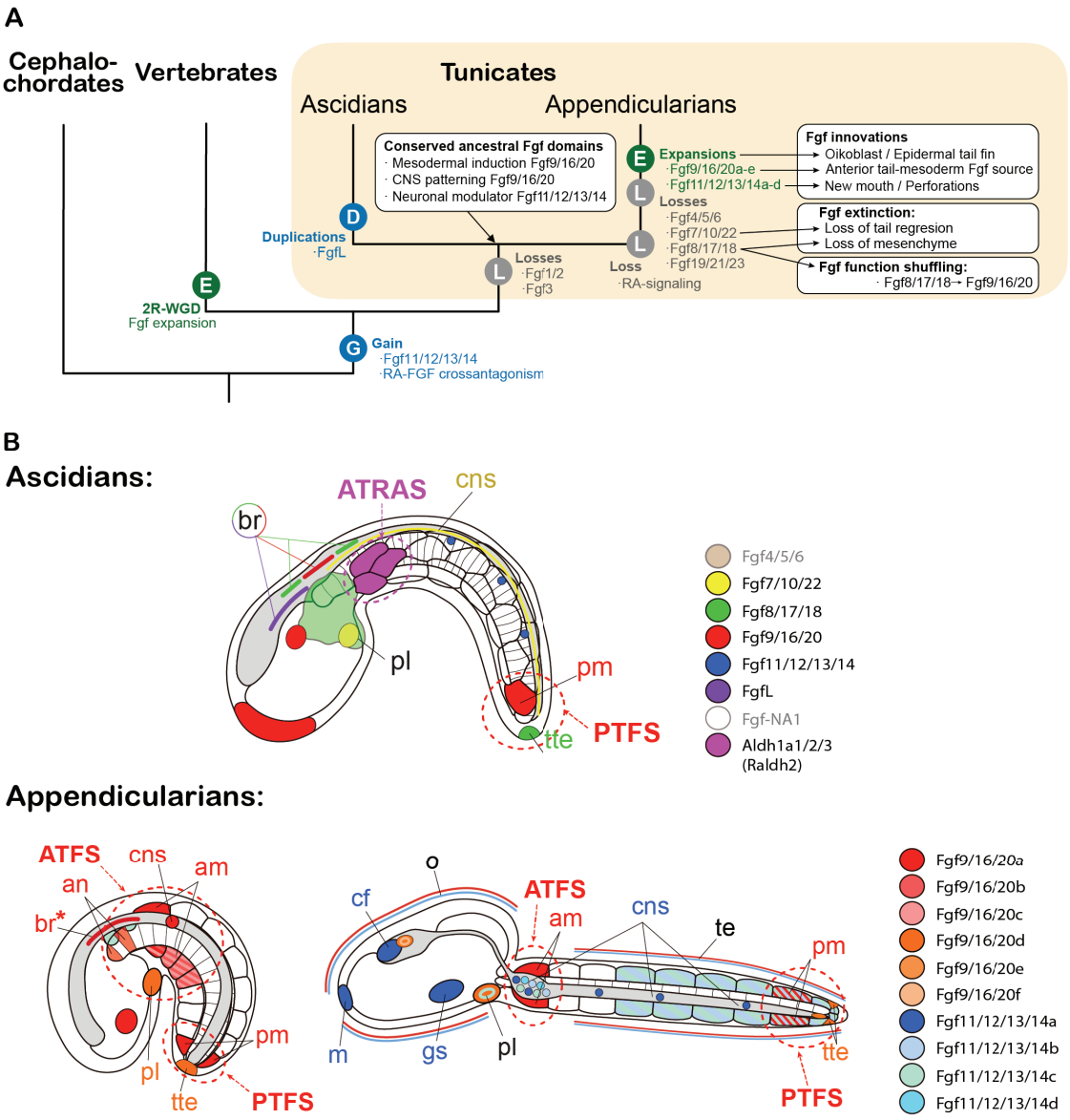


Figure 4.2. Evolutionary scenario of Fgf subfamilies in tunicates. (A) Evolutionary tree of chordate subphyla indicating main events of losses (L), gains (G), duplications (D) or expansions by burst of duplications (E), as well main associated patterns of conservation, innovation, function shuffling and extinction of Fgf expression domains. 2R-WGD two rounds of whole genome duplication **(B)** Comparative schematic representation of the main expression domains of Fgf subfamily members between ascidians and appendicularians (tailbud stage in the left, and hatchling in the right) . Distinct colours are assigned to each Fgf as indicated in the figure legend. FgfNA1 has not been described in ascidians, and Fgf4/5/6 has been detected maternally and widely throughout development with no obvious tissue-specific domains. Asterisk denotes that expression belong to a slight earlier expression than the one represented in the figure. Ascidians are characterized by the presence of a PTFS (posterior-tail FGF source) and an ATRAS (Anterior-tail RA source), while in appendicularians the loss of RA-signalling might be related to the innovation of an ATFS (anterior-tail FGF source) considering the gain of conserved RA-FGF antagonistic action at the base of olfactores Abbreviations: am: anterior notochord, am: anterior muscle, br: brain, cf: ciliary funnel, cns: central nervous system, gs: gill slit, m: mouth, o: oikoplastic epithilium, pl: placode, pm: posterior muscle, te: tail epidermis, tte: terminal tail epidermis.

Overall, the evolution of the Fgf family in appendicularians represents a paradigmatic example of what could be referred as “less, but more”, in which massive gene losses, but also extensive duplications, result in the loss or conservation of ancestral *Fgf* expression domains and in the innovation of new expression domains. Interestingly, many of these innovations, as well as some of the losses, can be related to the transition from an ancestral ascidian-like biphasic lifestyle to the fully free-living lifestyle that characterize appendicularians. Our results on the characterization of the Fgf family in *O. dioica*, presented in section 1, along with the putative evolutionary implications described in this section, are currently being prepared for publication under the title “Less, but more: Fgf gene family evolution in the appendicularian tunicate *Oikopleura dioica*”. The full draft of the manuscript is provided in **Annex 4**.

Future functional analyses are necessary to explore several key areas:

- The impact of recruiting *Fgf11/12/13/14* genes on the innovation of a new mouth opening.
- The role of Fgf recruitment in the patterning of the oikoblast and the innovation of the house.
- The role of Fgf in the development of the fin and its significance for the evolution of tail movements.
- The emergence of an anterior tail FGF source (ATFS) and its relationship to the evolutionary knockout of RA signaling in appendicularians.
- The implications of the loss of *Fgf7/10/22* for the loss of tail absorption and the absence of metamorphosis during the evolution of the fully free-living lifestyle of appendicularians.

These investigations will provide deeper insights into the evolutionary innovations that have shaped the unique developmental processes of appendicularians.

CONCLUSIONS

1. The complete catalogue of *Fgf* genes in the *Oikopleura dioica* lineage comprises 10 genes, which are conserved in a one-to-one orthology across the three cryptic species despite the unprecedented scrambling of their genomes.

2. Appendicularians have lost six out of the eight chordate Fgf subfamilies: Fgf1/2, Fgf3, Fgf4/5/6, Fgf7/10/22, Fgf8/17/18 and Fgf19/21/23, what makes them an evolutionary knockout (eKO) for these *Fgf* genes

3. The massive loss of Fgf subfamilies in appendicularians has been accompanied by an expansion of the two remaining ones, namely Fgf9/16/20 and Fgf11/12/13/14, rendering six Fgf9/16/20 and four Fgf11/12/13/14 novel paralogs in the O. dioica lineage.

4. Differences on gene structure, putative functional motifs, and developmental expression patterns of *O. dioica* Fgf9/16/20 and Fgf11/12/13/14 paralogs indicate functional diversification of the expanded Fgf subfamilies.

5. The *FgfR* gene has paralleled the gene expansion and diversification of Fgf9/16/20 and Fgf11/12/13/14 subfamilies in appendicularians, suggesting that instead of a "simplified" version of the Fgf signalling, the appendicularians Fgf-FgfR system depicts a lineage-specific entangling of the pathway.

6. The three main intracellular transduction pathways associated to the activation of the FgfR (i.e. MAPK, PLCγ/PKC and PI3K/AKT) have been mostly preserved in *O. dioica* and the expression of their components suggests an active role of the RTK signalling during development.

7. The loss of classical *Ras* genes and the *Spred* gene in *O. dioica* indicates structural modifications in the molecular mechanisms mediating the intracellular response to FgfR activation.

8. FGF signalling is essential for gastrulation and the determination of cellular lineages in *O. dioica*, as revealed by developmental failure, loss of tissue-specific marker gene expression, and downregulation of developmental genes upon systemic inhibition of the FgfR with SU5402 from early development.

9. The detection of zygotic *Fgf* expression in vegetal blastomeres at the 8-cell stage and the effects of early inhibition of the FGF signalling are compatible with its function as a timer for zygotic genome activation, as recently proposed in ascidians.

10. The FGF signalling pathway does not have an essential function in cardiac differentiation in *O. dioica*, as revealed by time-window specific FGF inhibitory treatments.

11. The broad epidermal expression of *Fgf9/16/20* and *Fgf11/12/13/14* paralogs in various regions of the house-building organ, the oikoblast, and specific areas of the developing fins suggests a massive recruitment of novel *Fgf* paralogs in the evolution of developmental mechanisms for appendicularian-specific adaptive innovations related to their fully free-living lifestyle.

12. The common presence of intron signatures and alternative 5' splice-variants in all *Fgf11/12/13/14* orthologs in appendicularians, ascidians, and vertebrates, but not in cephalochordates, suggests that this Fgf subfamily might be a synapomorphy of olfactores, likely related to the evolution of its intracellular function modulating neuronal excitability.

13. The expression of *Fgf11/12/13/14a* during the development of the ciliary funnel of the brain, the mouth opening, and the pharyngeal slits points to a role for this intracellular Fgf in the formation of ciliated epidermal perforations and suggests that the mouth of *O. dioica* might not be homologous to the oral siphon in ascidians.

14. The losses and expansions of *Fgf* genes in appendicularians have been accompanied by functional shuffling among lost genes and novel paralogs, as exemplified by the expression of *Fgf9/16/20d* at the tip of the tail in *O. dioica*, equivalent to the expression of *Fgf8/17/18* at the tip of the tail in ascidians.

15. The loss of *Fgf7/10/22* in appendicularians might be related to the retention of the tail throughout their life, as suggested by the fact that *Fgf7/10/22* knockouts in ascidians fail to regress the tail in their drastic metamorphosis.

16. The study of Fgf family evolution in appendicularians helps to better understand the role of gene loss as an evolutionary force and illustrates a paradigmatic example of "less, but more", where massive gene losses, but also extensive duplications, result in the overall conservation or loss of ancestral Fgf expression domains, functional shuffling among paralogs, and innovation of new expression domains.

BIBLIOGRAPHY

Akbulut, S., Reddi, A. L., Aggarwal, P., Ambardekar, C., Canciani, B., Kim, M. K. H., Hix, L., Vilimas, T., Mason, J., Basson, M. A., Lovatt, M., Powell, J., Collins, S., Quatela, S., Phillips, M., & Licht, J. D. (2010). Sprouty Proteins Inhibit Receptor-mediated Activation of Phosphatidylinositol-specific Phospholipase C. *Molecular Biology of the Cell*, *21*, 3487–3496. https://doi.org/10.1091/mbc.E10

Albalat, R., & Cañestro, C. (2016). Evolution by gene loss. *Nature Reviews. Genetics*, *17*(7), 379–391. https://doi.org/10.1038/nrg.2016.39

Almazán, A., Ferrández-Roldán, A., Albalat, R., & Cañestro, C. (2019). Developmental atlas of appendicularian Oikopleura dioica actins provides new insights into the evolution of the notochord and the cardio-paraxial muscle in chordates. *Developmental Biology*, *448*(2), 260–270. https://doi.org/10.1016/j.ydbio.2018.09.003

Aloe, L. (2004). Rita Levi-Montalcini: The discovery of nerve growth factor and modern neurobiology. In *Trends in Cell Biology* (Vol. 14, Issue 7, pp. 395–399). https://doi.org/10.1016/j.tcb.2004.05.011

Andrikou, C., & Hejnol, A. (2021). FGF signaling acts on different levels of mesoderm development within Spiralia. *Development*, *148*(10). https://doi.org/10.1242/dev.196089

Antoine, M., Reimers, K., Wirz, W., Gressner, A. M., Müller, R., & Kiefer, P. (2005). Fibroblast growth factor 3, a protein with a dual subcellular fate, is interacting with human ribosomal protein S2. *Biochemical and Biophysical Research Communications*, *338*(2), 1248–1255. https://doi.org/10.1016/j.bbrc.2005.10.079

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B., & Herrmann, B. G. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Developmental Cell*, *4*(3), 395–406. https://doi.org/10.1016/s1534-5807(03)00055-8

Ayada, T., Taniguchi, K., Okamoto, F., Kato, R., Komune, S., Takaesu, G., & Yoshimura, A. (2009). Sprouty4 negatively regulates protein kinase C activation by inhibiting phosphatidylinositol 4,5-biphosphate hydrolysis. *Oncogene*, *28*(8), 1076–1088. https://doi.org/10.1038/onc.2008.464

Babonis, L. S., & Martindale, M. Q. (2017). Phylogenetic evidence for the modular evolution of metazoan signalling pathways. In *Philosophical Transactions of the*

Royal Society B: Biological Sciences (Vol. 372, Issue 1713). Royal Society of London. https://doi.org/10.1098/rstb.2015.0477

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., … Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871–876. https://doi.org/10.1126/science.abj8754

Bai, D., Ueno, L., & Vogt, P. K. (2009). Akt-mediated regulation of NFκB and the essentialness of NFκB for the oncogenicity of PI3K and Akt. *International Journal of Cancer*, *125*(12), 2863–2870. https://doi.org/10.1002/ijc.24748

Barolo, S., & Posakony, J. W. (2002). Three habits of highly effective signaling pathways: Principles of transcriptional control by developmental cell signaling. In *Genes and Development* (Vol. 16, Issue 10, pp. 1167–1181). https://doi.org/10.1101/gad.976502

Bassham, S., Cañestro, C., & Postlethwait, J. H. (2008). Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. *BMC Biology*, *6*, 35. https://doi.org/10.1186/1741-7007-6-35

Bassham, S., & Postlethwait, J. (2000). Brachyury (T) expression in embryos of a larvacean urochordate, Oikopleura dioica, and the ancestral role of T. *Developmental Biology*, *220*(2), 322–332. https://doi.org/10.1006/dbio.2000.9647

Bassham, S., & Postlethwait, J. H. (2005). The evolutionary history of placodes: A molecular genetic investigation of the larvacean urochordate Oikopleura dioica. *Development*, *132*(19), 4259–4272. https://doi.org/10.1242/dev.01973

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., … Zhang, J. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Beenken, A., Eliseenkova, A. V., Ibrahimi, O. A., Olsen, S. K., & Mohammadi, M. (2012). Plasticity in Interactions of Fibroblast Growth Factor 1 (FGF1) N Terminus with FGF Receptors Underlies Promiscuity of FGF1. *Journal of Biological Chemistry*, *287*(5), 3067–3078. https://doi.org/10.1074/jbc.M111.275891

Bertrand, S., Aldea, D., Oulion, S., Subirana, L., de Lera, A. R., Somorjai, I., & Escriva, H. (2015). Evolution of the Role of RA and FGF Signals in the Control of Somitogenesis in Chordates. *PLOS ONE*, *10*(9), e0136587. https://doi.org/10.1371/journal.pone.0136587

Bertrand, S., Camasses, A., Somorjai, I., Belgacem, M. R., Chabrol, O., Escande, M. L., Pontarotti, P., & Escriva, H. (2011). Amphioxus FGF signaling predicts the acquisition

of vertebrate morphological traits. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9160–9165. https://doi.org/10.1073/pnas.1014235108

Bertrand, S., Campo-Paysaa, F., Camasses, A., Garcia-Fernandez, J., Escriva, H., García-Fernàndez, J., & Escrivà, H. (2009). Actors of the tyrosine kinase receptor downstream signaling pathways in amphioxus. *Evolution and Development*, *11*(1), 13–26. https://doi.org/10.1111/j.1525-142X.2008.00299.x

Bertrand, S., Iwema, T., & Escriva, H. (2014). FGF signaling emerged concomitantly with the origin of eumetazoans. *Molecular Biology and Evolution*, *31*(2), 310–318. https://doi.org/10.1093/molbev/mst222

Bertrand, V., Hudson, C., Caillol, D., Popovici, C., & Lemaire, P. (2003). Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. *Cell*, *115*(5), 615–627. https://doi.org/10.1016/S0092-8674(03)00928-0

Bliznina, A., Masunaga, A., Mansfield, M. J., Tan, Y., Liu, A. W., West, C., Rustagi, T., Chien, H. C., Kumar, S., Pichon, J., Plessy, C., & Luscombe, N. M. (2021). Telomere-to-telomere assembly of the genome of an individual Oikopleura dioica from Okinawa using Nanopore-based sequencing. *BMC Genomics*, *22*(1). https://doi.org/10.1186/s12864-021-07512-6

Böttcher, R. T., & Niehrs, C. (2005). Fibroblast Growth Factor Signaling during Early Vertebrate Development. *Endocrine Reviews*, *26*(1), 63–77. https://doi.org/10.1210/er.2003-0040

Bouwmeester, T. (2001). The Spemann-Mangold organizer: the control of fate specification and morphogenetic rearrangements during gastrulation in Xenopus. *The International Journal of Developmental Biology*, *45*(1), 251–258.

Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., Louis, A., Naville, M., Nitta, K. R., Piette, J., Reeves, W., Scornavacca, C., Simion, P., Vincentelli, R., Bellec, M., Aicha, S. Ben, Fagotto, M., Guéroult-Bellone, M., Haeussler, M., … Lemaire, P. (2018). ANISEED 2017: Extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. *Nucleic Acids Research*, *46*(D1), D718–D725. https://doi.org/10.1093/nar/gkx1108

Brunet, A., Bonni, A., Zigmond, M. J., Lin, M. Z., Juo, P., Hu, L. S., Anderson, M. J., Arden, K. C., Blenis, J., & Greenberg, M. E. (1999). Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell*, *96*(6), 857–868. https://doi.org/10.1016/s0092-8674(00)80595-4

Bryant, D. M., & Stow, J. L. (2005). Nuclear Translocation of Cell-Surface Receptors: Lessons from Fibroblast Growth Factor. *Traffic*, *6*(10), 947–953. https://doi.org/10.1111/j.1600-0854.2005.00332.x

Bülow, H. E., Boulin, T., & Hobert, O. (2004). Differential Functions of the C. elegans FGF Receptor in Axon Outgrowth and Maintenance of Axon Position. *Neuron*, *42*(3), 367–374. https://doi.org/10.1016/S0896-6273(04)00246-6

Campsteijn, C., Øvrebø, J. I., Karlsen, B. O., & Thompson, E. M. (2012). Expansion of cyclin D and CDK1 paralogs in Oikopleura dioica, a chordate employing diverse cell cycle variants. *Molecular Biology and Evolution*, *29*(2), 487–502. https://doi.org/10.1093/molbev/msr136

Cañestro, C., Bassham, S., & Postlethwait, J. H. (2003). Seeing chordate evolution through the Ciona genome sequence. *Genome Biology*, *4*(3), 208. https://doi.org/10.1186/gb-2003-4-3-208

Cañestro, C., Bassham, S., & Postlethwait, J. H. (2005). Development of the central nervous system in the larvacean Oikopleura dioica and the evolution of the chordate brain. *Developmental Biology*, *285*(2), 298–315. https://doi.org/10.1016/j.ydbio.2005.06.039

Cañestro, C., Catchen, J. M., Rodríguez-Marí, A., Yokoi, H., & Postlethwait, J. H. H. J. H. (2009). Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genetics*, *5*(5), e1000496. https://doi.org/10.1371/journal.pgen.1000496

Cañestro, C., & Postlethwait, J. H. (2007). Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev Biol*, *305*(2), 522–538. https://doi.org/10.1016/j.ydbio.2007.02.032

Cañestro, C., Postlethwait, J. H., Gonzàlez-Duarte, R., & Albalat, R. (2006). Is retinoic acid genetic machinery a chordate innovation? *Evolution & Development*, *8*(5), 394–406. https://doi.org/10.1111/j.1525-142X.2006.00113.x

Cañestro, C., Yokoi, H., Postlethwait, J. H. J., Canestro, C., Yokoi, H., Postlethwait, J. H. J., Cañestro, C., Yokoi, H., & Postlethwait, J. H. J. (2007). Evolutionary developmental biology and genomics. *Nature Reviews. Genetics*, *8*(12), 932–942. https://doi.org/10.1038/nrg2226

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Cebrià, F., Kobayashi, C., Umesono, Y., Nakazawa, M., Mineta, K., Ikeo, K., Gojobori, T., Itoh, M., Taira, M., Alvarado, A. S., & Agata, K. (2002). FGFR-related gene nou-darake restricts brain tissues to the head region of planarians. *Nature*, *419*(6907), 620–624. https://doi.org/10.1038/nature01042

Chen, H., Xu, C.-F., Ma, J., Eliseenkova, A. V, Li, W., Pollock, P. M., Pitteloud, N., Miller, W. T., Neubert, T. A., & Mohammadi, M. (2008). A crystallographic snapshot of tyrosine trans-phosphorylation in action. *Proceedings of the National Academy of Sciences*

of the United States of America, *105*(50), 19660–19665. https://doi.org/10.1073/pnas.0807752105

Chow, S. Y., Yu, C. Y., & Guy, G. R. (2009). Sprouty2 interacts with protein Kinase Cδ and disrupts phosphorylation of protein kinase D1. *Journal of Biological Chemistry*, *284*(29), 19623–19636. https://doi.org/10.1074/jbc.M109.021600

Clarke, T., Bouquet, J.-M., Fu, X., Kallesøe, T., Schmid, M., & Thompson, E. M. (2007). Rapidly evolving lamins in a chordate, Oikopleura dioica, with unusual nuclear architecture. *Gene*, *396*(1), 159–169. https://doi.org/10.1016/j.gene.2007.03.006

Cohen, S., Levi-Montalcini, R., & Hamburger, V. (1954). *A NERVE GROWTH-STIMULATING FACTOR ISOLATED FROM SARCOMAS 37 AND 180**.

Combarnous, Y., & Nguyen, T. M. D. (2020). Cell communications among microorganisms, plants, and animals: Origin, evolution, and interplays. In *International Journal of Molecular Sciences* (Vol. 21, Issue 21, pp. 1–22). MDPI AG. https://doi.org/10.3390/ijms21218052

Coulier, F., Pontarotti, P., Roubin, R., Hartung, H., Goldfarb, M., & Birnbaum, D. (1997). Of worms and men: an evolutionary perspective on the fibroblast growth factor (FGF) and FGF receptor families. *Journal of Molecular Evolution*, *44*(1), 43–56. https://doi.org/10.1007/pl00006120

Crump, J. G., Maves, L., Lawson, N. D., Weinstein, B. M., & Kimmel, C. B. (2004). An essential role for Fgfs in endodermal pouch formation influences later craniofacial skeletal patterning. *Development*, *131*(22), 5703–5716. https://doi.org/10.1242/dev.01444

Cunningham, T. J., & Duester, G. (2015). Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nature Reviews. Molecular Cell Biology*, *16*(2), 110–123. https://doi.org/10.1038/nrm3932

Cunningham, T. J., Zhao, X., Sandell, L. L., Evans, S. M., Trainor, P. A., & Duester, G. (2013). Antagonism between Retinoic Acid and Fibroblast Growth Factor Signaling during Limb Development. *Cell Reports*, *3*(5), 1503–1511. https://doi.org/10.1016/j.celrep.2013.03.036

D'Aniello, S., Irimia, M., Maeso, I., Pascual-Anaya, J., Jiménez-Delgado, S., Bertrand, S., & Garcia-Fernàndez, J. (2008). Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. *Molecular Biology and Evolution*, *25*(9), 1841–1854. https://doi.org/10.1093/molbev/msn132

Danks, G., Campsteijn, C., Parida, M., Butcher, S., Doddapaneni, H., Fu, B., Petrin, R., Metpally, R., Lenhard, B., Wincker, P., Chourrout, D., Thompson, E. M., & Manak, J. R. (2013). OikoBase: A genomics and developmental transcriptomics resource for the urochordate Oikopleura dioica. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1159

Datta, S. R., Dudek, H., Tao, X., Masters, S., Fu, H., Gotoh, Y., & Greenberg, M. E. (1997). Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. *Cell*, *91*(2), 231–241. https://doi.org/10.1016/s0092-8674(00)80405-5

Davidson, B. (2007). Ciona intestinalis as a model for cardiac development. *Semin Cell Dev Biol*, *18*(1), 16–26. https://doi.org/10.1016/j.semcdb.2006.12.007

Davidson, B., Shi, W., Beh, J., Christiaen, L., & Levine, M. (2006). FGF signaling delineates the cardiac progenitor field in the simple chordate, Ciona intestinalis. *Genes Dev*, *20*(19), 2728–2738. https://doi.org/20/19/2728 [pii] 10.1101/gad.1467706

De Meyts, P. (2015). Receptor Tyrosine Kinase Signal Transduction and the Molecular Basis of Signalling Specificity. In D. L. Wheeler (Ed.), *Receptor Tyrosine Kinases: Structure, Functions and Role in Human Disease* (pp. 51–76). Springer New York. https://doi.org/10.1007/978-1-4939-2053-2_4

de Miguel, C., Cruz, J., Martín, D., & Franch-Marro, X. (2020). Dual role of FGF in proliferation and endoreplication of Drosophila tracheal adult progenitor cells. *Journal of Molecular Cell Biology*, *12*(1), 32–41. https://doi.org/10.1093/jmcb/mjz055

Decker, B., Liput, M., Abdellatif, H., Yergeau, D., Bae, Y., Jornet, J. M., Stachowiak, E. K., & Stachowiak, M. K. (2020). Global Genome Conformational Programming during Neuronal Development Is Associated with CTCF and Nuclear FGFR1—The Genome Archipelago Model. *International Journal of Molecular Sciences*, *22*(1), 347. https://doi.org/10.3390/ijms22010347

Dehal, P., & Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol*, *3*(10), e314. https://doi.org/10.1371/journal.pbio.0030314

Delsuc, F., Brinkmann, H., Chourrout, D., & Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*, *439*(7079), 965–968. https://doi.org/10.1038/nature04336

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J.-M. M., Da Silva, C., Brinkmann, H., Mikhaleva, J., Olsen, L. C., Jubin, C., Cañestro, C., Bouquet, J.-M. M., Danks, G., Poulain, J., Campsteijn, C., Adamski, M., Cross, I., Yadetie, F., Muffato, M., Louis, A., … Chourrout, D. (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science*, *330*(6009), 1381–1385. https://doi.org/10.1126/science.1194167

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316–319. https://doi.org/10.1038/nbt.3820

Diez del Corral, R., & Morales, A. V. (2017). The Multiple Roles of FGF Signaling in the Developing Spinal Cord. *Frontiers in Cell and Developmental Biology*, *5*. https://doi.org/10.3389/fcell.2017.00058

Diez del Corral, R., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., & Storey, K. (2003). Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension. *Neuron*, *40*(1), 65–79. https://doi.org/S0896627303005658 [pii]

Diez del Corral, R., & Storey, K. G. (2004). Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. *Bioessays*, *26*(8), 857–869. https://doi.org/10.1002/bies.20080

Dorey, K., & Amaya, E. (2010). FGF signalling: diverse roles during early vertebrate embryogenesis. *Development*, *137*(22), 3731–3742. https://doi.org/137/22/3731 [pii] 10.1242/dev.037689

Du, Y., Hsu, J. L., Wang, Y.-N., & Hung, M.-C. (2015). Nuclear Functions of Receptor Tyrosine Kinases. In D. L. Wheeler & Y. Yarden (Eds.), *Receptor Tyrosine Kinases: Structure, Functions and Role in Human Disease* (pp. 77–109). Springer New York. https://doi.org/10.1007/978-1-4939-2053-2_5

Edvardsen, R. B., Seo, H. C., Jensen, M. F., Mialon, A., Mikhaleva, J., Bjordal, M., Cartry, J., Reinhardt, R., Weissenbach, J., Wincker, P., & Chourrout, D. (2005). Remodelling of the homeobox gene complement in the tunicate Oikopleura dioica. *Curr Biol*, *15*(1), R12-3. https://doi.org/10.1016/j.cub.2004.12.010

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*(3), 276–278. https://doi.org/10.1038/s41587-020-0439-x

Fan, T.-P., Ting, H.-C., Yu, J.-K., & Su, Y.-H. (2018). Reiterative use of FGF signaling in mesoderm development during embryogenesis and metamorphosis in the hemichordate Ptychodera flava. *BMC Evolutionary Biology*, *18*(1), 120. https://doi.org/10.1186/s12862-018-1235-9

Fang, X., Stachowiak, E. K., Dunham-Ems, S. M., Klejbor, I., & Stachowiak, M. K. (2005). Control of CREB-binding protein signaling by nuclear fibroblast growth factor receptor-1: a novel mechanism of gene regulation. *The Journal of Biological Chemistry*, *280*(31), 28451–28462. https://doi.org/10.1074/jbc.M504400200

Fernández, R., & Gabaldón, T. (2020). Gene gain and loss across the metazoan tree of life. *Nature Ecology & Evolution*, *4*(4), 524–533. https://doi.org/10.1038/s41559-019-1069-x

Ferrández-Roldán, A. (2021). *Deconstruction of the cardiopharyngeal gene regulatory network in appendicularians, a paradigmatic study of Oikopleura dioica as an evolutionary knockout model*. University of Barcelona.

Ferrández-Roldán, A., Fabregà-Torrus, M., Sánchez-Serna, G., Duran-Bello, E., Joaquín-Lluís, M., Bujosa, P., Plana-Carmona, M., Garcia-Fernàndez, J., Albalat, R., &

Cañestro, C. (2021). Cardiopharyngeal deconstruction and ancestral tunicate sessility. *Nature*, *599*(7885), 431–435. https://doi.org/10.1038/s41586-021-04041-w

Ferrández-Roldán, A., Martí-Solans, J., Cañestro, C., & Albalat, R. (2019). Oikopleura dioica: An Emergent Chordate Model to Study the Impact of Gene Loss on the Evolution of the Mechanisms of Development. In W. Tworzydlo & S. Bilinski (Eds.), *Evo-Devo: Non-model Species in Cell and Developmental Biology.* (Vol. 68, pp. 63–105). Springer, Cham. https://doi.org/10.1007/978-3-030-23459-1_4

Fujii, S., Nishio, T., & Nishida, H. (2008). Cleavage pattern, gastrulation, and neurulation in the appendicularian, Oikopleura dioica. *Dev Genes Evol*, *218*(2), 69–79. https://doi.org/10.1007/s00427-008-0205-4

Furdui, C. M., Lew, E. D., Schlessinger, J., & Anderson, K. S. (2006). Autophosphorylation of FGFR1 Kinase Is Mediated by a Sequential and Precisely Ordered Reaction. *Molecular Cell*, *21*(5), 711–717. https://doi.org/10.1016/j.molcel.2006.01.022

García-Alonso, L., Romani, S., & Jiménez, F. (2000). The EGF and FGF Receptors Mediate Neuroglian Function to Control Growth Cone Decisions during Sensory Axon Guidance in Drosophila. *Neuron*, *28*(3), 741–752. https://doi.org/10.1016/S0896-6273(00)00150-1

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press. https://doi.org/10.1385/1-59259-890-0:571

Gerhart, J. (1999). 1998 Warkany Lecture: Signaling pathways in development. In *Teratology* (Vol. 60, Issue 4, pp. 226–239). https://doi.org/10.1002/(SICI)1096-9926(199910)60:4<226::AID-TERA7>3.0.CO;2-W

Gerhart, J., & Kirschner, M. (1997). *Cells, Embryos, and Evolution*. Blackwell Science, Inc.

Gil-Gálvez, A., Jimenez-Gancedo, S., Perez-Posada, A., Franke, M., Acemel, R. D., Lin, C.-Y., Chou, C., Su, Y.-H., Yu, J.-K., Bertrand, S., Schubert, M., Escrivá, H., Tena, J. J., & Gómez-Skarmeta, J. L. (2022). *Gain of gene regulatory network interconnectivity at the origin of vertebrates*. https://doi.org/10.1073/pnas

Goetz, R., & Mohammadi, M. (2013). Exploring mechanisms of FGF signalling through the lens of structural biology. *Nature Reviews Molecular Cell Biology*, *14*(3), 166–180. https://doi.org/10.1038/nrm3528

Goldfarb, M. (2001). Signaling By Fibroblast Growth Factors: The Inside Story. *Science's STKE*, *2001*(106). https://doi.org/10.1126/stke.2001.106.pe37

Goldfarb, M. (2005). Fibroblast growth factor homologous factors: Evolution, structure, and function. *Cytokine & Growth Factor Reviews*, *16*(2), 215–220. https://doi.org/10.1016/j.cytogfr.2005.02.002

Green, P. J., Walsh, F. S., & Doherty, P. (1996). Promiscuity of fibroblast growth factor receptors. *BioEssays*, *18*(8), 639–646. https://doi.org/10.1002/bies.950180807

Gudernova, I., Vesela, I., Balek, L., Buchtova, M., Dosedelova, H., Kunova, M., Pivnicka, J., Jelinkova, I., Roubalova, L., Kozubik, A., & Krejci, P. (2016). Multikinase activity of fibroblast growth factor receptor (FGFR) inhibitors SU5402, PD173074, AZD1480, AZD4547 and BGJ398 compromises the use of small chemicals targeting FGFR catalytic activity for therapy of short-stature syndromes. *Human Molecular Genetics*, *25*(1), 9–23. https://doi.org/10.1093/hmg/ddv441

Guijarro-Clarke, C., Holland, P. W. H., & Paps, J. (2020). Widespread patterns of gene loss in the evolution of the animal kingdom. *Nature Ecology & Evolution*, *4*(4), 519–523. https://doi.org/10.1038/s41559-020-1129-2

Hébert, J. M. (2011). FGFs: neurodevelopment's Jack-of-all-trades – how do they do it? *Frontiers in Neuroscience*, *5*. https://doi.org/10.3389/fnins.2011.00133

Heenan, P., Zondag, L., & Wilson, M. J. (2016). Evolution of the Sox gene family within the chordate phylum. *Gene*, *575*(2). https://doi.org/10.1016/j.gene.2015.09.013

Holland, L. Z. (2016). Tunicates. *Current Biology*, *26*(4), R146–R152. https://doi.org/https://doi.org/10.1016/j.cub.2015.12.024.

Horie, R., Hazbun, A., Chen, K., Cao, C., Levine, M., & Horie, T. (2018). Shared evolutionary origin of vertebrate neural crest and cranial placodes. *Nature*, *560*(7717), 228–232. https://doi.org/10.1038/s41586-018-0385-7

Hosp, J., Sagane, Y., Danks, G., & Thompson, E. M. (2012). The evolving proteome of a complex extracellular matrix, the Oikopleura house. *PloS One*, *7*(7), e40172. https://doi.org/10.1371/journal.pone.0040172

Huang, B., Li, X., Tu, X., Zhao, W., Zhu, D., Feng, Y., Si, X., & Chen, J.-G. (2018). OTX1 regulates cell cycle progression of neural progenitors in the developing cerebral cortex. *Journal of Biological Chemistry*, *293*(6), 2137–2148. https://doi.org/10.1074/jbc.RA117.001249

Hudson, C., Sirour, C., & Yasuo, H. (2016). Co-expression of Foxa.a, Foxd and Fgf9/16/20 defines a transient mesendoderm regulatory state in ascidian embryos. *ELife*, *5*. https://doi.org/10.7554/eLife.14692

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Imai, K. S., Hino, K., Yagi, K., Satoh, N., & Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a

comprehensive understanding of gene networks. *Development*, *131*(16), 4047–4058. https://doi.org/10.1242/dev.01270

Imai, K. S., Satoh, N., & Satou, Y. (2002a). Early embryonic expression of FGF4/6/9 gene and its role in the induction of mesenchyme and notochord in Ciona savignyi embryos. *Development (Cambridge, England)*, *129*(7), 1729–1738. https://doi.org/10.1242/dev.129.7.1729

Imai, K. S., Satoh, N., & Satou, Y. (2002b). Region specific gene expressions in the central nervous system of the ascidian embryo. *Mech Dev*, *119 Suppl*, S275-7. https://doi.org/10.1016/S0925-4773(03)00128-X

Imai, K. S., Stolfi, A., Levine, M., & Satou, Y. (2009). Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. *Development*, *136*(2), 285–293. https://doi.org/dev.026419 [pii] 10.1242/dev.026419

Itoh, N., & Ornitz, D. M. (2004). Evolution of the Fgf and Fgfr gene families. In *Trends in Genetics* (Vol. 20, Issue 11, pp. 563–569). https://doi.org/10.1016/j.tig.2004.08.007

Itoh, N., & Ornitz, D. M. (2011). Fibroblast growth factors: from molecular evolution to roles in development, metabolism and disease. *Journal of Biochemistry*, *149*(2), 121–130. https://doi.org/10.1093/jb/mvq121

Jacob, F. (1977). Evolution and tinkering. *Science*, *196*(4295), 1161–1166. https://doi.org/10.1126/science.860134

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Käll, L., Krogh, A., & Sonnhammer, E. L. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, *338*(5), 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016

Kamei, S., Yajima, I., Yamamoto, H., Kobayashi, A., Makabe, K. W., Yamazaki, H., Hayashi, S. I., & Kunisada, T. (2000). Characterization of a novel member of the FGFR family, HrFGFR, in Halocynthia roretzi. *Biochemical and Biophysical Research Communications*, *275*(2), 503–508. https://doi.org/10.1006/bbrc.2000.3334

Kawazoe, T., & Taniguchi, K. (2019). The Sprouty/Spred family as tumor suppressors: Coming of age. In *Cancer Science* (Vol. 110, Issue 5, pp. 1525–1535). Blackwell Publishing Ltd. https://doi.org/10.1111/cas.13999

Kazanietz, M. G., & Cooke, M. (2024). Protein kinase C signaling “in” and “to” the nucleus: Master kinases in transcriptional regulation. *Journal of Biological Chemistry*, *300*(3), 105692. https://doi.org/10.1016/j.jbc.2024.105692

Keduka, E., Kaiho, A., Hamada, M., Watanabe-Takano, H., Takano, K., Ogasawara, M., Satou, Y., Satoh, N., & Endo, T. (2009). M-Ras evolved independently of R-Ras and its neural function is conserved between mammalian and ascidian, which lacks classical Ras. *Gene*, *429*(1–2), 49–58. https://doi.org/10.1016/j.gene.2008.10.001

Khamsi, R. (2004). Chickens join the genome club. *Nature*. https://doi.org/10.1038/news041206-8

Kim, H. J., & Bar-Sagi, D. (2004). Modulation of signalling by sprouty: A developing story. In *Nature Reviews Molecular Cell Biology* (Vol. 5, Issue 6, pp. 441–450). https://doi.org/10.1038/nrm1400

Kim, K., Gibboney, S., Razy-Krajka, F., Lowe, E. K., Wang, W., & Stolfi, A. (2020). Regulation of Neurogenesis by FGF Signaling and Neurogenin in the Invertebrate Chordate Ciona. *Frontiers in Cell and Developmental Biology*, *8*, 477. https://doi.org/10.3389/fcell.2020.00477

Kirov, A., Al-Hashimi, H., Solomon, P., Mazur, C., Thorpe, P. E., Sims, P. J., Tarantini, F., Kumar, T. K. S., & Prudovsky, I. (2012). Phosphatidylserine externalization and membrane blebbing are involved in the nonclassical export of FGF1. *Journal of Cellular Biochemistry*, *113*(3), 956–966. https://doi.org/10.1002/jcb.23425

Kishi, K., Hayashi, M., Onuma, T. A., & Nishida, H. (2017). Patterning and morphogenesis of the intricate but stereotyped oikoplastic epidermis of the appendicularian, Oikopleura dioica. *Developmental Biology*, *428*(1). https://doi.org/10.1016/j.ydbio.2017.06.008

Kourakis, M. J., & Smith, W. C. (2007). A conserved role for FGF signaling in chordate otic/atrial placode formation. *Dev Biol*, *312*(1), 245–257. https://doi.org/S0012-1606(07)01376-0 [pii] 10.1016/j.ydbio.2007.09.020

Krause, M., Dent, E. W., Bear, J. E., Loureiro, J. J., & Gertler, F. B. (2003). Ena/VASP Proteins: Regulators of the Actin Cytoskeleton and Cell Migration. *Annual Review of Cell and Developmental Biology*, *19*, 541–564. https://doi.org/10.1146/annurev.cellbio.19.050103.103356

Kugler, J. E., Kerner, P., Bouquet, J. M., Jiang, D., & Di Gregorio, A. (2011). Evolutionary changes in the notochord genetic toolkit: a comparative analysis of notochord genes in the ascidian Ciona and the larvacean Oikopleura. *BMC Evol Biol*, *11*(1), 21. https://doi.org/1471-2148-11-21 [pii] 10.1186/1471-2148-11-21

Kumano, G., & Nishida, H. (2007). Ascidian embryonic development: an emerging model system for the study of cell fate specification in chordates. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *236*(7), 1732–1747. https://doi.org/10.1002/dvdy.21108

Kumar, V., Goutam, R. S., Park, S., Lee, U., & Kim, J. (2021). Functional Roles of FGF Signaling in Early Development of Vertebrate Embryos. *Cells*, *10*(8), 2148. https://doi.org/10.3390/cells10082148

Laezza, F., Lampert, A., Kozel, M. A., Gerber, B. R., Rush, A. M., Nerbonne, J. M., Waxman, S. G., Dib-Hajj, S. D., & Ornitz, D. M. (2009). FGF14 N-terminal splice variants differentially modulate Nav1.2 and Nav1.6-encoded sodium channels. *Molecular and Cellular Neurosciences*, *42*(2), 90–101. https://doi.org/10.1016/j.mcn.2009.05.007

Lemaire, P. (2011). Evolutionary crossroads in developmental biology: The tunicates. *Development*, *138*(11), 2143–2152. https://doi.org/10.1242/dev.048975

Lemaire, P., & Piette, J. (2015). Tunicates: Exploring the sea shores and roaming the open ocean. A tribute to Thomas Huxley. In *Open Biology* (Vol. 5, Issue 6). Royal Society of London. https://doi.org/10.1098/rsob.150053

Lemmon, M. A., & Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. *Cell*, *141*(7), 1117–1134. https://doi.org/10.1016/j.cell.2010.06.011

Leptin, M., & Affolter, M. (2004). Drosophila Gastrulation: Identification of a Missing Link. *Current Biology*, *14*(12), R480–R482. https://doi.org/10.1016/j.cub.2004.06.016

Lin, H., Lu, P., Zhou, M., Wu, F., Weng, L., Meng, K., Yang, D., Li, S., Jiang, C., & Tian, H. (2019). Purification of recombinant human fibroblast growth factor 13 in E. coli and its molecular mechanism of mitogenesis. *Applied Microbiology and Biotechnology*, *103*(17), 7017–7027. https://doi.org/10.1007/s00253-019-09973-y

Lo, T.-W., Branda, C. S., Huang, P., Sasson, I. E., Goodman, S. J., & Stern, M. J. (2008). Different isoforms of the C. elegans FGF receptor are required for attraction and repulsion of the migrating sex myoblasts. *Developmental Biology*, *318*(2), 268–275. https://doi.org/10.1016/j.ydbio.2008.03.026

Lu, J., Wu, T., Zhang, B., Liu, S., Song, W., Qiao, J., & Ruan, H. (2021). Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Communication and Signaling*, *19*(1), 60. https://doi.org/10.1186/s12964-021-00741-y

Luo, L. Y., & Hahn, W. C. (2015). Oncogenic Signaling Adaptor Proteins. In *Journal of Genetics and Genomics* (Vol. 42, Issue 10, pp. 521–529). Institute of Genetics and Developmental Biology. https://doi.org/10.1016/j.jgg.2015.09.001

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. N., Lopez, R., & Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkae241

Maier, E. C., & Whitfield, T. T. (2014). RA and FGF Signalling Are Required in the Zebrafish Otic Vesicle to Pattern and Maintain Ventral Otic Identities. *PLoS Genetics*, *10*(12), e1004858. https://doi.org/10.1371/journal.pgen.1004858

Manning, B. D., & Toker, A. (2017). AKT/PKB Signaling: Navigating the Network. *Cell*, *169*(3), 381–405. https://doi.org/10.1016/j.cell.2017.04.001

Martínez-Morales, P. L., Diez del Corral, R., Olivera-Martínez, I., Quiroga, A. C., Das, R. M., Barbas, J. A., Storey, K. G., & Morales, A. V. (2011). FGF and retinoic acid activity gradients control the timing of neural crest cell emigration in the trunk. *Journal of Cell Biology*, *194*(3), 489–503. https://doi.org/10.1083/jcb.201011077

Martí-Solans, J., Belyaeva, O. V. O. V., Torres-Aguila, N. P. N. P., Kedishvili, N. Y. N. Y., Albalat, R., & Cañestro, C. (2016). Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a Chordate. *Molecular Biology and Evolution*, *33*(9), 2401–2416. https://doi.org/10.1093/molbev/msw118

Martí-Solans, J., Ferrández-Roldán, A., Godoy-Marín, H., Badia-Ramentol, J., Torres-Aguila, N. P., Rodríguez-Marí, A., Bouquet, J. M., Chourrout, D., Thompson, E. M., Albalat, R., & Cañestro, C. (2015). Oikopleura dioica culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis (New York, N.Y. : 2000)*, *53*(1), 183–193. https://doi.org/10.1002/dvg.22800

Martí-Solans, J., Godoy-Marín, H., Diaz-Gracia, M., Onuma, T. A., Nishida, H., Albalat, R., & Cañestro, C. (2021). Massive Gene Loss and Function Shuffling in Appendicularians Stretch the Boundaries of Chordate Wnt Family Evolution. *Frontiers in Cell and Developmental Biology*, *9*, 700827. https://doi.org/10.3389/fcell.2021.700827

Masunaga, A., Mansfield, M. J., Tan, Y., Liu, A. W., Bliznina, A., Barzaghi, P., Hodgetts, T. L., Ferrández-Roldán, A., Cañestro, C., Onuma, T. A., Plessy, C., & Luscombe, N. M. (2022). The cosmopolitan appendicularian Oikopleura dioica reveals hidden genetic diversity around the globe. *Marine Biology*, *169*(12). https://doi.org/10.1007/s00227-022-04145-5

Matus, D. Q., Thomsen, G. H., & Martindale, M. Q. (2007). FGF signaling in gastrulation and neural development in Nematostella vectensis, an anthozoan cnidarian. *Development Genes and Evolution*, *217*(2), 137–148. https://doi.org/10.1007/s00427-006-0122-3

McClintock, J. M., Carlson, R., Mann, D. M., & Prince, V. E. (2001). Consequences of Hox gene duplication in the vertebrates: an investigation of the zebrafish Hox paralogue group 1 genes. *Development*, *128*(13), 2471–2484.

Meister, L., Escriva, H., & Bertrand, S. (2022). Functions of the FGF signalling pathway in cephalochordates provide insight into the evolution of the prechordal plate. *Development*, *149*(10). https://doi.org/10.1242/dev.200252

Mercader, N., Leonardo, E., Piedra, M. E., Martínez-A, C., Ángeles Ros, M., & Torres, M. (2000). Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development*, *127*(18), 3961–3970. https://doi.org/10.1242/dev.127.18.3961

Mikhaleva, Y., Skinnes, R., Sumic, S., Thompson, E. M., & Chourrout, D. (2018). Development of the house secreting epithelium, a major innovation of tunicate larvaceans, involves multiple homeodomain transcription factors. *Developmental Biology*, *443*(2), 117–126. https://doi.org/10.1016/j.ydbio.2018.09.006

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/nar/gkaa913

Mistry, N., Harrington, W., Lasda, E., Wagner, E. J., & García-Blanco, M. A. (2003). Of urchins and men: Evolution of an alternative splicing unit in fibroblast growth factor receptor genes. *RNA*, *9*(2), 209–217. https://doi.org/10.1261/rna.2470903

Miyakawa, K., Hatsuzawa, K., Kurokawa, T., Asada, M., Kuroiwa, T., & Imamura, T. (1999). A hydrophobic region locating at the center of fibroblast growth factor-9 is crucial for its secretion. *The Journal of Biological Chemistry*, *274*(41), 29352–29357. https://doi.org/10.1074/jbc.274.41.29352

Miyakawa, K., & Imamura, T. (2003). Secretion of FGF-16 requires an uncleaved bipartite signal sequence. *Journal of Biological Chemistry*, *278*(37), 35718–35724. https://doi.org/10.1074/jbc.M300690200

Miyazaki, Y., Nishida, H., & Kumano, G. (2007). *Brain induction in ascidian embryos is dependent on juxtaposition of FGF9 / 16 / 20-producing and -receiving cells*. https://doi.org/10.1007/s00427-006-0129-9

Mohammadi, M., McMahon, G., Sun, L., Tang, C., Hirth, P., Yeh, B. K., Hubbard, S. R., & Schlessinger, J. (1997). Structures of the Tyrosine Kinase Domain of Fibroblast Growth Factor Receptor in Complex with Inhibitors. *Science*, *276*(5314), 955–960. https://doi.org/DOI: 10.1126/science.276.5314.955

Mok, G. F., Cardenas, R., Anderton, H., Campbell, K. H. S., & Sweetman, D. (2014). Interactions between FGF18 and retinoic acid regulate differentiation of chick embryo limb myoblasts. *Developmental Biology*, *396*(2), 214–223. https://doi.org/10.1016/j.ydbio.2014.10.004

Nagatomo, K., & Fujiwara, S. (2003). Expression of Raldh2, Cyp26 and Hox-1 in normal and retinoic acid-treated Ciona intestinalis embryos. *Gene Expr Patterns*, *3*(3), 273–277. https://doi.org/10.1016/S1567-133X(03)00051-6

Naville, M., Henriet, S., Warren, I., Sumic, S., Reeve, M., Volff, J. N., & Chourrout, D. (2019). Massive Changes of Genome Size Driven by Expansions of Non-autonomous Transposable Elements. *Current Biology*, *29*(7), 1161–1168. https://doi.org/10.1016/j.cub.2019.01.080

Neben, C. L., Lo, M., Jura, N., & Klein, O. D. (2019). Feedback regulation of RTK signaling in development. In *Developmental Biology* (Vol. 447, Issue 1, pp. 71–89). Elsevier Inc. https://doi.org/10.1016/j.ydbio.2017.10.017

Newton, A. C. (2010). Protein kinase C: poised to signal. *American Journal of Physiology-Endocrinology and Metabolism*, *298*(3), E395–E402. https://doi.org/10.1152/ajpendo.00477.2009

Nguyen Ba, A. N., Pogoutse, A., Provart, N., & Moses, A. M. (2009). NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*, *10*, 202. https://doi.org/10.1186/1471-2105-10-202

Nishida, H. (1987). Cell Lineage Analysis in Ascidian Embryos by Intracellular Injection of a Tracer Enzyme III. Up to the Tissue Restricted Stage. In *DEVELOPMENTAL BIOLOGY* (Vol. 121).

Nishida, H. (2008). Development of the appendicularian Oikopleura dioica: culture, genome, and cell lineages. *Dev Growth Differ*, *50 Suppl 1*(SUPPL. 1), S239-56. https://doi.org/10.1111/j.1440-169X.2008.01035.x

Nutt, S. L., Dingwell, K. S., Holt, C. E., & Amaya, E. (2001). *Xenopus Sprouty2 inhibits FGF-mediated gastrulation movements but does not affect mesoderm induction and patterning*. https://doi.org/10.1101/gad

Ohno, S. (1970). *Evolution by gene duplication*. Springer-Verlag.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733-45. https://doi.org/10.1093/nar/gkv1189

Olivera-Martinez, I., Harada, H., Halley, P. A., & Storey, K. G. (2012). Loss of FGF-Dependent Mesoderm Identity and Rise of Endogenous Retinoid Signalling Determine Cessation of Body Axis Elongation. *PLoS Biology*, *10*(10), e1001415. https://doi.org/10.1371/journal.pbio.1001415

Olivera-Martinez, I., & Storey, K. G. (2007). Wnt signals provide a timing mechanism for the FGF-retinoid differentiation switch during vertebrate body axis extension. *Development*, *134*(11), 2125–2135. https://doi.org/10.1242/dev.000216

Olsen, L. C., Kourtesis, I., Busengdal, H., Jensen, M. F., Hausen, H., & Chourrout, D. (2018). Evidence for a centrosome-attracting body like structure in germ-soma segregation

during early development, in the urochordate Oikopleura dioica. *BMC Developmental Biology*, *18*(1), 4. https://doi.org/10.1186/s12861-018-0165-5

Olsen, S. K., Garbi, M., Zampieri, N., Eliseenkova, A. V., Ornitz, D. M., Goldfarb, M., & Mohammadi, M. (2003). Fibroblast Growth Factor (FGF) Homologous Factors Share Structural but Not Functional Homology with FGFs. *Journal of Biological Chemistry*, *278*(36), 34226–34236. https://doi.org/10.1074/jbc.M303183200

Olsnes, S., Klingenberg, O., & Wie¸dłocha, A. (2003). Transport of Exogenous Growth Factors and Cytokines to the Cytosol and to the Nucleus. *Physiological Reviews*, *83*(1), 163–182. https://doi.org/10.1152/physrev.00021.2002

Onuma, T. A., Hayashi, M., Gyoja, F., Kishi, K., Wang, K., & Nishida, H. (2020). A chordate species lacking Nodal utilizes calcium oscillation and Bmp for left–right patterning. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(8), 4188–4198. https://doi.org/10.1073/pnas.1916858117

Ornitz, D. M. (2000). FGFs, heparan sulfate and FGFRs: complex interactions essential for development. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *22*(2), 108–112. https://doi.org/10.1002/(SICI)1521-1878(200002)22:2<108::AID-BIES2>3.0.CO;2-M

Ornitz, D. M., & Itoh, N. (2015). The fibroblast growth factor signaling pathway. *Wiley Interdisciplinary Reviews: Developmental Biology*, *4*(3), 215–266. https://doi.org/10.1002/wdev.176

Ornitz, D. M., & Itoh, N. (2022). New developments in the biology of fibroblast growth factors. *WIREs Mechanisms of Disease*, *14*(4), e1549. https://doi.org/10.1002/wsbm.1549

Oulion, S., Bertrand, S., & Escriva, H. (2012). Evolution of the FGF Gene Family. *International Journal of Evolutionary Biology*, *2012*, 1–12. https://doi.org/10.1155/2012/298147

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., & Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. *European Journal of Cell Biology*, *97*(6), 422–441. https://doi.org/10.1016/j.ejcb.2018.06.003

Pablo, J. L., & Pitt, G. S. (2016). Fibroblast Growth Factor Homologous Factors: New Roles in Neuronal Health and Disease. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *22*(1), 19–25. https://doi.org/10.1177/1073858414562217

Pallarès-Albanell, J., Ortega-Flores, L., Senar-Serra, T., Ruiz, A., Abril, J. F., Rossello, M. & Almudi, I. Gene regulatory dynamics during the development of a paleopteran insect, the mayfly *Cloeon dipterum*. bioRxiv 2024.05.14.594094. https://doi.org/10.1101/2024.05.14.594094

Papp, B., Notebaart, R. A., Pal, C., & Pál, C. (2011). Systems-biology approaches for predicting genomic evolution. *Nat Rev Genet*, *12*(9), 591–602. https://doi.org/nrg3033 [pii] 10.1038/nrg3033

Paschaki, M., Schneider, C., Rhinn, M., Thibault-Carpentier, C., Dembélé, D., Niederreither, K., & Dollé, P. (2013). Transcriptomic analysis of murine embryos lacking endogenous retinoic acid signaling. *PloS One*, *8*(4), e62274. https://doi.org/10.1371/journal.pone.0062274

Pasini, A., Manenti, R., Rothbacher, U., & Lemaire, P. (2012). Antagonizing retinoic acid and FGF/MAPK pathways control posterior body patterning in the invertebrate chordate Ciona intestinalis. *PLoS One*, *7*(9), e46193. https://doi.org/10.1371/journal.pone.0046193 PONE-D-12-17849 [pii]

Pennati, A., Jakobi, M., Zeng, F., Ciampa, L., & Rothbächer, U. (2024). Optimizing CRISPR/Cas9 approaches in the polymorphic tunicate Ciona intestinalis. *Developmental Biology*, *510*, 31–39. https://doi.org/10.1016/j.ydbio.2024.03.003

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650–1667. https://doi.org/10.1038/nprot.2016.095

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science : A Publication of the Protein Society*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Pires-daSilva, A., & Sommer, R. J. (2003). The evolution of signalling pathways in animal development. In *Nature Reviews Genetics* (Vol. 4, Issue 1, pp. 39–49). https://doi.org/10.1038/nrg977

Plessy, C., Mansfield, M. J., Bliznina, A., Masunaga, A., West, C., Tan, Y., Liu, A. W., Grašič, J., del Río Pisula, M. S., Sánchez-Serna, G., Fabrega-Torrus, M., Ferrández-Roldán, A., Roncalli, V., Navratilova, P., Thompson, E. M., Onuma, T., Nishida, H., Cañestro, C., & Luscombe, N. M. (2024). Extreme genome scrambling in marine planktonic *Oikopleura dioica* cryptic species. *Genome Research*. https://doi.org/10.1101/gr.278295.123

Plotnikov, A. N., Eliseenkova, A. V., Ibrahimi, O. A., Shriver, Z., Sasisekharan, R., Lemmon, M. A., & Mohammadi, M. (2001). Crystal Structure of Fibroblast Growth Factor 9 Reveals Regions Implicated in Dimerization and Autoinhibition. *Journal of Biological Chemistry*, *276*(6), 4322–4329. https://doi.org/10.1074/jbc.M006502200

Pollak, D. D., Minh, B. Q., Cicvaric, A., & Monje, F. J. (2014). A novel Fibroblast Growth Factor Receptor family member promotes neuronal outgrowth and synaptic plasticity in Aplysia. *Amino Acids*, *46*(11), 2477–2488. https://doi.org/10.1007/s00726-014-1803-2

Popovici, C., Conchonaud, F., Birnbaum, D., & Roubin, R. (2004). Functional Phylogeny Relates LET-756 to Fibroblast Growth Factor 9. *Journal of Biological Chemistry*, *279*(38), 40146–40152. https://doi.org/10.1074/jbc.M405795200

Popovici, C., Fallet, M., Marguet, D., Birnbaum, D., & Roubin, R. (2006). Intracellular trafficking of LET-756, a fibroblast growth factor of C. elegans, is controlled by a balance of export and nuclear signals. *Experimental Cell Research*, *312*(9), 1484–1495. https://doi.org/10.1016/j.yexcr.2006.01.012

Popovici, C., Roubin, R., Coulier, F., & Birnbaum, D. (2005). An evolutionary history of the FGF superfamily. *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, *27*(8), 849–857. https://doi.org/10.1002/bies.20261

Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J. K., Benito-Gutierrez, E. L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J. J., Grigoriev, I. V, Horton, A. C., de Jong, P. J., Jurka, J., Kapitonov, V. V, … Rokhsar, D. S. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature*, *453*(7198), 1064–1071. https://doi.org/10.1038/nature06967

Putnam, N. H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V. V, Jurka, J., Genikhovich, G., Grigoriev, I. V, Lucas, S. M., Steele, R. E., Finnerty, J. R., Technau, U., Martindale, M. Q., & Rokhsar, D. S. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*, *317*(5834), 86–94. https://doi.org/10.1126/science.1139158

Razy-Krajka, F., Gravez, B., Kaplan, N., Racioppi, C., Wang, W., & Christiaen, L. (2018). An FGF-driven feed-forward circuit patterns the cardiopharyngeal mesoderm in space and time. *ELife*, *7:e29656.* https://doi.org/10.7554/eLife.29656.001

Rebscher, N., Deichmann, C., Sudhop, S., Fritzenwanker, J. H., Green, S., & Hassel, M. (2009). Conserved intron positions in FGFR genes reflect the modular structure of FGFR and reveal stepwise addition of domains to an already complex ancestral FGFR. *Development Genes and Evolution*, *219*(9–10), 455–468. https://doi.org/10.1007/s00427-009-0309-5

Rees, J. M., Palmer, M. A., & Gillis, J. A. (2024). Fgf signalling is required for gill slit formation in the skate, Leucoraja erinacea. *Developmental Biology*, *506*, 85–94. https://doi.org/10.1016/j.ydbio.2023.11.008

Reilly, J. F., & Maher, P. A. (2001). Importin beta-mediated nuclear import of fibroblast growth factor receptor: role in cell proliferation. *The Journal of Cell Biology*, *152*(6), 1307–1312. https://doi.org/10.1083/jcb.152.6.1307

Rentzsch, F., Fritzenwanker, J. H., Scholz, C. B., & Technau, U. (2008). FGF signalling controls formation of the apical sensory organ in the cnidarian Nematostella vectensis. *Development*, *135*(10), 1761–1769. https://doi.org/10.1242/dev.020784

Revest, J. M., DeMoerlooze, L., & Dickson, C. (2000). Fibroblast growth factor 9 secretion is mediated by a non-cleaved amino- terminal signal sequence. *Journal of Biological Chemistry*, *275*(11), 8083–8090. https://doi.org/10.1074/jbc.275.11.8083

Ribes, V., Le Roux, I., Rhinn, M., Schuhbaur, B., & Dollé, P. (2009). Early mouse caudal development relies on crosstalk between retinoic acid,Shh and Fgf signalling pathways. *Development*, *136*(4), 665–676. https://doi.org/10.1242/dev.016204

Roskoski, R. (2012). ERK1/2 MAP kinases: Structure, function, and regulation. *Pharmacological Research*, *66*(2), 105–143. https://doi.org/10.1016/j.phrs.2012.04.005

Röttinger, E., Saudemont, A., Duboc, V., Besnardeau, L., McClay, D., & Lepage, T. (2008). FGF signals guide migration of mesenchymal cells, control skeletal morphogenesis [corrected] and regulate gastrulation during sea urchin development. *Development (Cambridge, England)*, *135*(2), 353–365. https://doi.org/10.1242/dev.014282

Sarabipour, S., & Hristova, K. (2016). Mechanism of FGF receptor dimerization and activation. *Nature Communications*, *7*(1), 10262. https://doi.org/10.1038/ncomms10262

Satoh, N. (2003). The ascidian tadpole larva: comparative molecular development and genomics. *Nat Rev Genet*, *4*(4), 285–295. https://doi.org/10.1038/nrg1042

Satou, Y. (2020). A gene regulatory network for cell fate specification in Ciona embryos. In *Current Topics in Developmental Biology* (Vol. 139, pp. 1–33). Academic Press Inc. https://doi.org/10.1016/bs.ctdb.2020.01.001

Satou, Y., Imai, K. S., & Satoh, N. (2002). Fgf genes in the basal chordate Ciona intestinalis. *Development Genes and Evolution*, *212*(9), 432–438. https://doi.org/10.1007/s00427-002-0266-8

Satou, Y., Sasakura, Y., Yamada, L., Imai, K. S., Satoh, N., & Degnan, B. (2003). A genomewide survey of developmentally relevant genes in Ciona intestinalis. *Development Genes and Evolution*, *213*(5–6), 254–263. https://doi.org/10.1007/s00427-003-0317-9

Schäfer, T., Zentgraf, H., Zehe, C., Brügger, B., Bernhagen, J., & Nickel, W. (2004). Unconventional Secretion of Fibroblast Growth Factor 2 Is Mediated by Direct Translocation across the Plasma Membrane of Mammalian Cells. *Journal of Biological Chemistry*, *279*(8), 6244–6251. https://doi.org/10.1074/jbc.M310500200

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. *Cell*, *103*(2), 211–225. https://doi.org/10.1016/s0092-8674(00)00114-8

Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., Linhardt, R. J., & Mohammadi, M. (2000). Crystal Structure of a Ternary FGF-FGFR-Heparin Complex Reveals a Dual Role for Heparin in FGFR Binding and Dimerization. *Molecular Cell*, *6*(3), 743–750. https://doi.org/10.1016/S1097-2765(00)00073-3

Schmahl, J., Kim, Y., Colvin, J. S., Ornitz, D. M., & Capel, B. (2004). Fgf9 induces proliferation and nuclear localization of FGFR2 in Sertoli precursors during male sex determination. *Development (Cambridge, England)*, *131*(15), 3627–3636. https://doi.org/10.1242/dev.01239

Schoorlemmer, J., & Goldfarb, M. (2002). Fibroblast Growth Factor Homologous Factors and the Islet Brain-2 Scaffold Protein Regulate Activation of a Stress-activated Protein Kinase. *Journal of Biological Chemistry*, *277*(51), 49111–49119. https://doi.org/10.1074/jbc.M205520200

Seo, H. C., Kube, M., Edvardsen, R. B., Jensen, M. F., Beck, A., Spriet, E., Gorsky, G., Thompson, E. M., Lehrach, H., Reinhardt, R., & Chourrout, D. (2001). Miniature genome in the marine chordate Oikopleura dioica. *Science*, *294*(5551), 2506. https://doi.org/10.1126/science.294.5551.2506

Seo, H. C., Edvardsen, R. B., Maeland, A. D., Bjordal, M., Jensen, M. F., Hansen, A., Flaat, M., Weissenbach, J., Lehrach, H., Wincker, P., Reinhardt, R., & Chourrout, D. (2004). Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica. *Nature*, *431*(7004), 67–71. https://doi.org/10.1038/nature02709

Sheng, Z., Liang, Y., Lin, C.-Y., Comai, L., & Chirico, W. J. (2005). Direct Regulation of rRNA Transcription by Fibroblast Growth Factor 2. *Molecular and Cellular Biology*, *25*(21), 9419–9426. https://doi.org/10.1128/MCB.25.21.9419-9426.2005

Shi, W., Peyrot, S. M., Munro, E., & Levine, M. (2009). FGF3 in the floor plate directs notochord convergent extension in the Ciona tadpole. *Development (Cambridge, England)*, *136*(1), 23–28. https://doi.org/10.1242/dev.029157

Shimauchi, Y., Murakami, S. D., & Satoh, N. (2001). FGF signals are involved in the differentiation of notochord cells and mesenchyme cells of the ascidian Halocynthia roretzi. *Development*, *128*(14), 2711–2721. https://doi.org/https://doi.org/10.1242/dev.128.14.2711

Shiotsugu, J., Katsuyama, Y., Arima, K., Baxter, A., Koide, T., Song, J., Chandraratna, R. A. S., & Blumberg, B. (2004). Multiple points of interaction between retinoic acid and FGF signaling during embryonic axis formation. *Development*, *131*(11), 2653–2667. https://doi.org/10.1242/dev.01129

Simeone, A. (1998). Otx1 and Otx2 in the development and evolution of the mammalian brain. *The EMBO Journal*, *17*(23), 6790–6798. https://doi.org/10.1093/emboj/17.23.6790

Sirbu, I. O., Zhao, X., & Duester, G. (2008). Retinoic acid controls heart anteroposterior patterning by down-regulating Isl1 through the Fgf8 pathway. *Developmental Dynamics : An Official Publication of the American Association of Anatomists*, *237*(6), 1627–1635. https://doi.org/10.1002/dvdy.21570

Sivak, J. M., Petersen, L. F., & Amaya, E. (2005). FGF signal interpretation is directed by sprouty and spred proteins during mesoderm formation. *Developmental Cell*, *8*(5), 689–701. https://doi.org/10.1016/j.devcel.2005.02.011

Sluzalska, K. D., Slawski, J., Sochacka, M., Lampart, A., Otlewski, J., & Zakrzewska, M. (2021). Intracellular partners of fibroblast growth factors 1 and 2 - implications for functions. *Cytokine & Growth Factor Reviews*, *57*, 93–111. https://doi.org/10.1016/j.cytogfr.2020.05.004

Smallwood, P. M., Munoz-Sanjuan, I., Tong, P., Macke, J. P., Hendry, S. H., Gilbert, D. J., Copeland, N. G., Jenkins, N. A., & Nathans, J. (1996). Fibroblast growth factor (FGF) homologous factors: new members of the FGF family implicated in nervous system development. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(18), 9850–9857. https://doi.org/10.1073/pnas.93.18.9850

Sochacka, M., Opalinski, L., Szymczyk, J., Zimoch, M. B., Czyrek, A., Krowarsch, D., Otlewski, J., & Zakrzewska, M. (2020). FHF1 is a bona fide fibroblast growth factor that activates cellular signaling in FGFR-dependent manner. *Cell Communication and Signaling*, *18*(1), 69. https://doi.org/10.1186/s12964-020-00573-2

Somorjai, I., Martí-Solans, J., Diaz-Gracia, M., Nishida, H., Imai, K. S., Escrivà, H., Cañestro, C., & Albalat, R. (2018). Wnt evolution and function shuffling in liberal and conservative chordate genomes. *Genome Biology*, *19*(1), 98. https://doi.org/10.1186/s13059-018-1468-3

Søviknes, A. M., & Glover, J. C. (2008). *Continued Growth and Cell Proliferation Into Adulthood in the Notochord of the Appendicularian Oikopleura dioica*.

Stach, T. (2007). Ontogeny of the appendicularian Oikopleura dioica (Tunicata, Chordata) reveals characters similar to ascidian larvae with sessile adults. *Zoomorphology*, *126*(3), 203–214. https://doi.org/10.1007/s00435-007-0041-5

Stach, T., & Turbeville, J. M. (2002). Phylogeny of Tunicata inferred from molecular and morphological characters. *Mol Phylogenet Evol*, *25*(3), 408–428. https://doi.org/10.1016/S1055-7903(02)00305-6

Stach, T., Winter, J., Bouquet, J.-M. M., Chourrout, D., & Schnabel, R. (2008). Embryology of a planktonic tunicate reveals traces of sessility. *Proc Natl Acad Sci U S A*, *105*(20), 7229–7234. https://doi.org/10.1073/pnas.0710196105

Stachowiak, M. K., Birkaya, B., Aletta, J. M., Narla, S. T., Benson, C. A., Decker, B., & Stachowiak, E. K. (2015). "Nuclear FGF receptor-1 and CREB binding protein: an integrative signaling module". *Journal of Cellular Physiology*, *230*(5), 989–1002. https://doi.org/10.1002/jcp.24879

Stathopoulos, A., Tam, B., Ronshaugen, M., Frasch, M., & Levine, M. (2004). pyramus and thisbe: FGF genes that pattern the mesoderm of Drosophila embryos. *Genes & Development*, *18*(6), 687–699. https://doi.org/10.1101/gad.1166404

Stolfi, A., & Christiaen, L. (2012). Genetic and genomic toolbox of the chordate Ciona intestinalis. *Genetics*, *192*(1), 55–66. https://doi.org/10.1534/genetics.112.140590

Stolfi, A., Ryan, K., Meinertzhagen, I. A., & Christiaen, L. (2015). Migratory neuronal progenitors arise from the neural plate borders in tunicates. *Nature*, *527*(7578), 371–374. https://doi.org/10.1038/nature15758

Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., & Schwede, T. (2020). QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics*, *36*(6), 1765–1771. https://doi.org/10.1093/bioinformatics/btz828

Technau, U. (2020). Gastrulation and germ layer formation in the sea anemone Nematostella vectensis and other cnidarians. *Mechanisms of Development*, *163*, 103628. https://doi.org/10.1016/j.mod.2020.103628

Technau, U., & Scholz, C. B. (2003). Origin and evolution of endoderm and mesoderm. *The International Journal of Developmental Biology*, *47*(7–8), 531–539.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, *40*(7), 1023–1025. https://doi.org/10.1038/s41587-021-01156-3

Teven, C. M., Farina, E. M., Rivas, J., & Reid, R. R. (2014). Fibroblast growth factor (FGF) signaling in development and skeletal diseases. *Genes and Diseases*, *1*(2), 199–213. https://doi.org/10.1016/j.gendis.2014.09.005

Thisse, B., & Thisse, C. (2005). Functions and regulations of fibroblast growth factor signaling during embryonic development. *Developmental Biology*, *287*(2), 390–402. https://doi.org/10.1016/j.ydbio.2005.09.011

Tolkin, T., & Christiaen, L. (2012). Development and Evolution of the Ascidian Cardiogenic Mesoderm. In *Current Topics in Developmental Biology* (Vol. 100, pp. 107–142). Academic Press Inc. https://doi.org/10.1016/B978-0-12-387786-4.00011-7

Torres-Águila, N. P., Martí-Solans, J., Ferrández-Roldán, A., Almazán, A., Roncalli, V., D'Aniello, S., Romano, G., Palumbo, A., Albalat, R., & Cañestro, C. (2018). Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate Oikopleura dioica. *Communications Biology*, *1*, 121. https://doi.org/10.1038/s42003-018-0127-2

Treen, N., Chavarria, E., Weaver, C. J., Brangwynne, C. P., & Levine, M. (2023). An FGF timer for zygotic genome activation. *Genes and Development*, *37*(3–4), 80–85. https://doi.org/10.1101/gad.350164.122

Treen, N., Yoshida, K., Sakuma, T., Sasaki, H., Kawai, N., & Yamamoto, T. (2014). *Tissue-specific and ubiquitous gene knockouts by TALEN electroporation provide new approaches to investigating gene function in Ciona*. *December 2013*, 481–487. https://doi.org/10.1242/dev.099572

Troedsson, C., Bouquet, J.-M., Skinnes, R., Acuna, J.-L., Zech, K., Frischer, M. E., & Thompson, E. M. (2009). Regulation of filter-feeding house components in response to varying food regimes in the appendicularian, Oikopleura dioica. *Journal of Plankton Research*, *31*(12), 1453–1463. https://doi.org/10.1093/plankt/fbp085

Tulin, S., & Stathopoulos, A. (2010). Extending the family table: Insights from beyond vertebrates into the regulation of embryonic development by FGFs. *Birth Defects Research Part C: Embryo Today: Reviews*, *90*(3), 214–227. https://doi.org/10.1002/bdrc.20182

Turner, N., & Grose, R. (2010). Fibroblast growth factor signalling: from development to cancer. *Nature Reviews Cancer*, *10*(2), 116–129. https://doi.org/10.1038/nrc2780

Veeman, M. T., Newman-Smith, E., El-Nachef, D., & Smith, W. C. (2010). The ascidian mouth opening is derived from the anterior neuropore: reassessing the mouth/neural tube relationship in chordate evolution. *Developmental Biology*, *344*(1), 138–149. https://doi.org/10.1016/j.ydbio.2010.04.028

Villanueva, S., Glavic, A., Ruiz, P., & Mayor, R. (2002). Posteriorization by FGF, Wnt, and Retinoic Acid Is Required for Neural Crest Induction. *Developmental Biology*, *241*(2), 289–301. https://doi.org/10.1006/dbio.2001.0485

Wagner, E., & Levine, M. (2012). FGF signaling establishes the anterior border of the Ciona neural tube. *Development*, *139*(13), 2351–2359. https://doi.org/dev.078485 [pii] 10.1242/dev.078485

Wang, C., Wang, C., Hoch, E. G., & Pitt, G. S. (2011). Identification of novel interaction sites that determine specificity between fibroblast growth factor homologous factors and voltage-gated sodium channels. *The Journal of Biological Chemistry*, *286*(27), 24253–24263. https://doi.org/10.1074/jbc.M111.245803

Wang, K., Dantec, C., Lemaire, P., Onuma, T. A., & Nishida, H. (2017). Genome-wide survey of miRNAs and their evolutionary history in the ascidian, Halocynthia roretzi. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-3707-5

Wang, K., Omotezako, T., Kishi, K., Nishida, H., & Onuma, T. A. (2015). Maternal and zygotic transcriptomes in the appendicularian, Oikopleura dioica: novel protein-encoding genes, intra-species sequence variations, and trans-spliced RNA leader. *Development Genes and Evolution*, *225*(3), 149–159. https://doi.org/10.1007/s00427-015-0502-7

Wang, K., Tomura, R., Chen, W., Kiyooka, M., Ishizaki, H., Aizu, T., Minakuchi, Y., Seki, M., Suzuki, Y., Omotezako, T., Suyama, R., Masunaga, A., Plessy, C., Luscombe, N. M., Dantec, C., Lemaire, P., Itoh, T., Toyoda, A., Nishida, H., & Onuma, T. A. (2020). A genome database for a Japanese population of the larvacean *Oikopleura dioica*. *Development, Growth & Differentiation*, *62*(6), 450–461. https://doi.org/10.1111/dgd.12689

Ward, N., & Moreno-Hagelsieb, G. (2014). Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PloS One*, *9*(7), e101850. https://doi.org/10.1371/journal.pone.0101850

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Waterhouse, A. M., Studer, G., Robin, X., Bienert, S., Tauriello, G., & Schwede, T. (2024). The structure assessment web server: for proteins, complexes and more. *Nucleic Acids Research*, *52*(W1), W318–W323. https://doi.org/10.1093/nar/gkae270

Wennerberg, K., Rossman, K. L., & Der, C. J. (2005). The Ras superfamily at a glance. *Journal of Cell Science*, *118*(5), 843–846. https://doi.org/10.1242/jcs.01660

Wilson, V., Olivera-Martinez, I., & Storey, K. G. (2009). Stem cells, signals and vertebrate body axis extension. *Development*, *136*(10), 1591–1604. https://doi.org/10.1242/dev.021246

Wipf, P., & Halter, R. J. (2005). Chemistry and biology of wortmannin. *Organic and Biomolecular Chemistry*, *3*(11), 2053–2061. https://doi.org/10.1039/b504418a

Wu, Q.-F., Yang, L., Li, S., Wang, Q., Yuan, X.-B., Gao, X., Bao, L., & Zhang, X. (2012). Fibroblast Growth Factor 13 Is a Microtubule-Stabilizing Protein Regulating Neuronal Polarization and Migration. *Cell*, *149*(7), 1549–1564. https://doi.org/10.1016/j.cell.2012.04.046

Xu, R., Ori, A., Rudd, T. R., Uniewicz, K. A., Ahmed, Y. A., Guimond, S. E., Skidmore, M. A., Siligardi, G., Yates, E. A., & Fernig, D. G. (2012). Diversification of the structural determinants of fibroblast growth factor-heparin interactions: implications for binding specificity. *The Journal of Biological Chemistry*, *287*(47), 40061–40073. https://doi.org/10.1074/jbc.M112.398826

Yasuo, H., & Hudson, C. (2007). FGF8/17/18 functions together with FGF9/16/20 during formation of the notochord in Ciona embryos. *Developmental Biology*, *302*(1), 92–103. https://doi.org/10.1016/j.ydbio.2006.08.075

Žárský, V., & Tachezy, J. (2015). Evolutionary loss of peroxisomes - not limited to parasites. *Biology Direct*, *10*(1). https://doi.org/10.1186/s13062-015-0101-6

Zhang, X., Bao, L., Yang, L., Wu, Q., & Li, S. (2012). Roles of intracellular fibroblast growth factors in neural development and functions. *Science China. Life Sciences*, *55*(12), 1038–1044. https://doi.org/10.1007/s11427-012-4412-x

Zhao, X., & Duester, G. (2009). Effect of retinoic acid signaling on Wnt/β-catenin and FGF signaling during body axis extension. *Gene Expression Patterns*, *9*(6), 430–435. https://doi.org/10.1016/j.gep.2009.06.003

Zhao, X., Sirbu, I. O., Mic, F. A., Molotkova, N., Molotkov, A., Kumar, S., & Duester, G. (2009). Retinoic Acid Promotes Limb Induction through Effects on Body Axis Extension but Is Unnecessary for Limb Patterning. *Current Biology*, *19*(12), 1050–1057. https://doi.org/10.1016/j.cub.2009.04.059

Annex 1

1. **Supplementary Tables**
2. **Supplementary Figures**
3. **Supplementary Material**

1. Supplementary Tables

Supplementary Table 1. Genes cloned in this project. Primers used, length of the insert, DNA used as a template, and genome used for the design of the primers.

Gene	Forward primer	Reverse primer	Length	Template	Genome used for design
Fgf11/12/13/14a	5'CAATGAGCAAACGACGAAAGAGTTTTG	5'TTAAACAGCTTGTTCTCCAGTTACTTC	599 bp	cDNA BAR	NOR
Fgf11/12/13/14b	5'ATGGTTTCGCGCTTATTCTCATGTGC	5'TCAGACATTCTTTGAACCATTGAACTC	615 bp	cDNA BAR	NOR
Fgf11/12/13/14c	5'GATGGCTAGTTCCGAATCTGGAG	5'GTCGGTATTTTTAGCATTTGAGTTG	357 bp	cDNA BAR	NOR
Fgf11/12/13/14d	5'GAGCTACAGTGCCGAAGTGG	5'CCAGATTCCGGATTCTTGAATTCTAT	320 bp	cDNA BAR	NOR
Fgf9/16/20a	5'GCAAGACAGGATTCCACCTAG	5'GCGGCAGGAACTGCGACGAC	357 bp	cDNA BAR	NOR
Fgf9/16/20b	5'GTCCAGCTTTACGCTCGAAC	5'GACGCAGAGAAGCTTGTAGAG	406 bp	cDNA BAR	NOR
Fgf9/16/20c	5'GAACTATCAAAAATGCGCAGACA	5'GAACACCGTGGCTTAAAAGTCG	420 bp	cDNA BAR	NOR
Fgf9/16/20d	5'GTTGTCGTTGCACCTGTTGA	5'GTCTGCCGAAGCGTTTTTTC	454 bp	gDNA BAR	BAR
Fgf9/16/20e	5'GCCAACCCTTATCCAGCGAA	5'CCTCGTCGACATCGGACATT	423 bp	gDNA BAR	BAR
Fgf9/16/20f	5'GAGCAATCGTAGCCCCAGAA	5'GGCGGGAATATGAGCCTTTT	428 bp	gDNA BAR	BAR
FgfRa	5'GCTCGCTTTCTGCCATCTTC	5'GAACCCACTGGCAATGAATTTTC	581 bp	cDNA BAR	NOR
FgfRb	5'GGTGATCTGCTCAAGTTTCTGC	5'GTTATCTCGTGACCATGAACATCAA	617 bp	cDNA BAR	NOR
FgfRc	5'GTTGACATAATTGATGAGGGTTTCT	5'GAGTTCAAGTGTGGTAGAATCTC	674 bp	cDNA BAR	NOR
Erk1/2	5'GAAGGAGCCTACGGCATAG	5'GCTAGAATACATCCGACAGAC	563 bp	cDNA BAR	NOR
Mek1/2	5'GGGAAGCTCCACGAGTCTAT	5'GTCGGCGAGAATGTCGATGA	423 bp	cDNA BAR	BAR
PLCy	5'GACGACAATGACGAGCACGA	5'GCCCATCTCTGTCACTTCTGA	709 bp	cDNA BAR	BAR
PI3KCA	5'CTATACTGGCACCTCCGGC	5'TGACATGCCGTCGAGAAACTC	784 bp	cDNA BAR	BAR
PDK	5'ATGACCAGGCTAGTAGCTGG	5'GAGCGTTTTGCCCTCTTCG	786 bp	cDNA BAR	BAR
AKT	5'ACACGAAACGACGGCTTG	5'GTCTGTCAAGGGCGTTGAAATC	747 bp	cDNA BAR	BAR
Sprouty	5'GAGAAGAAGTCTACCCGACTG	5'CACTTGCAGTTGTGTGGACAC	681 bp	gDNA BAR	BAR

Supplementary Table 2. Reciprocal Best Blast Hit (RBBH) approach using *Ciona robusta Fgf* genes as original queries. For each query *Fgf*, the top 5 tblastn hits on the *O. dioica* Bar2_p4 genome were considered, regardless of their e-values or bit score. Subsequently, *O. dioica* sequences were used as queries against the *C. robusta* gene models on the ANISEED database. Hits that returned a *C.robusta Fgf* gene were considered for further validation. Hits that returned a *C.robusta Fgf* gene as RBBH but did not pass validation are shaded in yellow. These were usually short sequences, with stop codons in all reading frames, that resulted in no significant hits or hits in non-Fgf proteins when launched with blastp against the nr NCBI database.

Query: Cro_FGF7/10/22 (NP_001027763.1 / KH.S406.19.v1.A.SL1-1)

Subject (HIT)	Score	E-Value	CroRBBH	RBBH e-value
Chr2:9021974..9022282	51.2 bits (121)	5,00E-07	KH2012:KH.L28.8.v1.A.nonSL2-1	6,00E-26
Chr2:9747408..9747506	35.4 bits (80)	0.069	KH2012:KH.L28.8.v1.A.nonSL2-1	2,00E-07
PAR:6453273..6452996	28.9 bits (63)	0.70	KH2012:KH.L28.8.v1.A.ND1-1	3,00E-11
XSR:5323838..5324017*	30.8 bits (68)	2.0	KH2012:KH.S406.19.v1.A.SL1-1	2.5
Chr2:1581664..1581512	30.4 bits (67)	2.5	KH2012:KH.C10.496.v1.A.SL1-1	7,00E-08

Query: Cro_FGF8/17/18 (NP_001027648.1 / KH.C5.5.v2.A.SL3-1)

Subject (HIT)	Score	E-Value	CroRBBH	RBBH e-value
Chr2:9021977..9022246	39.3 bits (90)	0.009	KH2012:KH.L28.8.v1.A.nonSL2-1	8,00E-22
Chr2:10861742..10861560	32.3 bits (72)	1.2	KH2012:KH.C2.119.v1.A.ND1-1	14
PAR:7170104..7169982*	31.6 bits (70)	2.0	KH2012:KH.C5.5.v1.A.nonSL1-1	0.048
PAR:11285242..11285382	30.0 bits (66)	6.1	KH2012:KH.L28.8.v1.A.ND1-1	6,00E-11
Chr1:9816004..9815840*	29.6 bits (65)	8.0	KH2012:KH.C5.5.v2.A.SL3-1	0.42

Query: Cro_FGF11/12/13/14 (NP_001027733.1 / KH.L28.8.v1.A.nonSL2-1)

Subject (HIT)	Score	E-Value	CroRBBH	RBBH e-value
Chr2:9021767..9022393	142 bits (358)	5,00E-38	KH2012:KH.L28.8.v1.A.nonSL2-1	3,00E-53
PAR:11285101..11285548	92.0 bits (227)	7,00E-21	KH2012:KH.L28.8.v1.A.nonSL2-1	3,00E-30
Chr2:9746991..9747509	87.0 bits (214)	4,00E-19	KH2012:KH.L28.8.v1.A.nonSL2-1	4,00E-24
PAR:6453315..6453127	79.7 bits (195)	1,00E-16	KH2012:KH.L28.8.v1.A.nonSL2-1	3,00E-19
Chr2:4655988..4656318	33.5 bits (75)	0.004	KH2012:KH.C2.125.v2.A.ND1-1	1,00E-11

Query: Cro_FGFL (NP_001027650.1 / KH.C14.70.v1.A.SL1-1)

Subject (HIT)	Score	E-Value	CroRBBH	RBBH e-value
Chr1:7718066..7717929*	30.4 bits (67)	7.5	KH2012:KH.C14.70.v1.A.SL1-1	0.031
PAR:6453294..6453145	30.4 bits (67)	9.0	KH2012:KH.L28.8.v1.A.ND1-1	7,00E-13
PAR:2775938..2776108	29.3 bits (64)	17	KH2012:KH.S854.2.v2.A.nonSL4-1	0.50
Chr1:6401928..6401845*	28.5 bits (62)	27	KH2012:KH.C14.70.v1.A.SL2-1	0.10
Chr2:9021974..9022123	28.1 bits (61)	38	KH2012:KH.L28.8.v1.A.nonSL2-1	1,00E-14

Query: Cro_FGF4/5/6 (NP_001027747.1 / KH.S615.3.v1.A.ND1-1)

Subject (HIT)	Score	E-Value	CroRBBH	RBBH e-value
PAR:11285245..11285382	34.7 bits (78)	0.087	KH2012:KH.L28.8.v1.A.ND1-1	6,00E-11
PAR:4424163..4424041	31.6 bits (70)	0.78	KH2012:KH.C5.357.v1.A.ND1-1	4,00E-07
PAR:76070..75978*	29.3 bits (64)	4.7	KH2012:KH.S615.3.v2.A.ND2-1	0.010
PAR:10102401..10102514	28.5 bits (62)	7.0	KH2012:KH.C9.36.v2.A.SL2-1	3,00E-05
PAR:540826..540957*	28.5 bits (62)	7.8	KH2012:KH.S615.3.v1.A.ND1-1	1.8

Query: Cro_FGF9/16/20 (NP_001027649.1 / KH.C2.125.v1.R.nonSL3-1)

Subject (HIT)	Score	E-Value	CroRBBH	RBBH e-value
Chr2:4655803..4656303	60.5 bits (145)	8,00E-10	KH2012:KH.C2.125.v1.A.nonSL4-1	1,00E-12
Chr2:9021908..9022285	58.5 bits (140)	4,00E-09	KH2012:KH.L28.8.v1.A.nonSL2-1	1,00E-37
PAR:6453312..6453139	50.8 bits (120)	1,00E-06	KH2012:KH.L28.8.v1.A.nonSL2-1	2,00E-17
PAR:11285137..11285397	48.9 bits (115)	5,00E-06	KH2012:KH.L28.8.v1.A.ND1-1	2,00E-19
PAR:7534356..7534066	32.0 bits (71)	0.87	KH2012:KH.C11.116.v3.R.ND1-1	1,50E+01

Query: Cro_FGFNA1 (NP_001106724.1 / KH.C1.697.v1.A.SL1-1)

Subject (HIT)	Score	E-Value	CroRBBH	RBBH e-value
PAR:6408239..6408586*	33.5 bits (75)	0.54	KH2012:KH.C1.697.v1.A.SL1-1	5,00E-08
XSR:8548211..8548303*	31.6 bits (70)	2.1	KH2012:KH.C1.697.v1.A.SL1-1	0.008
XSR:10751936..10752019	29.6 bits (65)	9.4	-	>1000
XSR:2076351..2076491*	29.3 bits (64)	13	KH2012:KH.C1.697.v1.A.SL1-1	0.48
Chr2:12856579..12856665*	28.5 bits (62)	21	KH2012:KH.C1.697.v2.A.SL1-2	0.62

Supplementary Table 3. List of sequences used for Fgf phylogenetic analyses.

Sequence	Locus / Acc. Number	Source (Genome/Database)	Annotation
Odi_BAR_FGF11/12/13/14a	Chr2:9020505..9022448	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF11/12/13/14b	Chr2:9744211..9747677	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF11/12/13/14c	PAR:11284383..11285677	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF11/12/13/14d	PAR:6451928..6453947	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF9/16/20a	Chr2:4655663..4656440	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF9/16/20b	PAR:7533901..7534442	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF9/16/20c	PAR:1413001..1413553	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF9/16/20d	Chr2:5643393..5644002	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF9/16/20e	PAR:11389135..11389712	Bar2_p4 (Plessy et al. 2024)	This work
Odi_BAR_FGF9/16/20f	PAR:3507954..3508448	Bar2_p4 (Plessy et al. 2024)	This work
Odi_OSA_FGF11/12/13/14a	Chr2:6485097..6486582	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF11/12/13/14b	Chr2:9655601..9658256	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF11/12/13/14c	PAR:12145637..12146869	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF11/12/13/14d	PAR:8502402..8504255	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF9/16/20a	Chr2:4057249..4057933	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF9/16/20b	PAR:7751594..7752147	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF9/16/20c	PAR:1133127..1133695	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF9/16/20d	Chr2:8844967..8845581	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF9/16/20e	PAR:12053422..12053999	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OSA_FGF9/16/20f	PAR:2031273..2031859	OSKA2016v1.9 (Wang et al. 2020, Plessy et al. 2024)	This work
Odi_OKI_FGF11/12/13/14a	chr2:10022444..10024786	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF11/12/13/14b	chr2:9472351..9474932	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF11/12/13/14c	PAR:11065158..11066460	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF11/12/13/14d	PAR:12086239..12087762	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF9/16/20a	chr2:1954143..1954870	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF9/16/20b	PAR:8320805..8321373	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF9/16/20c	PAR:1213449..1213990	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF9/16/20d	chr2:6688144..6688740	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF9/16/20e	PAR:10502928..10503533	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Odi_OKI_FGF9/16/20f	PAR:2497535..2498083	OKI2018_I69_1.0 (Bliznina et al. 2021)	This work
Oal_FGF11/12/13/14a	SCLG01003638.1(2146..3772)	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF11/12/13/14b	SCLG01000154.1(26305..23610)	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF11/12/13/14c	SCLG01000635.1(58689..60938)*Cris	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF9/16/20a	SCLG01000066.1(479605..478374)	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF9/16/20b	SCLG01016498.1(710763..710290)	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF9/16/20d	SCLG01000041.1(758504..758923)- SCLG01000041.1a	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF9/16/20e	SCLG01000041.1(727238..726792)- SCLG01000041.1b	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF9/16/20f	SCLG01000451.1(53450..53004)	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF9/16/20g	SCLG01000774.1(2599..2153)	SCLG00000000 (Naville et al. 2019)	This work
Oal_FGF9/16/20i	SCLG01000111.1(226965..226519)	SCLG00000000 (Naville et al. 2019)	This work
Ova_FGF11/12/13/14a	SCLH01001993.1(34382..26125)	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF11/12/13/14b	SCLH01001253.1(213599..215262) *Cris	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF11/12/13/14c	SCLH01001575.1(41537..39332)*Cris	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF11/12/13/14d	SCLH01001575.1(59521..57456)*Cris	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF11/12/13/14e	SCLH01000807.1(252111..285090)*Cris	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF9/16/20a	SCLH01001123.1(82290.. 80973)	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF9/16/20b	SCLH01000034.1(264516..264974)	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF9/16/20c	SCLH01077972.1(355..11)+SCLH01051512.1(201..100)	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF9/16/20d	SCLH01000146.1(101156..121829)	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF9/16/20e	SCLH01000065.1(612752..612348)	SCLH00000000 (Naville et al. 2019)	This work
Ova_FGF9/16/20f	SCLH01013944.1(551..1018)	SCLH00000000 (Naville et al. 2019)	This work
Cro_FGF7/10/22	NP_001027763.1	NCBI	
Cro_FGF4/5/6	NP_001027747.1	NCBI	
Cro_FGF8/17/18	NP_001027648.1	NCBI	
Cro_FGF9/16/20	NP_001027649.1	NCBI	
Cro_FGF11/12/13/14	NP_001027733.1	NCBI	
Cro_FGFNA1	NP_001106724.1	NCBI	
Cro_FGFL	NP_001027650.1	NCBI	
Csa_FGF7/10/22	Cisavi.CG.ENS81.R13.2500097-2507381	ANISEED	
Csa_FGF8/17/18	Cisavi.CG.ENS81.R48.3103937-3116122	ANISEED	
Csa_FGF9/16/20	Cisavi.CG.ENS81.R17.698380-701770	ANISEED	
Csa_FGF11/12/13/14	Cisavi.CG.ENS81.R5.136106-138806	ANISEED	
Csa_FGFL	Cisavi.CG.ENS81.R26.343852-348323	ANISEED	
Csa_FGFNA1	R281(56947..49903)	ANISEED	This work
Pma_FGF7/10/22	CAB3245885.1	NCBI	
Pma_FGF4/5/6	S393(71640..65442)	ANISEED	This work
Pma_FGF8/17/18	CAB3245909.1	NCBI	
Pma_FGF9/16/20	Phmamm.CG.MTP2014.S128.g03805	ANISEED	
Pma_FGF11/12/13/14	Phmamm.CG.MTP2014.S509.g09581	ANISEED	
Pma_FGFNA1	CAB3245899.1	NCBI	

Supplementary Table 3. List of sequences used for Fgf phylogenetic analyses (continued).

Pma_FGFL	Phmamm.CG.MTP2014.S438.g08737	ANISEED	
Pfu_FGF7/10/22	Phfumi.CG.MTP2014.S3447.g05713+S4935(1530..1787)	ANISEED	This work
Pfu_FGF4/5/6	S727(13590..20032)	ANISEED	This work
Pfu_FGF8/17/18	S5210(1850..11653)+S18119(419..3158)+S13667(1..411)	ANISEED	This work
Pfu_FGF9/16/20	Phfumi.CG.MTP2014.S4550.g06431_edited	ANISEED	This work
Pfu_FGF11/12/13/14	Phfumi.CG.MTP2014.S3513.g05763+S11320(772..677)	ANISEED	This work
Pfu_FGFNA1	Phfumi.CG.MTP2014.S176.g01305(1-315aa)+S176(10288..9839)	ANISEED	This work
Pfu_FGFL	Phfumi.CG.MTP2014.S421.g02003	ANISEED	
Mocci_FGF7/10/22	Moocci.CG.ELv1_2.S238505.g05376	ANISEED	
Mocci_FGF4/5/6	S292116(5193..3385)+S360025(3816..3622)	ANISEED	This work
Mocci_FGF8/17/18	Moocci.CG.ELv1_2.S636971.g28201	ANISEED	
Mocci_FGF9/16/20	Moocci.CG.ELv1_2.S543820.g20921	ANISEED	
Mocci_FGF11/12/13/14	Moocci.CG.ELv1_2.S634979.g27905	ANISEED	
Mocci_FGFNA1-1	Moocci.CG.ELv1_2.S217951.g04788 +Moocci.CG.ELv1_2.S217951.g04785 +Moocci.CG.ELv1_2.S217951.g04787	ANISEED	This work
Mocci_FGFL	Moocci.CG.ELv1_2.S209253.g04531	ANISEED	
Moccu_FGF4/5/6	Mooccu.CG.ELv1_2.S253739.g09510	ANISEED	
Moccu_FGF7/10/22	Mooccu.CG.ELv1_2.S484508.g25097.01.p +Mooccu.CG.ELv1_2.S484508.g25098	ANISEED	This work
Moccu_FGF8/17/18	Mooccu.CG.ELv1_2.S686284.g43068	ANISEED	
Moccu_FGF9/16/20	Mooccu.CG.ELv1_2.S691170.g43610	ANISEED	
Moccu_FGF11/12/13/14	S705330(9793..8064)	ANISEED	This work
Moccu_FGFNA1	Mooccu.CG.ELv1_2.S651054.g39804 +Mooccu.CG.ELv1_2.S651054.g39805	ANISEED	This work
Mocul_FGF7/10/22	Moocul.CG.ELv1_2.S129652.g15182	ANISEED	
Mocul_FGF4/5/6	Moocul.CG.ELv1_2.S27543.g01266	ANISEED	
Mocul_FGF8/17/18	Moocul.CG.ELv1_2.S124626.g14593	ANISEED	
Mocul_FGF9/16/20	Moocul.CG.ELv1_2.S45502.g02447	ANISEED	
Mocul_FGF11/12/13/14	Moocul.CG.ELv1_2.S90886.g07225	ANISEED	
Mocul_FGFNA1	S89803(33483..37482)	ANISEED	This work
Mocul_FGFL	Moocul.CG.ELv1_2.S45171.g02411	ANISEED	
Bsc_FGF4/5/6	Boschl.CG.Botznik2013.chr9.g66473	ANISEED	
Bsc_FGF7/10/22a	Boschl.CG.Botznik2013.chr13.g54819	ANISEED	
Bsc_FGF7/10/22b	Boschl.CG.Botznik2013.chrUn.g61218	ANISEED	
Bsc_FGF9/16/20	chrUn(164340903..164349310)	ANISEED	This work
Bsc_FGF11/12/13/14	chr12(4003012..4002797) +chrUn(164004427..164004624) +chrUn(183004203..183004286)	ANISEED	This work
Bsc_FGFLa	Boschl.CG.Botznik2013.chrUn.g30482	ANISEED	
Bsc_FGFLb	Boschl.CG.Botznik2013.chrUn.g66060 +chrUN(1664611..1664186)	ANISEED	This work
Ble_FGF4/5/6	Boleac.CG.SB_v3.S479.g10408	ANISEED	
Ble_FGF9/16/20	Boleac.CG.SB_v3.S267.g06313	ANISEED	
Ble_FGF11/12/13/14	Boleac.CG.SB_v3.S75.g13823	ANISEED	
Ble_FGFL	Boleac.CG.SB_v3.S46.g10142	ANISEED	
Hau_FGF7/10/22a	Haaura.CG.MTP2014.S36.g00962	ANISEED	
Hau_FGF4/5/6	S2073(7085..3743)	ANISEED	This work
Hau_FGF8/17/18	Haaura.CG.MTP2014.S734.g06455+S3958(5147..6262)	ANISEED	This work
Hau_FGF9/16/20	Haaura.CG.MTP2014.S412.g04899	ANISEED	
Hau_FGF11/12/13/14	Haaura.CG.MTP2014.S1675.g08731	ANISEED	
Hau_FGFL	Haaura.CG.MTP2014.S28.g00796	ANISEED	
Hro_FGF7/10/22	Harore.CG.MTP2014.S111.g12376	ANISEED	
Hro_FGF4/5/6	S22(428024..431485)	ANISEED	This work
Hro_FGF8/17/18	Harore.CG.MTP2014.S348.g00092	ANISEED	
Hro_FGF9/16/20	Harore.CG.MTP2014.S81.g15688	ANISEED	
Hro_FGF11/12/13/14	Harore.CG.MTP2014.S56.g11365	ANISEED	
Hro_FGFL	Harore.CG.MTP2014.S65.g02939	ANISEED	
Bla_FGF1/2	ACF17006.1	NCBI	
Bla_FGF8/17/18	ACF17009.1	NCBI	
Bla_FGF9/16/20	ACF17010.1	NCBI	
Bla_FGFA	ACF17007.1	NCBI	
Bla_FGFB	ACF17008.1	NCBI	
Bla_FGFC	ACF17012.1	NCBI	
Bla_FGFD	ADU32860.1	NCBI	
Bla_FGFE	ACF17011.1	NCBI	
Bfl_FGF1/2	XP_035681331.1	NCBI	
Bfl_FGF8/17/18	XP_035674357.1	NCBI	
Bfl_FGF9/16/20	XP_035682407.1	NCBI	
Bfl_FGFA	XP_035666983.1	NCBI	
Bfl_FGFB	XP_035674550.1	NCBI	

Supplementary Table 3. List of sequences used for Fgf phylogenetic analyses (continued).

Bla_FGFC	ACF17012.1	NCBI	
Bla_FGFD	ADU32860.1	NCBI	
Bla_FGFE	ACF17011.1	NCBI	
Bfl_FGF1/2	XP_035681331.1	NCBI	
Bfl_FGF8/17/18	XP_035674357.1	NCBI	
Bfl_FGF9/16/20	XP_035682407.1	NCBI	
Bfl_FGFA	XP_035666983.1	NCBI	
Bfl_FGFB	XP_035674550.1	NCBI	
Bfl_FGFC	XP_035675192.1	NCBI	
Bfl_FGFD	XP_035675604.1	NCBI	
Bfl_FGFE	XP_035675347.1	NCBI	
Bbe_FGF1/2	A0A6P4Z6M6	UniProtKB	
Bbe_FGF8/17/18b	A0A6P4YWZ5	UniProtKB	
Bbe_FGF8/17/18a	A0A6P4YAI9	UniProtKB	
Bbe_FGF9/16/20	A0A6P5AZ54	UniProtKB	
Bbe_FGFA	A0A6P4ZXM1	UniProtKB	
Bbe_FGFB	A0A6P5AHS3	UniProtKB	
Bbe_FGFC	A0A6P4Z0T1	UniProtKB	
Bbe_FGFDa	NW_017804009.1[472696..480959]	NCBI	This work
Bbe_FGFDb	NW_017803872.1[143128..151359]	NCBI	This work
Bbe_FGFE	A0A6P4ZJL6	UniProtKB	
Dre_FGF1	NP_957054.1	NCBI	
Dre_FGF2	NP_997988.1	NCBI	
Dre_FGF3	NP_571366.1	NCBI	
Dre_FGF4	NP_571710.1	NCBI	
Dre_FGF5	NP_001009561.1	NCBI	
Dre_FGF6a	NP_001001398.2	NCBI	
Dre_FGF6b	NP_001009563.1	NCBI	
Dre_FGF7	NP_001007762.1	NCBI	
Dre_FGF8	NP_571356.2	NCBI	
Dre_FGF10a	NP_878290.1	NCBI	
Dre_FGF10b	NP_001039323.1	NCBI	
Dre_FGF11	NP_001012380.1	NCBI	
Dre_FGF12	NP_001071250.1	NCBI	
Dre_FGF13	NP_001007400.1	NCBI	
Dre_FGF14	NP_001012382.1	NCBI	
Dre_FGF16	NP_001035497.1	NCBI	
Dre_FGF17a	NP_878276.1	NCBI	
Dre_FGF17b	NP_999973.1	NCBI	
Dre_FGF18a	NP_001013282.1	NCBI	
Dre_FGF18b	NP_001012379.1	NCBI	
Dre_FGF19	NP_001012246.2	NCBI	
Dre_FGF20a	NP_001032180.1	NCBI	
Dre_FGF20b	NP_001034261.1	NCBI	
Dre_FGF21	NP_001038789.1	NCBI	
Dre_FGF22	NP_001035184.1	NCBI	
Dre_FGF23	NP_001009564.2	NCBI	
Dre_FGF24	NP_878291.2	NCBI	
Hsa_FGF1	NP_001341882.1	NCBI	
Hsa_FGF2	NP_001348594.1	NCBI	
Hsa_FGF3	NP_005238.1	NCBI	
Hsa_FGF4	NP_001998.1	NCBI	
Hsa_FGF5	NP_004455.2	NCBI	
Hsa_FGF6	NP_066276.2	NCBI	
Hsa_FGF7	NP_002000.1	NCBI	
Hsa_FGF8	NP_149354.1	NCBI	
Hsa_FGF9	NP_002001.1	NCBI	
Hsa_FGF10	NP_004456.1	NCBI	
Hsa_FGF11	NP_004103.1	NCBI	
Hsa_FGF12	NP_066360.1	NCBI	
Hsa_FGF13	NP_004105.1	NCBI	
Hsa_FGF14	NP_004106.1	NCBI	
Hsa_FGF16	NP_003859.1	NCBI	
Hsa_FGF17	NP_003858.1	NCBI	
Hsa_FGF18	NP_003853.1	NCBI	
Hsa_FGF19	NP_005108.1	NCBI	
Hsa_FGF20	NP_062825.1	NCBI	
Hsa_FGF21	NP_061986.1	NCBI	
Hsa_FGF22	NP_065688.1	NCBI	
Hsa_FGF23	NP_065689.1	NCBI	

Supplementary Table 3. List of sequences used for Fgf phylogenetic analyses (continued).

Mmu_FGF1	NP_034327.1	NCBI	
Mmu_FGF2	NP_032032.1	NCBI	
Mmu_FGF3	NP_032033.2	NCBI	
Mmu_FGF4	NP_034332.2	NCBI	
Mmu_FGF5	NP_034333.1	NCBI	
Mmu_FGF6	NP_034334.1	NCBI	
Mmu_FGF7	NP_032034.1	NCBI	
Mmu_FGF8	NP_034335.1	NCBI	
Mmu_FGF9	NP_038546.2	NCBI	
Mmu_FGF10	NP_032028.1	NCBI	
Mmu_FGF11	NP_001349552.1	NCBI	
Mmu_FGF12	NP_898887.1	NCBI	
Mmu_FGF13	NP_034330.2	NCBI	
Mmu_FGF14	NP_034331.2	NCBI	
Mmu_FGF15	NP_032029.1	NCBI	
Mmu_FGF16	NP_085117.2	NCBI	
Mmu_FGF17	NP_032030.1	NCBI	
Mmu_FGF18	NP_032031.1	NCBI	
Mmu_FGF20	NP_085113.2	NCBI	
Mmu_FGF21	NP_064397.1	NCBI	
Mmu_FGF22	NP_075793.1	NCBI	
Mmu_FGF23	NP_073148.1	NCBI	
Gga_FGF1	NP_990511.1	NCBI	
Gga_FGF2	NP_990764.1	NCBI	
Gga_FGF3	NP_990658.1	NCBI	
Gga_FGF4	NP_001026717.3	NCBI	
Gga_FGF5	XP_040525587.1	NCBI	
Gga_FGF6	XP_040516881.1	NCBI	
Gga_FGF7	NP_001012543.1	NCBI	
Gga_FGF8	NP_001012785.2	NCBI	
Gga_FGF9	NP_989730.1	NCBI	
Gga_FGF10	NP_990027.1	NCBI	
Gga_FGF12	NP_990219.1	NCBI	
Gga_FGF13	NP_001001743.2	NCBI	
Gga_FGF14	NP_990108.1	NCBI	
Gga_FGF16	NP_001038115.1	NCBI	
Gga_FGF18	NP_990045.1	NCBI	
Gga_FGF19	NP_990005.2	NCBI	
Gga_FGF20	XP_040526058.2	NCBI	
Gga_FGF22	XP_025000207.1	NCBI	
Gga_FGF23	XP_040516887.1	NCBI	
Xtr_FGF1	NP_001136293.1	NCBI	
Xtr_FGF2	NP_001017333.1	NCBI	
Xtr_FGF3	NP_001008154.1	NCBI	
Xtr_FGF4	NP_001136294.1	NCBI	
Xtr_FGF5	XP_002934055.1	NCBI	
Xtr_FGF6	NP_001136295.1	NCBI	
Xtr_FGF7	NP_001011366.1	NCBI	
Xtr_FGF8	NP_001008163.1	NCBI	
Xtr_FGF9	XP_002938621.1	NCBI	
Xtr_FGF10	NP_001016169.1	NCBI	
Xtr_FGF11	XP_031755504.1	NCBI	
Xtr_FGF12	NP_001093754.1	NCBI	
Xtr_FGF13	NP_001072569.1	NCBI	
Xtr_FGF14	NP_001136296.1	NCBI	
Xtr_FGF16	XP_031747368.1	NCBI	
Xtr_FGF19	NP_001136297.1	NCBI	
Xtr_FGF20	NP_001137399.1	NCBI	
Xtr_FGF22	NP_001137396.1	NCBI	
Xtr_FGF23a	XP_002940351.1	NCBI	
Xtr_FGF23b	XP_002940347.2	NCBI	
Lch_FGF1	XP_006003277.1	NCBI	
Lch_FGF2	XP_005994763.1	NCBI	
Lch_FGF3	XP_005996298.1	NCBI	
Lch_FGF4	XP_005996280.1	NCBI	
Lch_FGF5	XP_005999825.1	NCBI	
Lch_FGF6	XP_005989333.1	NCBI	
Lch_FGF7	XP_005998372.1	NCBI	
Lch_FGF8	XP_006003512.1	NCBI	
Lch_FGF9	XP_006007650.1	NCBI	

Supplementary Table 3. List of sequences used for Fgf phylogenetic analyses (continued).

Lch_FGF10	XP_006005973.1	NCBI	
Lch_FGF11	XP_006005745.1	NCBI	
Lch_FGF12	XP_014341857.1	NCBI	
Lch_FGF13	XP_005990389.1	NCBI	
Lch_FGF14	XP_005994553.1	NCBI	
Lch_FGF16	XP_006000297.1	NCBI	
Lch_FGF17	XP_005992664.1	NCBI	
Lch_FGF18	XP_006002502.1	NCBI	
Lch_FGF19	XP_005996281.1	NCBI	
Lch_FGF20	XP_006000150.1	NCBI	
Lch_FGF21	XP_014352523.1	NCBI	
Lch_FGF22	XP_005992028.1	NCBI	
Lch_FGF23	XP_005989331.1	NCBI	
Lch_FGF24	XP_006012032.1	NCBI	

Supplementary Table 4. Checking remnants of poly(A) tail in *O. dioica Fgf9/16/20* intron less paralogs.

	Adenine % downstream of STOP codon			
	50 nt	**100 nt**	**150 nt**	**200nt**
Fgf9/16/20a	34%	36%	34,70%	34%
Fgf9/16/20b	34%	42%	41,30%	41%
Fgf9/16/20c	20%	32%	32%	27,50%
Fgf9/16/20d	30%	41%	36%	36%
Fgf9/16/20e	26%	33%	30,70%	Overlaps gene
Fgf9/16/20f	10%	23%	27,30%	27,50%

Supplementary Table 5. Compared Fgf identity and similarity between *H. sapiens* paralogs and *O. dioica* paralogs.

***Homo sapiens* Fgfs - Whole protein sequences**

	FGF1	FGF2	FGF3	FGF4	FGF5	FGF6	FGF7	FGF10	FGF22	FGF8	FGF17	FGF18	FGF9	FGF16	FGF20	FGF11	FGF12	FGF13	FGF14	FGF19	FGF21	FGF23
FGF1		53,8%	22,1%	23,3%	20,7%	20,4%	26,9%	24,6%	27,6%	13,4%	15,7%	20,9%	28,4%	28,2%	31,6%	21,8%	16,7%	19,4%	21,0%	23,1%	22,8%	16,6%
FGF2	63,9%		26,3%	30,3%	23,6%	29,2%	26,9%	24,8%	29,4%	19,7%	16,8%	17,7%	28,7%	25,2%	27,8%	20,0%	20,7%	16,4%	17,0%	23,6%	21,2%	17,2%
FGF3	33,7%	35,8%		21,8%	29,4%	23,8%	24,9%	29,1%	30,4%	25,4%	22,5%	22,6%	28,9%	28,7%	29,9%	21,5%	20,3%	21,9%	25,4%	25,3%	27,0%	22,8%
FGF4	30,6%	40,8%	36,1%		29,0%	50,0%	22,6%	20,0%	30,1%	25,9%	22,0%	20,5%	29,0%	28,4%	33,5%	30,1%	24,7%	21,8%	22,7%	19,7%	20,7%	16,5%
FGF5	30,0%	35,8%	42,7%	40,3%		30,2%	27,1%	27,4%	21,1%	22,9%	17,1%	22,5%	28,1%	28,2%	28,6%	24,1%	20,1%	23,7%	28,0%	16,4%	19,3%	21,5%
FGF6	32,6%	40,1%	35,8%	62,1%	40,6%		25,2%	25,7%	29,2%	23,2%	22,9%	19,5%	32,2%	28,8%	32,5%	22,7%	21,4%	23,1%	20,7%	18,6%	22,1%	15,6%
FGF7	40,6%	42,8%	37,9%	38,5%	40,1%	37,0%		46,9%	38,6%	20,7%	23,6%	25,3%	27,4%	28,8%	26,7%	23,7%	19,7%	21,1%	19,4%	22,6%	18,7%	14,7%
FGF10	39,8%	38,3%	40,3%	34,9%	40,3%	39,5%	64,9%		39,5%	22,3%	20,8%	21,2%	31,3%	30,5%	27,3%	21,5%	22,5%	21,3%	20,8%	20,7%	20,0%	16,4%
FGF22	40,6%	45,6%	43,3%	42,1%	33,7%	43,4%	56,3%	57,2%		19,0%	22,4%	16,9%	33,6%	34,3%	31,6%	26,3%	19,1%	18,0%	21,7%	26,2%	23,3%	17,1%
FGF8	31,5%	33,5%	34,9%	37,1%	34,6%	35,8%	37,4%	35,4%	31,7%		55,0%	49,4%	22,8%	20,5%	23,8%	21,1%	17,6%	19,9%	17,2%	22,3%	18,3%	17,8%
FGF17	31,7%	32,7%	36,3%	34,0%	29,4%	36,0%	39,2%	35,2%	34,1%	67,9%		51,3%	21,0%	21,9%	21,7%	20,7%	13,9%	18,7%	21,6%	21,3%	16,1%	17,8%
FGF18	36,5%	28,8%	33,3%	30,7%	34,9%	29,7%	41,0%	34,4%	28,1%	63,9%	69,2%		24,7%	22,0%	21,6%	22,2%	20,8%	24,0%	21,8%	17,9%	14,0%	15,3%
FGF9	41,2%	38,8%	42,5%	39,5%	40,3%	45,9%	44,8%	48,7%	48,3%	35,2%	36,5%	37,4%		69,6%	69,6%	31,1%	24,7%	24,3%	24,6%	19,6%	23,8%	17,4%
FGF16	38,4%	32,9%	41,1%	41,2%	42,6%	43,3%	44,1%	48,1%	47,8%	35,1%	37,8%	35,2%	79,0%		62,0%	25,7%	22,3%	25,9%	26,5%	24,2%	26,1%	17,0%
FGF20	43,4%	37,7%	41,3%	43,1%	38,8%	45,3%	39,4%	43,0%	42,2%	37,7%	34,5%	29,9%	80,8%	74,5%		30,3%	23,2%	26,0%	24,6%	21,9%	21,9%	20,1%
FGF11	34,9%	29,8%	36,0%	40,2%	39,0%	37,3%	40,9%	35,9%	38,6%	31,0%	31,7%	33,9%	46,2%	40,6%	46,9%		56,0%	52,2%	55,6%	20,0%	18,3%	16,1%
FGF12	26,9%	34,1%	36,3%	33,9%	35,8%	35,9%	36,8%	36,7%	31,6%	31,7%	26,4%	33,6%	41,6%	37,6%	39,3%	70,0%		65,5%	71,5%	19,5%	20,3%	18,6%
FGF13	32,8%	24,2%	35,7%	29,7%	40,5%	35,4%	36,2%	34,3%	30,5%	31,6%	30,9%	31,4%	36,3%	42,1%	39,6%	64,9%	80,6%		67,6%	17,7%	24,2%	19,2%
FGF14	32,7%	25,5%	43,4%	31,7%	43,7%	35,2%	30,6%	36,1%	35,7%	26,9%	36,6%	35,7%	39,3%	41,4%	38,4%	67,3%	84,3%	82,6%		19,6%	23,5%	15,7%
FGF19	32,4%	32,8%	35,9%	31,6%	25,0%	32,2%	32,1%	33,5%	38,4%	33,0%	32,8%	26,7%	30,6%	36,5%	32,5%	32,1%	30,6%	28,4%	30,2%		34,3%	27,1%
FGF21	28,8%	30,9%	38,8%	31,2%	29,7%	33,2%	32,5%	33,6%	34,8%	31,3%	28,6%	23,9%	32,4%	33,2%	30,0%	26,8%	32,2%	35,2%	35,3%	46,8%		24,4%
FGF23	24,2%	24,7%	35,6%	23,7%	32,8%	23,2%	23,1%	26,2%	28,0%	28,6%	28,0%	23,9%	23,0%	23,0%	28,0%	25,6%	29,3%	29,7%	26,6%	38,7%	37,1%	

Identity

Similarity

***Oikopleura dioca* Fgfs - Whole protein sequences**

	FGF11/12/13/14a	FGF11/12/13/14b	FGF11/12/13/14c	FGF11/12/13/14d	FGF9/16/20a	FGF9/16/20b	FGF9/16/20c	FGF9/16/20d	FGF9/16/20e	FGF9/16/20f
FGF11/12/13/14a		51,50%	34,80%	33,90%	19,80%	15,10%	18,10%	15,00%	14,40%	16,20%
FGF11/12/13/14b	71,40%		31,50%	32,60%	21,70%	5,70%	18,60%	17,10%	13,60%	14,60%
FGF11/12/13/14c	57,60%	50,70%		30,70%	20,00%	14,40%	16,40%	17,50%	15,20%	15,40%
FGF11/12/13/14d	53,20%	48,50%	45,60%		14,30%	16,10%	15,30%	13,10%	5,70%	13,40%
FGF9/16/20a	38,30%	39,60%	38,60%	30,20%		18,70%	26,20%	23,70%	19,60%	23,90%
FGF9/16/20b	29,30%	9,50%	32,20%	27,70%	33,50%		27,60%	21,90%	22,50%	32,80%
FGF9/16/20c	30,60%	29,00%	36,00%	27,50%	38,50%	56,50%		21,20%	22,20%	29,30%
FGF9/16/20d	28,80%	29,40%	30,30%	30,50%	40,50%	43,20%	38,10%		61,60%	23,50%
FGF9/16/20e	25,10%	22,80%	27,60%	13,90%	35,90%	42,70%	38,90%	79,30%		21,10%
FGF9/16/20f	25,50%	25,70%	22,90%	27,60%	39,00%	49,40%	55,70%	42,10%	38,30%	

Identity

Similarity

***O. dioica* Fgf9/16/20- FGF domain**

	FGF9/16/20a	FGF9/16/20b	FGF9/16/20c	FGF9/16/20f	FGF9/16/20d	FGF9/16/20e
FGF9/16/20a		23.7%	27.7%	25.9%	32.1%	26.8%
FGF9/16/20b	43.9%		30.6%	36.6%	24.1%	25.2%
FGF9/16/20c	43.8%	61.9%		30.8%	21.2%	22.4%
FGF9/16/20f	41.5%	56.0%	60.2%		27.4%	22.7%
FGF9/16/20d	52.9%	51.1%	37.2%	48.1%		63.4%
FGF9/16/20e	47.1%	49.6%	40.8%	43.9%	79.4%	

Identity

Similarity

***H. sapiens* Fgf9/16/20- FGF domain**

	FGF9	FGF16	FGF20
FGF9		87.8%	86.3%
FGF16	95.0%		80.6%
FGF20	95.0%	92.8%	

Identity

Similarity

Supplementary Table 6. Compared FgfR identity and similarity between *H. sapiens* paralogs and *O. dioica* paralogs. Percentages were inferred separately for the intracellular TK domain and the extracellular portion of the FgfRs.

Intracellular TK domain

	Hsa_FGFR1	Hsa_FGFR2	Hsa_FGFR3	Hsa_FGFR4
Hsa_FGFR1	100%	84.1%	78.3%	67.5%
Hsa_FGFR2	92.8%	100%	82.4%	72.1%
Hsa_FGFR3	88.7%	90.9%	100%	74.9%
Hsa_FGFR4	78.7%	83.6%	83.1%	100%

Identity (lower triangle) / Similarity (upper triangle)

	OdiBAR_FGFRa	OdiBAR_FGFRb	OdiBAR_FGFRc
OdiBAR_FGFRa	100%	35.0%	32.6%
OdiBAR_FGFRb	50.7%	100%	31.9%
OdiBAR_FGFRc	50.2%	49.6%	100%

Identity (lower triangle) / Similarity (upper triangle)

Extracellular Portion

	Hsa_FGFR1	Hsa_FGFR2	Hsa_FGFR3	Hsa_FGFR4
Hsa_FGFR1	100%	62.6%	50.7%	45.2%
Hsa_FGFR2	75.1%	100%	54.3%	46.7%
Hsa_FGFR3	65.1%	65.7%	100%	51.0%
Hsa_FGFR4	61.9%	62.4%	61.4%	100%

Identity (lower triangle) / Similarity (upper triangle)

	OdiBAR_FGFRa	OdiBAR_FGFRb	OdiBAR_FGFRc
OdiBAR_FGFRa	100%	14.5%	13.1%
OdiBAR_FGFRb	24.4%	100%	19.3%
OdiBAR_FGFRc	26.3%	34.0%	100%

Identity (lower triangle) / Similarity (upper triangle)

Supplementary Table 7. Interspecies FgfR identity and similarity. Percentages were inferred separately for the intracellular TK domain (only *O. dioica* cryptic species) and the extracellular portion of the FgfRs. Species included: *Oikopleura dioica* (Odi), *Homo sapiens* (Hsa), *Ciona robusta* (Cro), *Branchiostoma floridae* (Bfl), *Mus musculus* (Mmu), *Gallus gallus* (Gga) and *Lepisosteus oculatus* (Loc).

Intracellular TK domain - *O. dioica* cryptic species (BAR - OSA/OKI)

	Odi_FGFRa_BAR		Odi_FGFRb_BAR		Odi_FGFRc_BAR	
	Identity	Similarity	Identity	Similarity	Identity	Similarity
Odi_FGFRa_OSA	98.6%	99.7%	99.0%	99.3%	90.7%	96.3%
Odi_FGFRa_OKI	94.3%	98.0%	96.1%	99.3%	78.4%	90.3%

Extracellular portion - *O. dioica* cryptic species (BAR - OSA/OKI)

	Odi_FGFRa_BAR		Odi_FGFRb_BAR		Odi_FGFRc_BAR	
	Identity	Similarity	Identity	Similarity	Identity	Similarity
Odi_FGFRa_OSA	90.2%	94.4%	87.3%	95.0%	79.8%	89.5%
Odi_FGFRa_OKI	67.6%	83.0%	63.1%	79.0%	66.4%	81.5%

Extracellular portion of the FgfRs - Various chordate FgfR

	Cro__FGFR		Bfl__FGFR	
	Identity	Similarity	Identity	Similarity
Odi_FGFRa	16.7%	29.0%	16.8%	28.8%
Odi_FGFRb	14.2%	27.2%	13.6%	22.0%
Odi_FGFRc	12.4%	23.7%	9.6%	17.3%
Hsa_FGFR1	33.1%	46.3%	38.4%	54.2%
Hsa_FGFR2	15.0%	20.2%	39.0%	54.3%
Hsa_FGFR3	30.4%	39.7%	38.9%	50.3%
Hsa_FGFR4	31.1%	39.8%	39.6%	52.9%

Extracellular portion of the FgfRs - Vertebrate FgfR

	Hsa_FGFR1		Hsa_FGFR2	
	Identity	Similarity	Identity	Similarity
Mmu_FGFR1	97.9%	98.7%	91.2%	93.4%
Gga_FGFR1	81.6%	89.7%	76.4%	83.8%
Loc_FGFR1	56.5%	72.8%	63.4%	73.6%

	Hsa_FGFR3		Hsa_FGFR4	
	Identity	Similarity	Identity	Similarity
Mmu_FGFR3	90.1%	92.5%	88.1%	91.3%
Gga_FGFR3	59.3%	69.8%	65.4%	75.2%
Loc_FGFR3	54.1%	66.9%	32.2%	38.8%

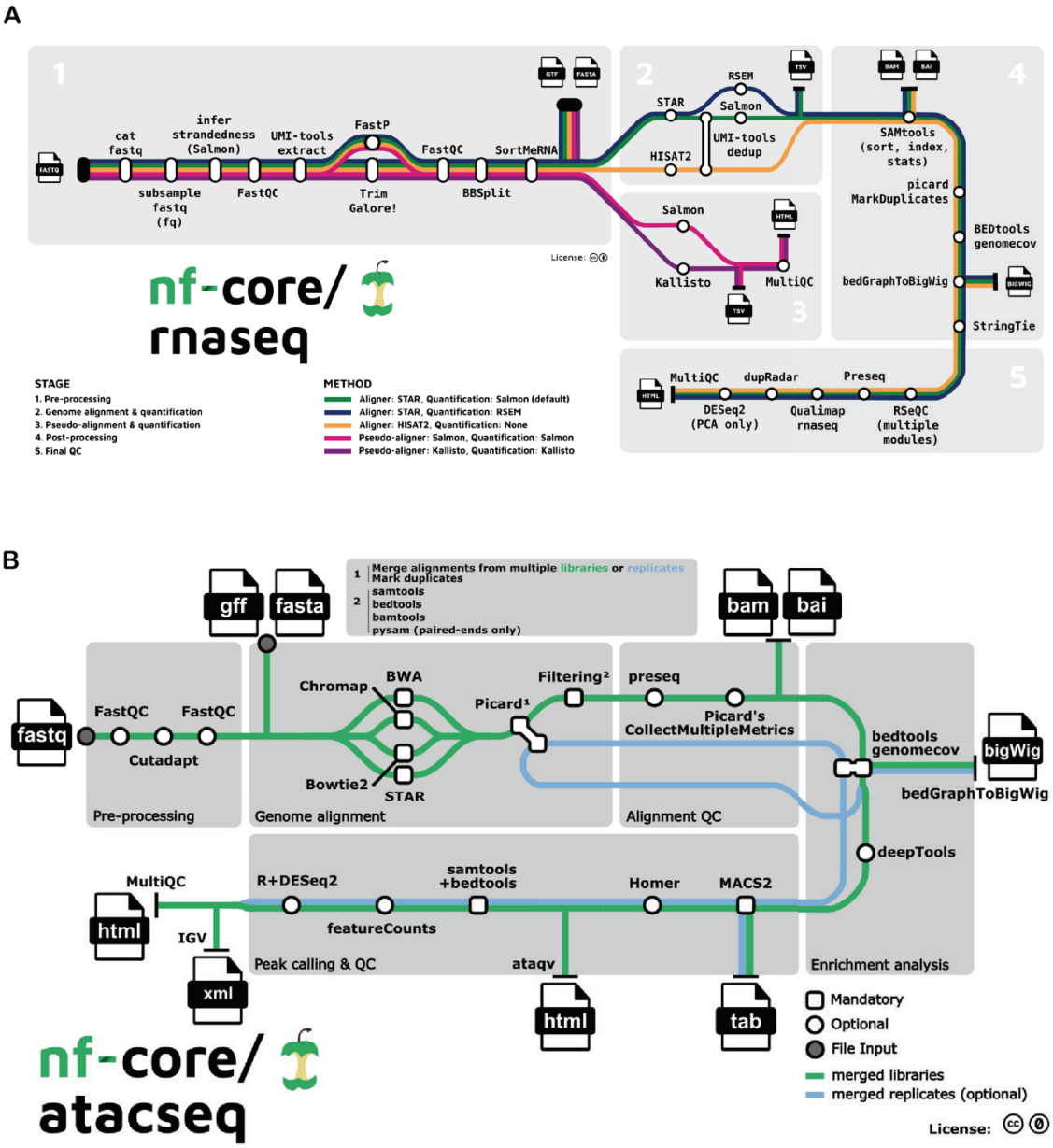
Supplementary Table 8. Gene expression matrix values for *FgfR* genes, extracted from the OikoBase (http://oikoarrays.biology.uiowa.edu) (Danks et al., 2013). Expression units are not specified in the source.

		Developmental time																	
Gene	Annotation ID	Oocyte	2-8 cells	1HPF	Tailbud	Hatched	Early tadpole	Tailshift	Day 1	Day 2	Day 3	Day 4	Day 5	Trunk	Testis	Ovary	Day 2 dense	Day 3 dense	Day 4 dense
FGFRa	GSOIDG00008404001	5692	5560	5572	1895	1031	2088	2452	1271	1495	1630	765	1014	0	0	2618	1524	1224	1913
FGFRb	GSOIDG00009004001	1821	1828	2347	1487	468	0	680	353	310	462	0	0	0	0	1196	0	0	0
FGFRc	GSOIDG00010261001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

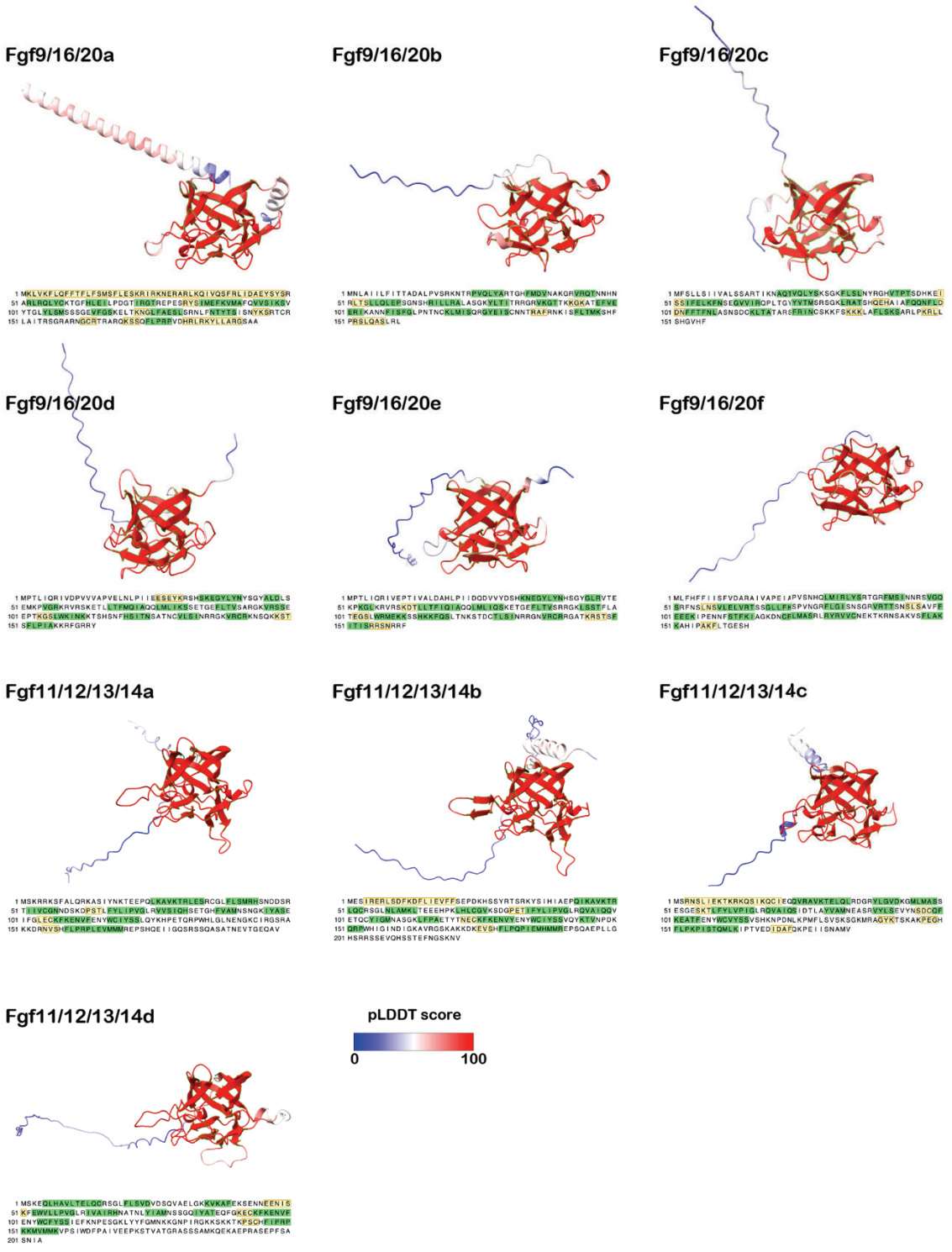
Supplementary Table 9. Search of genes involved in intracellular transduction pathways. The human gene was first used to identify the ortholog in *C. robusta*, and subsequently human and ascidian genes were used as queries to search the ortholog in *O. dioica*. Hyphens indicate genes not found through the RBBH method.

Protein	*Homo sapiens*	*Ciona robusta*	*Oikopleura dioica* Bar2_p4
MEK1/2	NP_002746.1 NP_109587.1	KH.L147.22	Chr1:8350574..8352253
ERK	NP_620407.1 NP_002737.2	KH.L153.20	Chr2:11200708..11202156
KSR	NP_055053.1 NP_775869.3	KH.L8.14	XSR:11422609..11425606
RAF	NP_004324.2	KH.L18.20	PAR:6700456..6702922
H-Ras/K-Ras/N-Ras	NP_005334.1 NP_203524.1 NP_002515.1	-	-
M-Ras	NP_001078518.1	KH.L172.2	-
R-Ras2	NP_036382.2	KH.C2.451	-
Rap	NP_056461.1 NP_002877.2	KH.C5.109 KH.L34.27	PAR:7509105..7509833 PAR:9345190..9346366 PAR:11017881..11018568
Ral-A	NP_005393.2	KH.S1011.2	Chr1:4625667..4626462
Rheb	NP_005605.1 NP_653194.1	KH.C7.785	XSR:9376352..9377513
SOS	NP_005624.2 NP_008870.2	KH.L13.16	PAR:5887094..5891476
GRB2	NP_002077.1	KH.C8.67	Chr2:11402227..11404073
FRS2	NP_006645.3 NP_006644.1	KH.C14.464	-
AKT	NP_005154.2 NP_001617.1 NP_005456.1	KH.C4.30 KH.C4.709	PAR:13181373..13184938
PDK	NP_002604.1	KH.C7.462 KH.C10.71	PAR:7667935..7669757
PK3CA	NP_006209.2	KH.L60.15	XSR:773873..776523
GAB1	NP_997006.1	-	-
PKC	NP_002728.1	KH.C3.45	XSR:5195954..5198247
PLCy	NP_002651.2	KH.C3.493	XSR:9027039..9031200

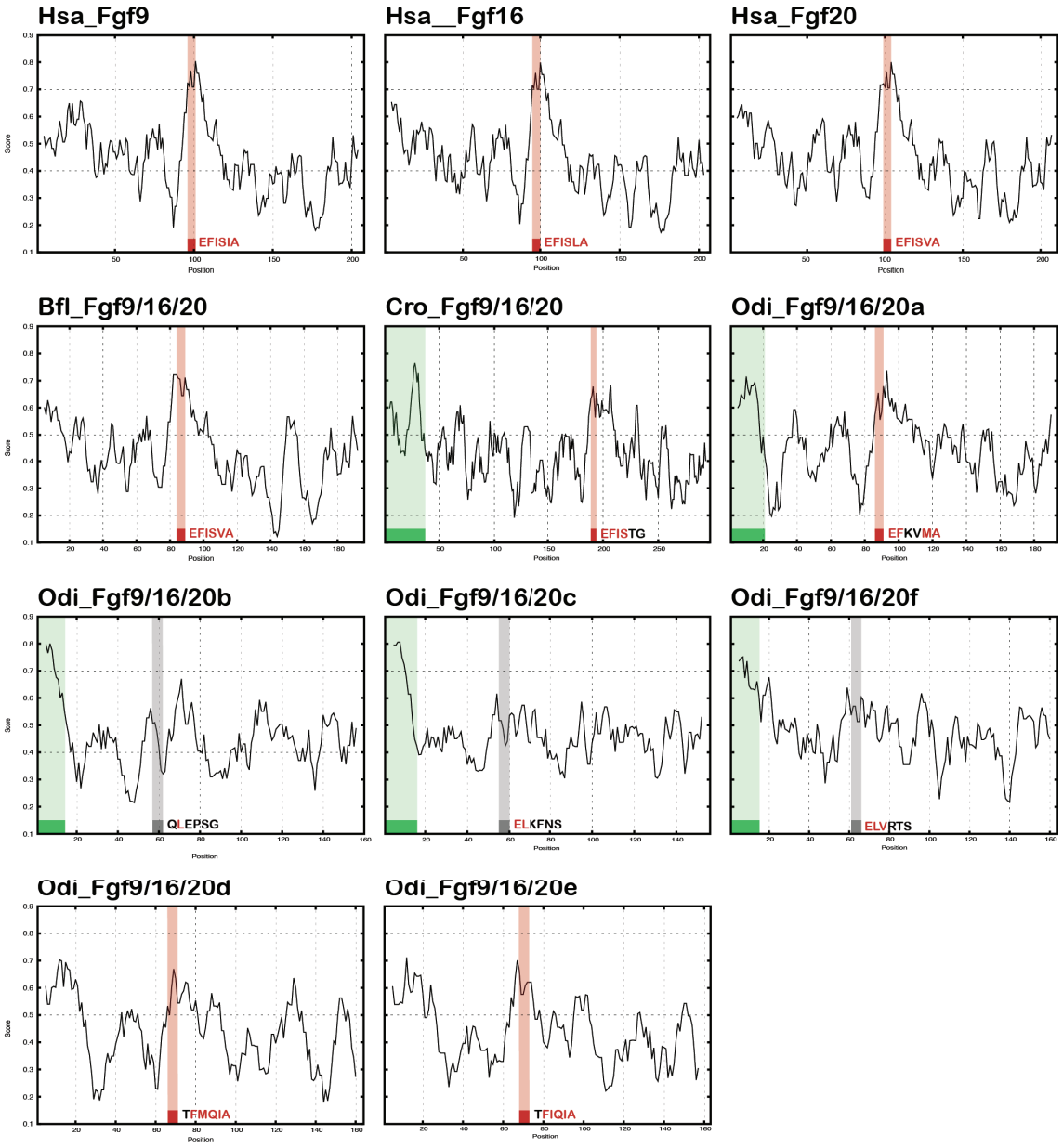
2. Supplementary Figures



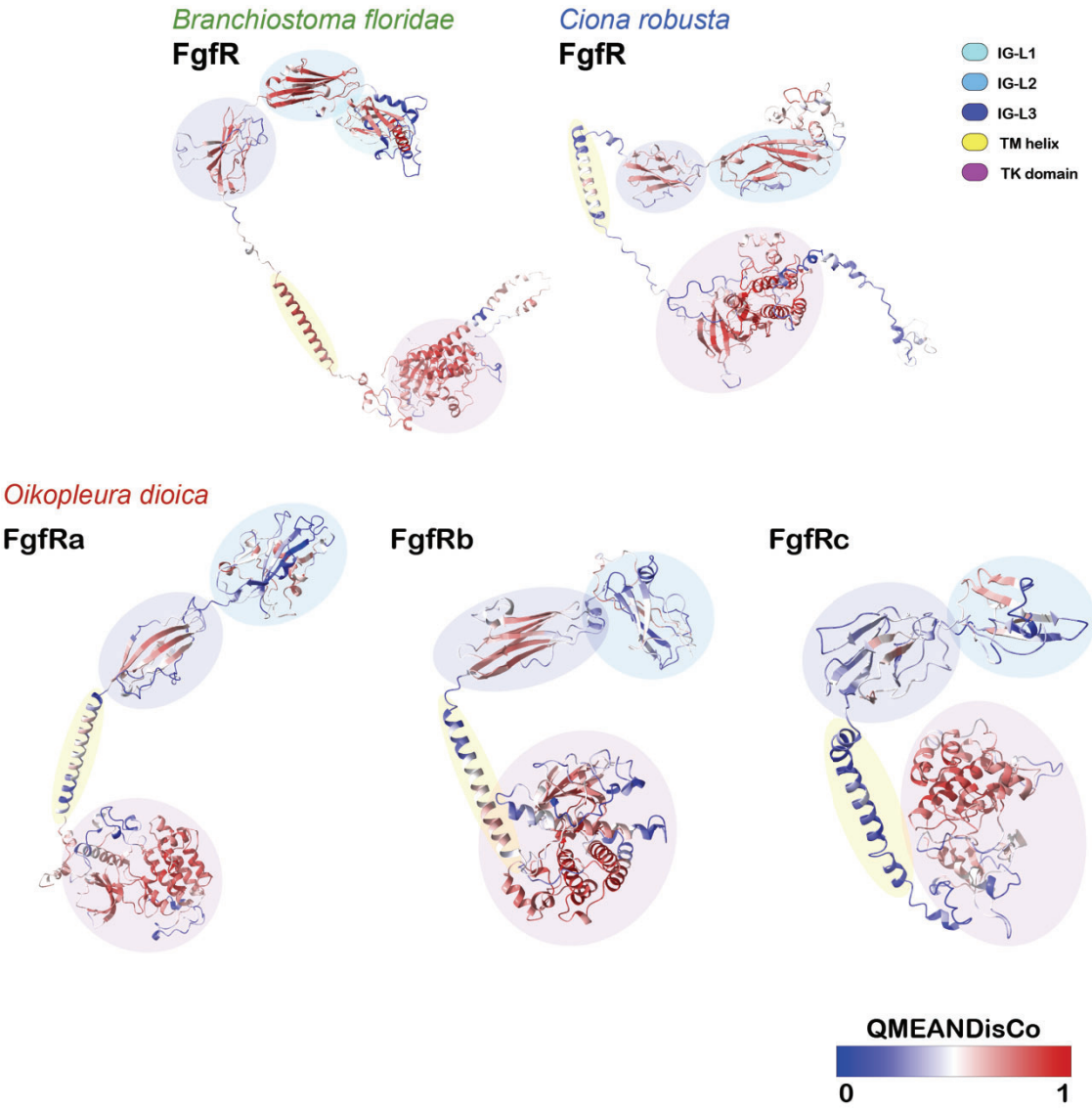
Supplementary Figure 1. (A) nf-core/rnaseq workflow. Extracted from https://nf-co.re/rnaseq/3.14.0/ **(B) nf-core/atacseq workflow.** Extracted from https://nf-co.re/atacseq/2.1.2/



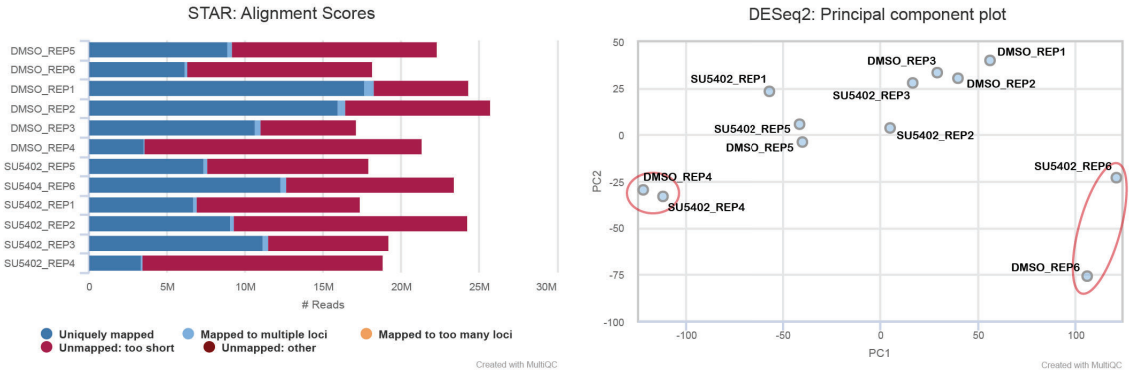
Supplementary Figure 2. Three-dimensional models for O. dioica Fgf proteins. Models are colored according to their predicted local distance difference test (pLDDT) score. Predicted β-sheets are highlighted in light green in the models as well as in the protein sequences. Predicted α-helices are highlighted in light yellow in the protein sequences. All Fgf11/12/13/14 paralogs models and sequences correspond to the proximal isoform.



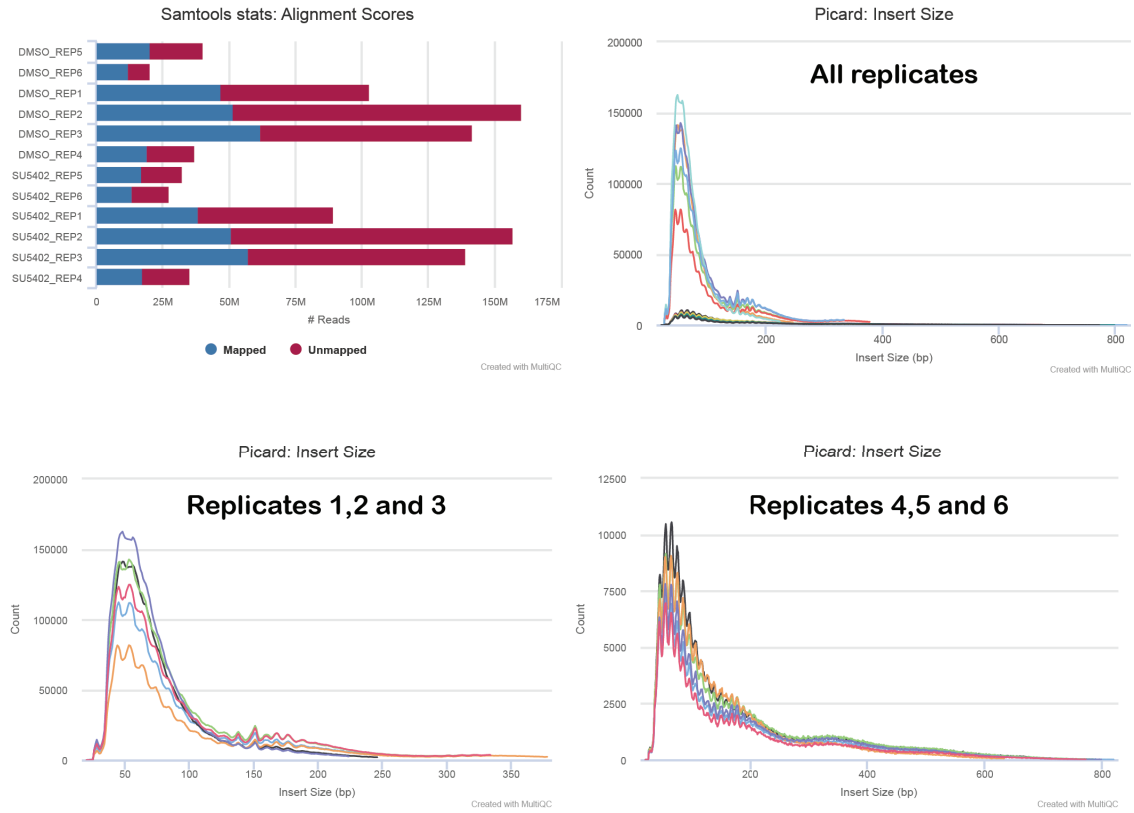
Supplementary Figure 3. Hydropathy profile of chordate Fgf9/16/20 proteins. The graphs show the Kyte-Doolittle hydropathy score (y-axis) along the length of each protein sequence for Homo sapiens (Hsa), Branchiostoma floridae (Bfl), Ciona robusta (Cro), and Oikopleura dioica (Odi) Fgf9/16/20 orthologs. Scores are scaled to 1 for comparison purposes. Green squares and shaded areas indicate the presence and position of a cleavable signal peptide (SP) as predicted by Phobius. Red squares and shaded areas mark the position of well-conserved EFISIA motifs. Gray squares and shaded areas represent the positions of degraded EFISIA motifs. The amino acid letters in the EFISIA motif of each protein are colored red if they are conserved with the vertebrate counterpart or if they have been substituted by an aminoacid with similar physicochemical properties.



Supplementary Figure 4. Three-dimensional models of the Fgf receptors. *O. dioica* FgfR models are based on the protein sequences of the annotations derived from this project. *B. floridae* FgfR model is based on the protein sequence in XP_035673320.1. *C. robusta* FgfR model is based on the protein sequence in NP_001037820.1. Models are coloured according to their local QMEANDisCo score along the protein sequence. Coloured shades in the background mark the different domains and motifs that build each protein according to the colour code in the legend.



Supplementary Figure 5. RNA-seq alignment scores and PCA plot. Replicates marked with a red circle In the PCA plot were discarded for DGE analyses.



Supplementary Figure 6. ATAC-seq alignment scores and insert

3. Supplementary Material

3.1. WMISH protocol for *Oikopleura dioica* samples

*All steps before coloration are performed in 200µL Eppendorf tubes and the volume used of each solution is 200µL per reaction/tube in every step. Coloration is performed in glass bowls and the volume used of TMNT/Coloration solution is 400µL per reaction/bowl.

*In all steps, the supernatant is removed softly with a P200 and, in order to avoid losing embryos, the tip is emptied on a petri dish and checked under the stereomicroscope. If embryos were taken along with the supernatant, they are put back into the reaction tube using a P2.

Solution	Reference / Components
PBS	Sigma-Aldrich #P5493
PBST	Tween-20 0.2% in PBS
MABT	Maleic acid 100 mM NaCl 150 mM Tween-20 0.1% Adjust pH to 7.5
MABT-SS	Add to MABT: 0.25% w/v Bovine Serum Albuminum (BSA) 5% v/v sheep serum
SSC20x	Sigma-Aldrich #S6639
SSCT	Tween-20 0.2% in SSC
Prehybridisation solution	Formamide 50% Heparine 1% SSC20x 25% EDTA pH 8 5 mM Yeast RNA 0.5 mg/mL 50x Denhardt's solution 2% Tween-20 0.75%
Denhardt's solution	Ficoll 1% Polyvinylpyrrolidone 1% BSA 1%
TMNT	Tris-HCl pH 9.5 0.1 M MgCl2 50 mM NaCl 50 mM Tween-20 0.5‰
NBT	Roche #11383213001
BCIP	Roche #11383221001
Colouration solution	NBT 0.2% + BCIP 0.2% in TMNT
Reduction solution	DTT (PanReac AppliChem #A3668) 50mM in PBST

Day 1 (*RNAse free*)

*Embryos stored in 70% EtOH at -20ºC are transferred to 200 µL Eppendorf tubes.

*Prepare before starting: DTT 50mM in PBST; Triethanolamine 100mM + Tween-20 0.25%; 2.5 and 5 µl/ml acetic anhydride solutions (do not shake, mix softly with the micropipette); Prehybridization solution.

1. **Rehydration**:
   - (1) Wash 1:1 EtOH:PBS-T – RT – 5'
   - (2) Wash PBS-T – RT – 5'
2. **Reduction solution treatment**
   - DTT 50mM in PBST – RT – 10'
3. **(2x) Wash triethanolamine 100mM (pH8)** – RT – 5'

* Triethanolamine makes embryos look transparent (difficult to notice under the stereomicroscope)

4. **Acetic anhydride treatment** (embryos recover the stringency here)

a. 2.5µl/ml acetic anhydride in triethanolamine 100 mM + Tween-20 0.25% – RT – 5'
b. 5µl/ml acetic anhydride in triethanolamine 100 mM + Tween-20 0.25% – RT – 5'

5. (2x) Wash PBST – RT – 5'

6. Prehybridization (Pre-hyb) at 63°C and 300rpm ~3h or more

* Formamide in the Pre-hyb solution makes embryos look transparent (difficult to notice)

7. Hybridization:

a. Add the probe to prehybridization solution (now Hyb) in 1:300 dilution (for DIG-labelled probes ranging 100-200 ng/µL in RNA concentration)
c. Denaturalize the probe: Incubate the Hyb solution 85°C - 5' followed by 4°C – 5'
b. Replace the Pre-hyb solution with the Hyb solution and incubate overnight at 63°C

* Optimal dilution of the probe depends on the original concentration of the probe, on the intensity of the target gene expression and on the fixed embryos themselves. 1:300 is a good starting point for RNA probes ranging 100-200 ng/µL, but it may need to be adjusted. The hybridization temperature and the number and accuracy of the washes with SSCT solution on day 2 play also an important role.

Day 2

*Prepare before starting: MABT-SS (Blocking solution); 2x, 0.2x and 0.1x SSCT solutions.

*2x and 0,2x SSCT solutions are preheated and kept at 65ºC before use.

8. Post-Hybridization washes - 65°C (rise the temperature of the Thermomixer)

a. (1x) Remove 100µL Hyb solution and add 100µL preheated 2xSSCT – 65ºC – 15'
b. (2x) Wash with preheated 2xSSCT – 65ºC – 15'
c. (2x) Wash with preheated 0.2xSSCT – 65ºC – 10'
d. (1x) Wash with 0.1xSSCT – RT – 5'

9. (2x) Wash with MABT – RT – 10'

10. Blocking: Incubate in MABT-SS solution – RT – 30'-2h

11. Antibody. Incubate with Anti-DIG-AP (Roche #11093274910) diluted 1:2500 in MABT-SS overnight at 4°C

Day 3

*Prepare before starting: TMNT; Coloration solution (TMNT + NBT + BCIP) – protect from light exposure.

12. Antibody washes: 2h of washes with MAB-T every 15' (8 washes) – RT

13. Wash with 200 µL TMNT and transfer the embryos from the tube to a glass bowl containing 200 µL TMNT (400 µL in total) – RT – 5´

14. COLORATION. Remove TMNT and add 400 µL coloration solution. Let the embryos dye at room temperature and protected from light exposure. Change coloration solution twice a day until they get dyed.

*Coloration reaction incubation time can vary form a few minutes to more than one week, although generally most of the reactions proceed between ~2h and ~2 days. If the reaction is going too slow or too fast, you can play with the incubation temperature and with the concentration of NBT and BCIP in the coloration solution.

STOP COLORATION

- (3) Rinse with Tween-20 0.1% in H_2O – RT – No incubation time (do the three washes in a row, sample by sample)
- (2) Wash with PBST – RT – 5´
- Keep in glycerol 80% in PBS – embryos can be stored long term at 4ºC

3.2. ATAC-seq protocol for *Oikopleura dioica* samples

Reagents, Stock solutions and Buffers:

Reagents and materials:

*Those marked with a (*) require a stock/working solution preparation prior to their use

- **PBS 10x** (*) (Sigma #P5493), stored at RT – for Rinse Buffer and Transposition Reaction Mix preparation.
- **Nuclease-free H_2O** (Sigma #95284, H_2O DEPC)
- **Protease from *Streptomyces griseus* (=Actinase E, Pronase E)** (*) (Sigma, P8811) – Stored at -20ºC in powder format – for chemical dechorionatiion of prehatchling embryos.
- **Tris-HCl UltraPure™ 1M, pH 7.5** (ThermoFisher, catalog #15567-027) – Stored at 4ºC. – for Mother Buffer and TD Buffer preparation.
- **NaCl** (*) (Sigma #S3014), stored at RT – for Mother Buffer preparation
- **1M MgCl2** (Sigma, catalog #M1028), stored at RT – for Mother Buffer preparation
- **cOmplete™, EDTA-free Protease Inhibitor Cocktail**(*) (Roche, #11873580001) – Stored at 4ºC in tablets format – for Rinse Buffer and Resuspension Buffer preparation
- **N,N-Dimethylformamide** (Sigma, D4551-250ML), stored at RT – for TD Buffer preparation.
- **Detergents:**
  - **Nonidet® P 40 Substitute (NP-40) 10% in aqueous solution** (VWR-G Biosciences, MFCD00132851), stored at 4ºC – for Lysis Buffer and Transposition Reaction Mix preparation
  - **Digitonin**(*) (Promega, catalog # G9441), stored at -20ºC – for Lysis Buffer and Transposition Reaction Mix preparation
  - **Tween-20 10%** (Sigma/Roche, catalog # 11332465001), stored at 4ºC – for Lysis Buffer, Wash Buffer and Transposition Reaction Mix preparation .
- **Tn5 Tagment DNA Enzyme 1 (TDE1)** – Custom, for Transposition Reaction Mix Preparation
- **Primers** – for library generation PCR and qPCR. The kind of primers depends on the indexing requirement of the sequencing technique that will be used to sequence the libraries (single index / combinatorial index / unique dual index).
- **NEBNext High-Fidelity 2X PCR Master Mix** (NEB, catalog #M0541S) – for library generation PCR.
- **SYBR Green I** – for qPCR quantification.
- **Zymo DNA Clean and Concentrator -5 Kit** (Zymo Research, #D4003), stored at RT – for Library Purification
- **Hoechst 33342 solution** (*) (ThermoFisher #62249) – for nuclei staining and quantification.
- **Neubauer chamber** – for nuclei quantification

Stock/Working solutions:

- **PBS 1x stock solution:** Dilute 5 mL of PBS (10x conc.) in 45 mL of nuclease-free water.
- **5M NaCl stock solution:** Dissolve 14.61 gr of NaCl in 50 mL of nuclease-free water.
- **Actinase-E solution:** Prepare a 0.05% p/v concentrated solution in Sterilized Sea Water (SSW) and keep it at 4ºC until use. DO NOT VORTEX, MIX IT WITH THE MICROPIPPETE. *First, check that the balance used to weight such small masses is precise enough measuring the wight of 1 μL of water (= 1 mg).
  - **1 mL:** Dissolve 0,5 mg of Actinase E in 1 mL SSW.
  - **2 mL:** Dissolve 1 mg of Actinase E in 2 mL SSW.
  - **5 mL:** Dissolve 2,5 mg of Actinase E in 5 mL SSW.
- **cOmplete™ protease inhibitor stock solution (25x conc.):**
  Dissolve one cOmplete EDTA-free tablet in 2 ml nuclease-free H_2O. The stock solution (25x) can be stored at 2 to 8 °C for 1 to 2 weeks, or at least 12 weeks at -15 to -25 °C.
- **Digitonin stock solution (1% conc.):**
  Dilute 1:1 with water to make a 1% working stock, aliquot and store at -20ºC for up to 6 months. Do not freeze/thaw more than 5 times.

Buffers and reaction mix:

☐ **Rinse Buffer** (PBS + Proteases inhibitor)
- **1mL:** 960 μL PBS 1x + 40 μL cOmplete™ stock solution (25x)
- **5 mL:** 4,8 mL PBS 1x + 200 μL cOmplete™ stock solution (25x)
- **10 mL:** 9,6 mL PBS 1x + 400 μL cOmplete™ stock solution (25x)
- **50 mL:** 48 mL PBS 1x + 2 mL cOmplete™ stock solution (25x)

☐ **Mother Buffer -** base of Resuspension, Lysis and Wash Buffer. Can be stored at room temperature long term.

Reagent	Final Concentration	Volume for 50 mL	Volume for 10 mL
1M Tris-HCl pH 7.5	10 mM	500 μL	100 μL
5M NaCl	10 mM	100 μL	20 μL
1M MgCl2	3 mM	150 μL	30 μL
Nuclease-free water	NA	49.25 ml	9.85 mL

☐ **Resuspension Buffer** (Mother Buffer + Proteases inhibitor) – prepare and cool it before use.
- **1 mL:** 960 μL Mother Buffer + 40 μL cOmplete™ stock solution (25x)
- **5 mL:** 4,8 mL Mother Buffer + 200 μL cOmplete™ stock solution (25x)
- **10 mL:** 9,6 mL Mother Buffer + 400 μL cOmplete™ stock solution (25x)

☐ **Lysis Buffer** (Mother Buffer + Detergents)

Reagent	Final Concentration	Volume for 50 μL	Volume for 1 mL
Tween-20 10%	0.1% v/v	0,5 μL	10 μL

NP-40 10%	0.1% v/v	0,5 µL	10 µL
Digitonin 1% (*stock solution)	0.01% v/v	0,5 µL	10 µL
Mother Buffer	NA	48,5 µL	970 µL

☐ **Wash Buffer** (Mother Buffer + Tween-20)

- **1 mL:** 990 µL Mother Buffer + 10 µL Tween-20 10% (final 0.1% v/v)
- **5 mL:** 4,95 mL Mother Buffer + 50 µL Tween-20 10% (final 0.1% v/v)
- **10 mL:** 9,9 mL Mother Buffer + 100 µL Tween-20 10% (final 0.1% v/v)

☐ **TD Buffer (2x conc.) (Tagment DNA Buffer) -** 2x TD Buffer may be kept at −20ºC up to 6 months.

Reagent	Final Concentration	Volume for 5 mL	Volume for 10 mL
1M Tris-HCl pH 7.5	20 mM	100 µL	200 µL
1M MgCl2	10 mM	50 µL	100 µL
N,N-Dimethylformamide (DMF)	20% v/v	1 mL	2 mL
Nuclease-free water	NA	3,85 mL*	7,7 mL*

* **To prepare 10 ml of TD buffer:** mix 200 µl of Tris–HCl 1M pH 7.5, 100 µl MgCl2 1 M, and 6 ml of water. Before the addition of DMF, check pH and adjust it to 7.6 with 100% acetic acid. Finally, add 2 ml N,N-dimethylformamide and increase the volume up to 10 ml with water. Sterilize the solution by filtration using a 0.2-µm cellulose acetate syringe filter.

☐ **Transposition Reaction Mix –** prepare right before use

50 µL per reaction (10.000 – 50.000 nuclei):

25 µl - 2X TD Buffer (Tagment DNA Buffer)
16.5 µl - 1X PBS
0.5 µl - 10% Tween-20 (final 0.1% v/v)
0.5 µl - 1% Digitonin (final 0.01% v/v)
X µl - Tn5 Transposase (Tagment DNA Enzyme 1)
X µl - nuclease-free H2O

Vf = 50 µl

Protocol:

Notes:
Reference ATACseq protocols are based on the *Buenrostro et al.* papers, which are based on using 50.000 cells (=nuclei) for the transposition reaction. Typically, 50.000-100.000 cells yield the best results, but as few as 5.000 have been reported to work.

1. In vitro fertilization and culture:

- Make an estimation of the number of embryos that will be needed to reach the appropriate number of cells in each experiment. The estimation should be based on the developmental stage at which the cell lysis is going to be performed and the number of cells that *O. dioica* embryos are expected to have at that determined developmental stage.
- Throughout the procedure, many cells will be accidentally lost (washes, stuck to the tubes or to the tips…) and a portion of them will be intentionally taken to count the number of cells in the sample, so it's better to overestimate the number of embryos needed.
  P.e: For having 10.000 cells, ~150 64c embryos would be needed. We'll take at least 200.
  P.e: For having 10.000 cells, ~50 ITB embryos (~200c) would be needed. We'll take at least 200.

1. Transfer the embryos to an Eppendorf tube taking as less volume of SSW as possible. Let them precipitate.
   *For post-hatch larvae, use a glass Pasteur pipette to avoid losing many embryos since they get stuck to plastic pipette tips. If too much SSW is carried along and embryos don't precipitate, centrifuge the tube 30"- 60" at 600 g.
2. **Actinase-E treatment** – optional for pre-hatch embryos to weaken the chorion. Remove supernatant and resuspend the embryos in 100 µL of **Actinase-E solution** (0.05% p/v in SSW) pipetting 2-3 times. Incubate 90" at RT while the embryos precipitate to the bottom of the tube. If embryos don't precipitate fast enough, do a centrifuge pulse (~10-15") at 600 g. Remove as much supernatant as possible and quickly proceed to rinsing the embryos.

2. Cell lysis and nuclei obtain - Perform steps 4 to 9 in the 4ºC cold chamber

1. Resuspend the embryos in 1 ml of **Rinse Buffer** and immediately centrifuge the tube 1' at 600g.
2. Remove the supernatant and resuspend in 1 mL of **cold Resuspension Buffer.**
3. Centrifuge 5 min at 1500 g (=4000 rpm) at 4°C (pre-chilled) to pellet the embryos.
4. Remove 900 µL of the supernatant (100 µL left). Then aspirate carefully the remaining 100 µL with a P200 to avoid taking the cell pellet.
   *If necessary, aspirate the last 10 µL with a P10.
5. Resuspend pellet in 100 µL of **Lysis Buffer** by pipetting three times.
6. Incubate the cell lysis reaction on ice for 3 minutes.
   *For pre-hatch embryos inside the chorion, pipette softly and constantly with a

P200 checking the tube under the stereomicroscope until the solution becomes clear (normally ~2').

7. After lysis, add 1 mL of **Wash Buffer** and invert the tube 3 times to mix.
8. Centrifuge 10 min at 1500 rcf (=4000 rpm) at 4°C (pre-chilled) to pellet the nuclei. *Prepare the Transposition Reaction Mix during the centrifugation step.
9. Remove the supernatant (cytoplasm) and keep the pellet (nuclei) with two pipetting steps, as in step 4.

3. Transposition:

10. Nuclei were resuspended in 50 μL of Transposition Reaction Mix by pipetting 6 times.
11. Transposition reactions were incubated at 37ºC for 20min in a thermomixer with shaking at 1000 rpm.
12. Clean up the reaction with Zymo DNA clean and concentrator 5 column. (You can stock the DNA at -20 ºC by adding 250u DNA Binding Buffer up to 2 weeks). Elute DNA in 16 μl elution buffer and store at -20 ºC until amplification.

4. Library preparation:

Primers:

Ad1_noMX:	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19_AGGTTGGG	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT
Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

13. Partial PCR:

+25 uM Primer Ad1 2.5 ul
+25 uM Primer Ad2 2.5 ul
+2x NEBNext Master Mix 25 ul
+Transposed Sample 20 ul

Cycling Conditions

One cycle of:

72°C 5 min

98°C 30 sec

Then 5 cycles of:

98°C 10 sec

63°C 30 sec

72°C 1 min

4ºC – during the qPCR

- Take an aliquote of 5 μl of each PCR product for the qPCR and keep the rest at 4ºC.

14. qPCR:

+5 μl partial PCR

+2 μl primers UDI μM

+5 μl NEBNext High-Fidelity 2X PCR Master Mix

+0,15 μl 100x SYBR Green I

+ 2,85 μl H_2O DEPC

Vf = 15 μl

Cycling Conditions

One cycle of:

98°C 30 sec

Then 20 cycles of:

98°C 10 sec

63°C 30 sec

72°C 1 min

- To calculate the additional number of cycles needed for each library, plot linear Rn versus cycle and determine the cycle number (*N*) that corresponds to one-third of the maximum fluorescent intensity.

15. Final PCR: put the tubes from the partial PCR back in the thermocycler and run *N* number of cycles for each library.

16. Purify the final PCR reaction using a Zymo DNA Clean and Concentrator-5 Kit and elute in 20 μl H_2O DEPC.

5. Library QC

Annex 2

Extreme genome scrambling in marine planktonic *Oikopleura dioica* cryptic species

Plessy C, Mansfield MJ, Bliznina A, Masunaga A, West C, Tan Y, Liu AW, Grašič J, Del Río Pisula MS, **Sánchez-Serna G**, Fabrega-Torrus M, Ferrández-Roldán A, Roncalli V, Navratilova P, Thompson EM, Onuma T, Nishida H, Cañestro C, Luscombe NM.

Genome Res. 2024 Apr 25;34(3):426-440. doi: 10.1101/gr.278295.123. PMID: 38621828; PMCID: PMC11067885.

Extreme genome scrambling in marine planktonic *Oikopleura dioica* cryptic species

Charles Plessy,[1,9] Michael J. Mansfield,[1,9] Aleksandra Bliznina,[1,10] Aki Masunaga,[1] Charlotte West,[1,11] Yongkai Tan,[1] Andrew W. Liu,[1] Jan Grašič,[1] María Sara del Río Pisula,[1] Gaspar Sánchez-Serna,[2,3] Marc Fabrega-Torrus,[2,3] Alfonso Ferrández-Roldán,[2,3] Vittoria Roncalli,[2,3,12] Pavla Navratilova,[4,5] Eric M. Thompson,[5,6] Takeshi Onuma,[7,8] Hiroki Nishida,[8] Cristian Cañestro,[2,3] and Nicholas M. Luscombe[1]

[1]*Genomics and Regulatory Systems Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Onna-son, Okinawa 904-0495, Japan;* [2]*Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona (UB), Barcelona 08028, Spain;* [3]*Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona 08028, Spain;* [4]*Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany, 779 00 Olomouc, Czech Republic;* [5]*Sars International Centre, University of Bergen, Bergen N-5008, Norway;* [6]*Department of Biological Sciences, University of Bergen, Bergen N-5020, Norway;* [7]*Faculty of Science, Kagoshima University, Kagoshima 890-0065, Japan;* [8]*Department of Biological Sciences, Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan*

Genome structural variations within species are rare. How selective constraints preserve gene order and chromosome structure is a central question in evolutionary biology that remains unsolved. Our sequencing of several genomes of the appendicularian tunicate *Oikopleura dioica* around the globe reveals extreme genome scrambling caused by thousands of chromosomal rearrangements, although showing no obvious morphological differences between these animals. The breakpoint accumulation rate is an order of magnitude higher than in ascidian tunicates, nematodes, *Drosophila,* or mammals. Chromosome arms and sex-specific regions appear to be the primary unit of macrosynteny conservation. At the microsyntenic level, scrambling did not preserve operon structures, suggesting an absence of selective pressure to maintain them. The uncoupling of the genome scrambling with morphological conservation in *O. dioica* suggests the presence of previously unnoticed cryptic species and provides a new biological system that challenges our previous vision of speciation in which similar animals always share similar genome structures.

[Supplemental material is available for this article.]

The concept of "reference genome" for each species comes from the notion that genomic structural variations and chromosomal rearrangements within species are rare, which is a fundamental aspect sustaining projects such as the Earth Biogenome Project (EBP) (Damas et al. 2021). It is widely accepted that, in each species, the distribution and order of genes on chromosomes are not random, as changes in gene order are likely to affect the regulation of gene expression, and in humans, it has been intimately associated with a variety of diseases, including cancer (Li et al. 2020). How evolution acts on the preservation or variation of gene order within species, or even between closely related organisms, remains poorly understood.

Comparisons of distantly related groups of metazoans have revealed gene linkages within chromosomes that have been preserved for more than half a billion years (Simakov et al. 2022). The conservation of gene linkage is a feature referred to as "conserved synteny," from the Greek meaning "same ribbon," which describes homologous genes that colocate, independently of order, within a single chromosome (Passarge et al. 1999). Differences in the scale and extent of synteny conservation have led to the concepts of micro- and macrosynteny. Microsynteny (also known as "collinearity" in genomics) refers to the conservation of gene content and order within sets of tightly linked orthologous genes. Generally, closely related species tend to possess greater conservation of microsynteny, and for this reason, it can even be used to clarify phylogenies (Drillon et al. 2020; Pereira-Santana et al. 2020). Although microsynteny is generally weakly conserved in distantly related species, the remnants of ancient linkage karyotype groups can be detected at the chromosome scale; the conservation of genes on chromosomes that can be traced back to an ancestral karyotype is reflected in the concept of macrosynteny, examples of which include the chromosomal conservation that can be traced back to the last common ancestor of metazoans (Simakov et al. 2022). The most famous example of

[9]These authors contributed equally to this work.
Present addresses: [10]Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK; [11]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge CB10 1SD, UK; [12]Integrative Marine Ecology Department, Stazione Zoologica Anton Dohrn, 80121 Naples, Italy
Corresponding authors: charles.plessy@oist.jp, canestro@ub.edu
Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.278295.123. Freely available online through the *Genome Research* Open Access option.

34:426–440 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/24; www.genome.org

conserved microsynteny in animals is the *Hox* cluster, which contains genes that regulate axial patterning during embryogenesis and whose ancestry can be traced back to the origin of bilaterian animals hundreds of millions of years ago (for a recent review, see Wanninger 2024). There are many other examples of highly conserved microsyntenies across metazoans, in many cases related to the functional constraints imposed by *cis*-regulatory elements on the coordinated transcription of nearby genes. This includes genomic regulatory blocks (GRBs), within which the action of conserved noncoding elements allows the coordinated expression of genes in a local genomic neighborhood (Hurst et al. 2004; Engström et al. 2007; Irimia et al. 2012; Rowley and Corces 2018). Thus, evaluating the conservation and loss of synteny can provide important information for generating testable hypotheses related to gene regulation, genome biology, and evolution.

The loss of synteny can be provoked by genome rearrangements, such as chromosome translocations related to unequal recombination, or by chromosome fragment mobilization owing to transposon activity. Both of these processes can result in changes of gene order and the reallocation of genes to different genomic neighborhoods. The accumulation over time of many rearrangements results in genome "scrambling," a concept that in linguistics refers to language syntaxes that permit changes in word order without altering the meaning of a sentence. Scrambling has been used to describe the patterns of synteny loss in genomic comparisons of distantly related species, such as fugu and humans, whose genome organization has significantly diverged over hundreds of millions of years (Aparicio et al. 2002). However, fundamental questions remain, such as how evolutionary forces act to constrain or accelerate the rate of rearrangement or how phenotypic differences could be related to rearrangements. Addressing these problems is difficult at large time scales and genetic distances.

Chromosomes and chromosomal rearrangements have been a classic topic of discussion in evolutionary biology, in support of Darwin's theory of natural selection and the origin of species (Darwin 1859; Dobzhansky 1937; Goldschmidt 1940), for the enormous potential that these chromosomal changes can become raw material for evolution, enabling populations to quickly isolate from each other and facilitating the rapid evolution of adaptations to sudden environmental variations. The description of chromosomal rearrangements associated with adaptive phenotypes (Joron et al. 2011; Lamichhaney et al. 2016) has brought renewed attention to this area, especially given the advent of sequence technologies that enable high-quality, telomere-to-telomere, chromosome-scale genome assemblies across the tree of life, such as those produced under the umbrella of the EBP (for review, see Damas et al. 2021). Lepidopterans provide one of the most enigmatic examples of how genome rearrangements are responsible for the speciation and adaptations such as mimicry (Joron et al. 2011; Hill et al. 2019; de Vos et al. 2020), mostly owing to reciprocal translocation and repeated events of fusion and fission among a highly dynamic number of chromosomes between species.

To better understand the phenomenon of genome scrambling, we study the zooplanktonic appendicularian tunicate *Oikopleura dioica*. *O. dioica* has the smallest nonparasitic animal genome reported to date (Seo et al. 2004; Denoeud et al. 2010; Wang et al. 2020a; Bliznina et al. 2021). This genome reduction appears to be the result of a drastic process of compaction involving a reduction in repeat content (~15%) (Henriet et al. 2015), as well as numerous gene losses (for a review, see Ferrández-Roldán et al. 2021). *O. dioica*'s karyotype comprises three chromosome pairs (Körner 1952; Liu et al. 2020): two acrocentric autosomes and an acrocentric X and Y sex chromosome containing a long pseudoautosomal region (PAR) connected to sex-specific regions by a ribosomal DNA locus (Denoeud et al. 2010; Bliznina et al. 2021). The Y-specific region is repeat-rich and gene-poor and differs from all other genomic regions. Chromosome contact analysis of *O. dioica* suggests that there is relatively little interaction between the arms of individual chromosomes or sex-specific regions, which corresponds with the "type-I" genome architecture reported by Hoencamp et al. (2021). In *O. dioica*, a significant fraction of genes is densely packed in a head-to-tail configuration and transcribed in polycistronic mRNAs, forming operons, which are processed by the addition of a *trans*-spliced leader RNA (Ganot et al. 2004), similar to the operons seen in other eukaryotic taxa (Van der Ploeg 1986; Stover and Steele 2001; Blumenthal and Gleason 2003; Zayas et al. 2005; Zhang et al. 2007; Zeller 2010). In contrast to bacterial operons, in which cotranscribed genes tend to be functionally related, in *O. dioica* the functions of genes in operons are more loosely related, with a trend toward housekeeping, cell cycle, translation, and germline functions (Zeller 2010; Danks et al. 2015; Wang et al. 2015). How operons might relate to genome scrambling is not known. At the same time, genome compaction in *O. dioica* also appears to have been accompanied by a drastic loss of conserved microsynteny compared with other chordate genomes, including the disintegration of the paradigmatic *Hox* cluster (Seo et al. 2004).

Our recent study of cross-fertility and molecular markers revealed that *O. dioica* sampled from the Japanese Seto inland sea (Osaka University laboratory strain), from the subtropical island of Okinawa, Japan (OIST laboratory strain), and from the Mediterranean Catalan coast (University of Barcelona laboratory strain) (Fig. 1A) were reproductively isolated and showed high genetic distance. Despite this, we were unable to identify reliable morphological characteristics that could be used in the field to distinguish these samples without the need for crossing experiments or DNA sequencing, suggesting that these are cryptic species (Masunaga et al. 2022). A collegial discussion on a possible taxonomy update is currently taking place in the tunicate scientific community, and to facilitate a consensual conclusion, as well as for the sake of simplicity in this paper, we will refer to these populations as "lineages" named after the location in which specimens have been collected or from which laboratory cultures have originated. The telomere-to-telomere genome assembly of *O. dioica* from Okinawa (Bliznina et al. 2021) further implied the existence of differences in gene organization compared with genome sequences obtained from individuals sampled in Osaka (Wang et al. 2020a) and Bergen (Norway) (Denoeud et al. 2010), but this possibility could not be investigated properly without chromosome-scale assemblies for all lineages. Worse, the apparent level of synteny compared with chromosomes from the Okinawa lineage was variable between contigs. Here, using chromosome-scale genome assemblies, we report a substantial degree of genomic rearrangement between *O. dioica* lineages, describing the genomic features that underlie this genomic scrambling and laying the foundations toward making *O. dioica* an attractive system to study the loss of conserved synteny in the absence of obvious phenotypic differences.

Results

Pan-oceanic genome assemblies of *O. dioica*

We generated chromosome-scale genome assemblies of *O. dioica* specimens from Barcelona (BAR) and Osaka (OSA), which were

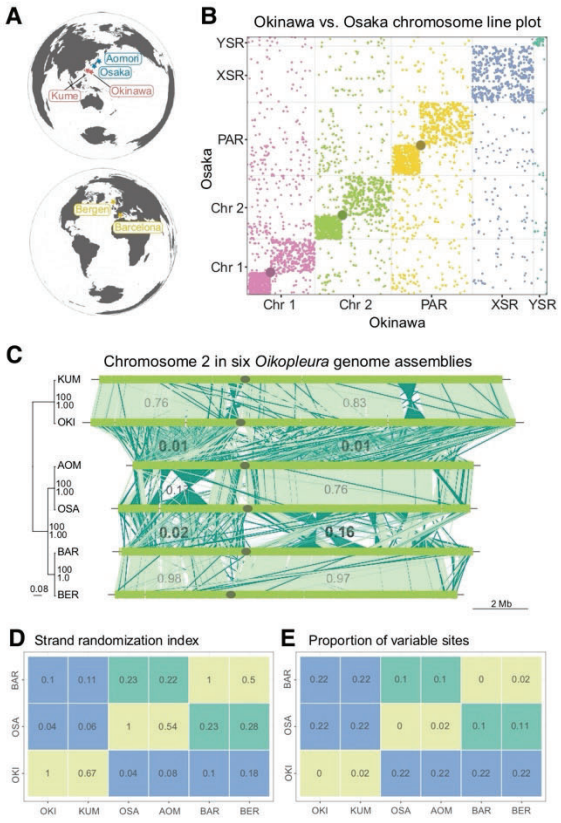


Figure 1. Extensive genomic rearrangement in *Oikopleura dioica* lineages. (*A*) Geographical map locating the origin of the lineages and assemblies. (*B*) Line plot representation of the whole-genome alignment between the Okinawa and Osaka genomes. Each chromosome is plotted in a different color that will identify them in the following figures. A gray dot is overlaid at the position of centromeres. (*C*, *left*) Maximum likelihood phylogenetic tree of 5162 single-copy orthologous genes common to all *O. dioica* genomes. The tree is midpoint-rooted with clade support values indicating bootstrap values from RAxML and Bayesian posterior probability from MrBayes. Branch lengths are proportional to the estimated number of substitutions per nucleotide site. (*Right*) Pairwise comparisons of Chromosome 2 between *Oikopleura* genomes (names abridged with their first three letters). Dark green indicates plus/minus-strand alignments; the gray ellipse, the centromere. The numbers indicate the scrambling index computed for a given arm pair. (*D*,*E*) Scrambling index (*D*) and proportion of variable sites (*E*) across all single-copy orthologous nucleotide sequences for same-lineage (green), Osaka–Barcelona (yellow), and Okinawa–other (red) pairs of genomes.

added to our published assembly from an individual from Okinawa (OKI) (Table 1). We validated each reference assembly using an additional contig-level assembly chosen or generated according to sequence similarity (respectively, Bergen [BER], Aomori [AOM], and Kume [KUM]) (Table 1; Fig. 1A). The animals from Kume were cross-fertilized with the Okinawa laboratory strain, and all three pairs displayed similarity scores (see below) in the same order of magnitude. Together with geographical proximity and the results presented below, we refer to each pair of cross-validating assemblies as belonging to the same lineage or clade. We assembled the Barcelona genome using a similar procedure to the Okinawan genome, including the use of chromosome conformation information (Hi-C libraries) to aid scaffolding. A Hi-C contact map (Supplemental Fig. S1) showed that the chromosome arms and the sex-specific regions had few interactions with each other, and the assembly graph connected the sex-specific regions to the PAR's long arm through ribosomal DNA repeats. Moreover, we have constructed a new Osaka genome assembly by scaffolding the OSKA2016 assembly (Wang et al. 2020a) with long Nanopore reads that we sequenced from single individuals from the same laboratory strain. To ensure consistency (Weisman et al. 2022), we generated updated annotations for all genomes using a common automated pipeline, including repeat masking and gene prediction steps, which provide a robust set of annotations that facilitate inter-species comparisons (Table 1).

The scrambled genomes of *O. dioica*

To investigate the evolution of the chromosomes in *O. dioica*, we developed a reproducible, standardized pipeline to compute the optimal set of one-to-one local alignments in a pair of genomes using the LAST software (Frith and Kawaguchi 2015; Mitsuhashi et al. 2020) and the Nextflow workflow system (Di Tommaso et al. 2017). The all-by-all pairwise genome alignments revealed an unexpected level of genomic rearrangement, and the most extreme case of scrambling was observed in the OKI–OSA (Fig. 1B) and OKI–BAR (Supplemental Fig. S2) comparisons. The line plot comparing the whole-genome sequences of *O. dioica* from OSA and OKI revealed a striking pattern, with little to no conservation of collinear DNA segments on any chromosome (Fig. 1B). Multichromosome line plots comparing *O. dioica* from all three lineages further revealed that the genome scrambling phenomenon was common among all compared genomes (Fig. 1C). In general, the extent of genome scrambling was proportional to and increased with genetic distance (Fig. 1C). Within-lineage comparisons showed little scrambling, with large, intact collinear segments of DNA visible (Fig. 1C; Supplemental Fig. S2). These observations suggested that genome scrambling was therefore a common evolutionary characteristic in *O. dioica* genomes.

To quantify the degree of scrambling between any pair of genomes and to determine how scrambling might relate to other measures of genetic distance, we created a "scrambling index," which measures the degree of strand randomization and, thus, the loss of collinearity between aligned regions. A scrambling index value approaching one indicates that most aligned bases have the same orientation (i.e., plus-to-plus or minus-to-minus); scrambling index values approaching zero indicate that either alignment orientation is equally frequent (i.e., plus-to-minus and vice-versa) (Fig. 1C). Computation of the scrambling index for each genome pair (Fig. 1D) yielded high values for within-lineage comparisons, allowing us to rule out technological biases introduced by different sequencing technologies (Table 1). The smallest scrambling indices were obtained for comparisons of the Okinawa lineage to other lineages (Fig. 1D). Comparisons between the Osaka and Barcelona lineages also yielded intermediate scrambling index values (near 0.2), which was congruent with the intermediate degree of scrambling observed in line plot comparisons (Fig. 1C). Each pair's scrambling index value was proportional to the proportion of variable sites among the single-copy ortholog nucleotide sequences (Fig. 1E).

Impact of genome scrambling on macrosynteny conservation in *O. dioica*

Line plots between Osaka and Okinawa showed that ~94% of all rearrangements were intra-chromosomal, whereas inter-chromosomal rearrangements were rare (Fig. 1B). Within each chromosome, rearrangements tended to occur within arms or the sex-

Table 1. Details on each of the assemblies used in this study

Name	Version	Region	Lineage name	Breeding	Length	Scaffold N50	AT-richness	Genes	BUSCO score	Technology	Level	BioSample	Reference
Okinawa	OKI2018_I69_1.0	Okinawa, Ryūkyū archipelago	Okinawa	Laboratory	64,281,565	17,092,476	59%	17,291	64%	Nanopore + Illumina + HiC	Chromosomal	SAMEA7282646	Bliznina et al. 2021
Kume	KUM-M3-7f	Kume, Ryūkyū archipelago	Okinawa	Wild	64,653,574	2,969,719	59%	16,852	66%	Nanopore	Scaffold	SAMEA111279290	This paper
Osaka	OSKA2016v1.9	Honshū, Japan inland sea	Osaka	Laboratory	56,625,162	15,521,227	58%	15,720	63%	PacBio + Illumina	Chromosomal	SAMD00227923	Wang et al. 2020a; this paper
Aomori	AOM-5-5f	Honshū, Japan, northeast Pacific coast	Osaka	Wild	56,753,784	6,419,763	59%	15,224	65%	Nanopore	Scaffold	SAMEA111279288	This paper
Barcelona	Bar2_p4	Spain, Mediterranean sea coast	Barcelona	Laboratory	55,793,437	13,545,857	60%	14,272	64%	Nanopore + Illumina + HiC	Chromosomal	SAMEA111279286	This paper
Bergen	OdB3	Norway, north Atlantic coast	Barcelona	Laboratory	70,471,451	395,387	57%	17,113	60%	Sanger	Scaffold	SAMEA2272014	Denoeud et al. 2010

(Name) Name of the genome reflecting provenance; (version) character string uniquely identifying the used version of the assembly file; (region) where the animal or the founders of its laboratory line was isolated; (breeding) wild isolate or laboratory culture line; (length) assembly sequence length in base pairs; (scaffold N50) half of the genome is covered by scaffolds of at least this length; (AT richness) a percentage of A's and T's in the sequence; (genes) number of genes in the annotation used in this study; (BUSCO score) detected percentage of the metazoan set of Benchmark Unique Single-Copy Orthologs; (technology) sequencing and assembly technologies used; (level) chromosomal or scaffold; (BioSample) identifier in the DDBJ, EBI, and NCBI databases; and (reference) peer-reviewed manuscripts describing the already published genome sequences.

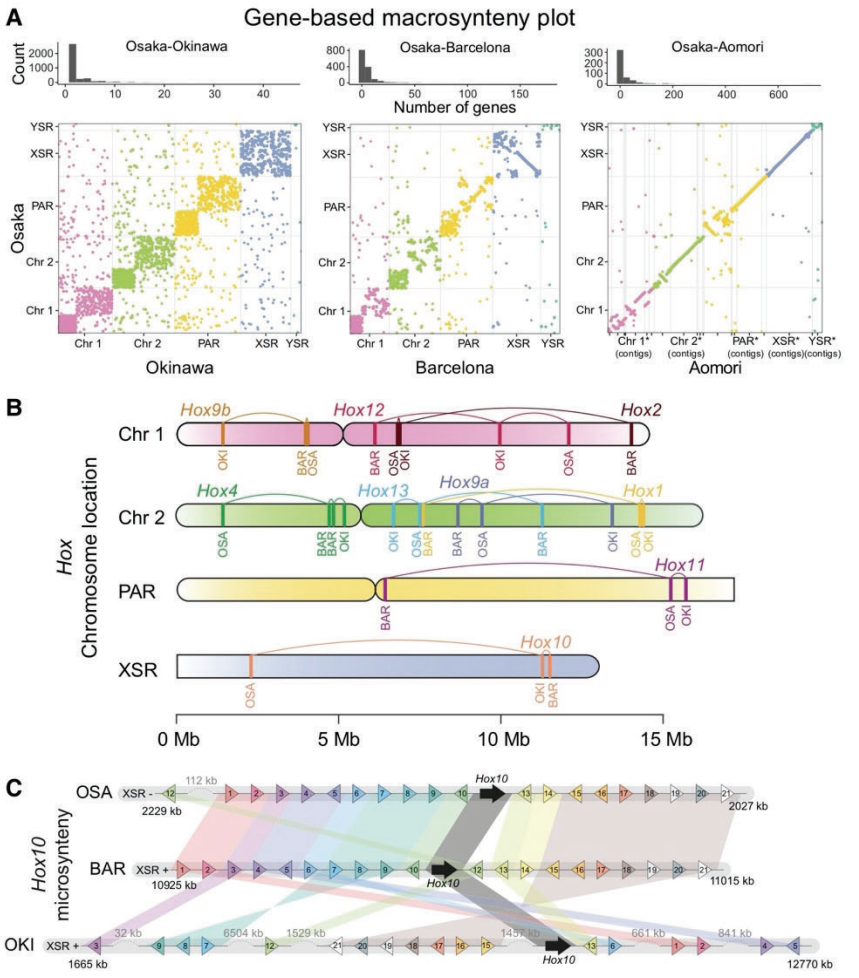


Figure 2. The preservation of orthologous synteny blocks gradually decreases with increasing evolutionary distance in *O. dioica*. (*A*, *top*) Histogram of the number of orthologous genes per syntenic region in pairs of genomes. (*Bottom*) Dot plots indicating the coordinates of genes belonging to the same orthogroup in pairs of genomes. (*B*) Comparative chromosome mapping of the *Hox* genes in the genomes of *O. dioica* from Osaka (OSA), Barcelona (BAR), and Okinawa (OKI). (*C*) Comparative microsynteny conservation of the block of the next 10 genes at each side of the *Hox10* genes in the same three genomes.

specific regions (~99%), which for the sake of simplicity we will also refer to as "arms." To investigate the impact of genome scrambling on the evolution of synteny blocks, we compared gene-order conservation across lineages. We computed 5162 groups of single-copy orthologs present in the six genomes, and visualized them with strand-independent macrosynteny dot plots, which showed the positions of the same gene in a pair of genomes (Fig. 2A; Supplemental Fig. S3). This confirmed that gene-order rearrangements were mostly restricted to homologous arms (Fig. 2A), and confirmed that some inter-chromosomal translocations observed at the whole-genome level involved whole-gene translocations. The number of orthologs per synteny block decreased with increasing genetic distance, with a maximum of 44 for Osaka versus Okinawa, a maximum of 174 for Osaka versus Barcelona, and a maximum of 714 for Osaka versus Aomori (Fig. 2A; Supplemental Fig. S3).

As case studies, we next mapped the chromosomal locations of several genes associated with characteristic gene clusters (*Hox*, *Fgf*, and *Myosins*) (Fig. 2B; Supplemental Figs. S4, S5). Microsynteny conservation of the *Hox* cluster has been shown to be essential for embryonic development and axial patterning in vertebrates, but the disintegration of the *Hox* cluster in *O. dioica* from Bergen suggested that it may not be essential in *O. dioica* (Seo et al. 2004). We mapped all *Hox* genes of *O. dioica* in all our assembled genomes, corroborating that the full catalog of *Hox* genes in *O. dioica* is reduced to six genes from which all central *Hox* genes (*Hox3* to *Hox7*) have been lost. Comparison of the position of *Hox* orthologs revealed that multiple changes in gene order must have occurred, although all have been maintained within the same chromosome arm. Mapping of all six Fgf genes previously reported in *O. dioica* (Oulion et al. 2012) revealed a similar pattern of gene movement (Supplemental Fig. S4). On the other hand, chromosome mapping of the eight myosin heavy chain class II genes presented a different pattern, whereby orthologs occasionally seemed to move more freely, including inter-arm and inter-chromosomal translocations (Supplemental Fig. S4). We also inspected patterns of microsynteny in these gene families by examining their 10 nearest neighboring genes up- and downstream (Fig. 2C; Supplemental Fig. S5). In general, gene families in Barcelona and Osaka showed far greater conservation of microsynteny with each other than either does with Okinawa. These examples revealed different degrees of microsynteny conservation, ranging from near-complete conservation of entire blocks (e.g., *MyhF*, *MyhG*, *Fgf11/12/13/14a*, *Fgf11/12/13/14b*, and *Hox1*) to situations in which a block has seemingly fragmented into many small pieces (e.g., *Fgf9/16/20a* and *Hox10* or *Hox12*). Based on our examination of chromosome mapping for different conserved gene families in *O. dioica*, the position of a gene in one lineage had little predictive power for the position or orientation of that gene in other lineages.

Genome scrambling moves short functional regions

We next identified "breakpoint regions" to search for the molecular breakpoints responsible for scrambling synteny blocks. First, we identified collinear alignments, defining them as adjacent alignments in the same orientation in both genomes. We termed the regions flanked by these collinear alignments "bridge regions." We then defined "collinear regions" as successions of collinear alignments and bridge regions. The "breakpoint regions" were therefore the remaining unaligned regions, for which there was no one-to-one correspondence in a pair of genomes, and always correspond to an interruption of collinearity. Lastly, we termed aligned regions that were not collinear to anything as "isolated alignments"

(Fig. 3A). Although breakpoint regions tended to be short (0.32 ± 5.1 kbp, $n = 8821$ for the Okinawa–Osaka comparison), they covered a considerable fraction of the genome (~23.5%) (Fig. 3B). Three reasons may explain the lack of alignability in breakpoint regions: (1) so many mutations accrued in these regions that they exceed the limits of detectable sequence similarity; (2) repeats were the target or the cause of the breaks; or (3) the mechanism involved the loss of DNA.

To determine how the phenomenon of scrambling related to functional genomic regions, we studied the frequencies of coding and conserved noncoding elements (a proxy for potential regulatory regions) (Tan et al. 2019) and repeats at the boundaries of the four nonoverlapping classes of genome segments (Fig. 3C). The alignments' boundaries with breakpoint regions tended to coincide with exon start positions, as well as with intron stop positions to a lesser extent (Fig. 3C). Isolated alignments were less frequently part of operons. In terms of noncoding elements, repeats were depleted in isolated alignments, whereas conserved noncoding elements were enriched, with a peak downstream from the alignment start position, consistent with previously reported patterns of erosion (Royo et al. 2011). Breakpoint regions were the least likely to be found within genes. Bridge regions occurred mainly in genic regions, with strong enrichment for introns, which is consistent with the high intron turnover reported earlier (Edvardsen et al. 2004; Denoeud et al. 2010), and repeats (which may be intronic) upstream of collinear alignments; bridge regions were also most frequently associated with operons. Altogether, the most marked changes in the frequency of genomic elements between classes were related to the frequency of protein-coding features, with the exception of operons, which showed modest changes in frequency at the edges of aligned and breakpoint or bridge regions.

Genome scrambling does not preserve operon structure

We next assessed the conservation of operons within the chromosome-scale assemblies of *O. dioica*. Operons may impose some limitations to rearrangements in synteny blocks. For example, a single-gene inversion in the middle of a three-gene operon could result in expression defects by decoupling that gene from its primary regulatory elements. The number of operons per chromosome-scale assembly ranged between 2379 and 3124, representing between 6653 and 9543 operonic genes (Fig. 4A). Only a small number of operons preserved homologous genes across Okinawa, Osaka, and Barcelona (Fig. 4A), and this finding remained true when considering alternative criteria for operon equivalence (Supplemental Fig. S6). Among protein-coding genetic elements—operons, genes, and exons—operons were the most likely to overlap breakpoint regions. In the Okinawa–Osaka genome pair, 616 out of 1281 operons overlapped a breakpoint (48%), whereas 5294 out of 17,291 genes (30%) and 16,787 out of 106,811 exons (15%) overlapped one (Fig. 4B). Further, large and small operons were both affected by scrambling (Fig. 4C). Detailed comparison of operon microsynteny revealed examples of operons with complete conservation located on the same chromosome for Okinawa, Osaka, and Barcelona (Fig. 4F). Other examples showed the conservation of an operon following a translocation of some operonic genes to a new location (Fig. 4D–E). In some cases, an operon rearrangement involved duplication and translocation of a large portion of an operon into a new chromosome (Fig. 4G). Although operons were rarely conserved between lineages in general, operonic genes from one genome were significantly more likely to be operonic in a second genome across all within-lineage pairs ($P \ll 0.001$, chi-squared $\gg 4420.2$, d.f. = 3). Overall, our data revealed an absence of strong selective constraints to strictly maintain operon structure between lineages, suggesting operons are prone to be impacted by genome scrambling.

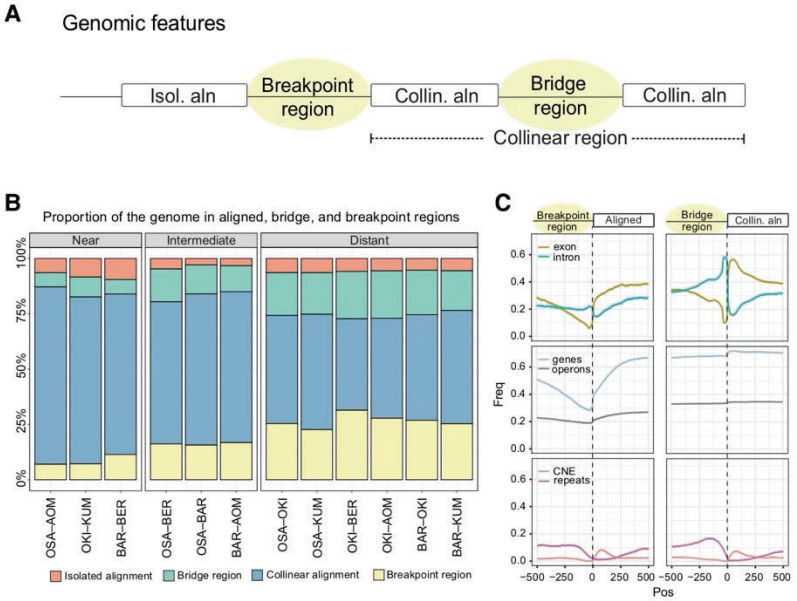


Figure 3. Properties of genomic alignments. (*A*) We divided the aligned and unaligned regions of the genome into four categories according to their participation in collinear regions. Collinear regions are defined as an uninterrupted succession of alignments that are on the same chromosome strand and in the same order in both genomes. (*B*) Proportion of the four categories in different alignment pairs, grouped by evolutionary distance. (*C*) Enrichment of genomic features at the boundary between breakpoint or bridge regions and aligned regions in the Okinawa–Osaka comparison (for other pairs, see Supplemental Fig. S7). (CNE) Conserved noncoding elements.

Genome scrambling and chromosomal evolution

Given that our data pointed to chromosome arms as the primary scale of macrosynteny conservation, to better understand chromosomal evolution in *O. dioica*, we investigated the distribution of breakpoint regions, operon sizes, and mutation rate at the chromosome level (Fig. 5A–D). Our analysis revealed that short chromosome arms consistently showed four different qualities compared with long chromosome arms: (1) short arms showed a higher relative frequency of breakpoints; (2) short arms contained shorter genes and shorter operons; (3) genes on short arms overlapped breakpoint regions at a higher rate (~50% vs. ~20%; $P \ll 0.001$, chi-squared $\gg 109.6$, d.f. = 2); and (4) genes on short arms showed elevated d_N/d_S values. In all cases, the XSR showed patterns comparable to long arms. Our analysis also revealed that these features also consistently varied across chromosome arms, differing between the centers of chromosome arms and subtelomeric or

pericentromeric regions. As reported for the Okinawa genome (Bliznina et al. 2021), repeat density increased whereas gene and operon density decreased in subtelomeric and pericentromeric regions for the Osaka and Barcelona genomes (Supplemental Fig. S8). The co-occurrence in short arms of an increase in repeat content, a greater frequency of breakpoint regions, and elevated d_N/d_S values together implied that repeat-related rearrangements could play a role in generating or maintaining structural variations that yielded nonrecombining loci, leading to more rapid accumulation of point mutations and substitutions.

Evolutionary framework of the unprecedented genome scrambling in *O. dioica*

To relate the rate of scrambling to evolutionary distance, we estimated a species tree and divergence times for *O. dioica* using orthologs common to chordates (Fig. 6; Supplemental Table S1). We used relatively unconstrained priors for the nodes within the appendicularians owing to a lack of supporting fossil evidence. As well, because *O. dioica* is among the fastest-evolving animals known (Berná et al. 2012), we attempted to reduce the effects of heterotachy through careful ortholog selection, curation, alignment trimming, and comparison of multiple replicates (Supplemental Table S1). The resulting phylogenetic tree supported the existence of at least three independent lineages of *O. dioica*, which were estimated to have shared a last common ancestor about 25 million years ago (Mya). This split represented the divergence between the Okinawa lineage and other lineages, and a more recent divergence time of ~7.3 Mya was estimated for the split between the Osaka and Barcelona lineages.

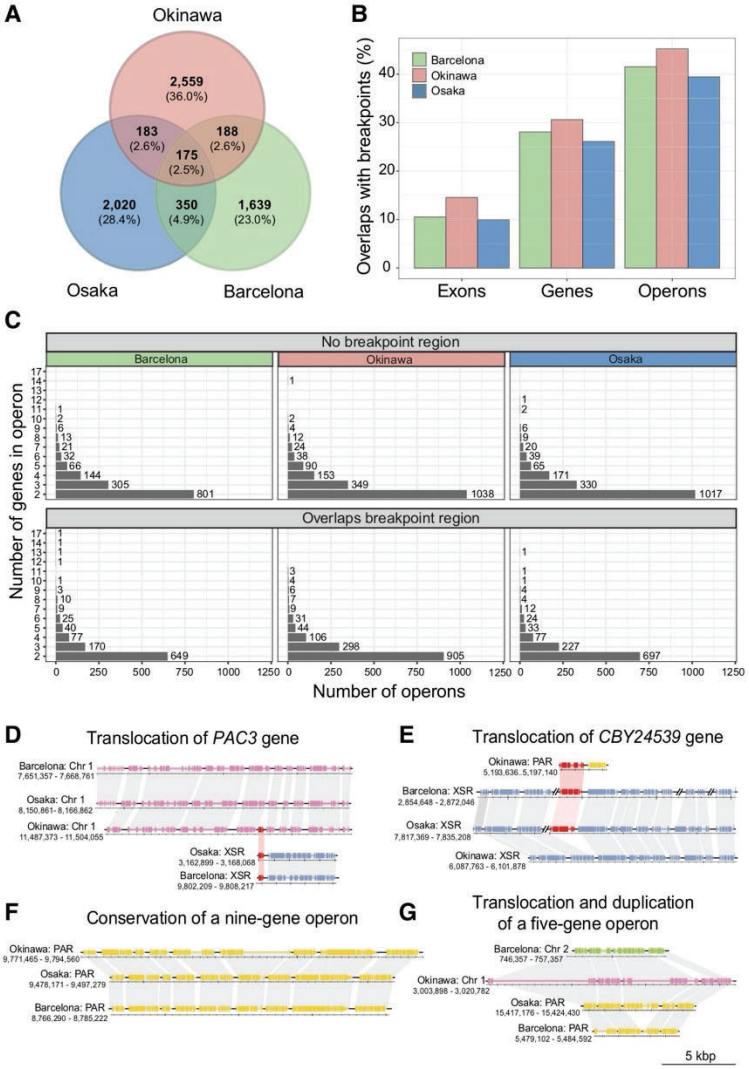


Figure 4. Conservation of operons in *O. dioica* lineages using the chromosome-level genomes as representatives. (*A*) Number of shared and unique operons across the chromosome assemblies representing each lineage. (*B*) The proportion of protein-coding genetic elements that overlap a breakpoint region for each genome. (C) Size distribution of operons that overlap or do not overlap a breakpoint region. (*D*,*E*) Translocation of the genes *PAC3* (*D*) and *CBY24539* (*E*; putative activin type I receptor), belonging to different operons in Okinawa lineages and the other lineages. (*F*) The nine-gene operon reported by Ganot et al. (2004) is conserved in Osaka, Barcelona, and Okinawa. (*G*) An example of an operon that has been translocated to different chromosomes in each species and duplicated in the Barcelona genome.

Using these divergence time estimates, we calculated that the breakpoint accumulation for *O. dioica* lies between six and 25 breakpoint regions per megabase pair per million years (Fig. 7A,E). To better contextualize this result, we estimated the same value for comparisons of ascidian tunicates using two isolates of *Ciona intestinalis*, *Ciona robusta* (differing only by pigmentation) (Caputi et al. 2007) and *Ciona savignyi* (a known example of scrambling on a long divergence time of ~100 Mya) (Fig. 7B; Satou et al. 2019), and found that breakpoint accumulation in *O. dioica* is up to an order of magnitude higher (Fig. 7E; Supplemental Fig. S9). This is also several orders of magnitude greater than the reported rate for comparisons of mammals (Damas et al. 2022); using our pipeline, the number of breakpoints between *Pan troglodytes* and *Bos taurus* yielded approximately 0.7 breakpoint regions per megabase pair per million years, based on a conservative divergence estimate of 62 Mya (Delsuc et al. 2018). A similar figure (about 0.6) was found comparing the karyotype-derived muntjac deer *Muntiacus muntjak* to its close relative *Muntiacus reevesi* (~5 Mya divergence) (Supplemental Fig. S1, line plot; Mudd et al. 2020). Further, to relate our results to other invertebrates with short generation times, we also computed these values for near, intermediate, and distantly related species of *Drosophila*, in which scrambling was reported earlier (Fig. 7C; Suvorov et al. 2022), and *Caenorhabditis*, which also contain *trans*-spliced operons (Fig. 7D). Importantly, between the effects of heterotachy, the potential for ortholog misidentification, and misalignment, the divergence time estimates for the splits between the *O. dioica* lineages were more likely to be overestimated than underestimated, in

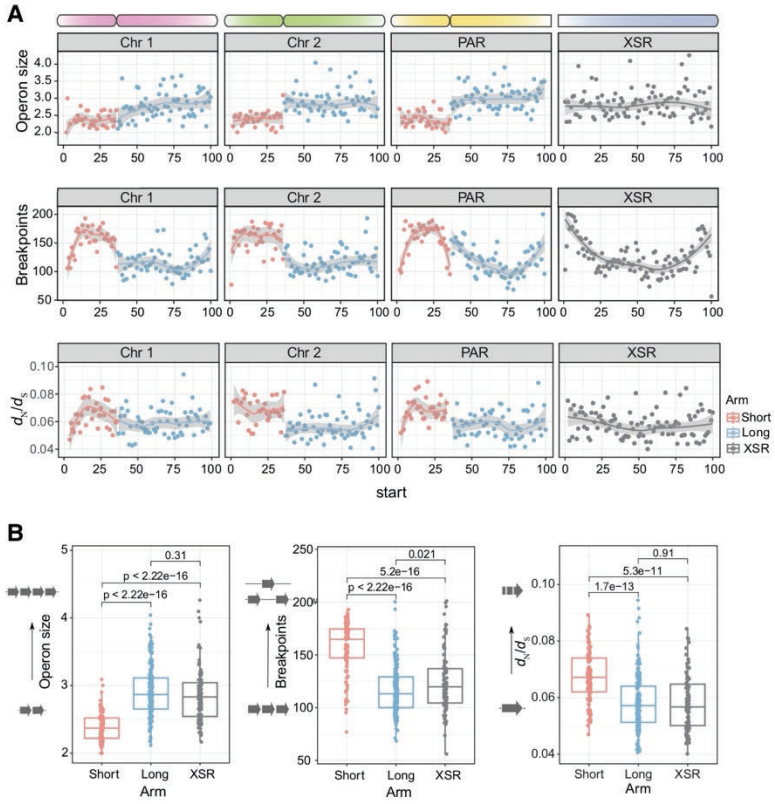


Figure 5. Genome-wide patterns of genomic feature density. (*A*) The mean values for various genomic features (*y*-axis) versus chromosomal location by the percentage of each chromosome's length (*x*-axis). Each bin is the average of the three chromosomal assemblies representing each lineage. Two regions of the chromosomes show characteristic differences in feature distribution: The first difference can be seen between short and long chromosome arms, and the second difference is between the centers and edges of chromosome arms. (*B*) Short and long arms show significant differences in operon size, the number of breakpoint regions, and d_N/d_S ratios (Wilcoxon rank-sum test).

which case the rate of chromosomal rearrangements would be even greater than we have computed. In conclusion, based on these metrics, all *O. dioica* lineages showed a distinctly greater rate of scrambling than any other group of animals (Fig. 7E).

Discussion

Genome scrambling and speciation

Our study design, combined with the recent divergence times estimated for *O. dioica*, allowed us to study genome scrambling on a finer timescale than was reached by previous studies. Despite the relatively small evolutionary distances between the *O. dioica* lineages used here, we discovered thousands of breakpoints, which may be an order of magnitude higher than other ascidian tunicates or flies with similar divergence times and may be more comparable to species pairs that diverged hundreds of millions of years ago (Fig. 7; *Drosophila* 12 Genomes Consortium 2007; Hane et al. 2011; Albertin et al. 2022; Damas et al. 2022). The phylogeny we estimated suggests that the three lineages (KUM + OKI, AOM + OSA, BAR + BER) may correspond to three distinct cryptic species, which was corroborated by our analyses of marker genes (Masunaga et al. 2022), but we ask the reader to refrain from using their names as species names to let the taxonomical discussion proceed to an optimal solution. Our molecular clock analysis suggests that the Osaka and Barcelona lineages shared a common ancestor ~7 Mya, which diverged from the Okinawa lineage >20 Mya. The genetic and environmental factors that might have contributed to cryptic speciation in this clade are unknown. It is tempting to speculate that the extreme rate of rearrangement in *O. dioica* could accelerate sympatric speciation through the formation of reproductively incompatible subpopulations within an area, even in marine environments lacking physical geographic boundaries. Further extensive sequencing of *O. dioica* genomes around the globe and surveys of intrapopulation genetic variation are needed to validate this hypothesis. Sampling other appendicularian species will be necessary to explore if genome scrambling is present beyond *O. dioica*, as it could be a hallmark of the evolution of appendicularian genomes, perhaps related to the high rate of gene loss in the clade (Ferrández-Roldán et al. 2021).

The case of *O. dioica* has two qualities that run contrary to typical biological intuitions: The similarity of morphology would not have predicted significant differences in genome structure, and the significant differences of genome structure would not have predicted similarity of morphology. As such, it shows that genetic distances and even taxonomic ranks may be insufficient to predict the amount of information that could be gained by sequencing a given organism's genome, which is particularly relevant for large-scale genome sequencing projects. Conversely, it provides a clear example of a case in which substantial differences in genome structure do not result in easily determined distinguishing characteristics (synapomorphies) that could be useful for taxonomic purposes. Although *O. dioica* may represent a unique challenge (and opportunity) for taxonomists and evolutionary biologists, we believe that difficulties in understanding the relationship between genome conservation and what defines a biological species will become increasingly common in the postgenomic era, as clades across the tree of life continue to be sequenced. Thus, our results serve as a reminder that translating results between different scientific fields cannot solely rely on raw data but requires interdisciplinary cooperation and expertise.

Mechanisms of genome scrambling and impact on gene regulation

Between the high divergence times between the *O. dioica* lineages, the elevated rate of evolution in *O. dioica*, and the complex nature of the identified rearrangements, we were unable to identify precise molecular breakpoints that could be explained by a simple and specific genetic mechanism. Future comparisons between less distant *O. dioica* lineages, or even within populations, might help us to better understand the mechanisms responsible for this

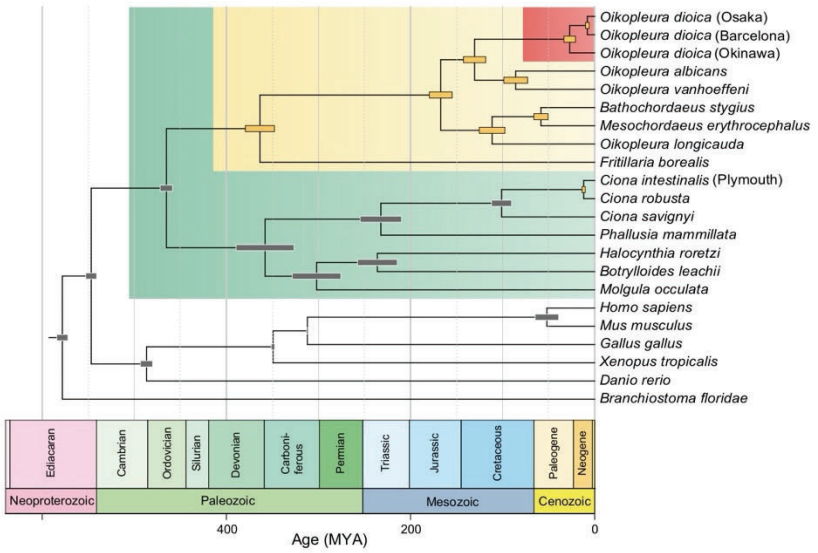


Figure 6. Time-scaled phylogenetic tree including several appendicularians, tunicates, and vertebrates, including *O. dioica*, based on 177 single-copy orthologous protein sequences. The different clades of *O. dioica* lineages were estimated to have shared a common ancestor ~25 Mya (95% HPD: 12–41 Mya). The Osaka and Barcelona lineages were estimated to have diverged more recently, ~7 Mya (95% HPD: 5–10 Mya).

massive genome scrambling. It is tempting, though, to speculate that the loss of the canonical nonhomologous end-joining (NHEJ) DNA repair pathway in *O. dioica* might have created synergies that act to promote scrambling. For instance, the alternative microhomology-based pathway (MMEJ), which was shown to be active in experimentally induced lesions in *O. dioica* (Deng et al. 2018), is slower than other repair mechanisms (Fu et al. 2021), which might allow for greater chromatin movement to occur before the repair of a double-stranded break. Cut-and-paste transposons that use the MMEJ pathway may also act as a source of microhomologies that could facilitate repair by MMEJ. The low repeat content of *O. dioica* genomes might therefore be a reflection of genomic instability that also causes scrambling. Although *O. dioica* genomes seem to be repeat-sparse, a relatively small number of interspersed repeats is sufficient to facilitate rearrangements through repair mechanisms such as homologous recombination. Scrambling in *O. dioica* seems to correlate with phylogenetic distance and divergence time. Parsimoniously, the mechanisms underlying scrambling are more likely to involve the gradual and ongoing accumulation of rearrangements rather than the result of one or more dramatic lineage-specific rearrangement events.

The genome of *O. dioica* is not as well annotated as those of humans or mice, and the significant genomic rearrangements between different lineages of *O. dioica* complicate the comparison of epigenomic or transcriptomic data across these lineages. Consequently, our knowledge of elements like enhancers, promoters, or topologically associating domains (TADs) and their similarities across lineages is too limited for practical use in current research. This issue is being addressed in ongoing projects that are generating data from various laboratory strains concurrently. Despite these challenges, operons can still be inferred through the proximity of predicted coding sequences, allowing their use in this study.

The operon structures that control the transcription of neighboring genes in *O. dioica* are rarely identically conserved between lineages. Two properties of *O. dioica* operons could be related to this observation: (1) the expression levels of operonic genes are not strongly correlated (Danks et al. 2015), and (2) the functional categories of operonic genes are not necessarily correlated. Together, these observations suggest that the operon structure in *O. dioica* need not solely or primarily be related to the regulation of transcription. On the contrary, the presence of the operon transcriptional system could act to decrease the necessity for genes to retain their own promoters, by allowing them to freely insert into other operons with their own transcriptional machinery. Indeed, we identified several lines of evidence suggesting that operon-switching can occur. As such, an operon system such as exists in *O. dioica* might in fact help to maintain gene expression in the context of genome scrambling. Although operons may facilitate genome scrambling, they may not directly cause scrambling; operons are found in the short-lived nematode *Caenorhabditis elegans* (Blumenthal and Gleason 2003) without a marked difference in the rate of scrambling compared with *Drosophila*. The fact that *O. dioica* operons scramble underlines the profound difference between bacterial operons and eukaryotic operons that use *trans*-splicing and calls for further investigations in *O. dioica* to better understand how genome scrambling can affect the regulation and evolutionary dynamics of operons in eukaryotes.

In conclusion, our results reveal an unprecedented degree of genome scrambling among what was considered a single cosmopolitan *O. dioica* species but which, according to our findings, may represent multiple cryptic species around the globe. In contrast to lepidopterans, in which speciation is intimately linked to highly dynamic evolution in the number of chromosomes (Joron et al. 2011; Hill et al. 2019; de Vos et al. 2020), the karyotype of *O. dioica* remains constant between lineages. Despite massive genome scrambling that drastically changes gene order and disintegrates microsynteny, the lineages that we studied do not show obvious morphological differences (Masunaga et al. 2022); they share similar ecological niches throughout the world; and laboratory cultures seeded by local samples are used as an animal model internationally without previously noticing physiological differences (Bouquet et al. 2009; Martí-Solans et al. 2015; Masunaga et al. 2020). This apparent uncoupling of conservation of morphology from conservation of genome structure—perhaps the first such example of this phenomenon among animals, and certainly among the chordates—has important implications for investigating genotype–phenotype relationships in other species.

Methods

Sampling, genome sequencing, genome assembly, and scaffolding

We extracted high-molecular-weight DNA from one individual ("Bar2") from the Barcelona laboratory strain (Martí-Solans et al. 2015) using a modified salting-out protocol (Masunaga et al. 2022), sequenced it on MinION sequencer Mk1B (Oxford

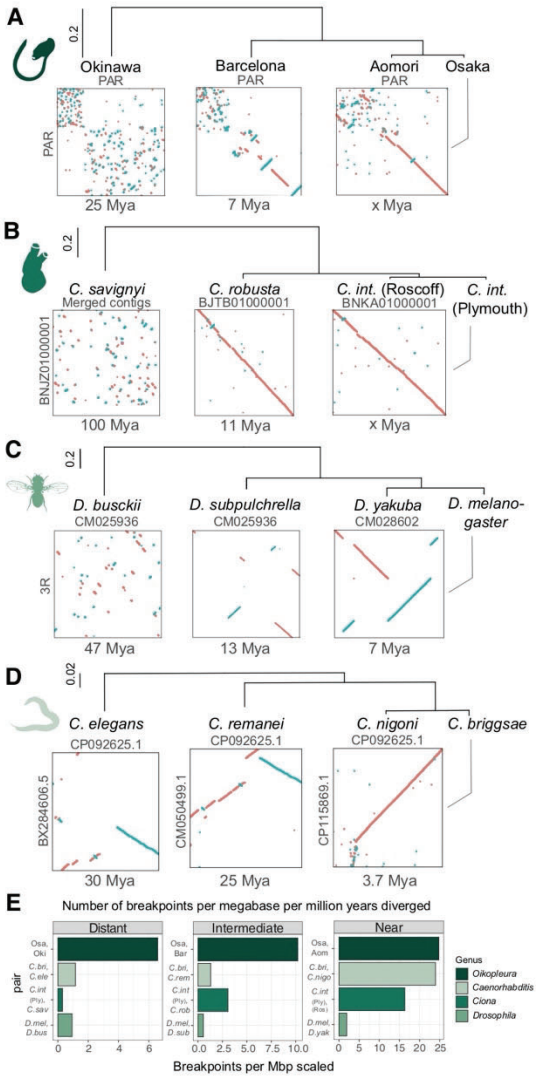


Figure 7. Genome scrambling in 10 × 10-Mbp regions for pairs of genomes in *O. dioica* (*A*), *Ciona* (*B*), *Drosophila* (*C*), and *Caenorhabditis* (*D*). Scaffold names are indicated in gray. (*E*) The number of breakpoint regions per megabase aligned per million years diverged for different animal clades.

Nanopore Technologies [ONT]) using a SQK-LSK109 kit (ONT) following the manufacturer's instructions and base-called it with the Guppy software (ONT) version 4.4.2 using the Rerio model res_dna_r941_min_crf_v031 (https://github.com/nanoporetech/rerio). The shortest reads were discarded until remaining data reached 7×10^9 nt, using filtlong software (https://github.com/rrwick/Filtlong), resulting in a read N50 >30,000 nt. We then assembled the genome using Flye software (Kolmogorov et al. 2019) version 2.8.2-b1689 with the --min-overlap 10,000 parameter using a custom Nextflow pipeline (see Data access). To ensure one-to-one correspondence between assemblies, we removed alternative haplotype sequences using the purge_dups tool (Guan et al. 2020). However, as a single removal step was not efficient enough, we used an iterative approach in which haplotigs were flagged with purge_dups, reads were aligned uniquely to contigs with LAST and last-split, and then reads aligning to purged haplotigs were removed before restarting the whole assembly process. Iterations were stopped after purge_dups stopped discovering haplotigs, and an assembly was selected that provided the best tradeoff between contiguity (which typically increased during the first iterations) and a low number of duplicated single-copy orthologs (Supplemental Table S2). The contigs were then polished with Pilon 1.22 (Walker et al. 2014) using short-read sequences from the same individual (available from the NCBI BioProject [https://www.ncbi.nlm.nih.gov/bioproject/] under accession number PRJEB55052) and scaffolded using Hi-C data from the Bergen line at tailbud stage (available from the NCBI Sequence Read Archive [SRA; https://www.ncbi.nlm.nih.gov/sra] under accession number SRR14470734) using Juicer (Durand et al. 2016) and 3D-DNA (Dudchenko et al. 2017), as in the work of Bliznina et al. (2021). The correctness of the scaffolding was later assessed using the same tools with Hi-C data from the Barcelona line itself (Supplemental Fig. S1).

We sequenced the Kume and Aomori genomes using single animals isolated from wild populations (Masunaga et al. 2020) with the same method except that we base-called with Guppy version 5.0.11 and the Guppy model dna_r9.4.1_450 bps_sup and used Flye version 2.8.3-b1763 with the parameters --min-overlap 10,000 --extra-params assemble_ovlp_divergence = 0.04,repeat_graph_ovlp_divergence = 0.04,read_align_ovlp_divergence = 0.04, max_bubble_length = 800,000,use_minimizers = 1,minimizer_window = 5, and no scaffolding no polishing was performed.

We rescaffolded the OSKA2016 genome (Wang et al. 2020a) by merging scaffolds that were overlapped by long contigs from independent single individual genome draft Nanopore long-read assemblies from the same laboratory strain (SRA: SAMEA6864573 and BioProject: PRJEB55052). As a last resort, we arbitrarily merged some contigs to a chromosome arm based on synteny information. The resulting OSKA2016v1.9 assembly is described in more detail at GitHub (https://github.com/oist/LuscombeU_OSKA2016_rescaffolding).

We sequenced genomes exclusively from male animals because they simplify the assembly of the sex-specific regions, which are single copy in males.

For all genomes, we counted metazoan near-universal single-copy genes using the benchmarking universal single-copy orthologs (BUSCO) (Manni et al. 2021) tool version 5.2.1 and an AUGUSTUS model trained for annotating the OKI2018_I69_1.0 assembly (Hoff and Stanke 2019; Bliznina et al. 2021). Although this version of BUSCO appears to have a lower detection baseline compared with the v3 series that we used for the OKI2018_I69 genome assembly (64% vs. 73%) (Bliznina et al. 2021), the completeness of our new assemblies is consistent with the score of the OKI2018_I69 genome assembly for which we have previously shown high completeness (Bliznina et al. 2021). Finally, we removed unplaced scaffolds from all chromosomal assemblies.

Pairwise genome alignment and comparison

We aligned pairs of genomes using the same approach as previously described (Bliznina et al. 2021). In brief, we used the LAST software (Kiełbasa et al. 2011) to align a "query" genome to a "target" genome indexed with the YASS seed (Noé and Kucherov 2005) for long and weak similarities with parameters and a scoring matrix determined by LAST-TRAIN software (Hamada et al. 2017); filtered the resulting many-to-many set of alignment pairs with the last-split tool (Frith and Kawaguchi 2015), which searches for an optimal set of one-to-one local alignments; and finally removed alignments that include a significant amount of masked sequences with the last-postmask tool (Frith 2011). Our one-to-one alignments

share some features with the "chains" of Kent et al. (2003) but do not allow local inversions.

We parallelized this process in a Nextflow (Di Tommaso et al. 2017) workflow available at GitHub (https://github.com/oist/plessy_pairwiseGenomeComparison/tree/v5.1.0). To load the alignment coordinates in the R environment for statistical computing (R Core Team 2023), we wrote a package called GenomicBreaks (https://oist.github.io/GenomicBreaks/) using core Bioconductor libraries (Lawrence et al. 2013). In this package, the strand_randomisation_index function computes for each chromosome the absolute difference of the total length of opposite-strand alignments to the total length of same-strand alignments and divides the result by the total length of the aligned regions so that a value of one indicates that all alignments are same-strand, and a value of zero indicates that, overall, the orientations appear to be random. The average of the values obtained on each chromosome is then computed and weighted by the length of the chromosomes. For the computation of the breakpoint and bridge regions, we used the strictest definition of collinearity, in which it is interrupted by inversions (changes of alignment strand) and translocations (presence of one extra aligned region in one genome only) of any length. A copy of the software and the alignment files is archived (see Data access). A rendering of the R vignettes that we used to produce these visualizations is available at GitHub (https://oist.github.io/LuscombeU_OikScrambling/) and Zenodo (doi:10.5281/zenodo.10677221) and compiled as an interactive notebook of vignettes in Supplemental Data.

Pairwise comparison between the *O. dioica* genomes produced in this work and the *Oikopleura vanhoeffeni*, *Oikopleura longicauda*, and *Oikopleura albicans* genomes (Naville et al. 2019) were loaded in the CNEr package (Tan et al. 2019) to define conserved noncoding elements with a window size of 50 and an identity threshold of 48 (see Data access).

Repeat masking and gene annotation

For each genome, a custom library of repeats was created by merging outputs of three different software—RepeatModeler (Flynn et al. 2020) version 2.0.1, MITE-Hunter (Han and Wessler 2010) version 11–2011, and SINE_Finder (Wenke et al. 2011)—that were used as input for RepeatMasker (Smit et al. 2013) version 4.1.0. The repeats identified by homology searches were soft-masked in each assembly.

Gene models were predicted using AUGUSTUS (Stanke et al. 2006) v3.3.3 using the species model trained for OKI2018_I69 *O. dioica* (Bliznina et al. 2021). To produce more accurate annotations, transcripts aligned to genomes with BLAT (Kent 2002) version 36 were used as "hints." In cases in which an assembled transcriptome was not available, data from related individuals were used. In particular, the transcriptome assembly generated by Wang et al. (2015) was used for predicting genes in both OSKA2016v1.9 and AOM-5-5f genomes, whereas the transcriptome assembly generated by Bliznina et al. (2021) was used for reannotation of the OKI2018_I69 genome and annotation of the KUM-M3-7f genome. A Barcelona transcriptome assembly was used for gene prediction in the Bar2_p4 genome. The parameter "--allow_hinted_splicesites" was used with AUGUSTUS to allow the prediction of noncanonical splice sites (GAAG, GCAG, GGAG, GTCG, GTAA).

Operons were annotated for each species, defining an operon as a set of genes that follow each other on the same strand and are separated by an intergenic distance of at most 500 bp, as this definition produces distributions of operon lengths comparable to the one reported by Denoeud et al. (2010). For the operon conservation analysis in Figure 4, two kinds of gene equivalence were considered: genes equivalent by assignment to the same hierarchical orthogroup ("HOG") (see Supplemental Fig. S6A–C) or genes equivalent by assignment to the same orthogroup (OG) (Supplemental Fig. S6D–F) using OrthoFinder. Because HOGs can include one-to-one orthologs as well as paralogs, it is relatively permissive to regard genes in HOGs as equivalent; however, assignment to the same OG is even more permissive, as OGs often contain many HOGs and may represent entire gene families. Operon equivalence was also assessed in two ways: Operons considered equivalent when all genes of an operon from species 1 were equivalent to all genes of an operon from species 2 (exact) (Supplemental Fig. S6A,B,D–E). A second type of operon equivalence allowed for up to one gene to differ in operons of length three or greater (inexact) (Supplemental Fig. S6C,F). This means an operon consisting of genes ABC in species 1 would match to any/all operons containing the genes ABC, XBC, AXC, or ABX in species 2. Overall, the conclusion that operons are not conserved between species of *O. dioica* is unaffected by the intergenic distance used to define operons (Supplemental Fig. 6, cf. A and B, cf. D and E), gene equivalence criteria (A vs. D, B vs. E), or operon equivalence criteria (Supplemental Fig. 6, cf. B and C, cf. E and F).

Ortholog identification

Gene orthology was reconstructed using OrthoFinder (Emms and Kelly 2015, 2019) version 2.5.4 based on 26 proteomes spanning three subphylums of chordates. To improve orthology assignment within *O. dioica*, multiple tunicate species were included as recommended in the OrthoFinder tutorials (https://davidemms.github.io/). Gene predictions for six appendicularian genomes from Naville et al. (2019) and two geographically distinct *C. intestinalis* genomes (Plymouth and Roscoff) (Satou et al. 2021) were computed using a similar approach to *O. dioica*, including repeat-masking followed by gene prediction with AUGUSTUS version 3.3.3. Gene prediction used either the *O. dioica* or *Ciona* model, as other species lack publicly available gene annotations. The proteomes of other species were downloaded from UniProt: *Branchiostoma floridae* (UP000001554), *C. intestinalis* type "A" (*robusta*, UP000008144), *C. savignyi* (UP000007875), *Danio rerio* (UP000000437), *Xenopus tropicalis* (UP000008143), *Gallus gallus* (UP000000539), *Mus musculus* (UP000000589), and *Homo sapiens* (UP000005640). Four more tunicate species were included from the Aniseed database: *Botrylloides leachii*, *Halocynthia roretzi*, *Molgula oculata*, and *Phallusia mammillata*. To remove redundancy in the data set, protein sequences were clustered at 100% identity using CD-HIT (Li and Godzik 2006) version 4.8.1. Alternative haplotypes were removed from the Bergen *O. dioica* proteome, and only the longest isoforms per gene were used for the analysis. OrthoFinder was run with the parameters -M msa -T raxml-ng with the following fixed species tree to ensure that *O. dioica* sequences fall within the oikopleurid branch:

(((*Danio_rerio*,(*Xenopus_tropicalis*,((*Mus_musculus*,*Homo_sapiens*),*Gallus_gallus*))),(((*Molgula_oculata*,(*Halocynthia_roretzi*,*Botrylloides_leachii*)),((*Ciona_savignyi*,((*C_intesinalis_P*,*C_intestinalis_R*),*Ciona_robusta*)),*Phallusia_mammillata*)),(*Fritillaria_borealis*,((*Oikopleura_longicauda*,(*Mesochordaeus_erythrocephalus*,*Bathochordaeus_sp*)),((*Oikopleura_vanhoeffeni*,*Oikopleura_albicans*),((KUM-M3-7f,OKI2018_I69),((Bar2_p4,OdB3),(AOM-5-5f,OSKA2016v1.9)))))))),*Branchiostoma_floridae*).

The Hox protein sequences of the Bergen genome were used as reference Hox sequences for *O. dioica* (Seo et al. 2004). In general, Hox genes were assigned appropriate orthogroups by OrthoFinder, although Hox11 could not be identified within the Barcelona proteome and the *Hox9* model for Osaka had not been

spliced appropriately. Regardless, the identities of these genes were confirmed by alignment with the Bergen sequence as well as multiple sequence alignment with all orthologous family members followed by tree estimation with IQ-TREE (Nguyen et al. 2015) version 1.6.12.

Phylogenomics and divergence time estimation

A species tree for *O. dioica* was estimated using a concatenated alignment of 5162 single-copy orthologous nucleotide sequences common to all six *O. dioica* genomes. A maximum likelihood tree was estimated with RAxML (Stamatakis 2014) version 8.2.4 using the GTRCAT substitution model and the autoMRE bootstopping criterion. The same data were used to estimate a Bayesian tree using MrBayes 3.2.7 (Ronquist et al. 2012) with six gamma-distributed rate categories, the 4-×4-nt substitution model. The MCMC chain was computed with three runs, a maximum of 100,000,000 generations, and 25% burn-in, with automatic stoppage after the average standard deviation of the split frequencies was lower than 0.01 and a minimum split frequency of 0.10. For both maximum likelihood and Bayesian analyses, each ortholog was assigned a separate independent partition.

To estimate the divergence times of *O. dioica* lineages, we created a stringent and conservative set of single-copy orthologous protein sequences in accordance with recommended practices in phylogenomics (Philippe et al. 2017; Simion et al. 2020), acknowledging that heterotachy is particularly problematic in the case of *O. dioica* (Berná et al. 2012) and the difficulty of retrieving accurate ortholog sequences from larvacean genomes of variable completeness and contiguity. A set of single-copy orthologous protein sequences was extracted from the results of OrthoFinder, selecting proteins that were shared by 10 or more of the 26 species, yielding 555 ortholog candidates. Each candidate ortholog was aligned using PRANK (Löytynoja 2014) v.170427 and then trimmed with HmmCleaner (Di Franco et al. 2019), and a gene tree was estimated with RAxML (Stamatakis 2014) version 8.2.4, with 100 rapid bootstraps and a gamma model of rate heterogeneity with automatic model selection using PROTGAMMAAUTO. Each gene tree was compared with the later species tree with the ete3 toolkit and evaluated for congruence (Huerta-Cepas et al. 2016). A supermatrix (concatenated alignment) was constructed, and gene information content was assessed with MARE (https://bonn.leibniz-lib.de/en/research/research-centres-and-groups/mare) v0.1.2-rc, which reduced the number of orthologs to 177. The alignment supermatrix generated from these 177 genes (containing 60,630 aligned amino acid sites, including gaps) was used to estimate a species tree with RAxML using 100 rapid bootstrap replicates, the gamma model of rate heterogeneity, and automatic model selection for each gene as separate partitions. To estimate divergence times, BEAST1 (Suchard et al. 2018) v1.10.4 was used with the BEAGLE library (Ayres et al. 2012) with the following parameters: the birth–death tree density model (Gernhard 2008), a linked random local clock model (Drummond and Suchard 2010), an unlinked gamma-distributed rate heterogeneity with four categories for each partitioned gene, and the CTMC scale reference prior model (Ferreira and Suchard 2008). To estimate only divergence times, the tree topology was fixed to the species tree estimated by RAxML. Where possible, the divergence time estimates published by Delsuc et al. (2018) using their LN CAT-GTR + Γ_4 model were used as normally distributed priors on our tree with matching mean and standard deviation. Each node that did not correspond between the two studies, including the appendicularian proteomes that we annotated, uses uniformly distributed priors with a maximum age as the age of the tunicates, owing to a lack of suitable fossils to calibrate these nodes. The only exception was a normally distributed prior for the split between *C. intestinalis* and *C. robusta*, which used the value reported by Bouchemousse et al. (2016). To ensure the models had converged, Tracer (Rambaut et al. 2018) was used (v1.7.2), and further, three replicate analyses were performed using these parameters, taking the last 100 million steps after convergence for calculating statistics. The final resampled, combined metrics are reported in Supplemental Table S1. The maximum clade credibility tree with node heights summarized to the median is depicted in Figure 5, using the replicate with the best marginal likelihood estimated by generalized stepping-stone sampling. The R libraries ggtree (Yu 2020) version 3.2.1, treeio (Wang et al. 2020b) version 1.18.1, and deeptime (Hoffmann et al. 2022) version 0.2.2 were used for tree visualization.

d_N / d_S estimation

To generate d_N/d_S estimates for *O. dioica* genes, single-copy orthologous proteins common to all six *O. dioica* proteomes were assessed. Each orthologous protein was aligned using PRANK, and protein alignments were converted to codon alignments using PAL2NAL (Suyama et al. 2006) v14.1. Then, a global estimate for d_N/d_S was calculated using the CODEML program of the PAML package (Yang 1997, 2007) version 4.9j using the species tree estimated from all single-copy orthologs as the tree input file, as well as the FMutSel mutation-selection model (codonfreq = 7). Estimating a single d_N/d_S value for a gene family, irrespective of differences between sites or branches, is almost certain to underestimate d_N/d_S; although this is less powerful for identifying cases of positive selection, it is nonetheless suitable for roughly characterizing substitution patterns across genome as used in Figure 4. To support the estimates produced from global comparisons, maximum likelihood and Bayesian estimates for d_S, d_N, and d_N/d_S were also calculated for all pairs (using runmode = −2 and runmode = −3) and are depicted in Supplemental Figure S10, providing support for the relatively low d_N/d_S values reported by global estimates.

Data access

Raw Nanopore reads generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJEB55052. Software, alignments, and intermediate data are available at Zenodo (https://doi.org/10.5281/zenodo.10241527) and as Supplemental Code and Supplemental Data.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank the DNA Sequencing Section and the Scientific Computing and Data Analysis Section of the Research Support Division at OIST for their support; Martin Frith, Ferdinand Marlétaz, Jiashun Miao, and Thomas Bourguignon for critical comments; Cristina Frías Lopez for initial bioinformatic support on the Barcelona genome assembly; and Atsuo Nishino for providing the Aomori samples. This work was supported by OIST core funding and, in part, by grant NFR-FRIBIO 204891/F20 to E.M.T. from the Norwegian Research Council. M.J.M. acknowledges funding from the Japan Society for the Promotion of Science as a JSPS International Research Fellow (Luscombe Unit, Okinawa Institute of Science and Technology Graduate University). C.C. was funded by PID2019-110562GB-I00 and PID2022-141627NB-

I00 from the Spanish Ministerio de Ciencia e Innovación and by ICREA Acadèmia Ac2215698 and 2021-SGR00372 AGAUR, Generalitat de Catalunya; V.R. by 2017BP00139 AGAUR, Generalitat de Catalunya and 2019IRBio001 from IRBio, Universitat de Barcelona; G.S.-S. by a FPU18/02414 fellowship from Ministerio de Educación y cultura; M.F.-T. by colaboración-2015/16; M.F.-T. by a PREDOC2020/58 fellowship from Universitat de Barcelona; and A.F.-R. by MS12 Margarita Salas from Ministerio de Universidades (Spain). The sequencing of the *O. dioica* genome from Barcelona has been performed under the Catalan Initiative for the Earth Biogenome Project.

Author contributions: Conceptualization was by C.P. and N.M.L. Data curation was by A.B., C.P., M.J.M., and P.N. Formal analysis was by A.B., C.P., E.M.T., M.J.M., and P.N. Funding acquisition was by C.C., E.M.T., and N.M.L. Investigation was by A.B., A.M., A.W.L., C.C., C.P., C.W., G.S.-S., J.G., M.F.-T., M.J.M., M.S.R.P., N.M.L., P.N., T.O., and Y.T. Methodology was by A.F.-R., C.P., C.W., M.J.M., P.N., and V.R. Project administration was by C.P. and N.M.L. Resources were by H.N., E.M.T., and C.C. Software was by C.P. and M.J.M. Supervision was by C.P., E.M.T., and N.M.L. Validation was by C.P. and M.J.M. Visualization was by A.B., A.M., C.P., and M.J.M. Writing of the original draft was by A.B., C.P., M.J.M., and N.M.L. Review and editing were by C.P., E.M.T., M.J.M., P.N., C.C., and N.M.L.

References

Albertin CB, Medina-Ruiz S, Mitros T, Schmidbaur H, Sanchez G, Wang ZY, Grimwood J, Rosenthal JJC, Ragsdale CW, Simakov O, et al. 2022. Genome and transcriptome mechanisms driving cephalopod evolution. *Nat Commun* **13:** 2427. doi:10.1038/s41467-022-29748-w

Aparicio S, Chapman J, Stupka E, Putnam N, Chia J-M, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297:** 1301–1310. doi:10.1126/science.1072104

Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, Huelsenbeck JP, Ronquist F, Swofford DL, Cummings MP, et al. 2012. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst Biol* **61:** 170–173. doi:10.1093/sysbio/syr100

Berná L, D'Onofrio G, Alvarez-Valin F. 2012. Peculiar patterns of amino acid substitution and conservation in the fast evolving tunicate *Oikopleura dioica*. *Mol Phylogenet Evol* **62:** 708–717. doi:10.1016/j.ympev.2011.11.013

Bliznina A, Masunaga A, Mansfield MJ, Tan Y, Liu AW, West C, Rustagi T, Chien H-C, Kumar S, Pichon J, et al. 2021. Telomere-to-telomere assembly of the genome of an individual *Oikopleura dioica* from Okinawa using nanopore-based sequencing. *BMC Genomics* **22:** 222. doi:10.1186/s12864-021-07512-6

Blumenthal T, Gleason KS. 2003. *Caenorhabditis elegans* operons: form and function. *Nat Rev Genet* **4:** 112–120. doi:10.1038/nrg995

Bouchemousse S, Bishop JDD, Viard F. 2016. Contrasting global genetic patterns in two biologically similar, widespread and invasive *Ciona* species (Tunicata, Ascidiacea). *Sci Rep* **6:** 24875. doi:10.1038/srep24875

Bouquet J-M, Spriet E, Troedsson C, Otterå H, Chourrout D, Thompson EM. 2009. Culture optimization for the emergent zooplanktonic model organism *Oikopleura dioica*. *J Plankton Res* **31:** 359–370. doi:10.1093/plankt/fbn132

Caputi L, Andreakis N, Mastrototaro F, Cirino P, Vassillo M, Sordino P. 2007. Cryptic speciation in a model invertebrate chordate. *Proc Natl Acad Sci* **104:** 9364–9369. doi:10.1073/pnas.0610158104

Damas J, Corbo M, Lewin HA. 2021. Vertebrate chromosome evolution. *Annu Rev Anim Biosci* **9:** 1–27. doi:10.1146/annurev-animal-020518-114924

Damas J, Corbo M, Kim J, Turner-Maier J, Farré M, Larkin DM, Ryder OA, Steiner C, Houck ML, Hall S, et al. 2022. Evolution of the ancestral mammalian karyotype and syntenic regions. *Proc Natl Acad Sci* **119:** e2209139119. doi:10.1073/pnas.2209139119

Danks GB, Raasholm M, Campsteijn C, Long AM, Manak JR, Lenhard B, Thompson EM. 2015. Trans-splicing and operons in metazoans: translational control in maternally regulated development and recovery from growth arrest. *Mol Biol Evol* **32:** 585–599. doi:10.1093/molbev/msu336

Darwin C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London. https://en.wikisource.org/wiki/On_the_Origin_of_Species_%281859%29

Delsuc F, Philippe H, Tsagkogeorga G, Simion P, Tilak M-K, Turon X, López-Legentil S, Piette J, Lemaire P, Douzery EJP. 2018. A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol* **16:** 39. doi:10.1186/s12915-018-0499-2

Deng W, Henriet S, Chourrout D. 2018. Prevalence of mutation-prone microhomology-mediated end joining in a chordate lacking the c-NHEJ DNA repair pathway. *Curr Biol* **28:** 3337–3341.e4. doi:10.1016/j.cub.2018.08.048

Denoeud F, Henriet S, Mungpakdee S, Aury J-M, Da Silva C, Brinkmann H, Mikhaleva J, Olsen LC, Jubin C, Cañestro C, et al. 2010. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* **330:** 1381–1385. doi:10.1126/science.1194167

de Vos JM, Augustijnen H, Bätscher L, Lucek K. 2020. Speciation through chromosomal fusion and fission in Lepidoptera. *Philos Trans R Soc Lond B Biol Sci* **375:** 20190539. doi:10.1098/rstb.2019.0539

Di Franco A, Poujol R, Baurain D, Philippe H. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol Biol* **19:** 21. doi:10.1186/s12862-019-1350-2

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35:** 316–319. doi:10.1038/nbt.3820

Dobzhansky T. 1937. *Genetics and the origin of species*. Columbia University Press, New York.

Drillon G, Champeimont R, Oteri F, Fischer G, Carbone A. 2020. Phylogenetic reconstruction based on synteny block and gene adjacencies. *Mol Biol Evol* **37:** 2747–2762. doi:10.1093/molbev/msaa114

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450:** 203–218. doi:10.1038/nature06341

Drummond AJ, Suchard MA. 2010. Bayesian random local clocks, or one rate to rule them all. *BMC Biol* **8:** 114. doi:10.1186/1741-7007-8-114

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356:** 92–95. doi:10.1126/science.aal3327

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3:** 95–98. doi:10.1016/j.cels.2016.07.002

Edvardsen RB, Lerat E, Maeland AD, Flåt M, Tewari R, Jensen MF, Lehrach H, Reinhardt R, Seo H-C, Chourrout D. 2004. Hypervariable and highly divergent intron-exon organizations in the chordate *Oikopleura dioica*. *J Mol Evol* **59:** 448–457. doi:10.1007/s00239-004-2636-5

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16:** 157. doi:10.1186/s13059-015-0721-2

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20:** 238. doi:10.1186/s13059-019-1832-y

Engström PG, Ho Sui SJ, Drivenes O, Becker TS, Lenhard B. 2007. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res* **17:** 1898–1908. doi:10.1101/gr.6669607

Ferrández-Roldán A, Fabregà-Torrus M, Sánchez-Serna G, Duran-Bello E, Joaquín-Lluís M, Bujosa P, Plana-Carmona M, Garcia-Fernàndez J, Albalat R, Cañestro C. 2021. Cardiopharyngeal deconstruction and ancestral tunicate sessility. *Nature* **599:** 431–435. doi:10.1038/s41586-021-04041-w

Ferreira MAR, Suchard MA. 2008. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can J Stat* **36:** 355–368. doi:10.1002/cjs.5550360302

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci* **117:** 9451–9457. doi:10.1073/pnas.1921046117

Frith MC. 2011. Gentle masking of low-complexity sequences improves homology search. *PLoS One* **6:** e28819. doi:10.1371/journal.pone.0028819

Frith MC, Kawaguchi R. 2015. Split-alignment of genomes finds orthologies more accurately. *Genome Biol* **16:** 106. doi:10.1186/s13059-015-0670-9

Fu Y-W, Dai X-Y, Wang W-T, Yang Z-X, Zhao J-J, Zhang J-P, Wen W, Zhang F, Oberg KC, Zhang L, et al. 2021. Dynamics and competition of CRISPR-Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. *Nucleic Acids Res* **49:** 969–985. doi:10.1093/nar/gkaa1251

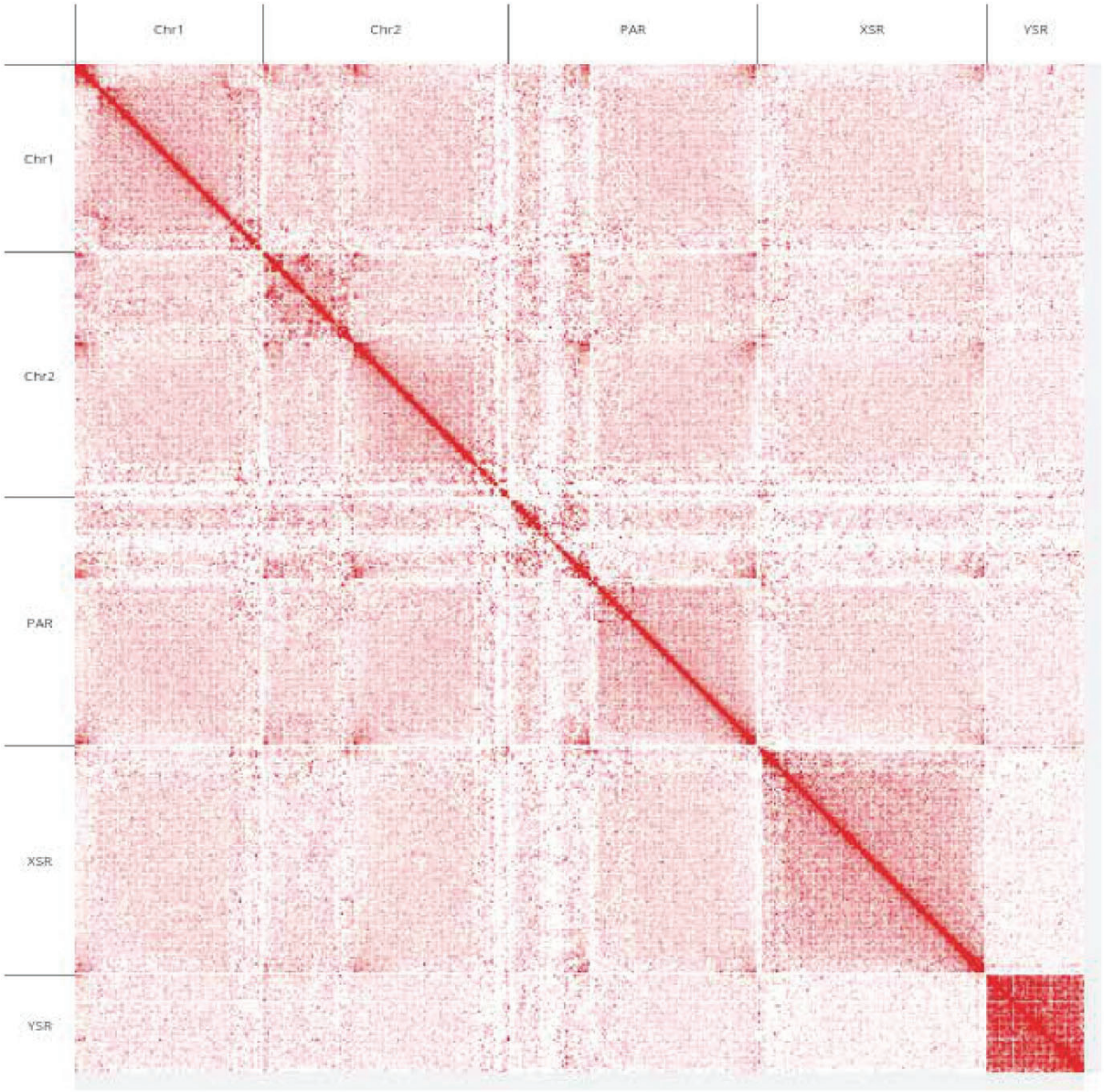
Ganot P, Kallesøe T, Reinhardt R, Chourrout D, Thompson EM. 2004. Spliced-leader RNA *trans* splicing in a chordate, *Oikopleura dioica*, with a compact genome. *Mol Cell Biol* **24:** 7795–7805. doi:10.1128/MCB.24.17.7795-7805.2004

Gernhard T. 2008. The conditioned reconstructed process. *J Theor Biol* **253:** 769–778. doi:10.1016/j.jtbi.2008.04.005
Goldschmidt R. 1940. *The material basis of evolution*. Yale University Press, New Haven, CT.
Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36:** 2896–2898. doi:10.1093/bioinformatics/btaa025
Hamada M, Ono Y, Asai K, Frith MC. 2017. Training alignment parameters for arbitrary sequencers with LAST-TRAIN. *Bioinformatics* **33:** 926–928. doi:10.1093/bioinformatics/btw742
Han Y, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res* **38:** e199. doi:10.1093/nar/gkq862
Hane JK, Rouxel T, Howlett BJ, Kema GHJ, Goodwin SB, Oliver RP. 2011. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol* **12:** R45. doi:10.1186/gb-2011-12-5-r45
Henriet S, Sumic S, Doufoundou-Guilengui C, Jensen MF, Grandmougin C, Fal K, Thompson E, Volff J-N, Chourrout D. 2015. Embryonic expression of endogenous retroviral RNAs in somatic tissues adjacent to the *Oikopleura* germline. *Nucleic Acids Res* **43:** 3701–3711. doi:10.1093/nar/gkv169
Hill J, Rastas P, Hornett EA, Neethiraj R, Clark N, Morehouse N, de la Paz Celorio-Mancera M, Cols JC, Dircksen H, Meslin C, et al. 2019. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. *Sci Adv* **5:** eaau3648. doi:10.1126/sciadv.aau3648
Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, Sedeño Cacciatore Á, Contessoto VG, van Heesbeen RGHP, van den Broek B, et al. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372:** 984–989. doi:10.1126/science.abe2218
Hoff KJ, Stanke M. 2019. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc Bioinformatics* **65:** e57. doi:10.1002/cpbi.57
Hoffmann M, Scherer M, Hempel T, Mardt A, de Silva B, Husic BE, Klus S, Wu H, Kutz N, Brunton SL, et al. 2022. Deeptime: a python library for machine learning dynamical models from time series data. *Mach Learn Sci Technol* **3:** 015009. doi:10.1088/2632-2153/ac3de0
Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol* **33:** 1635–1638. doi:10.1093/molbev/msw046
Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* **5:** 299–310. doi:10.1038/nrg1319
Irimia M, Tena JJ, Alexis MS, Fernandez-Miñan A, Maeso I, Bogdanović O, de la Calle-Mustienes E, Roy SW, Gómez-Skarmeta JL, Fraser HB. 2012. Extensive conservation of ancient microsynteny across metazoans due to *cis*-regulatory constraints. *Genome Res* **22:** 2356–2367. doi:10.1101/gr.139725.112
Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477:** 203–206. doi:10.1038/nature10341
Kent WJ. 2002. BLAT: the BLAST-like alignment tool. *Genome Res* **12:** 656–664. doi:10.1101/gr.229202
Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci* **100:** 11484–11489. doi:10.1073/pnas.1932072100
Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Res* **21:** 487–493. doi:10.1101/gr.113985.110
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37:** 540–546. doi:10.1038/s41587-019-0072-8
Körner WF. 1952. Untersuchungen über die Gehäusebildung bei Appendicularien (*Oikopleura dioicafol*). *Z Für Morphol Ökol Tiere* **41:** 1–53. doi:10.1007/BF00407623
Lamichhaney S, Fan G, Widemo F, Gunnarsson U, Thalmann DS, Hoeppner MP, Kerje S, Gustafson U, Shi C, Zhang H, et al. 2016. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat Genet* **48:** 84–88. doi:10.1038/ng.3430
Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9:** e1003118. doi:10.1371/journal.pcbi.1003118
Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22:** 1658–1659. doi:10.1093/bioinformatics/btl158
Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, Khurana E, Waszak S, Korbel JO, Haber JE, et al. 2020. Patterns of somatic structural variation in human cancer genomes. *Nature* **578:** 112–121. doi:10.1038/s41586-019-1913-9
Liu AW, Tan Y, Masunaga A, Bliznina A, West C, Plessy C, Luscombe NM. 2020. H3s28p antibody staining of Okinawan *Oikopleura dioica* suggests the presence of three chromosomes. *F1000Res* **9:** 780. doi:10.12688/f1000research.25019.2
Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. *Methods Mol Biol* **1079:** 155–170. doi:10.1007/978-1-62703-646-7_10
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* **38:** 4647–4654. doi:10.1093/molbev/msab199
Martí-Solans J, Ferrández-Roldán A, Godoy-Marín H, Badia-Ramentol J, Torres-Aguila NP, Rodríguez-Marí A, Bouquet JM, Chourrout D, Thompson EM, Albalat R, et al. 2015. *Oikopleura dioica* culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis* **53:** 183–193. doi:10.1002/dvg.22800
Masunaga A, Liu AW, Tan Y, Scott A, Luscombe NM. 2020. Streamlined sampling and cultivation of the pelagic cosmopolitan larvacean, *Oikopleura dioica*. *J Vis Exp* **160**. doi:10.3791/61279.
Masunaga A, Mansfield MJ, Tan Y, Liu AW, Bliznina A, Barzaghi P, Hodgetts TL, Ferrández-Roldán A, Cañestro C, Onuma TA, et al. 2022. The cosmopolitan appendicularian *Oikopleura dioica* reveals hidden genetic diversity around the globe. *Mar Biol* **169:** 157. doi:10.1007/s00227-022-04145-5
Mitsuhashi S, Ohori S, Katoh K, Frith MC, Matsumoto N. 2020. A pipeline for complete characterization of complex germline rearrangements from long DNA reads. *Genome Med* **12:** 67. doi:10.1186/s13073-020-00762-1
Mudd AB, Bredeson JV, Baum R, Hockemeyer D, Rokhsar DS. 2020. Analysis of muntjac deer genome and chromatin architecture reveals rapid karyotype evolution. *Commun Biol* **3:** 480. doi:10.1038/s42003-020-1096-9
Naville M, Henriet S, Warren I, Sumic S, Reeve M, Volff J-N, Chourrout D. 2019. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr Biol* **29:** 1161–1168.e6. doi:10.1016/j.cub.2019.01.080
Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32:** 268–274. doi:10.1093/molbev/msu300
Noé L, Kucherov G. 2005. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res* **33:** W540–W543. doi:10.1093/nar/gki478
Oulion S, Bertrand S, Escriva H. 2012. Evolution of the FGF gene family. *Int J Evol Biol* **2012:** 298147. doi:10.1155/2012/298147
Passarge E, Horsthemke B, Farber RA. 1999. Incorrect use of the term synteny. *Nat Genet* **23:** 387. doi:10.1038/70486
Pereira-Santana A, Gamboa-Tuz SD, Zhao T, Schranz ME, Vinuesa P, Bayona A, Rodríguez-Zapata LC, Castano E. 2020. Fibrillarin evolution through the tree of life: comparative genomics and microsynteny network analyses provide new insights into the evolutionary history of fibrillarin. *PLoS Comput Biol* **16:** e1008318. doi:10.1371/journal.pcbi.1008318
Philippe H, de Vienne DM, Ranwez V, Roure B, Baurain D, Delsuc F. 2017. Pitfalls in supermatrix phylogenomics. *Eur J Taxon* 283. doi:10.5852/ejt.2017.283
Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol* **67:** 901–904. doi:10.1093/sysbio/syy032
R Core Team. 2023. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna https://www.R-project.org/.
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61:** 539–542. doi:10.1093/sysbio/sys029
Rowley MJ, Corces VG. 2018. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19:** 789–800. doi:10.1038/s41576-018-0060-8
Royo JL, Maeso I, Irimia M, Gao F, Peter IS, Lopes CS, D'Aniello S, Casares F, Davidson EH, Garcia-Fernández J, et al. 2011. Transphyletic conservation of developmental regulatory state in animal evolution. *Proc Natl Acad Sci* **108:** 14186–14191. doi:10.1073/pnas.1109037108
Satou Y, Nakamura R, Yu D, Yoshida R, Hamada M, Fujie M, Hisata K, Takeda H, Satoh N. 2019. A nearly complete genome of *Ciona intestinalis* type A (*C. robusta*) reveals the contribution of inversion to chromosomal evolution in the genus *Ciona*. *Genome Biol Evol* **11:** 3144–3157. doi:10.1093/gbe/evz228
Satou Y, Sato A, Yasuo H, Mihirogi Y, Bishop J, Fujie M, Kawamitsu M, Hisata K, Satoh N. 2021. Chromosomal inversion polymorphisms in two sympatric ascidian lineages. *Genome Biol Evol* **13:** evab068. doi:10.1093/gbe/evab068
Seo H-C, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, et al. 2004. Hox cluster

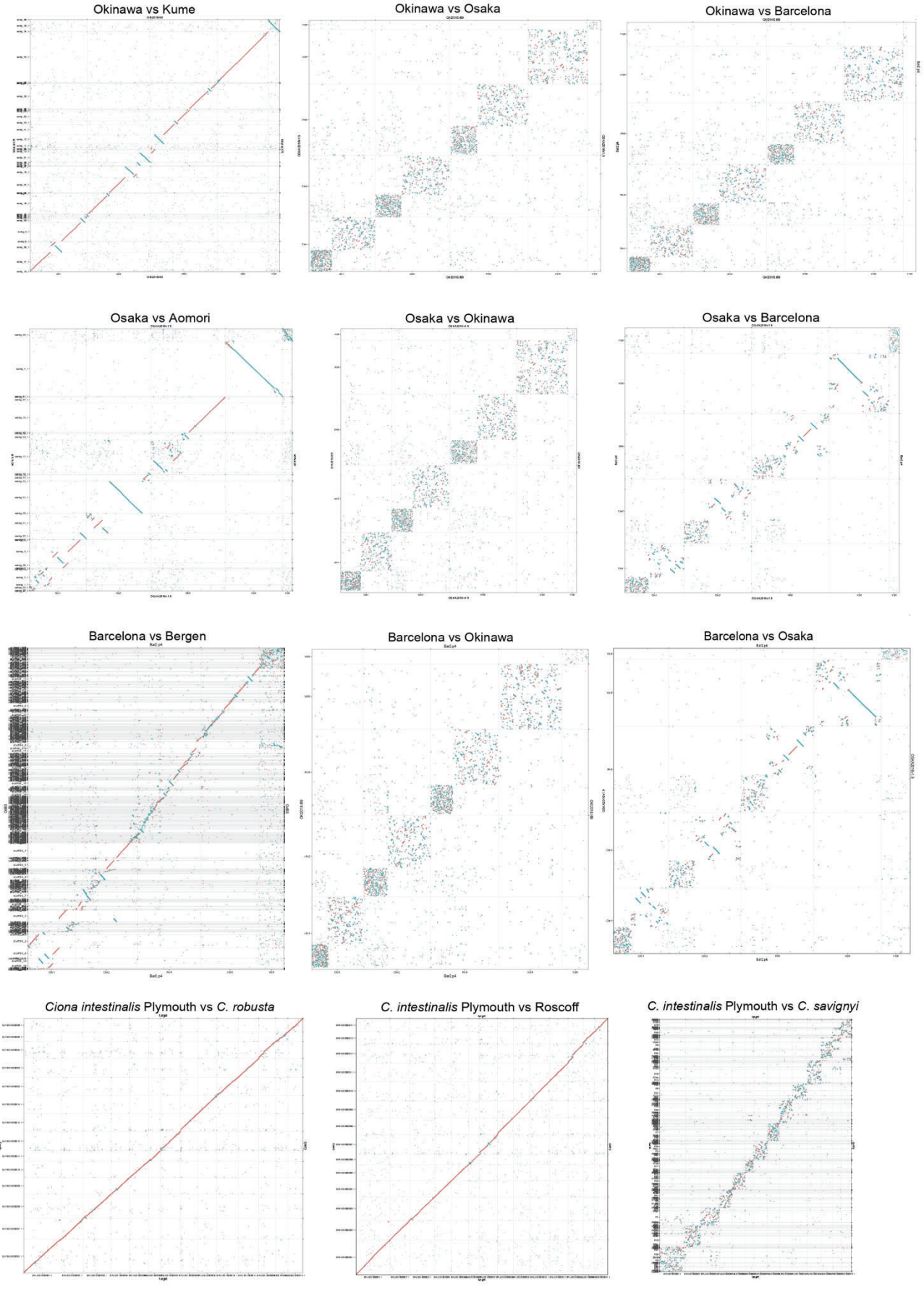
disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431:** 67–71. doi:10.1038/nature02709
Simakov O, Bredeson J, Berkoff K, Marletaz F, Mitros T, Schultz DT, O'Connell BL, Dear P, Martinez DE, Steele RE, et al. 2022. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci Adv* **8:** eabi5884. doi:10.1126/sciadv.abi5884
Simion P, Delsuc F, Philippe H. 2020. To what extent current limits of phylogenomics can be overcome? In *Phylogenetics in the genomic era* (ed. Scornavacca C, et al.), pp. 2.1:1–2.1:34. HAL, Paris. https://hal.science/hal-02535366
Smit A, Hubley R, Green P. 2013. RepeatMasker Open-4.0. http://www.repeatmasker.org.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313. doi:10.1093/bioinformatics/btu033
Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34:** W435–W439. doi:10.1093/nar/gkl200
Stover NA, Steele RE. 2001. Trans-spliced leader addition to mRNAs in a cnidarian. *Proc Natl Acad Sci* **98:** 5693–5698. doi:10.1073/pnas.101049998
Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* **4:** vey016. doi:10.1093/ve/vey016
Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell P, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr Biol* **32:** 111–123.e5. doi:10.1016/j.cub.2021.10.052
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34:** W609–W612. doi:10.1093/nar/gkl315
Tan G, Polychronopoulos D, Lenhard B. 2019. CNEr: a toolkit for exploring extreme noncoding conservation. *PLoS Comput Biol* **15:** e1006940. doi:10.1371/journal.pcbi.1006940
Van der Ploeg LH. 1986. Discontinuous transcription and splicing in trypanosomes. *Cell* **47:** 479–480. doi:10.1016/0092-8674(86)90608-2
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9:** e112963. doi:10.1371/journal.pone.0112963
Wang K, Omotezako T, Kishi K, Nishida H, Onuma TA. 2015. Maternal and zygotic transcriptomes in the appendicularian, *Oikopleura dioica*: novel protein-encoding genes, intra-species sequence variations, and *trans*-spliced RNA leader. *Dev Genes Evol* **225:** 149–159. doi:10.1007/s00427-015-0502-7
Wang K, Tomura R, Chen W, Kiyooka M, Ishizaki H, Aizu T, Minakuchi Y, Seki M, Suzuki Y, Omotezako T, et al. 2020a. A genome database for a Japanese population of the larvacean *Oikopleura dioica*. *Dev Growth Differ* **62:** 450–461. doi:10.1111/dgd.12689
Wang L-G, Lam TT-Y, Xu S, Dai Z, Zhou L, Feng T, Guo P, Dunn CW, Jones BR, Bradley T, et al. 2020b. *Treeio*: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol Biol Evol* **37:** 599–603. doi:10.1093/molbev/msz240
Wanninger A. 2024. Hox, homology, and parsimony: an organismal perspective. *Semin Cell Dev Biol* **152–153:** 16–23. doi:10.1016/j.semcdb.2023.01.007
Weisman CM, Murray AW, Eddy SR. 2022. Mixing genome annotation methods in a comparative analysis inflates the apparent number of lineage-specific genes. *Curr Biol* **32:** 2632–2639.e2. doi:10.1016/j.cub.2022.04.085
Wenke T, Döbel T, Sörensen TR, Junghans H, Weisshaar B, Schmidt T. 2011. Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. *Plant Cell* **23:** 3117–3128. doi:10.1105/tpc.111.088682
Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13:** 555–556. doi:10.1093/bioinformatics/13.5.555
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24:** 1586–1591. doi:10.1093/molbev/msm088
Yu G. 2020. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinformatics* **69:** e96. doi:10.1002/cpbi.96
Zayas RM, Bold TD, Newmark PA. 2005. Spliced-leader *trans*-splicing in freshwater planarians. *Mol Biol Evol* **22:** 2048–2054. doi:10.1093/molbev/msi200
Zeller RW. 2010. Computational analysis of *Ciona intestinalis* operons. *Integr Comp Biol* **50:** 75–85. doi:10.1093/icb/icq040
Zhang H, Hou Y, Miranda L, Campbell DA, Sturm NR, Gaasterland T, Lin S. 2007. Spliced leader RNA *trans*-splicing in dinoflagellates. *Proc Natl Acad Sci* **104:** 4618–4623. doi:10.1073/pnas.0700258104

Received July 19, 2023; accepted in revised form February 28, 2024.

Supplemental Figure S1: Contact map of a Hi-C library made from the Barcelona laboratory strain, aligned on the Barcelona chromosomal assembly.

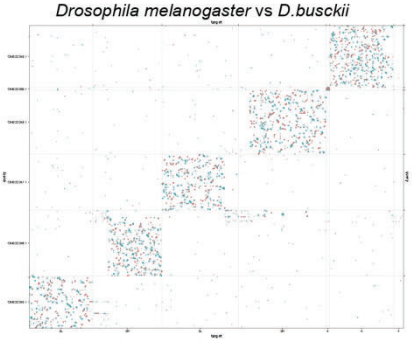


Supplemental Figure S2: Line plot representation of pairwise whole-genome alignments. Red: +/+ alignment. Blue: +/- alignment.

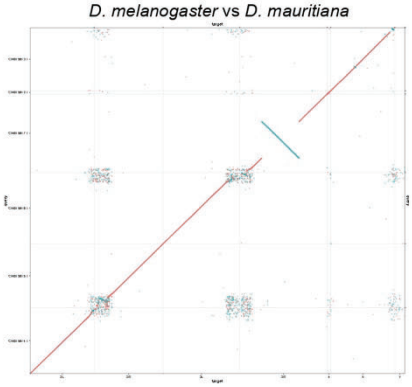


Supplemental Figure S2 (continued)

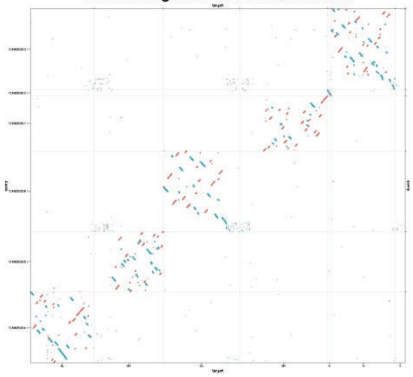
Drosophila melanogaster vs *D.busckii*



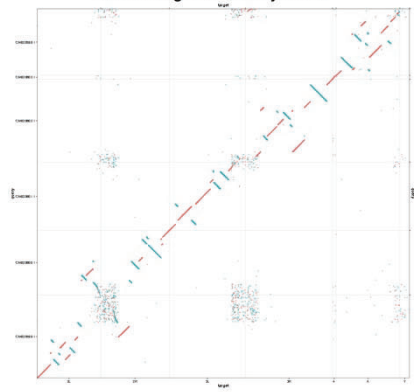
D. melanogaster vs *D. mauritiana*



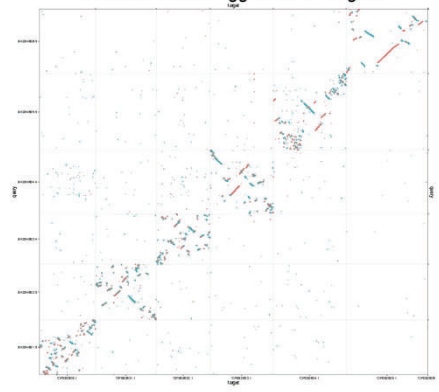
D. melanogaster vs *D. subobscura*



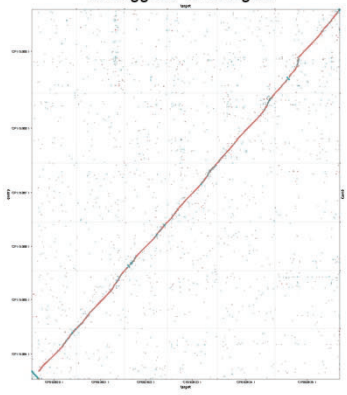
D. melanogaster vs *D. yakuba*



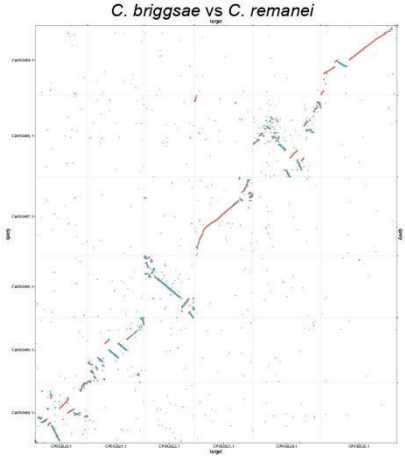
Caenorhabditis briggsae vs *C. elegans*



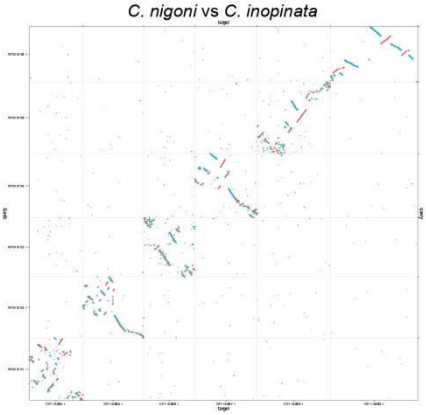
C. briggsae vs *C. nigoni*



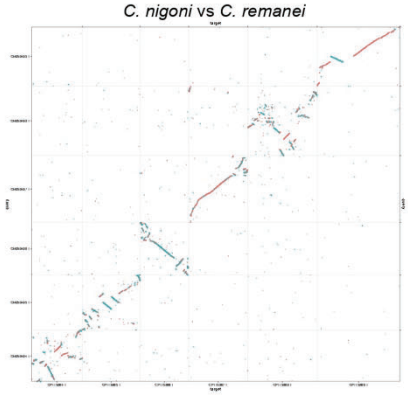
C. briggsae vs *C. remanei*



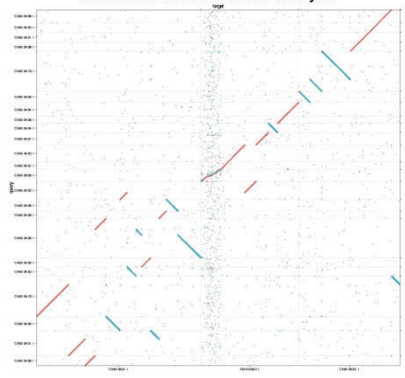
C. nigoni vs *C. inopinata*



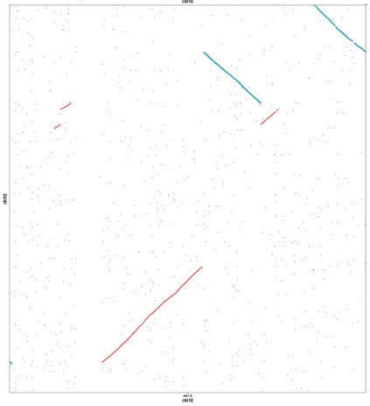
C. nigoni vs *C. remanei*



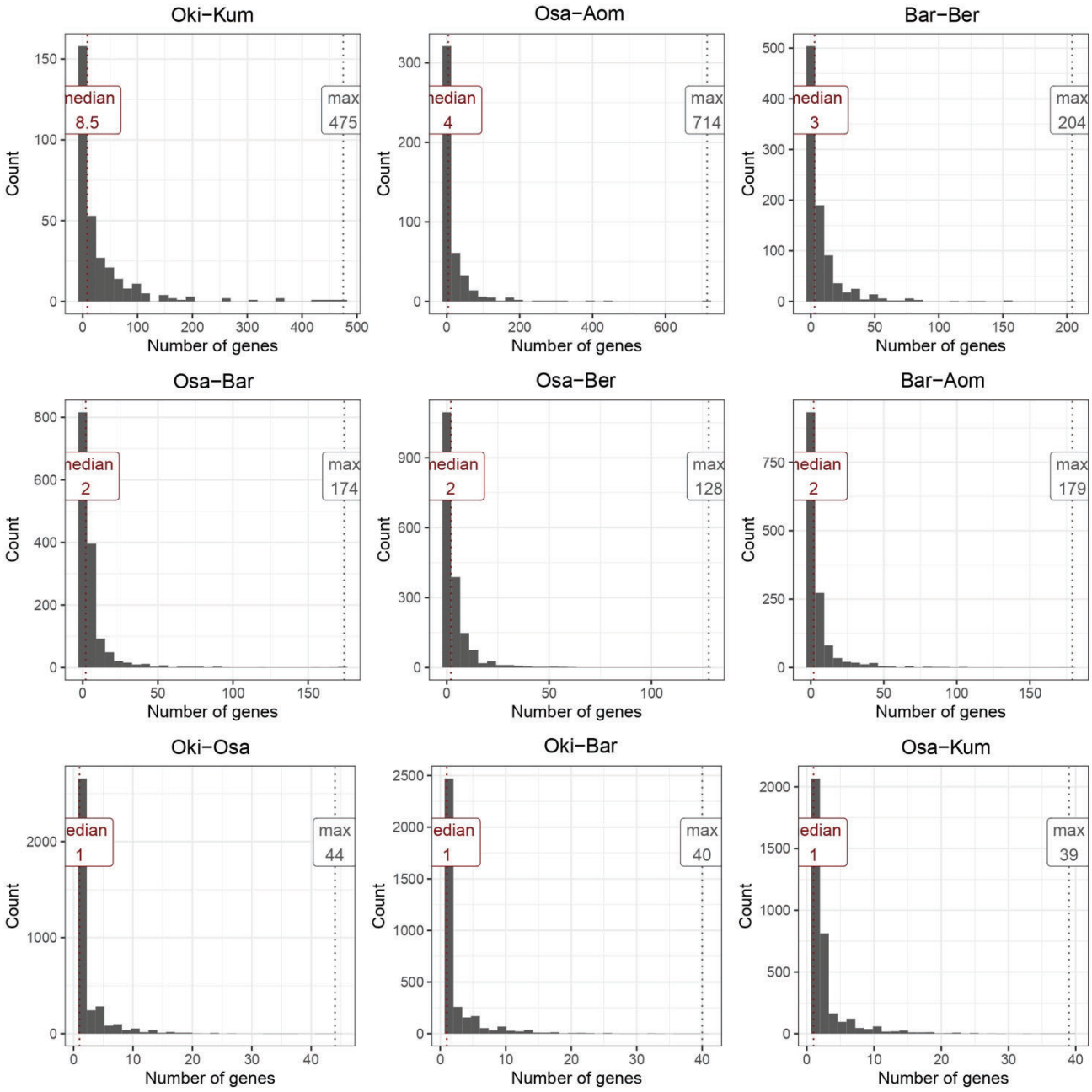
Muntiacus reevesi vs *M. muntjak*



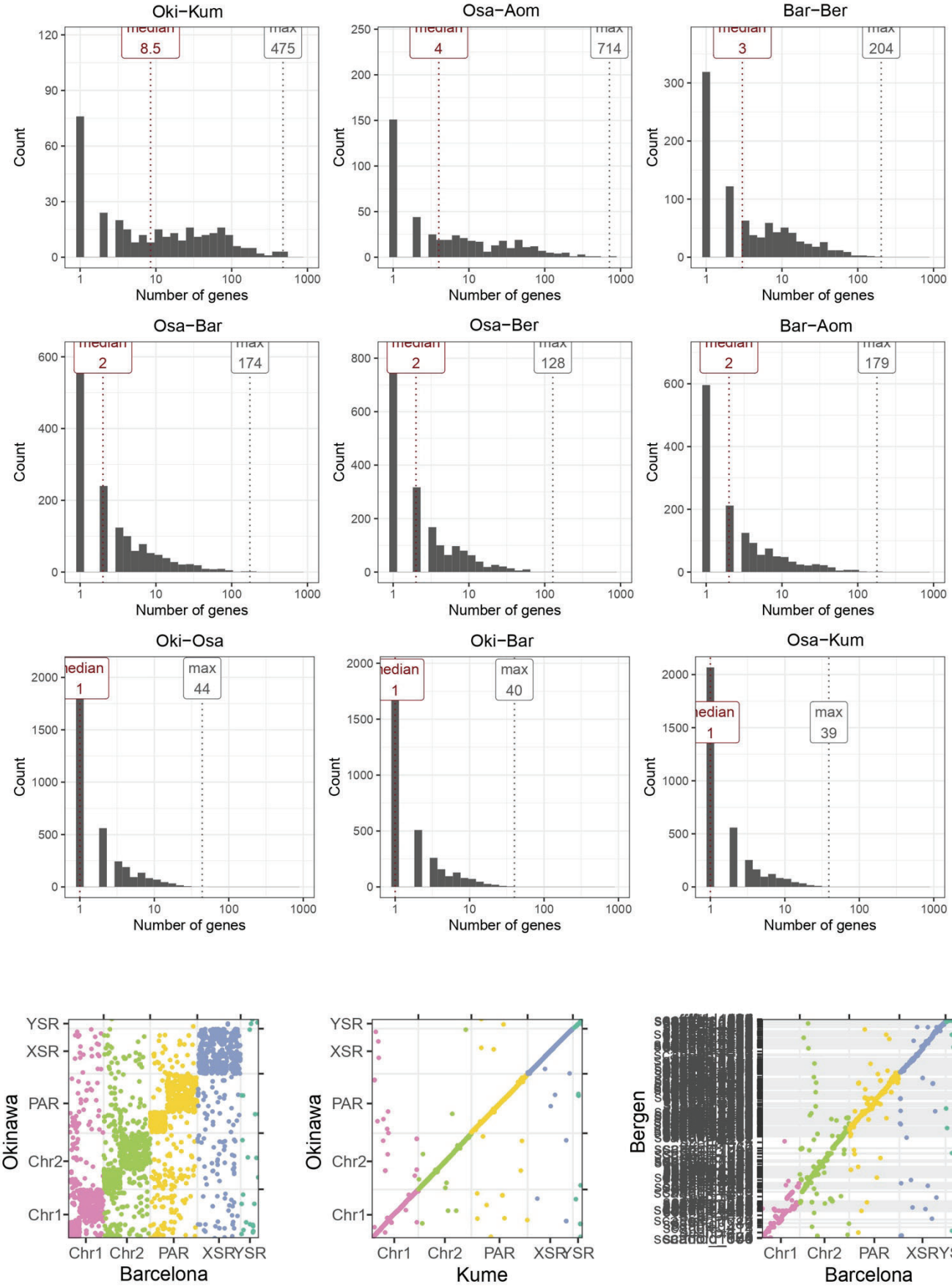
Homo sapiens chr18 vs *Mus musculus* chr18



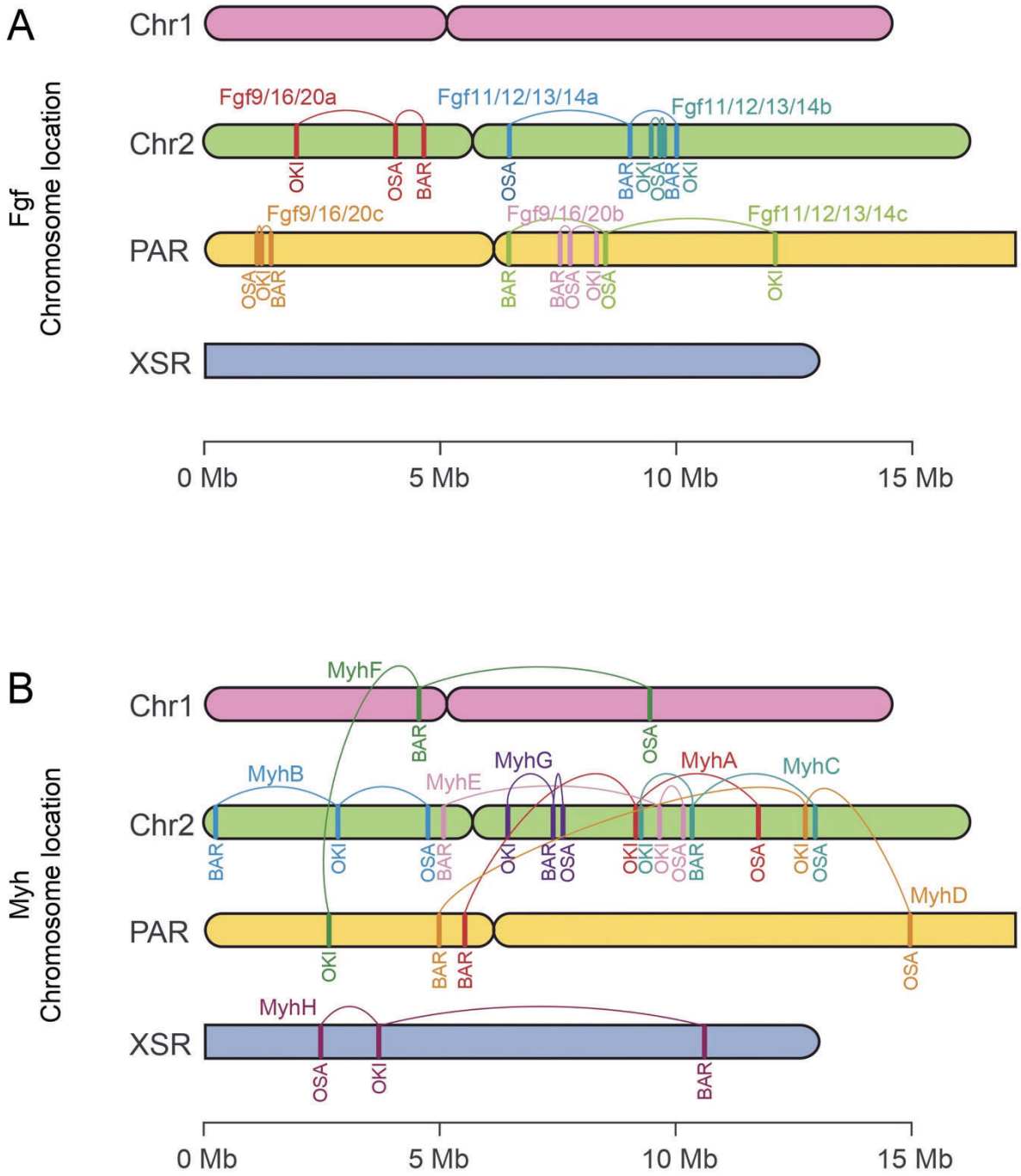
Supplemental Figure S3: Histogram of the number of orthologous genes per syntenic region in pairs of genomes (horizontal axis in natural or log scale), followed by dot-plot plots of homologous genes in pairs of genomes not displayed in the main Figure 2.



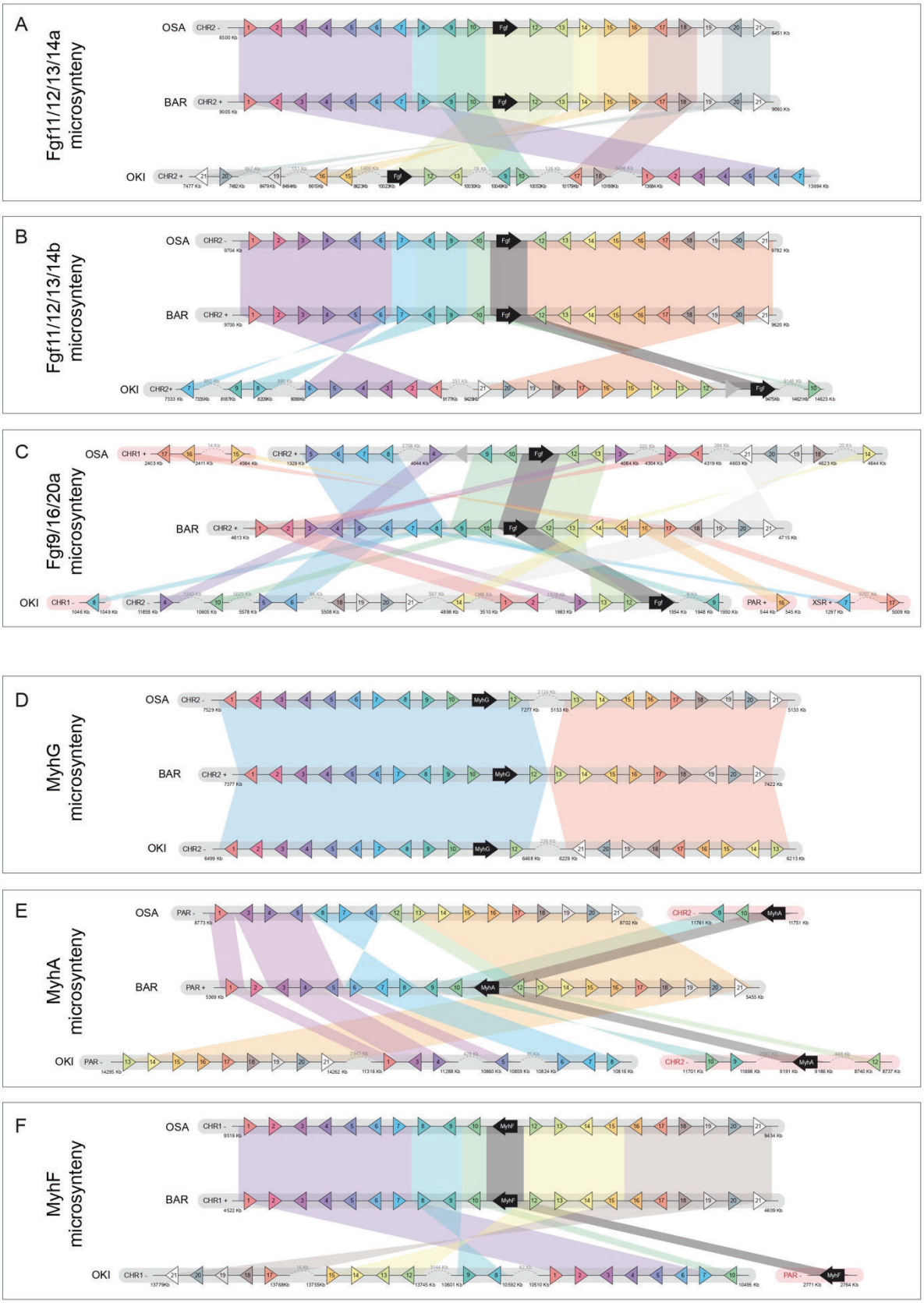
Supplemental Figure S3 (continued)



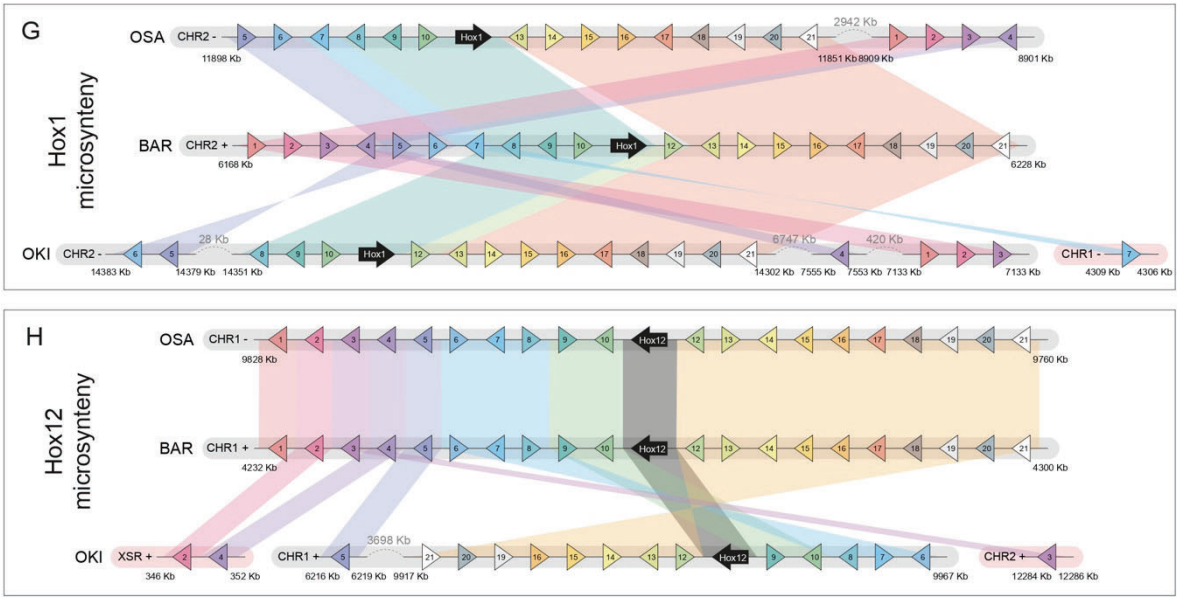
Supplemental Figure S4: Comparative chromosome mapping of the Fgf (A) and Myh genes in the genomes of *O. dioica* from Osaka (OSA), Barcelona (BAR) and Okinawa (OKI).



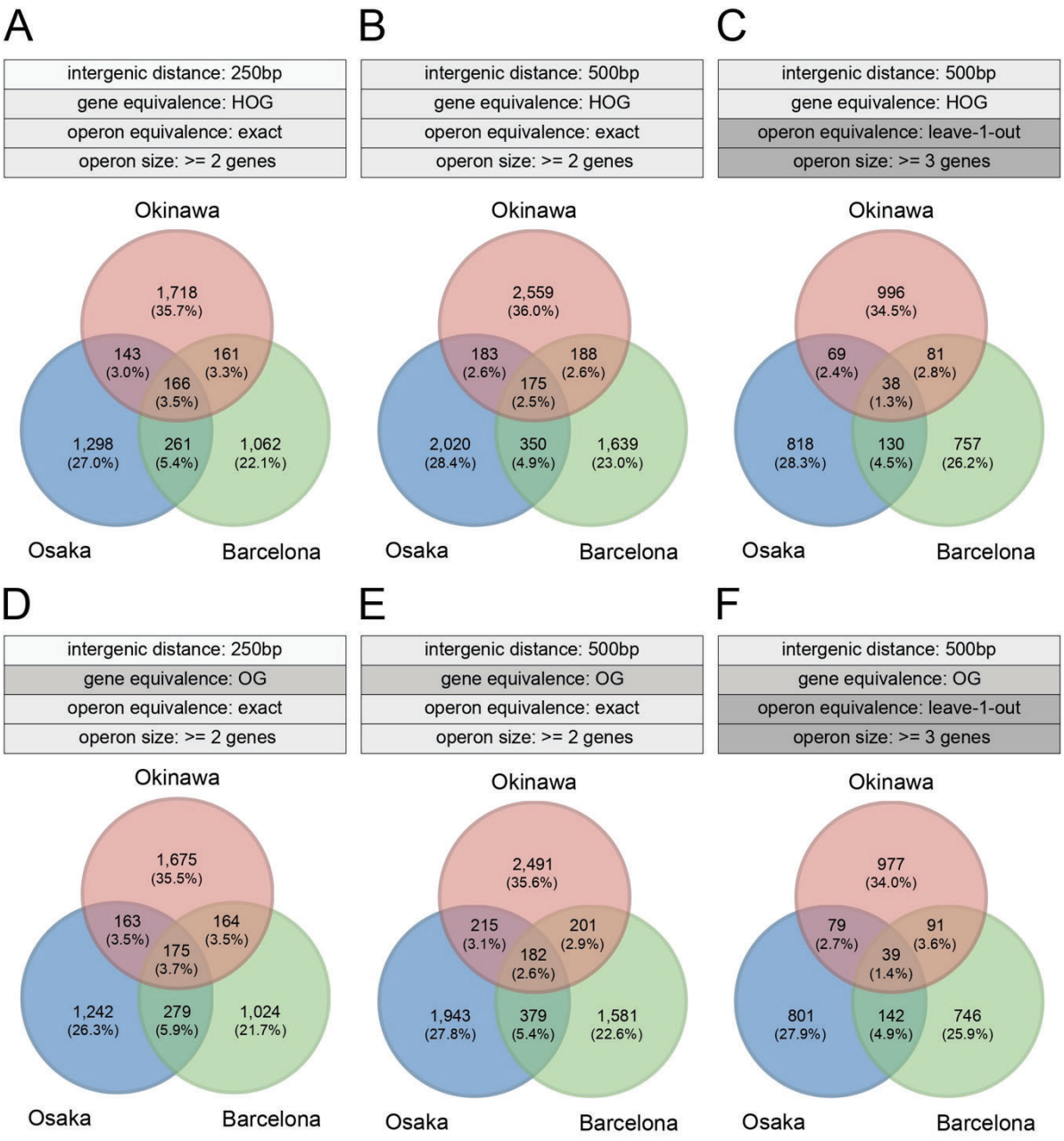
Supplemental Figure S5: Comparative microsynteny analysis of loci surrounding Fgf and Myosin gene family members in *O. dioica*. A: Fgf11/12/13/14a; B: Fgf11/12/13/14b; C: Fgf9/16/20) and Myh (D: MyhG; E: MyhA; F: MyhF). Species names are shortened as follows: Osaka (OSA), Barcelona (BAR) and Okinawa (OKI).



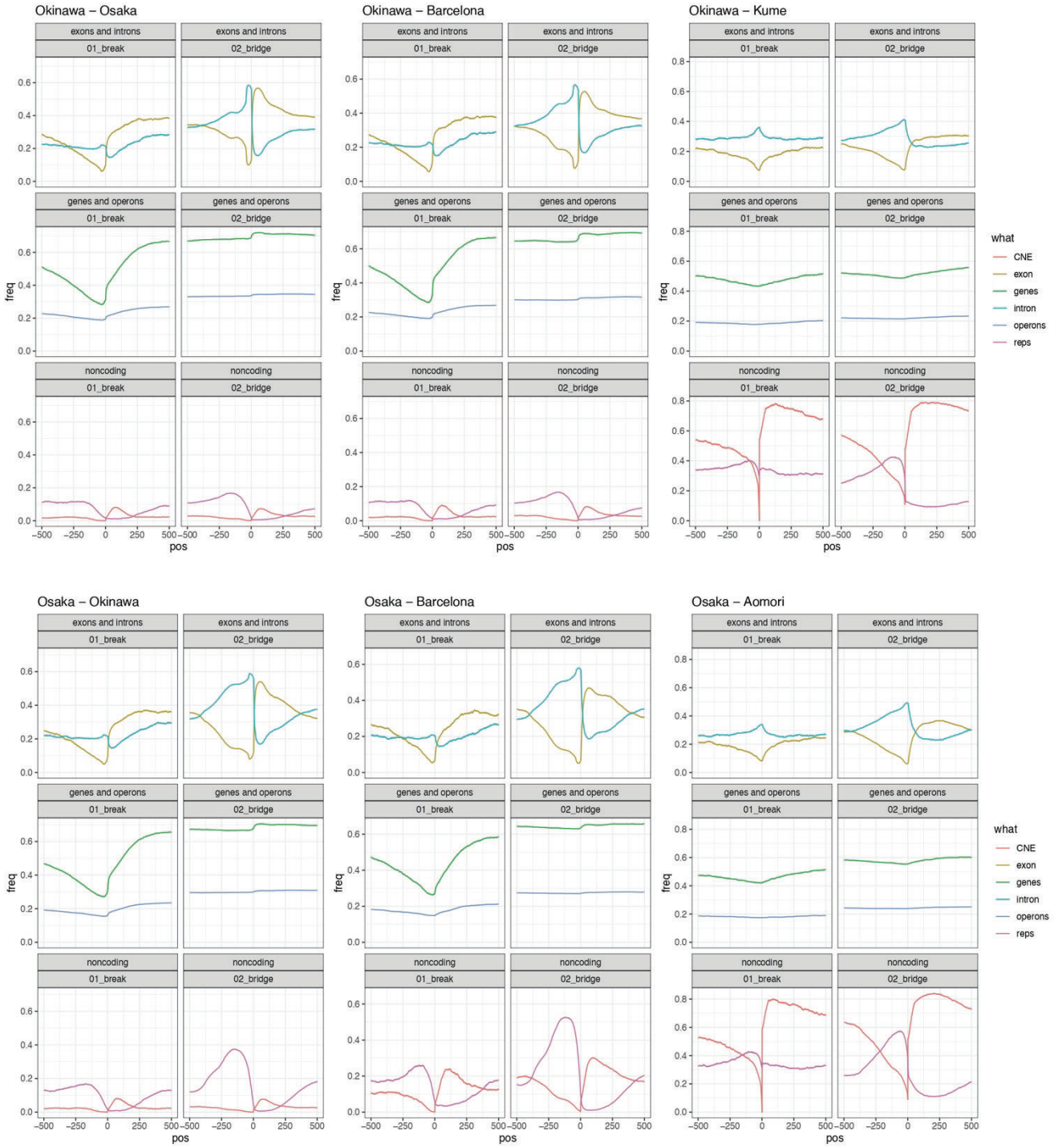
Supplemental Figure S5 (continued)



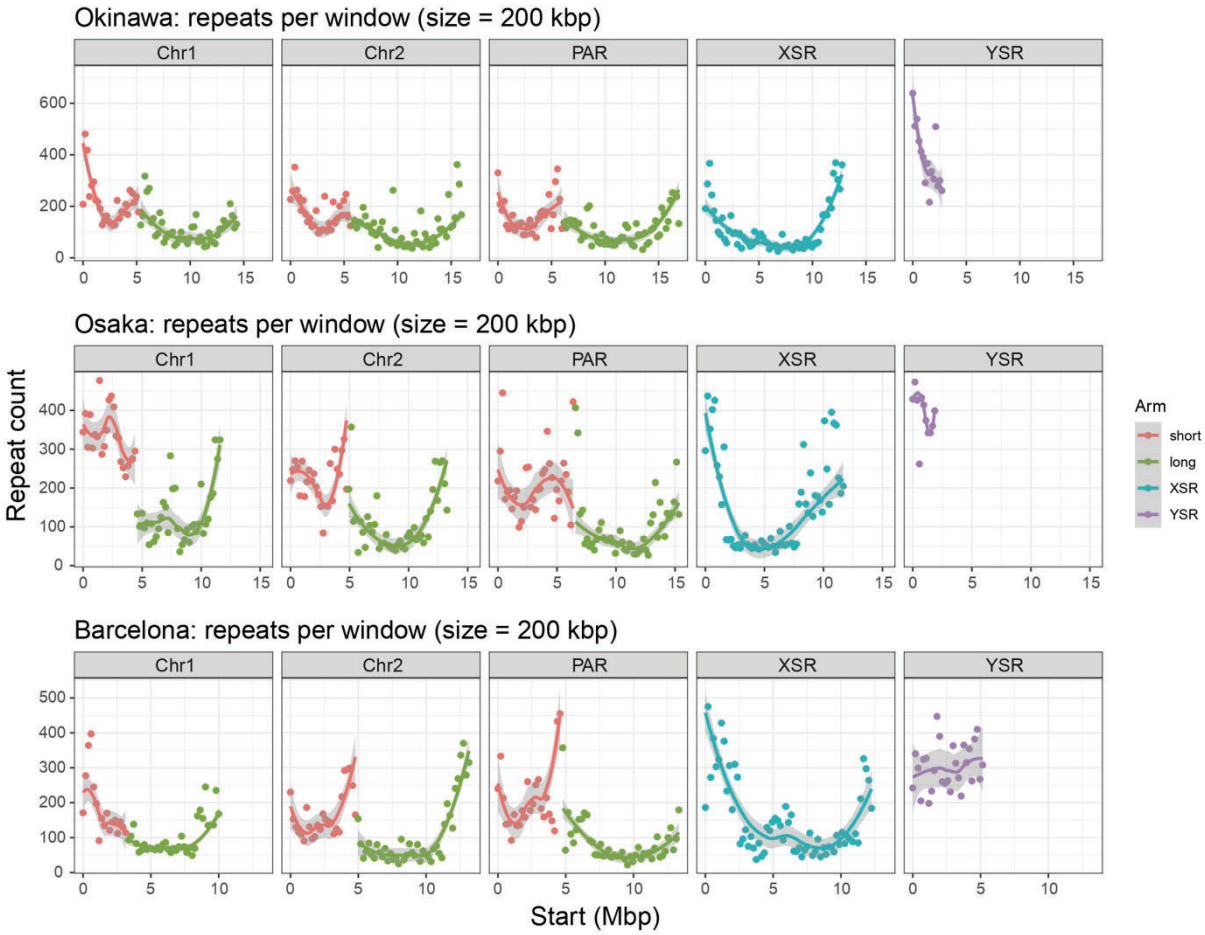
Supplemental Figure S6: Conservation of operons as defined by different matching criteria. We assessed operon conservation using sets of operons defined by different intergenic distances, gene equivalence criteria, and operon equivalence criteria.



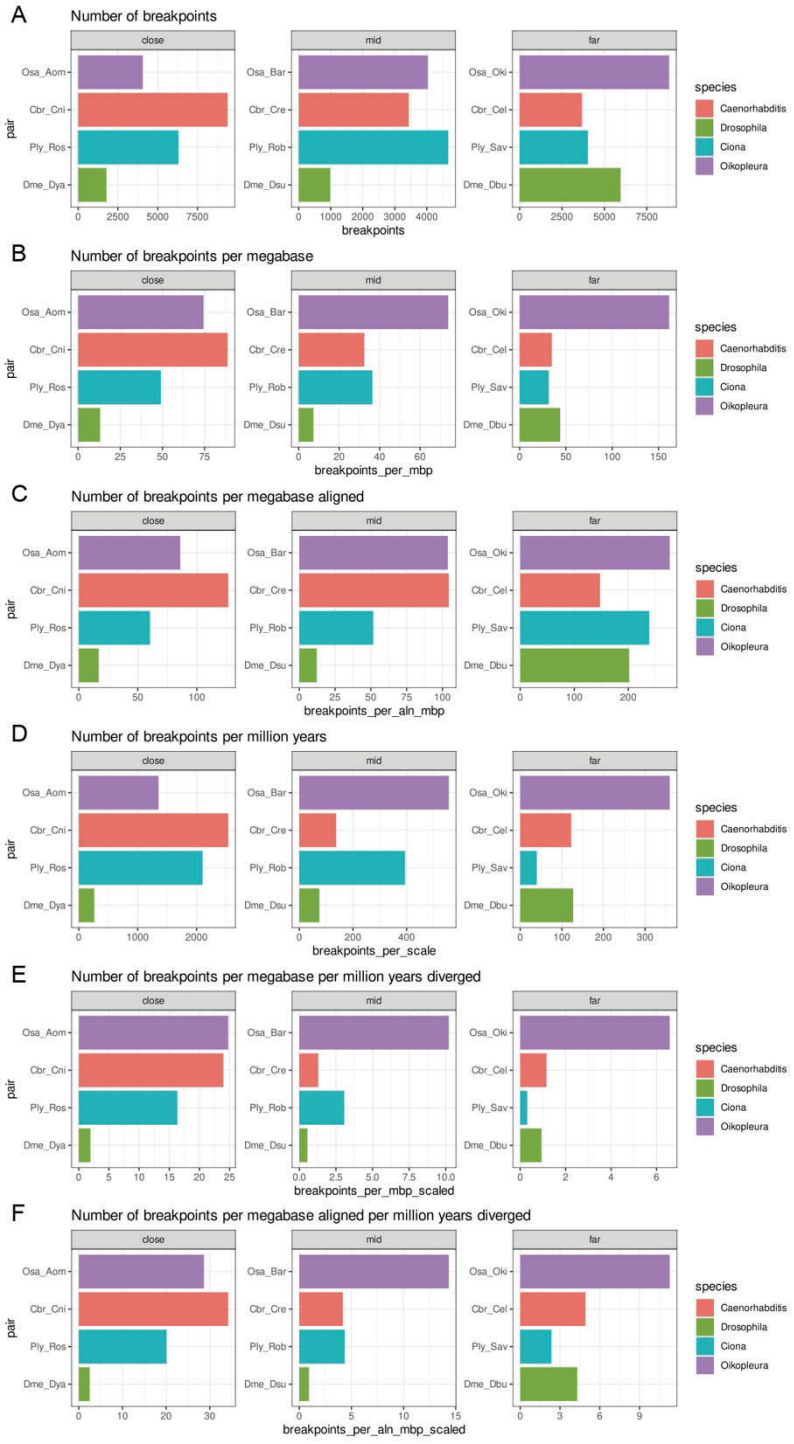
Supplemental Figure S7: Enrichment of genomic features at the boundary between breakpoint or bridge regions and aligned regions in various pairwise comparison (complement to Figure 3C). CNE: conserved non-coding elements. Reps: repeat elements.



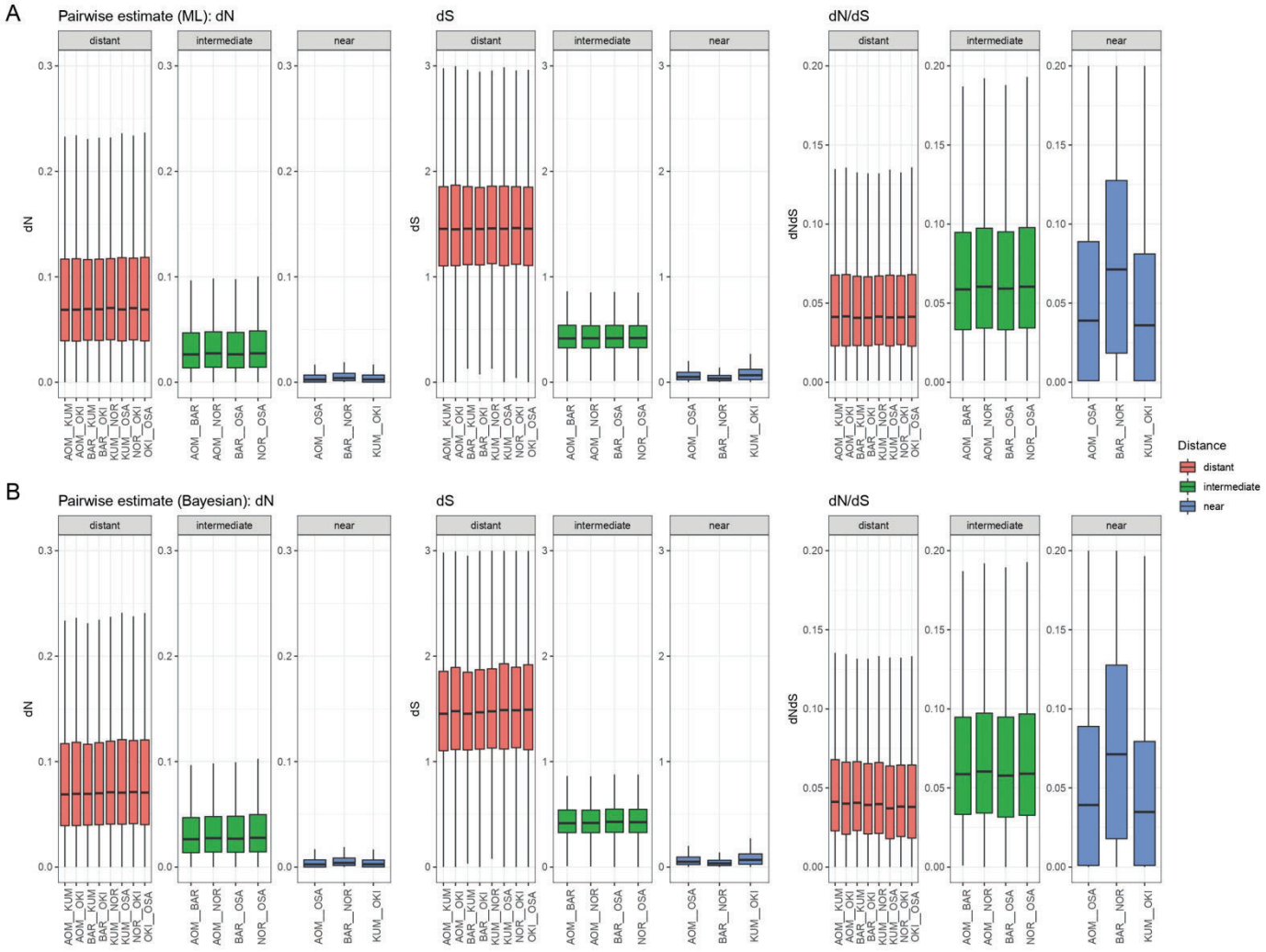
Supplemental Figure S8: Repeat density in *Oikopleura* genomes. The result for the Okinawan genome was originally reported in Bliznina et al. (2021), but is plotted here to facilitate comparisons with *O. dioica* from Osaka and Barcelona.



Supplemental Figure S9: Breakpoint accumulation in *Oikopleura*, *Ciona*, *Drosophila* and *Caenorhabditis* at close (near), mid (intermediary) and far (distant) evolutionary distance. The purpose of this figure is to provide extra context to the main text Figure 7 panel E and help the readers assess our choice for the normalisation. A) Number of breakpoints. B) Number of breakoints normalised by genome size, in megabases. C) Number of breakpoints normalised by genome size, excluding the regions that were not aligned (to take into account for instance that some *Drosophila* genomes assemblies contain very large centromeric regions). D) Number of breakpoints normalised by evolutionary distance, in million years. The evolutionary distance was estimated by a molecular clock (*Oikopleura*, main text Figure 5), or taken from the literature. E) Number of breakpoint normalised by genome size and evolutionary distance. This panel displays the same data as Figure 7E and is the direct output of R scripts (colors and bar orders were then edited in Figure 7E to match the other panels of the figure). F) Number of breakpoints normalised by alignment length (see C.) and evolutionary distance.



Supplemental Figure S10: Pairwise estimations for dN, dS, and dN/dS values for every genome pair. A) Maximum likelihood estimates and B) Bayesian estimates (calculated using the runmode = -2 and runmode = -3 settings in PAML).



Annex 3

Cardiopharyngeal deconstruction and ancestral tunicate sessility

Ferrández-Roldán A, Fabregà-Torrus M*, **Sánchez-Serna G***, Duran-Bello E, Joaquín-Lluís M, Bujosa P, Plana-Carmona M, Garcia-Fernàndez J, Albalat R, Cañestro C.

*Equal contribution

Nature. 2021 Nov;599(7885):431-435. doi: 10.1038/s41586-021-04041-w. Epub 2021 Nov 17. PMID: 34789899.

Article

Cardiopharyngeal deconstruction and ancestral tunicate sessility

https://doi.org/10.1038/s41586-021-04041-w

Received: 10 February 2021

Accepted: 17 September 2021

Published online: 17 November 2021

Check for updates

Alfonso Ferrández-Roldán[1,2], Marc Fabregà-Torrus[1,2,4], Gaspar Sánchez-Serna[1,2,4], Enya Duran-Bello[1,2], Martí Joaquín-Lluís[1,2], Paula Bujosa[1,2], Marcos Plana-Carmona[1,2], Jordi Garcia-Fernàndez[1,3], Ricard Albalat[1,2] & Cristian Cañestro[1,2]✉

A central question in chordate evolution is the origin of sessility in adult ascidians, and whether the appendicularian complete free-living style represents a primitive or derived condition among tunicates[1]. According to the 'a new heart for a new head' hypothesis, the evolution of the cardiopharyngeal gene regulatory network appears as a pivotal aspect to understand the evolution of the lifestyles of chordates[2–4]. Here we show that appendicularians experienced massive ancestral losses of cardiopharyngeal genes and subfunctions, leading to the 'deconstruction' of two ancestral modules of the tunicate cardiopharyngeal gene regulatory network. In ascidians, these modules are related to early and late multipotency, which is involved in lineage cell-fate determination towards the first and second heart fields and siphon muscles. Our work shows that the deconstruction of the cardiopharyngeal gene regulatory network involved the regressive loss of the siphon muscle, supporting an evolutionary scenario in which ancestral tunicates had a sessile ascidian-like adult lifestyle. In agreement with this scenario, our findings also suggest that this deconstruction contributed to the acceleration of cardiogenesis and the redesign of the heart into an open-wide laminar structure in appendicularians as evolutionary adaptations during their transition to a complete pelagic free-living style upon the innovation of the food-filtering house[5].

The discovery that the branching of cephalochordates is basal within chordates, and tunicates therefore are the sister group of vertebrates, provided aview of adult ancestral chordates as free-living organisms in contrast to the sessile ascidian-like lifestyle with alternating motile larva and sessile adults traditionally proposed by Garstang[6–8]. This novel view brought renewed interest in appendicularians, whose complete free-living style could parsimoniously represent the ancestral tunicate condition, considering their most accepted position as the sister group of the remaining tunicates[9–12] (although see ref. [13]). In contrast to ascidians, the development of the heart in appendicularians remains poorly understood, and therefore whether differences between the cardiopharyngeal gene regulatory networks (GRNs) of appendicularians and ascidians reflect adaptations to their different lifestyles remains unknown.

Cardiac developmental atlas and ontogeny

The open-wide laminar heart of appendicularians is considered the simplest chordate heart, consisting just of two layers, the myocardium and the pericardium, the former of which pumps against the stomach[14] (Fig. 1a). Here we provide a developmental atlas of the appendicularian heart and show that cardiogenesis in *Oikopleura dioica* is fast, spanning only 3.5 h from the early hatchling stage (5 h post-fertilization (hpf)), at which no morphological evidence of the cardiac primordium could yet be distinguished, until the late hatchling stage (8.5 hpf), when the heart began to beat (Supplementary Video 1).

Analysis of muscular *Actin 1* (*ActnM1*)[15] expression (in lieu of *Mesp*, which is the preferred precardiac marker in ascidians[16], but is absent in appendicularians (Fig. 1b)) integrated with data from 4D microscopy nuclear tracing[17] (Fig. 1c, Extended Data Fig. 1) identified B8.9 blastomere at the incipient tailbud stage as the first cardiac progenitor cell (CPC). Our analysis revealed that cardiac cells shared lineage with the first three anterior tail muscle cells, following the same ontogenetic origin as in ascidians[17] and therefore provided evidence that ascidian and appendicularian hearts are homologous.

Loss of cardiopharyngeal GRN early module

In vertebrates, the cardiopharyngeal field is the developmental domain that gives rise to the heart and branchiomeric muscles from a common pool of early cardiopharyngeal multipotent progenitors. After a binary-stepwise process of fate choices, early cardiopharyngeal progenitors give rise to the first and second heart fields and to branchiomeric muscles in the head and neck[2]. In ascidians, pharyngeal muscles (that is, siphon and longitudinal muscles) are considered homologous to vertebrate branchiomeric muscles, and their cardiopharyngeal GRN is

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain. [2]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain. [3]Institut de Biomedicina (IBUB), Universitat de Barcelona, Barcelona, Spain. [4]These authors contributed equally: Marc Fabregà-Torrus and Gaspar Sánchez-Serna. ✉e-mail: canestro@ub.edu

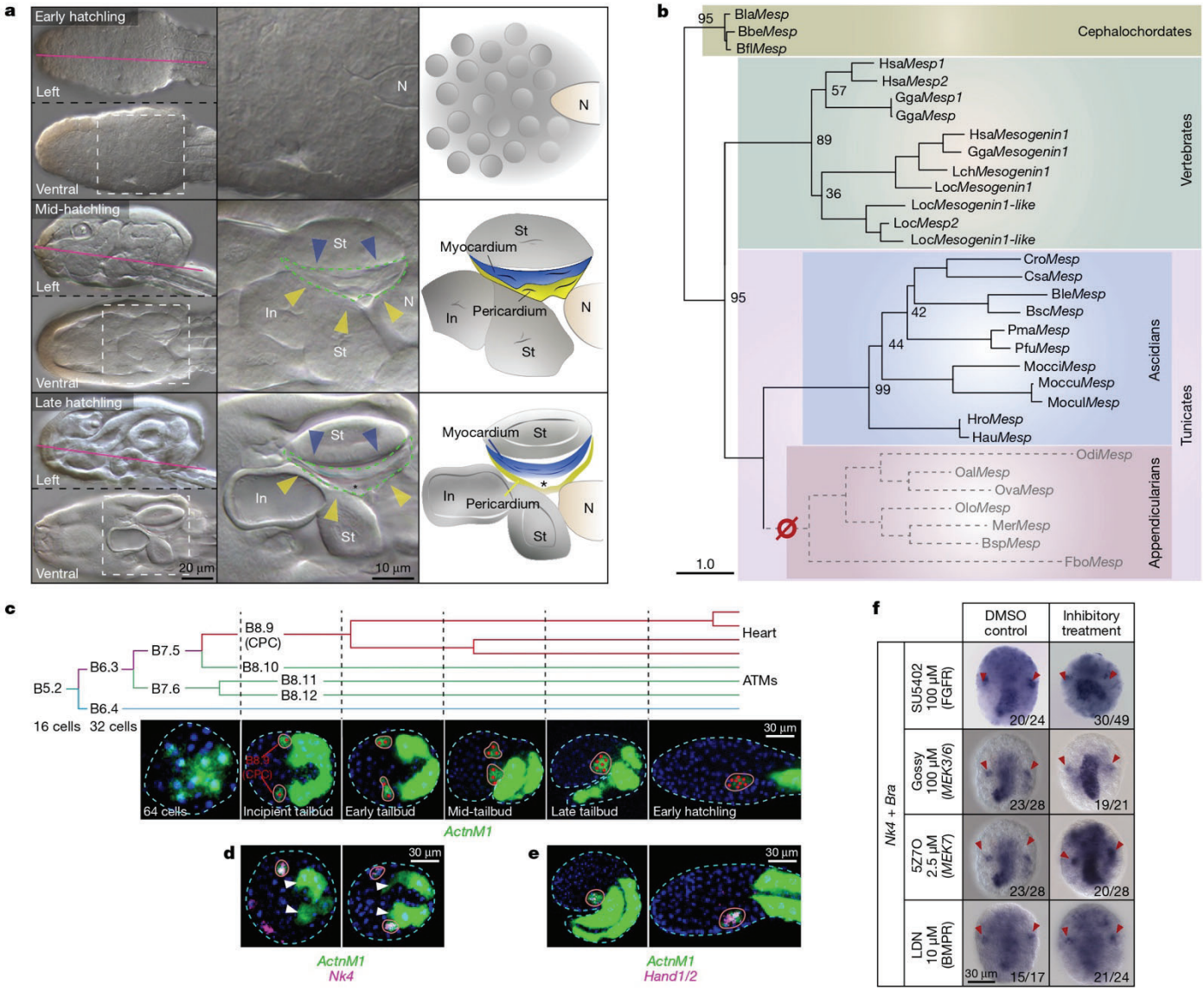


Fig. 1 | *O. dioica* cardiogenesis. a, Developmental cardiac atlas. In early hatchlings, no cardiac morphology is discernible near the notochord (N). In mid-hatchlings, the heart primordium is made of two layers—the myocardium (blue arrowheads) and the pericardium (yellow arrowheads)—located in the space between the stomachs (St), intestine (In) and notochord. In late hatchlings, the internal cardiac space (asterisk) has expanded, and the heart (which has asymmetrically relocated towards the left) begins to beat. Pink lines indicate ventral focal planes. Right, schematic representations. **b**, ML phylogenetic tree shows the loss of *Mesp* in appendicularians (red) and its presence in the rest of chordates (lost genes in grey; species abbreviations can be found in Supplementary Data 1). The scale bar indicates amino acid substitutions. Bootstrap values are shown. **c**, Integration of the cell lineage fate map reconstructed from 4D nuclear tracing (modified from ref. [16]) with confocal optical sections showing expression of *ActnM1* (green) and nuclear staining (blue) reveals B8.9 as the first CPC of the cardiac lineage (red lines) at the incipient tailbud stage (2 h 20 min post-fertilization), and its split from the anterior tail muscle (ATM) cell lineage (B8.10-B8.12; green lines) at the 64-cell stage, and the germline (B6.4; blue line) at the 32-cell stage (Extended Data Fig. 1). The red lines encircle CPCs (red dots). The dashed lines delineate embryos. **d**, Fluorescent whole-mount in situ hybridization (FWMISH) showed transient co-expression of *ActnM1* and *Nk4* in CPCs from incipient tailbud to mid-tailbud stages, but never in anterior tail muscles (white arrowheads). **e**, After *Nk4* downregulation, *ActnM1* and *Hand1/2* co-expressed in the CPCs at late tailbud and early hatchling stages. **f**, WMISH with the cardiac marker *Nk4* and the notochord marker *Brachyury* in DMSO control and treated embryos with inhibitors of the FGF–MEK and BMP signalling pathways (inhibited targets are in parenthesis; the number counts of the phenotypes out of the analysed embryos are also shown) show no effect of these signalling pathways on CPC (red arrowheads) determination nor the activation of the cardiogenic kernel. BMPR, BMP receptor; FGFR, FGF receptor.

highly conserved with vertebrates using a homologous binary-stepwise model[4,17–20]. Consistent with this model, the precardiac master regulator *Mesp* is expressed in multipotent pre-CPCs in both vertebrates and ascidians[21,22], followed by FGF–MAPK signalling mediated by ETS1/2 phosphorylation, and finally the activation of the cardiogenic kernel (that is, *Gata4/5/6*, *FoxF*, *Nk4* and *Hand1/2*) and BMP signalling[18,20,23–25].

Our genomic survey of seven appendicularian species and eleven ascidians revealed the absence of *Mesp*, *Ets1/2b*, *Gata4/5/6*, *Mek1/2* and *Hand-r* homologues in all analysed appendicularians, whereas they were present in all ascidians (Fig. 1b, Extended Data Figs. 2–8). Phylogenetic analyses suggested that the absence of these genes was probably due to ancestral gene losses that occurred at the base of the appendicularian lineage after its split from ascidians. Following the surprising absence of a homologue of *Mesp*, considering its precardiac master role in ascidians and vertebrates[21,22,26], we tested for the possibility of 'function shuffling' among *Mesp*-related basic helix–loop–helix

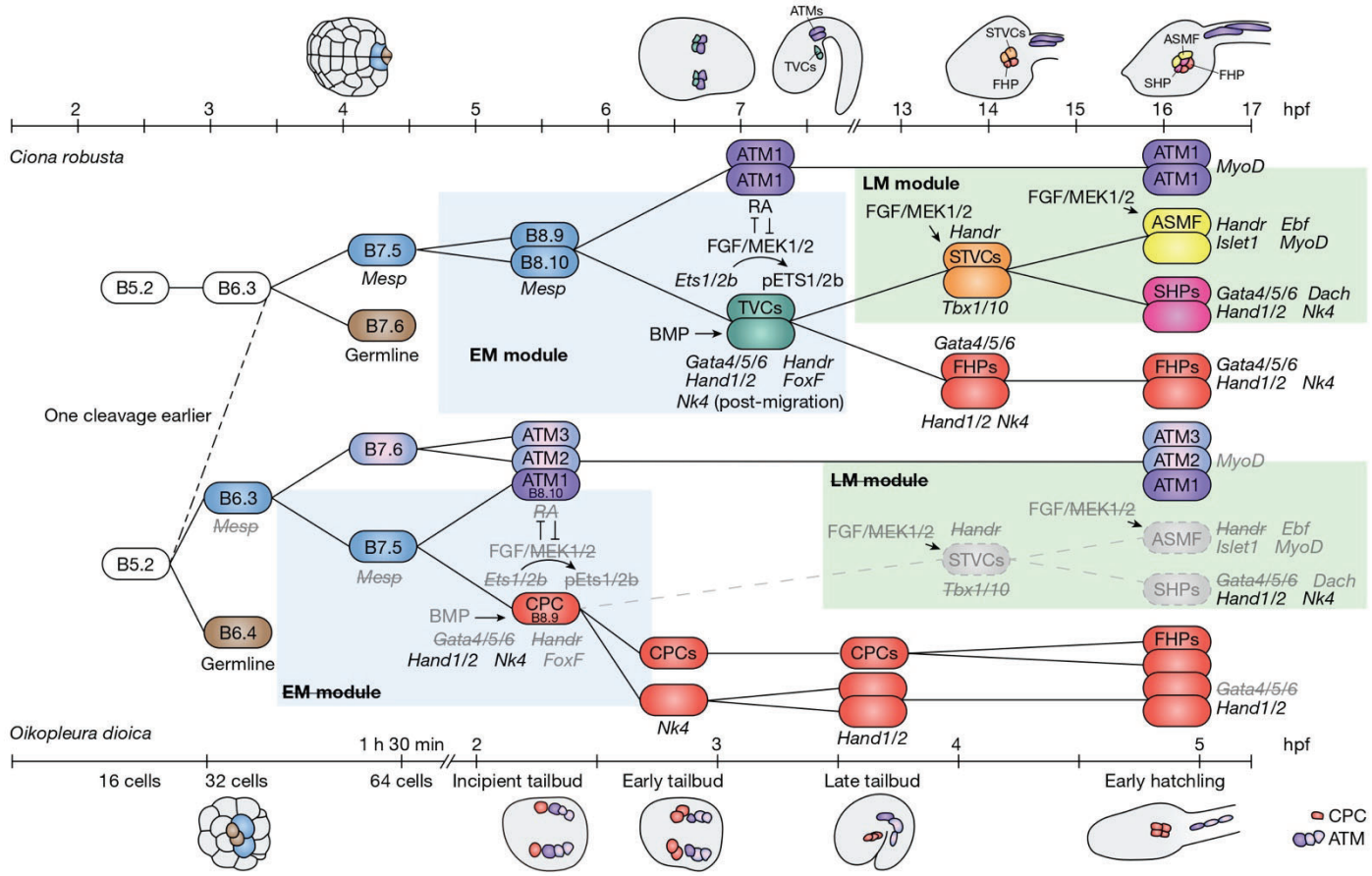


Fig. 2 | Comparison of the cardiopharyngeal cell lineage and GRN in ascidians and appendicularians. The determination of the cardiopharyngeal lineage (blue) and its split from the germline (brown) occurs one cleavage earlier in *O. dioica* than in the ascidian *C. robusta*. In contrast to *O. dioica*, in ascidians, the precardiac *Mesp*-positive cell B7.5 divides before the split of the anterior tail muscle lineage (ATM; purple) and the cardiopharyngeal lineage (blue). In *O. dioica*, the daughter cell of B7.5 rapidly activates the expression of the cardiogenic kernel (*Nk4* and *Hand1/2*) and becomes a CPC (red), whereas in ascidians, their counterpart TVCs (green) maintain a multipotent state, which will give rise to the first heart precursors (FHPs; red), second heart precursors (SHPs; pink) and atrial siphon muscle field (ASMF; yellow), the last two through intermediate secondary multipotent cells (STVCs; orange). The lack of *Dach* expression in the heart of *O. dioica* suggests the absence of a homologue of the ascidian second heart field. The numerous losses of cardiopharyngeal genes (grey strikethrough) and subfunctions (grey) highlight the deconstruction of the 'early' and 'late' ancestral multipotent GRN modules related with the early precardiac multipotency (EM module) of the TVCs in ascidians, and the late multipotency (LM module) of the STVCs and their derivatives such as pharyngeal muscles and second heart field, respectively. The developmental timelines depict the acceleration of cardiogenesis in *O. dioica* compared with ascidians, with the differentiation of the first CPC (red) as soon as 2.5 hpf in *O. dioica*.

(bHLH) genes. However, we did not observe any expression domain of *Math* or *Neurogenin* (the closest bHLH genes according to blast analysis) that were compatible with precardiac progenitors. We also tested for the possibility of function shuffling among appendicularian-specific duplications of paralogues closely related to the lost genes (that is, two *Ets1/2a* and four *Gata1/2/3*) as well as *MEK7* and *MEK3/6*, but again no tissue-specific expression domains compatible with CPCs were observed. We found homologues of *Nk4*, *Hand1/2* and *FoxF*. While *Nk4*, first, and then *Hand1/2* were sequentially expressed in the CPCs at incipient tailbud and mid-tailbud stages, respectively (Fig. 1d, e), no expression of *FoxF* was observed in the CPCs. These results showed that the first CPC (B8.9) resulting from the split of the anterior tail muscle lineage (B8.10) did not maintain a multipotent state (as their counterpart trunk ventral cells (TVCs) do in ascidians), but rapidly activated the expression of the cardiogenic kernel (*Nk4* and *Hand1/2*). In contrast to ascidians and vertebrates, this activation had become independent of *Mesp*, *Ets1/2*-mediated FGF signalling, *FoxF* and *Gata4/5/6* (ref.[27]). The co-elimination of *Mesp*, *Ets1/2b*, *MEK1/2*, *Gata4/5/6* and the loss of the cardiac subfunction of *FoxF* highlight the deconstruction of what can be considered an ancestral 'early multipotent' module that in ascidians and vertebrates is related to the early maintenance of the multipotent state of the precardiac progenitors[3], and consequently result in an accelerated cardiogenesis in *O. dioica* (Fig. 2).

The deconstruction of this early multipotent module suggested that the conserved cardiogenic roles of FGF–MAPK and BMP signalling in ascidians and vertebrates could have also been altered in appendicularians. To test this hypothesis, we performed inhibitory treatments against the FGF receptor, the surviving paralogues *MEK3/6* and *MEK7*, and BMP receptors, at different concentrations and time windows, and treated embryos were analysed by whole-mount in situ hybridization (Extended Data Fig. 9, Supplementary Data 2). Results revealed that the formation of the CPCs and the onset of *Nk4* expression were not altered in treated embryos in which FGF–MAPK or BMP signalling pathways had been inhibited (Fig. 1f). These results suggested that the determination and differentiation of the CPCs and the onset of the cardiogenic kernel had become independent of these two signalling pathways during the evolution of appendicularians.

Loss of cardiopharyngeal GRN late module

In ascidians, TVCs undertake a series of asymmetric cell divisions and regulatory transient secondary multipotent states (that is, STVCs) that

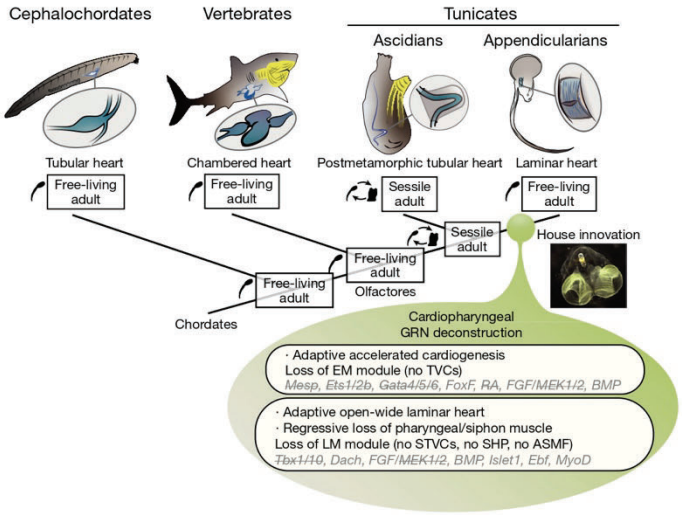


Fig. 3 | Evolutionary scenario of the deconstruction of the cardiopharyngeal GRN and acquisition of an adult free-living style in appendicularians. The evolutionary innovation of an accelerated cardiogenesis, an open-wide laminar heart and the loss of the pharyngeal/siphon muscle in appendicularians can be connected to the deconstruction of their cardiopharyngeal GRN. These three innovations accompanied and plausibly facilitated the transition from an ascidian-like ancestral condition alternating motile larva and sessile adults, to a complete free-living lifestyle of appendicularians upon their innovation of the house. The strikethrough and grey indicate loss of cardiopharyngeal genes and subfunctions, respectively.

not only give rise to the first and second heart precursors but also to the atrial siphon muscle founder cells[4]. We surveyed the appendicularian homologous genes encoding cardiopharyngeal transcription factors such as *Hand-r*, *Tbx1/10*, *Islet1* and *Ebf* (also known as *Coe*), which in ascidians become activated in an FGF–MAPK-dependent manner to determine the trajectory towards the atrial siphon muscles, including the activation of *MyoD* (also known as *Mrf*)[4,19,28] (Extended Data Fig. 10). In addition to the aforementioned absence of *Hand-r*, no *Tbx1/10* was present in any appendicularian, suggesting again an ancestral gene loss in the appendicularian lineage. We found single homologues for *Islet1*, *Ebf* and *MyoD*; the two former genes were expressed in the nervous system, and the latter in the oikoplastic epithelium, but no expression was found for any of these homologues in the trunk that could suggest the presence of a homologous tissue to the atrial or any other pharyngeal muscle. To test for the presence of a presumptive second heart field in *O. dioica*, we analysed the expression of the homologue of *Dach*, which in ascidians is activated by *Tbx1/10* in the absence of FGF–MAPK signalling, and is necessary to determine the identity of the second heart precursors[19]. Whole-mount in situ hybridization revealed that, while the single homologue of *Dach* was expressed in the nervous system, the endostyle and the trunk epidermis, no expression was detected in the heart, suggesting the absence of a second heart field homologue in *O. dioica* (Extended Data Fig. 10). In addition to *Dach*, our genome survey in *O. dioica* revealed the absence of 12 out of 25 genes that a recent single-cell transcriptomic analysis had revealed to be specific for the first or second heart field precursors in ascidians[19] (Supplementary Data 3). In summary, our results highlight that during the deconstruction of the cardiopharyngeal GRN, the loss of *Tbx1/10* and the loss of cardiopharyngeal subfunctions for *Dach*, *Islet1*, *Ebf* and *MyoD* might represent the loss of an ancestral 'late multipotent' module that is required for regulating secondary multipotency and differentiation of the second heart field and atrial muscle in ascidians, structures that appear to have been lost during the evolution of appendicularians[15] (Fig. 2).

Discussion

Homology and one cleavage earlier trend

Understanding the evolution of the cardiopharyngeal GRN in appendicularians and ascidians is key to infer the ancestral lifestyle of tunicates. Our work provides strong evidence supporting the homology between the hearts of ascidians and appendicularians despite their remarkable differences of morphology, developmental pace and physiology. A key difference is that the specification of the cardiopharyngeal cell lineage and its split from the tail muscle cell lineage occurs one cleavage earlier in appendicularians than in ascidians (Fig. 2). This 'one cleavage earlier' trend can be considered a general aspect of the evolution of the development of appendicularians, consistent with the observation already made by Delsman[29] that gastrulation in *O. dioica* occurred one cleavage earlier than in ascidians, and later corroborated by modern studies[16,30]. This trend has probably contributed to developmental acceleration, morphological simplification and reduction in cell number in this group of tunicates.

Deconstruction and evolutionary impact

One of the most striking findings of our work is the numerous losses of genes and subfunctions that highlight a process of deconstruction of the cardiopharyngeal GRN in appendicularians. The term 'deconstruction', originally coined in philosophy and later applied in literature, architecture, fashion and cookery, or even developmental biology[31], is not synonymous with destruction or homogenously distributed erosion, but instead it refers to the process of dismantling or breaking apart elements that traditionally are combined, and whose analysis facilitates the recognition of structural modules. Our EvoDevo work here, in agreement with the modular model for the control of heart cell identity proposed by Wang et al.[19], unveils the deconstruction of 'evolvable modules' of the cardiopharyngeal GRN, accompanied by developmental system drift and GRN rewiring during the evolution of the heart and pharyngeal muscle in appendicularians (Fig. 2). The loss of the 'early' and 'late' multipotent ancestral modules correlates with the loss of the multipotent states that in ascidians are maintained in the TVC and STVCs, respectively. The losses of these two ancestral modules can be connected to three evolutionary innovations that accompanied, and plausibly facilitated, the evolution from an ancestral sessile ascidian-like adult lifestyle to the pelagic fully free lifestyle of appendicularians: (1) an accelerated cardiogenesis, (2) the formation of an open-wide laminar heart, and (3) the loss of the siphon muscle (Fig. 3).

First, the accelerated cardiogenesis driven by the 'one-cleavage earlier' CPC specification and by the deconstruction of the GRN is probably the result of a primary adaptation to the faster development in appendicularians than in other tunicates. Moreover, accelerated cardiogenesis also enabled the heart to adaptatively begin beating as soon as 8.5 hpf in *O. dioica*—in contrast to a few days post-metamorphosis in ascidians—driving haemolymph circulation to be ready when juveniles inflate the first house (10 hpf) and begin pelagic filter feeding.

Second, the low number of cardiac cells (the myocardium is made of only six cells[14]) together with the apparent loss of the second heart field homologue in appendicularians is compatible with the transformation of an ascidian-like tubular heart into an open-wide laminar heart that beats against the stomach. Considering that haemolymph circulation in appendicularians is not only powered by the heart but also by tail movements[32], the adaptive innovation of a laminar cardiac structure plausibly offered a more efficient system to pump haemolymph waves propelled by the tail movements through an open-wide structure than through the less accessible space of a tubular ascidian-like heart.

Third, the loss of the 'body wall' and pharyngeal/siphon muscles in the trunk of *O. dioica*[15] can be considered the result of regressive evolution during the transition from a sessile ascidian-like to the pelagic style of appendicularians, in which their functions in sessile ascidians (siphon opening/closing and water squirting as a response to large

debris, predators, low tide, or the ejection of faeces or gametes) became useless upon the innovation of the house in appendicularians.

Future perspective

This work exemplifies how the study of gene loss and the application of the concept of deconstruction in evolutionary biology facilitates the recognition of modules and rewiring of the GRN to better understand the evolution of species. Our study, for instance, supports an evolutionary scenario in which the deconstruction of the cardiopharyngeal GRN was linked to regressive loss of features that characterize the ascidian-like sessile lifestyle such as the siphon muscles, and to the evolution of the accelerated cardiogenesis and the transformation to a laminar heart that could have been adaptatively selected during the transition of appendicularians to a pelagic complete free-living active style connected to the innovation of the house (Fig. 3). Our evidence, supporting the view that the last common tunicate ancestor had a biphasic lifestyle alternating motile larva and sessile adults[3,16], is compatible with the commonly accepted assumption that appendicularian branching is basal among tunicates, but it is also compatible with the possibility that appendicularians are phylogenetically related to some groups of ascidians (that is, Aplousobranchia[13]). Thus, our work provides a useful framework for future comparative studies of the cardiopharyngeal GRN among different tunicates, as well as for future efforts to clarify the potential neotenic origin of appendicularians and their phylogenetic relationship with other tunicates.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04041-w.

1. Satoh, N. in *Chordate Origins and Evolution* (ed. Satoh, N.) 17–30 (Academic, 2016).
2. Diogo, R. et al. A new heart for a new head in vertebrate cardiopharyngeal evolution. *Nature* **520**, 466–473 (2015).
3. Razy-Krajka, F. & Stolfi, A. Regulation and evolution of muscle development in tunicates. *Evodevo* **10**, 1–34 (2019).
4. Stolfi, A. et al. Early chordate origins of the vertebrate second heart field. *Science* **565**, 565–569 (2010).
5. Mikhaleva, Y., Skinnes, R., Sumic, S., Thompson, E. M. & Chourrout, D. Development of the house secreting epithelium, a major innovation of tunicate larvaceans, involves multiple homeodomain transcription factors. *Dev. Biol.* **443**, 117–126 (2018).
6. Garstang, W. The morphology of the Tunicata, and its bearings on the phylogeny of the Chrodata. *Quar. J. Micr. Sci.* **72**, 51–186 (1928).
7. Bourlat, S. J. et al. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006).
8. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
9. Swalla, B. J., Cameron, C. B., Corley, L. S. & Garey, J. R. Urochordates are monophyletic within the deuterostomes. *Syst. Biol.* **49**, 52–64 (2000).
10. Delsuc, F. et al. A phylogenomic framework and timescale for comparative studies of tunicates. *BMC Biol.* **16**, 39 (2018).
11. Kocot, K. M., Tassia, M. G., Halanych, K. M. & Swalla, B. J. Phylogenomics offers resolution of major tunicate relationships. *Mol. Phylogenet. Evol.* **121**, 166–173 (2018).
12. Braun, K., Leubner, F. & Stach, T. Phylogenetic analysis of phenotypic characters of Tunicata supports basal Appendicularia and monophyletic Ascidiacea. *Cladistics* **36**, 259–300 (2020).
13. Stach, T. Ontogeny of the appendicularian *Oikopleura dioica* (Tunicata, Chordata) reveals characters similar to ascidian larvae with sessile adults. *Zoomorphology* **126**, 203–214 (2007).
14. Nishida, H., Ohno, N., Caicci, F. & Manni, L. 3D reconstruction of structures of hatched larva and young juvenile of the larvacean *Oikopleura dioica* using SBF-SEM. *Sci. Rep.* **11**, 1–14 (2021).
15. Almazán, A., Ferrández-Roldán, A., Albalat, R. & Cañestro, C. Developmental atlas of appendicularian *Oikopleura dioica* actins provides new insights into the evolution of the notochord and the cardio-paraxial muscle in chordates. *Dev. Biol.* **448**, 260–270 (2019).
16. Stach, T., Winter, J., Bouquet, J.-M. M., Chourrout, D. & Schnabel, R. Embryology of a planktonic tunicate reveals traces of sessility. *Proc. Natl Acad. Sci. USA* **105**, 7229–7234 (2008).
17. Davidson, B. *Ciona intestinalis* as a model for cardiac development. *Semin. Cell Dev. Biol.* **18**, 16–26 (2007).
18. Christiaen, L., Stolfi, A. & Levine, M. BMP signaling coordinates gene expression and cell migration during precardiac mesoderm development. *Dev. Biol.* **340**, 179–187 (2010).
19. Wang, W. et al. A single-cell transcriptional roadmap for cardiopharyngeal fate diversification. *Nat. Cell Biol.* **21**, 674–686 (2019).
20. Racioppi, C., Wiechecki, K. A. & Christiaen, L. Combinatorial chromatin dynamics foster accurate cardiopharyngeal fate choices. *eLife* **8**, 1–33 (2019).
21. Lescroart, F. et al. Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nat. Cell Biol.* **16**, 829–840 (2014).
22. Satou, Y., Imai, K. S. & Satoh, N. The ascidian Mesp gene specifies heart precursor cells. *Development* **131**, 2533–2541 (2004).
23. Davidson, B., Shi, W., Beh, J., Christiaen, L. & Levine, M. FGF signaling delineates the cardiac progenitor field in the simple chordate, *Ciona intestinalis*. *Genes Dev.* **20**, 2728–2738 (2006).
24. Bernadskaya, Y. Y., Brahmbhatt, S., Gline, S. E., Wang, W. & Christiaen, L. Discoidin-domain receptor coordinates cell-matrix adhesion and collective polarity in migratory cardiopharyngeal progenitors. *Nat. Commun.* **10**, 57 (2019).
25. Shi, Y., Katsev, S., Cai, C. & Evans, S. BMP signaling is required for heart formation in vertebrates. *Dev. Biol.* **224**, 226–237 (2000).
26. Davidson, B., Shi, W. & Levine, M. Uncoupling heart cell specification and migration in the simple chordate *Ciona intestinalis*. *Development* **132**, 4811–4818 (2005).
27. Schachterle, W., Rojas, A., Xu, S.-M. & Black, B. L. ETS-dependent regulation of a distal Gata4 cardiac enhancer. *Dev. Biol.* **361**, 439–449 (2012).
28. Wang, W., Razy-Krajka, F., Siu, E., Ketcham, A. & Christiaen, L. NK4 antagonizes Tbx1/10 to promote cardiac versus pharyngeal muscle fate in the ascidian second heart field. *PLoS Biol.* **11**, e1001725 (2013).
29. Delsman, H. C. Contributions on the ontogeny of *Oikopleura dioica*. *Verch. Rijksinst. Onderz. Zee* **3**, 1–24 (1910).
30. Fujii, S., Nishio, T. & Nishida, H. Cleavage pattern, gastrulation, and neurulation in the appendicularian, *Oikopleura dioica*. *Dev. Genes Evol.* **218**, 69–79 (2008).
31. Hogan, B. Deconstructing the genesis of animal form. *Development* **131**, 2515–2520 (2004).
32. Fenaux, R. in *The Biology of Pelagic Tunicates* (ed. Bone, Q.) 25–34 (Oxford Univ. Press, 1998).

Methods

Biological material

O. dioica specimens were obtained from the Mediterranean coast of Barcelona (Catalonia, Spain). Culturing of *O. dioica* and embryo collections were performed as previously described[33]. This project did not involve any ethical issues related to informed consent, data protection issues, or humans. Experimentation on aquatic invertebrate animals such as the planktonic *Oikopleura dioica* is not subject to regulations regarding animal experimentation, because this applies only to vertebrate organisms (Real Decreto 223 14-3-1998, in Catalonia Ley 5/1995,DOGC2073,5172). Nevertheless, experimental procedures followed EU animal care guidelines and were approved by the Ethical Animal Experimentation Committee (CEEA-2009) of the University of Barcelona.

Genome database searches and phylogenetic analysis

Protein sequences from the tunicate *Ciona robusta* and the vertebrate *Homo sapiens* were used as queries in BLASTp and tBLASTn searches in genome databases of selected species: https://blast.ncbi.nlm.nih.gov/Blast.cgi for *Branchiostoma floridae*, *Branchiostoma belcheri*, *Branchiostoma lanceolatum, Gallus gallus*, *Lepisosteus oculatus* and *Latimeria chalumnae*; http://www.aniseed.cnrs.fr/ for the ascidian species[34] *Ciona savignyi*, *Phallusia fumigata*, *Phallusia mammillata*, *Halocynthia roretzi*, *Halocynthia aurantium*, *Botryllus schlosseri*, *Botryllus leachii*, *Molgula occulta*, *Molgula oculata* and *Molgula occidentalis*; and http://oikoarrays.biology.uiowa.edu/Oiko/ for *O. dioica*, and for six other appendicularian species with public genomes available in GeneBank (*Oikopleura albicans* SCLG01000000, *Oikopleura vanhoeffeni* SCLH01000000, *Oikopleura longicauda* SCLD01000000, *Mesochordaeus erythrocephalus* SCLF01000000, *Bathochordaeus stygius SCLE01000000* and *Fritillaria borealis* SDII01000000)[35]. The orthology between potential cardiac genes was initially assessed by blast reciprocal best hit (BRBH) and subsequently by phylogenetic analysis based on maximum likelihood (ML) inferences calculated with PhyML v3.0 and an automatic substitution model[36] using protein alignment generated by MUSCLE and reviewed manually with the package AliView v1.17.1 (ref.[37]). Species abbreviations and gene accession numbers are provided in Supplementary Data 1.

Cardiac lineage tracing using 4D microscopy

Cardiac lineage tracing was performed using Supplementary Video 2 from Stach et al.[16]. We followed cell divisions starting from B5.2 blastomere at the 16-cell stage until the late tailbud stage when the CPC divides from the first anterior tail muscle cell. Blastomere nomenclature follows that of Conklin for ascidians (vegetal blastomeres in capital letters, animal blastomeres in small letters, and blastomeres from the right underlined)[38].

Cloning and expression analysis

O. dioica genes were PCR amplified from cDNA obtained as previously described[39]. Then, they were cloned using the Topo TA Cloning Kit (K4530-20, Invitrogen) to synthesize antisense digoxigenin (DIG) and fluorescein (FITC) riboprobes for whole-mount in situ hybridization (WMISH)[39–41] and double fluorescent whole-mount in situ hybridization (FWMISH). Gene probes, forward primers, reverse primers, template, length and RNA-pol/digestion enzymes are as follows: Od*ActnM1* cross-hybridizing: 5′GTCCCCGCCATGTACGTCTG3′, 5′GCATCGGAATCGCTCGTTACCA3′, gDNA exon 2 partial, 389 bp, T3/NotI; Od*ActnM1* specific: 5′GATCGTCCACCGAAAGTGC3′, 5′GTCAGCAACTGTTTGAATATATTG3′, cDNA 3′ UTR, 351 bp, T7/PstI; Od*Brachyury*: 5′GGTTCGCACTGGATGAAACAGCC3′, 5′TATCCGTTCTGACACCAGTCGTTC3′, gDNA exon 3, 630 bp, T3/NotI; Od*Dach*: 5′GAGATGGATCCTGCGCAGC3′, 5′GTAAGTTTAAGAATATCCGAAAATCC3′, cDNA full length, 770 bp, T7/SpeI; Od*Ebf* (*COE*): 5′GAGATCATGTGTTCCCGATGTTG3′, 5′GTTGAGTGAAAGAAAACCTTGCT3′, cDNA exon 5 to exon 8, 578 bp, T3/NotI; Od*ERK*: 5′GAAGGAGCCTACGGCATAG3′, 5′GCTAGAATACATCCGACAGAC3′, cDNA exon 1 to exon 5, 563 bp, T3/NotI; Od*Ets1/2a1*: 5′GACGGCATTGATGGATTACAGCTAT3′, 5′GTTACTCTTCAGTCTCTGGCTC3′, cDNA exon 3/4 to exon 7, 480 bp, T3/NotI; Od*Ets1/2a2*: 5′GACTCTCTTTGCCATCAATCC3′, 5′GCCTTTTTTCGTCCAGCTAATG3′, cDNA exon 2 to exon 5, 728 bp, PCR4-TOPO,T3/NotI; Od*FoxF*: 5′GGCTGGAAGAATTCCGTCCG3′, 5′GAGCTGATTCGCATGGGCAGG3′, cDNA exon 3 to exon 7, 639 bp, T7/SpeI; Od*Gata1/2/3b*: 5′GCCTCTTCTGATTCGCCATTC3′, 5′GGAATGACTGTTTGGTGTTGG3′, cDNA exon1 to exon3, 791 bp, T3/Xbal; Od*Gata1/2/3d*: 5′GGGCAGAATATGAAAATGTATTTT3′, 5′GCTGACCGTCCGCTAGTC3′, cDNA fulllength, 1,086 bp, T7/BamHI; Od*Hand1/2*: 5′GATGGAGTTGAATTTGTATTCCGATC3′, 5′GATTCTTTTCTAATCAGATGGGCA3′, cDNA full length, 462 bp, T7/SpeI; Od*Islet*: 5′GGTCTCCGGTGATGAATTCCT3′, 5′GCTTTGTCGTAGGTTTAGGCCA3′, cDNA exon 4 to exon 8, 766 bp, T7/PstI; Od*Math6*: 5′GCCGAATTCCACACAACGAAG3′, 5′CAACGTTAGCAGGTATAGAATAG3′, cDNA exon 1/2 to 3′UTR, 602 bp, T7/SpeI; Od*MEK3/6*: 5′..AGTGACGAGCAACGACCCTC3′, 5′..ATGCATCTTTGCAGCTCTTCGT3′, cDNA exon 1 to exon 7, 1,034 bp, T3/NotI; Od*MEK7*: 5′..CCCCGTTCCTGTTCTAGGGTC3′, 5′..GGCCCGTTACAAAGACGCTG3′, cDNA exon 1 to exon 5, 842 bp, T7/SpeI; Od*MyoD*: 5′GTTACAAAATGACTATGACGGAAAC3′, 5′GGCTTCCAAGTTTCTTGACCAG3′, cDNA full length, 813 bp, T7/PstI; Od*Neurogenin*: 5′..GAGCACTTCTCCAAAAACAGAG3′, 5′.. GTTTTTACATTGTCGGAAATTCTG3′, cDNA fulllength, 833 bp, T3/NotI; Od*Nk4*: 5′GACCGAAAAATTACAACTATGAGC3′, 5′GCTGTAGCGCCGAGCTCAC3′, cDNA exon1 to exon3, 649 bp, T3/NotI.

For FWMISH, fixed embryos were rehydrated in PBT (PBS with 0.2% Tween-20), treated with 50 mM DTT in PBT (for 10 min at room temperature), washed in 0.1 M triethanolamine in PBT (2 × 5 min at room temperature), treated with two successive dilutions of acetic anhydride (0.25% and 0.5%) in 0.1 M triethanolamine (10 min at room temperature), and washed in PBT (2 × 5 min at room temperature). Prehybridization was carried out in a mixture of 50% formamide, 5× SSC, 0.1 mg/ml heparin, 0.15% Tween-20, 5 mM EDTA, 0.5 mg/ml yeast RNA and 1× Denhardt's reagent for 2 h at 63 °C. Then, hybridization was carried out in the same solution but adding the two probes at 0.5–1 ng/µl each, overnight at 63 °C. Next day, embryos were washed in successive dilution of SSC (2 × 10 min in 2× SSC/0.2% Tween-20 at 65 °C; 2 × 10 min in 0.2× SSC/0.2% Tween-20 at 65 °C; 1 × 5 min in 0.1× SSC/0.2% Tween-20 at room temperature) and 2 × 5 min in MABT (0.1 M maleic acid, 0.15 M NaCl and 0.1% Tween-20, pH 7.5). Blocking was performed by washing the embryos in a mixture of MABT, 2.5 mg/ml BSA and 5% sheep serum. Finally, anti-FITC-POD antibody (1:1,000 in blocking solution) (11426346910, Roche) was added to the samples for overnight incubation at 4 °C. The day after, samples were washed in MABT (eight times for 15 min at room temperature) and then in TNT (0.1 M Tris-HCl pH 7, 0.15 M NaCl and 0.3% Triton X-100) for 10 min. For the staining, embryos were incubated in TSA-tetramethylrhodamine (NEL742001KT, Perkin Elmer) for 10 min. Then, they were washed in TNT for 10 min, in PBT for 10 min, in 2% H_2O_2/PBT for 45 min, in PBT (two times for 5 min) and in MABT (two times for 5 min). Then, a second blocking step was performed followed by the addition of an anti-DIG-POD antibody (1:1,000 in blocking solution; 11207733910, Roche) that was incubated overnight at 4 °C. The morning after, as the previous day, the samples were washed and the coloration reaction was added, that in this case included TSA-FITC system-green (NEL741E001KT, Perkin Elmer) for 1 h 30 min. After coloration, embryos were washed in TNT (2 × 5 min at room temperature) and PBT (2 × 5 min at room temperature). Mounting was made in 80% glycerol/PBS with Hoechst-33342 1 µM (62249, Invitrogen) as previously described[15]. A confocal microscopy LSM880 (Zeiss) was used for imaging of samples and FIJI[42] was used to compose the confocal series and adjust the brightness and contrast.

Pharmacological treatments

For FGF receptor inhibition, animals were treated with 50 µM and 100 µM of SU5402 (SML0443, Merck) and AZD4547 (9403, BioVision)

from the two-cell stage (30 min post-fertilization (mpf)) and from 32-cell stage (70 mpf), respectively, to hatchling stage (4 hpf) in darkness. For *MEK3/6* inhibition, animals were treated with gossypetin (1176, Extrasynthese) 100 µM from the 2-cell to the 32-cell stage in darkness. For *MEK7* inhibition, animals were treated with 1 µM and 25 µM of 5Z-7-oxozeaenol (O9890, Merck) from the 2-cell stage to the 32-cell stage, respectively. For BMP inhibition, animals were treated with 10 µM of LDN (SML1119, Merck) and dorsomorphin (P5499, Merck) from the 2-cell stage and from the 32-cell stage until hatchling stage. To perform these treatments, eggs were pooled in 4 ml of SSW and fertilized with 200 µl of sperm dilution (the sperm of three males in 5 ml of SSW). At the desired time, embryos were transferred to a 3-mm Petri dish plate with 4 ml of treatment solution at 19 °C. Control embryos were incubated in DMSO 0.2% or 0.3% (v/v) depending on the concentration of the treatment. The effects of the treatments were scored by in situ hybridization. For tailbud embryos, we used cross-hybridizing *ActnM1*, *Nk4* and *Brachyury* probes[15,43], whereas for hatchling embryos, we used the specific *ActnM1* probe[15].

Statistics and reproducibility

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment. Descriptions of morphological features in live animals or expression domains obtained by WMISH or FWMISH were performed in at least five specimens (usually from 10 to 20) in each analysed developmental stage. WMISH and FWMISH were performed at least twice for each probed gene. Inhibitory treatments were performed at least twice for each condition, and numbers of phenotype counts out of the total number of analysed embryos are indicated in Fig. 1f, Extended Data Fig. 9 and Supplementary Data 2. The sex condition of embryos does not influence experimental design.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Accession numbers and URLs of databases from publicly available sources are provided in the Methods, Supplementary Information and Supplementary Data 1.

33. Martí-Solans, J. et al. *Oikopleura dioica* culturing made easy: a low-cost facility for an emerging animal model in EvoDevo. *Genesis* **53**, 183–193 (2015).
34. Brozovic, M. et al. ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. *Nucleic Acids Res.* **46**, D718–D725 (2018).
35. Naville, M. et al. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr. Biol.* **29**, 1161–1168 (2019).
36. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
37. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
38. Conklin, E. G. The organization and cell lineage of the ascidian egg. *J. Acad. Nat. Sci. Phila.* **13**, 1–119 (1905).
39. Martí-Solans, J. et al. Coelimination and survival in gene network evolution: dismantling the RA-signaling in a chordate. *Mol. Biol. Evol.* **33**, 2401–2416 (2016).
40. Bassham, S. & Postlethwait, J. Brachyury (T) expression in embryos of a larvacean urochordate, *Oikopleura dioica*, and the ancestral role of T. *Dev. Biol.* **220**, 322–332 (2000).
41. Cañestro, C. & Postlethwait, J. H. Development of a chordate anterior–posterior axis without classical retinoic acid signaling. *Dev. Biol.* **305**, 522–538 (2007).
42. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
43. Torres-Águila, N. P. et al. Diatom bloom-derived biotoxins cause aberrant development and gene expression in the appendicularian chordate *Oikopleura dioica*. *Commun. Biol.* **1**, 121 (2018).

Acknowledgements We thank all team members of the C.C. and R.A. laboratories for discussions; S. Artime for assistance with the animal facility; and L. Christiaen and A. Elewa for reading and commenting on the manuscript. C.C. was supported by BFU2016-80601-P and PID2019-110562GB-I00, R.A. by BIO2015-67358-C2-1-P and J.G.-F. by BFU2017-861152-P and PID2020-117820GB-I00 grants from Ministerio de Ciencia y Innovación (Spain). C.C., R.A. and J.G.-F. were also supported by grant 2017-SGR-1665 from Generalitat de Catalunya. A.F.-R. was supported by FPU14/02654, G.S.-S. by FPU18/02414, M.P.-C. by colaboración-2015/16, P.B. by colaboración-2016/17 and M.J.-L. by colaboración-2019/20 fellowships from Ministerio de Educación Cultura y Deporte. M.F.-T. was supported by a PREDOC2020/58 fellowship from the University of Barcelona. M.P.-C. was supported by PPL1415, A.F.-R by PPL1314 and P.B. by PPLB1617 from Asociación Española Contra el Cáncer (AECC).

Author contributions A.F.-R. carried out the cardiac developmental atlas, genome surveys, phylogenetic analyses, WMISH experiments, cell lineage mapping and BMP inhibitory treatments. M.F.-T. contributed to the *Gata* genome survey and FWMISH. G.S.-S. and P.B. contributed to the *Fgf/Mapk* genome survey and FGF inhibitory treatments. E.D.-B. contributed to the *Tbx* genome survey and WMISH. M.J.-L. contributed to the *Ets* and *Tbx* phylogenies. M.P.-C. contributed to *Islet* characterization. A.F.-R. interpreted the data and made the figures. C.C. conceptualized the project. J.G.-F. and R.A. provided resources. R.A. and C.C. supervised the experiments. C.C. and A.F.-R. wrote the manuscript. All authors commented on the manuscript and agreed to its final version.

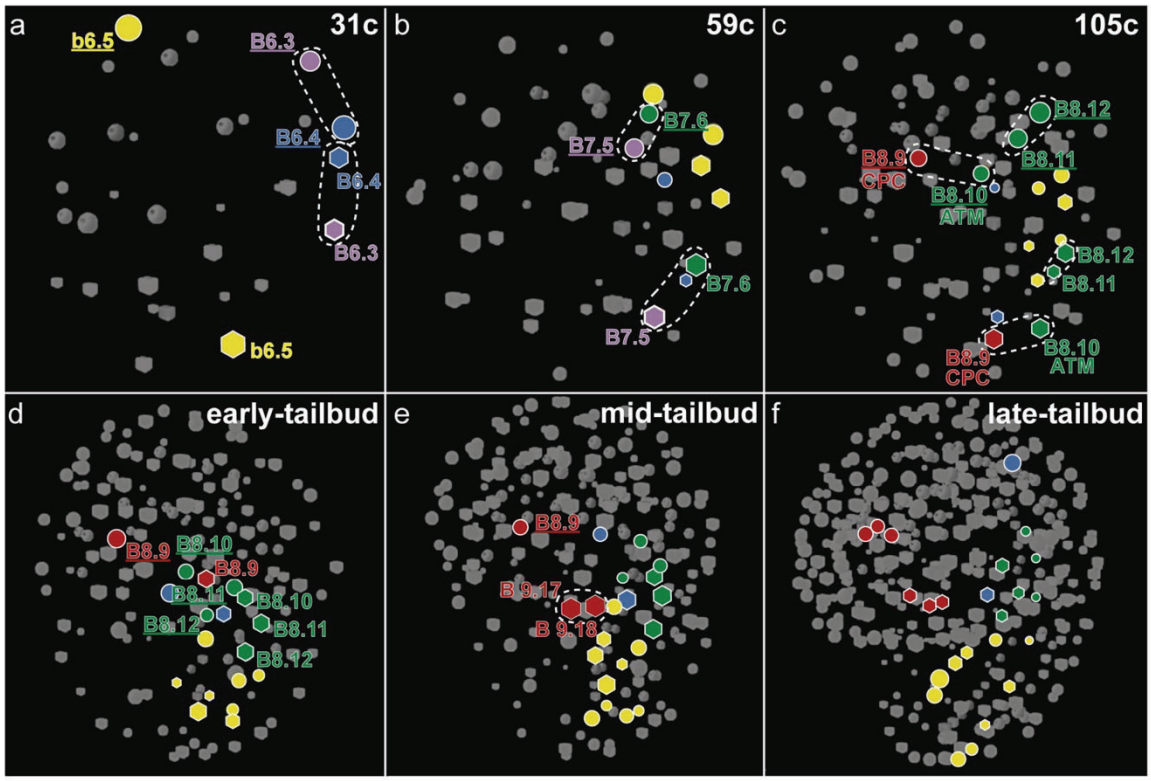
Competing interests The authors declare no competing interests.

Additional information

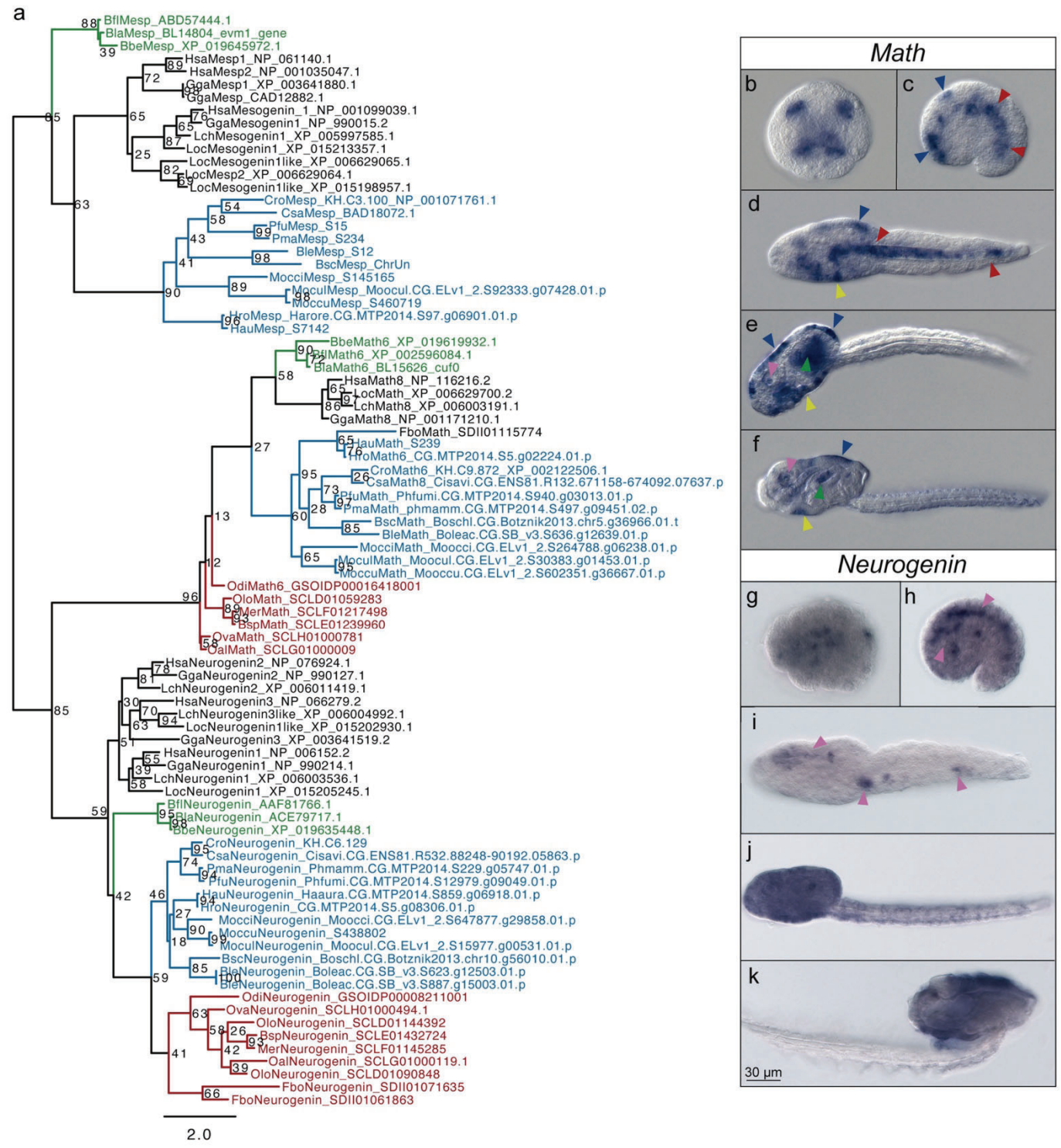
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-04041-w.

Correspondence and requests for materials should be addressed to Cristian Cañestro.

Peer review information *Nature* thanks Benoit Bruneau, Brad Davidson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

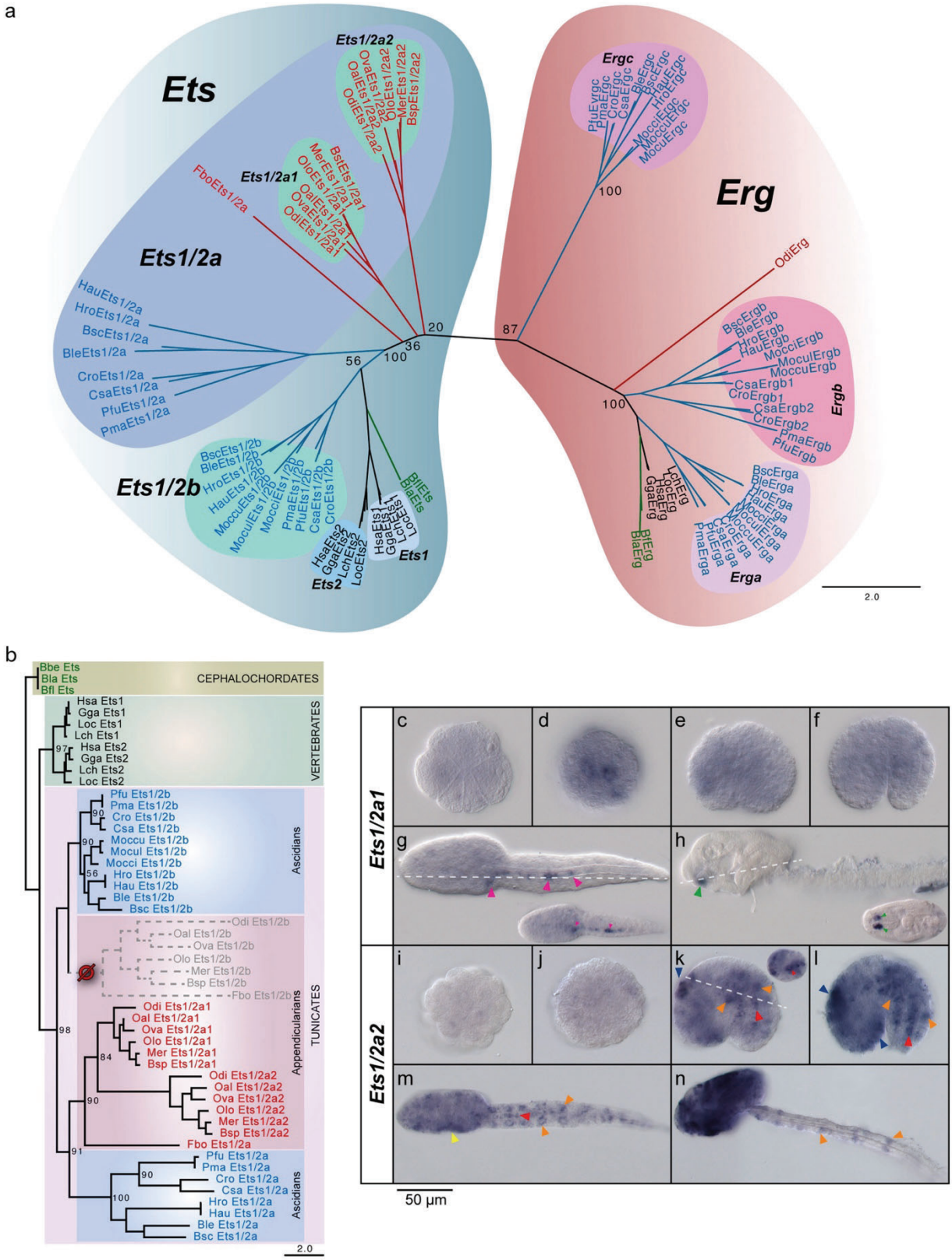


Extended Data Fig. 1 | 4D-reconstruction of a virtual cardiac cell tracing, based on nuclear position from the 30-cell stage to tailbud stages of *O. dioica* embryos (modified from Stach 2008)[16]. B8.9 appears as the first CPC. Blastomere nomenclature follows that of Conklin for ascidians (vegetal blastomeres in capital letters, animal blastomeres in small letters, and blastomeres from the right underlined)[38], and their fate are indicated in different colors: muscle+heart (purple), posterior tail muscle cells (yellow), anterior tail muscle cells (ATM, green), heart (red), germ-line (blue). Circles and hexagons represent blastomeres derived from right and left sides of the embryo, respectively. Dashes encircle sister cells resulting from a cell division.



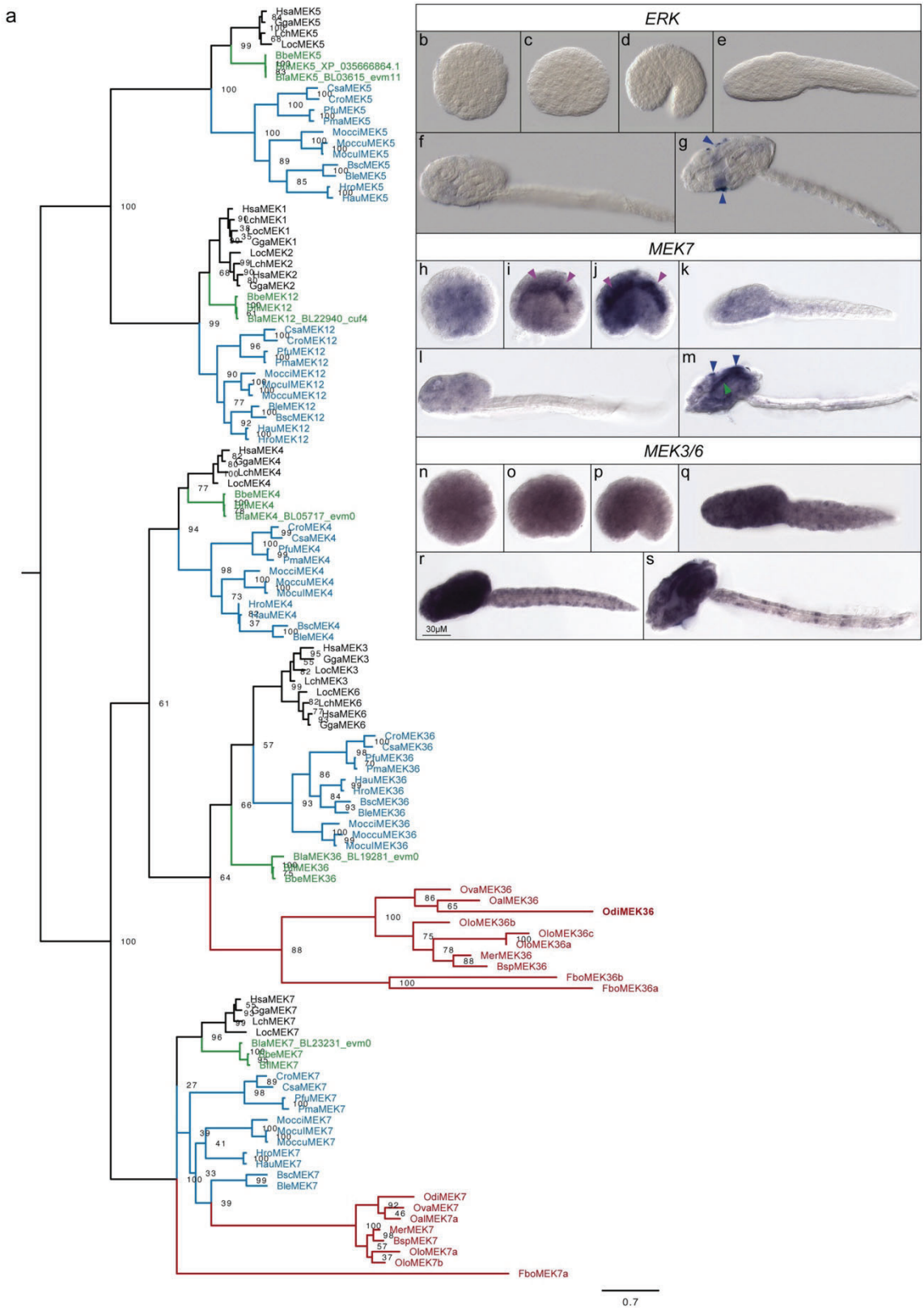
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Mesp ML phylogenetic tree and *Math* and *Neurogenin* expression. a, Unrooted phylogenetic tree, represented in a rectangular layout for the sake of clarity, showing the presence of bHLH homologs of *Neurogenin* and *Math* in appendicularians, but the absence of *Mesp*. The presence of *Mesp* in cephalochordates, vertebrates and all analyzed ascidians suggests an ancestral loss of *Mesp* at the base of the appendicularian lineage after its split from the lineage leading to ascidians. Bootstrap values are shown in the nodes. Scale bar indicates amino acid substitutions. Vertebrates (black): *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Ascidian tunicates (blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Mocci), *Molgula occulta* (Moccu), *Molgula oculata* (Mocul), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Appendicularian tunicates (red): *Bathochordaeus sp.* (Bsp), *Fritillaria borealis* (Fbo), *Mesochordaeus erythrocephalus* (Mer), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (green): *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla). **b**–**f**, Developmental expression pattern of *O. dioica Math* homolog. Whole mount in situ hybridization in different stages of *O. dioica* development showing expression in the notochord in tailbud and early-hatchling embryos (red arrowheads) (**c**, **d**), in epidermis (blue arrowheads) (**c**–**f**), in the rectum domain in hatchling stages (yellow arrowheads) (**d**–**f**), in later stages of neural system development (pink arrowheads) (**e**, **f**), and in later stages of digestive system development (green arrowheads) (**e**, **f**). **g**–**k**, Developmental expression pattern of *O. dioica Neurogenin* homolog. Whole mount in situ hybridization in different stages of *O. dioica* development shows that *Neurogenin* expression was restricted to nervous system in tailbud and early-hatchling stages (pink arrowheads) (**h**, **i**) but no expression was detected in any region compatible with cardiac function. Images from tailbud in advance correspond to left lateral views orientated anterior towards the left and dorsal towards the top.



Extended Data Fig. 3 | See next page for caption.

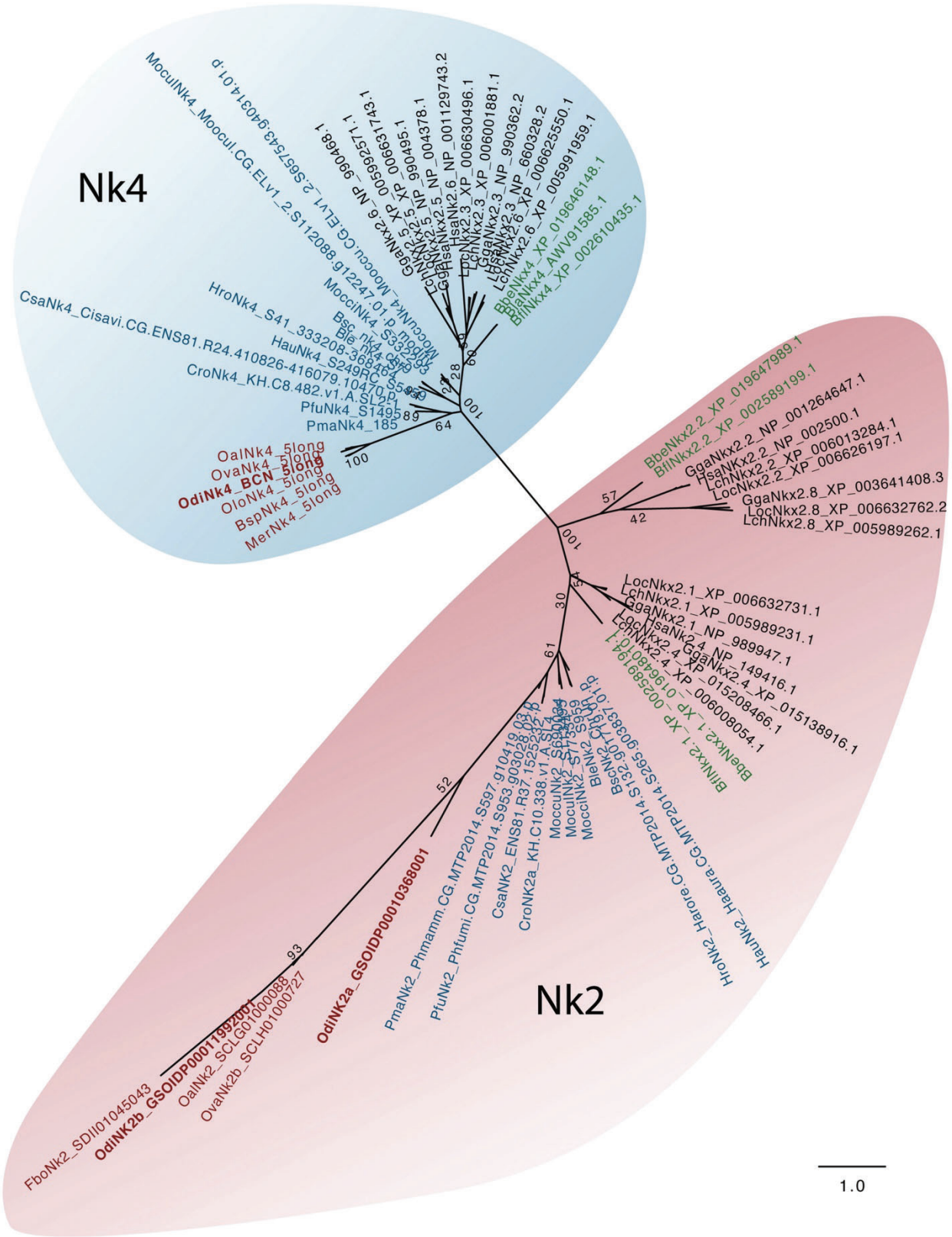
Extended Data Fig. 3 | Ets ML phylogenetic tree and expression. a, Unrooted phylogenetic tree of the *Ets* and *Erg* protein families showed a high bootstrap value separating both protein families what corroborated the existence of two *Ets1/2* genes in appendicularians. Scale bar indicates amino acid substitutions. Vertebrates (black): *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Ascidian tunicates (blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Mocci), *Molgula occulta* (Moccu), *Molgula oculata* (Mocul), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Appendicularian tunicates (red): *Bathochordaeus sp.* (Bsp), *Fritillaria borealis* (Fbo), *Mesochordaeus erythrocephalus* (Mer), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (green): *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla). **b**, Phylogenetic analysis of chordate *Ets1/2*, using cephalochordate sequences as outgroup, suggested that the two *Ets1/2* genes of appendicularians were co-orthologs to the ascidian *Ets1/2a*. **c**–**h**, Whole mount in situ hybridization of *O. dioica Ets1/2a1* did not show any clear expression before hatchling stages (**c**–**f**). In early-hatchling stage *Ets1/2a1* revealed expression in the migratory endodermal strand cells (pink arrowheads) (**g**). In late-hatchling the expression signal was restricted to the buccal gland (green arrowheads) (**h**). **i**, **j**, *Ets1/2a2* did not show expression until tailbud stage. **k**, **l**, In tailbud embryos, expression signal was detected in tail muscle cells (orange arrowheads), the notochord (red arrowheads) and the epidermis of the trunk (blue arrowheads). **m**, In early-hatchling expression signal continued in the tail muscle and the notochord and increased in the anal domain (yellow arrowhead). **n**, In late-hatchling stage, the *Ets1/2a2* expression covered the entire oikoplastic epithelium, and continued in the muscle cells of the tail. Large images from tailbud in advance correspond to left lateral views oriented anterior towards the left and dorsal towards the top. Inset images are dorsal views of optical cross sections at the levels of dashed lines.



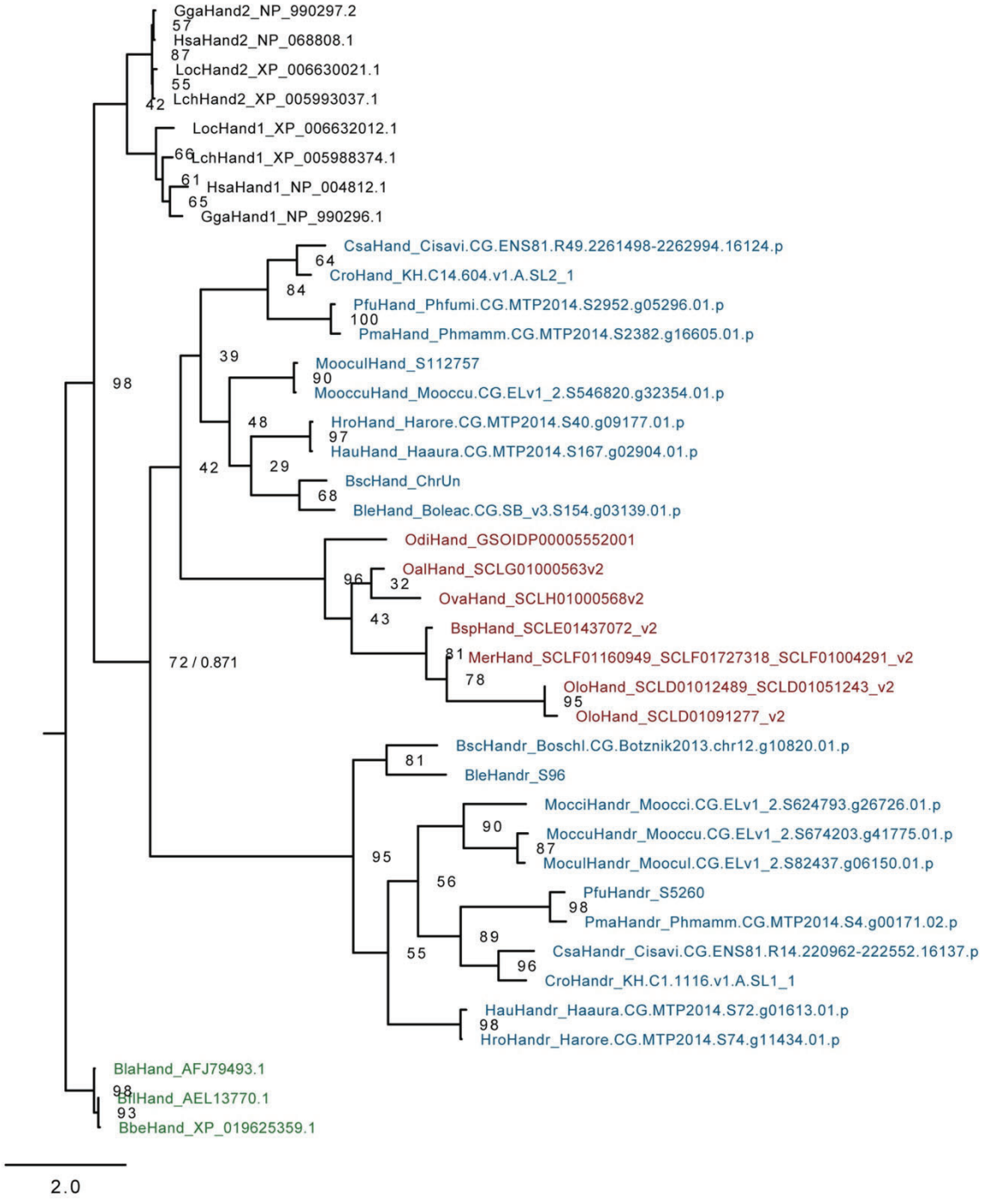
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | FGF/MAPK ML phylogenetic tree and expression. **a**, ML phylogenetic tree of the *MEK* subfamilies in chordates revealing the loss of the *MEK4*, *MEK5* and *MEK1/2* subfamilies in appendicularians, but the surviving of *MEK3/6* and *MEK7* subfamilies. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates (black): *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Ascidian tunicates (blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Mocci), *Molgula occulta* (Moccu), *Molgula oculata* (Mocul), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Appendicularian tunicates (red): *Bathochordaeus sp.* (Bsp), *Fritillaria borealis* (Fbo), *Mesochordaeus erythrocephalus* (Mer), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (green): *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla). **b**–**g**, Whole mount in situ hybridization of *ERK* homolog in different stages of *O. dioica* development did not detect expression in any studied stage (**b**–**f**) until late-hatchling when expression was detected in an specific central domain in the oikoplastic epithelium (blue arrowheads) (**g**). **h**–**m**, Whole mount in situ hybridization of *MEK7* homolog in *O. dioica* revealed expression in the developing neural tissue in tailbud stages (pink arrowheads) (**i**, **j**), and in the esophagus (green arrowhead) and the oikoplastic epithelium (blue arrowheads) in the late-hatchling stage (**m**). **n**–**s**, Whole mount in situ hybridization of *MEK3/6* homolog in different stages of *O. dioica* development did not show any obvious tissue specific expression domain in the trunk, but the signal was generalized, with the exception of muscle cells in the tail at late-hatchling stages. Images from tailbud in advanced correspond to left lateral views orientated anterior towards the left and dorsal towards the top.

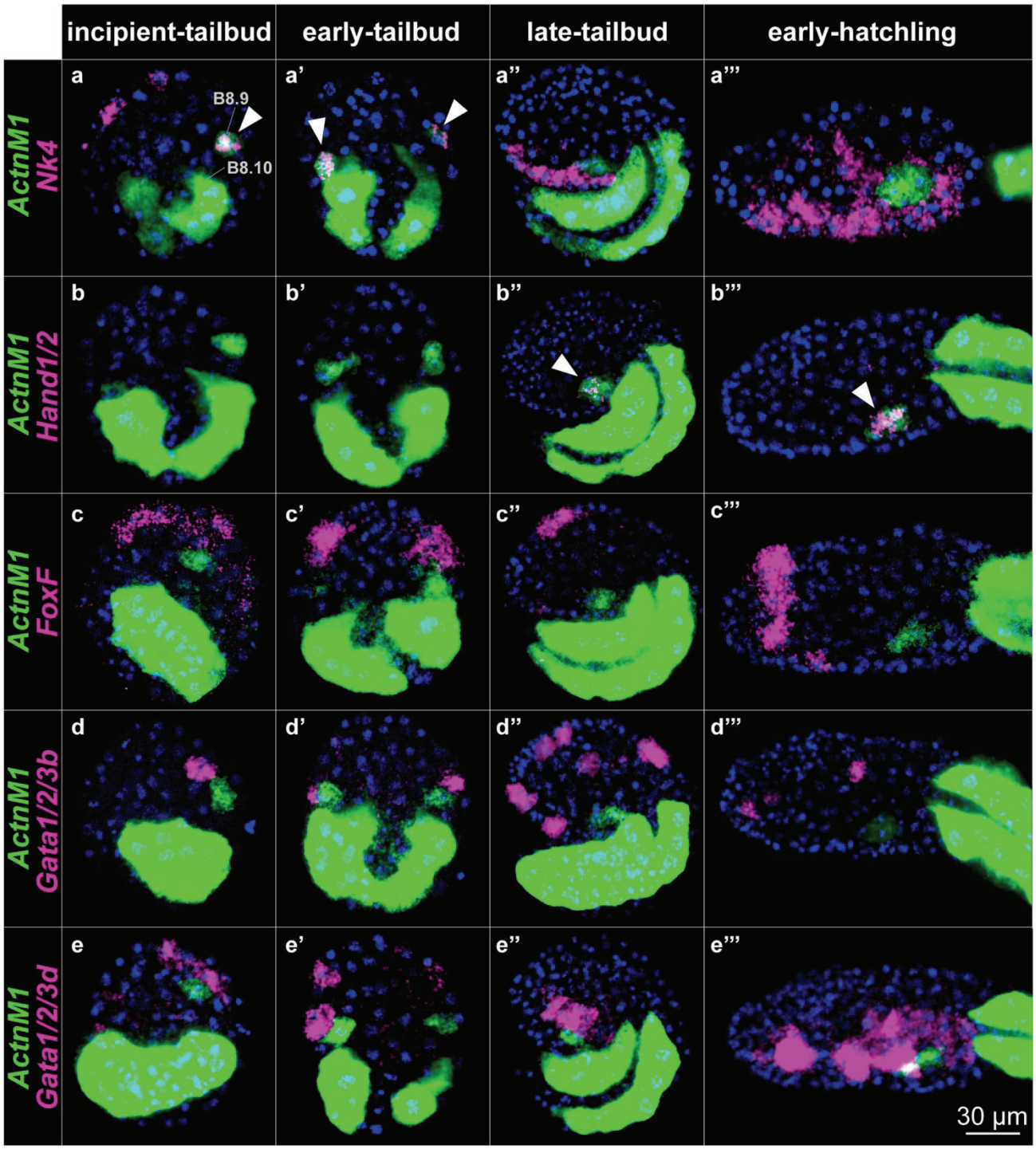
Extended Data Fig. 5 | Gata and FoxF ML phylogenetic trees in chordates. **a**, *Gata* ML phylogenetic tree reveals the loss of the *Gata4/5/6* in appendicularians, but the surviving and lineage specific duplications of *Gata1/2/3* in appendicularians. **b**, *FoxF* ML phylogenetic tree reveals the presence of an ortholog of *FoxF* in appendicularians. The sister *FoxQ* subfamily was used as outgroup to root the tree. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates (black): *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Ascidian tunicates (blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Mocci), *Molgula occulta* (Moccu), *Molgula oculata* (Mocul), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Appendicularian tunicates (red): *Bathochordaeus sp.* (Bsp), *Fritillaria borealis* (Fbo), *Mesochordaeus erythrocephalus* (Mer), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (Green): *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla).



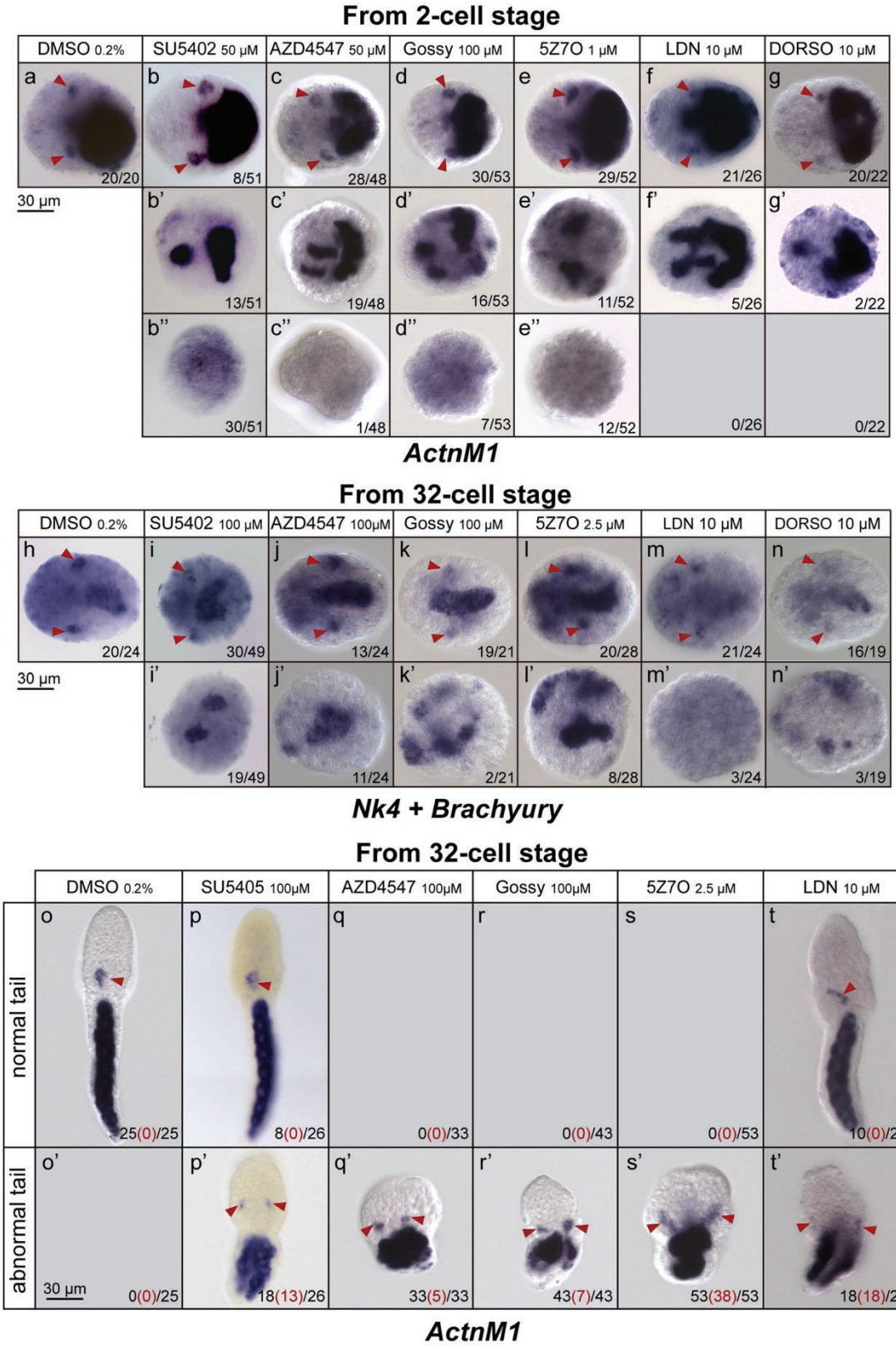
Extended Data Fig. 6 | NK ML phylogenetic tree in chordates reveals the presence of an ortholog of *Nk4* in appendicularians and two orthologs of the Nk2 subfamily. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates (black): *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Ascidian tunicates (blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Mocci), *Molgula occulta* (Moccu), *Molgula oculata* (Mocul), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Appendicularian tunicates (red): *Bathochordaeus sp.* (Bsp), *Fritillaria borealis* (Fbo), *Mesochordaeus erythrocephalus* (Mer), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (Green): *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla).



Extended Data Fig. 7 | Hand ML phylogenetic tree suggests that member of this family in *O. dioica* is homologous to ascidian *Hand1/2*. Despite the tree suggests that the second paralog of ascidian (*Hand-r*) arose by a duplication at the base of the tunicate clade, and therefore subsequently lost in appendicularians. The low node support –bootstrap and approximate likelihood-ratio test (aLRT)– and the presence of shared long amino acid domain rich in K between the *Hand1/2* and *Hand-r* in ascidians, but absent in appendicularians, do not allow us to discard the possibility that *Hand-r* was originated by a duplication within the ascidian lineage, and its basal branching in the tunicate clade is due to a long branch attraction phenomenon. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates (black): *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Ascidian tunicates (blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Mocci), *Molgula occulta* (Moccu), *Molgula oculata* (Mocul), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Appendicularian tunicates (red): *Bathochordaeus sp.* (Bsp), *Fritillaria borealis* (Fbo), *Mesochordaeus erythrocephalus* (Mer), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (green): *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla).

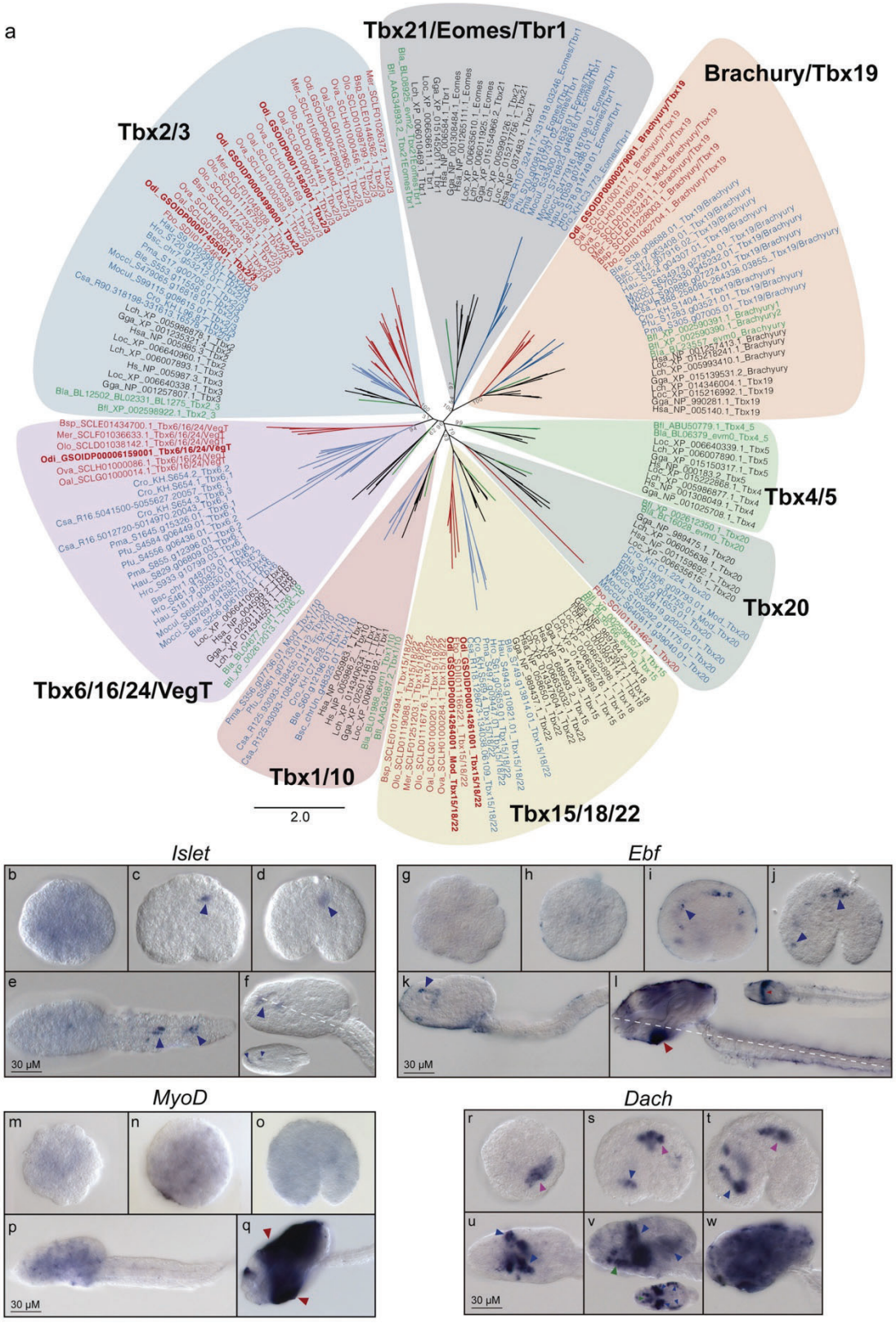


Extended Data Fig. 8 | Developmental coexpression patterns of *ActnM1* and potential cardiac transcription factors. Double fluorescent in situ hybridization of *ActnM1* with *Nk4*, *Hand1/2*, *FoxF*, *Gata1/2/3b* and *Gata1/2/3d*. *Nk4* expression signal was detected in ventral epidermis and the CPC (B8.9) from the incipient-tailbud stage (**a**) until the early-tailbud (**a'**). In later stages, we only detected expression in the epidermis, but not in the cardiac precursors (**a''**–**a'''**). *Hand1/2* was specifically expressed in the cardiac progenitors from late-tailbud to hatchling stages (**b''**–**b'''**). We did not detect expression of *FoxF*, *Gata1/2/3b* nor *Gata1/2/3d* in cardiac precursors, but they were expressed in different epidermal domains (**c**–**e'''**). The images correspond to the overlay of a stack of confocal sections with expression of the different genes. The small overlapping color in **e'''** is due to the overlay of the stack, and not to actual co-expression. White arrowheads indicate co-expression of *ActnM1* with the corresponding gene in cardiac progenitors. Incipient- and early-tailbud stages correspond to ventral views oriented anterior towards the top. Late-tailbud and early-hatchling stages correspond to lateral views oriented anterior towards the left and dorsal towards the top.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | FGF, MEK and BMP inhibition during heart development of *O. dioica*. **a**–**g'**, Whole mount in situ hybridization of *ActnM1* in DMSO-control (**a**) and treated embryos with inhibitors of FGFR (SU5402 and AZD4547), MEK3/6 (Gossypetin), MEK7 (5Z-7-Oxozeaenol) and BMP inhibitors (LDN and Dorsomorphin) from 2-cell stage up to early-tailbud stage (**b**–**g'**). Embryos treated with FGFR and MEK inhibitors affected gastrulation and caused abnormal phenotypes in which mesodermal derivatives showed either abnormal domains (**b'**–**e'**) or complete absence (**b''**–**e''**). However, those treated embryos that reached fairly normal incipient morphologies (**b**–**g**), showed the presence of CPCs (red arrowheads). **h**–**n'**, Whole mount in situ hybridization of *NK4+Brachyury* in DMSO-control (**h**) and treated embryos with FGFR, MEK and BMP inhibitors (**i**–**n'**) from 32-cell stage to early-tailbud stage. A majority of the treated embryos showed the *Nk4* expression in the CPCs (**i**–**n**), even in some with obvious abnormalities in the notochord (**i**). Only in embryos with severe abnormal morphologies or arrested, we could not distinguish the CPCs from other *Nk4* expression domains (**i'**–**n'**). **o**–**t'**, Whole mount in situ hybridization of *ActnM1* in DMSO-control (**o**) and treated embryos with FGFR, MEK and BMP inhibitors from 32-cell stage to early-hatchling stage (**p**–**t'**). Most of the treated embryos showed abnormal tails (**p'**–**t'**), in which the elongation and rotation had been affected. Moreover, while the CPCs had converged near the midline into a single cardiac field, we observed that in many embryos with tail malformations, the CPCs had not converged and were still bilaterally separated at the right and left sides of the trunk (red numbers in brackets). These results suggests that FGF/MEK/MAPK and BMP signaling pathways may be involved in tail elongation/rotation and late cardiac organogenesis. Tailbud embryos images correspond to dorsal views with anterior to the left. Hatchling images represent dorsal views with anterior to the top.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Tbx ML phylogenetic tree and *Islet*, *Ebf*, *MyoD* and *Dach* expression. a, ML phylogenetic tree of the Tbx subfamilies in chordates reveals the loss of *Tbx1/10* and *Tbx21/Eomes/Tbr1* subfamilies in appendicularians and the ancestral loss of *Tbx4/5* subfamily in tunicates. Scale bar indicates amino acid substitutions. Bootstrap values are shown in the nodes. Vertebrates (black): *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Lepisosteus oculatus* (Loc); Ascidian tunicates (blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Moocci), *Molgula occulta* (Mooccu), *Molgula oculata* (Moocul), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Ascidian appendicularians (red): *Bathochordaeus sp.* (Bsp), *Fritillaria borealis* (Fbo), *Mesochordaeus erythrocephalus* (Mer), *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura longicauda* (Olo), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (green): *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla). **b**–**w**, Whole mount in situ hybridization of *O. dioica Islet*, *Ebf*, *MyoD* and *Dach* homologs. 64-cell embryos did not showed expression of *Islet* (**b**) which was only detected in the developing nervous system from tailbud to hatchling embryos (**c**–**f**). *Ebf* (*COE*) did not show expression in early stages (**g**, **h**) but we detected expression in the nervous system from tailbud to mid-hatchling stage (**i**–**k**) and in the oikoplastic epithelium of late-hatchling embryos (**l**). We did not detect expression of *MyoD* from 32-cell to hatchling embryos (**m**–**p**). In late-hatchling embryos *MyoD* was expressed in the oikoplastic epithelium (**q**). *Dach* expression started at the 64-cell stage in the developing nervous system (pink arrowheads) and continued until late-tailbud stage (**r**–**t**). In tailbud stages, *Dach* started expressing in the trunk epidermis (blue arrowheads) which was maintained until late-hatchling stages when it was expressed in the whole oikoplastic epithelium (blue arrowheads) (**s**–**v**). In mid-hatchling stage, beside the epidermis, *Dach* expression was also detected in the endostyle (green arrowheads) (**w**). Large images from tailbud in advance correspond to left lateral views oriented anterior towards the left and dorsal towards the top. Inset images are dorsal views of optical cross sections at the levels of dashed lines. Pink arrowheads indicate the developing nervous system. Blue arrowheads indicate the oikoplastic epithelium.

Annex 4

Less, but more: new insights from appendicularians on chordate *Fgf* evolution and the divergence of tunicate lifestyles

Gaspar Sánchez-Serna, Jordi Badia-Ramentol, Paula Bujosa, Alfonso Ferrández-Roldán, Nuria P. Torres-Águila, Marc Fabregà-Torrus, Johannes N. Wibisana, Michael J. Mansfield, Charles Plessy, Nicholas M. Luscombe, Ricard Albalat and Cristian Cañestro*

bioRxiv 2024.08.30.610304; doi: https://doi.org/10.1101/2024.08.30.610304

Manuscript submitted for publication

Less, but more: new insights from appendicularians on chordate *Fgf* evolution and the divergence of tunicate lifestyles.

Gaspar Sánchez-Serna, Jordi Badia-Ramentol, Paula Bujosa, Alfonso Ferrández-Roldán, Nuria P. Torres-Águila, Marc Fabregà-Torrus, Johannes N. Wibisana, Michael J. Mansfield, Charles Plessy, Nicholas M. Luscombe, Ricard Albalat and Cristian Cañestro*

Gaspar Sánchez-Serna[1,2] <gasparsanchez@ub.edu> (0000-0003-3889-8648)
Jordi Badia-Ramentol[1,2,3] <jbadia@researchmar.net> (0000-0002-3587-0386)
Paula Bujosa[1,2,4] <paula.bujosa@irbbarcelona.org> (0000-0002-5945-6227)
Alfonso Ferrández-Roldán[1,2,5] <alfonsoferrandez@ub.edu> (0000-0003-0632-9838)
Nuria P. Torres-Águila[1,2] <ntorresag@ub.edu> (0000-0002-9502-9310)
Marc Fabrega-Torrus[1,2] <marcfabrega@ub.edu> (0000-0002-1446-5832)
Johannes N. Wibisana[6] <johannes.nicolaus@oist.jp> (0000-0002-2452-8702)
Michael J. Mansfield[6] <michael.mansfield@oist.jp> (0000-0003-4717-4721)
Charles Plessy[6] <charles.plessy@oist.jp> (0000-0001-7410-6295)
Nicholas M. Luscombe[6] <nicholas.luscombe@oist.jp> (0000-0001-5293-4778)
Ricard Albalat[1,2] <ralbalat@ub.edu> (0000-0003-0282-9595)
Cristian Cañestro* [1,2,] <canestro@ub.edu> (0000-0003-4623-8105)

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona (UB), Av. Diagonal 645, Barcelona 08028, Spain
[2] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain
[3] Current address: Cancer Research Program, Hospital del Mar Research Institute, C. del Dr. Aiguader, 88, Barcelona, 08003, Spain
[4] Current address: Cancer Science Unit, Institute for Research in Biomedicine, C/ Baldiri Reixac 10, Barcelona, 08028, Spain
[5] Current address: Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Passeig Maritim de la Barceloneta 37, Barcelona, 08003, Spain

[6] Genomics and Regulatory Systems Unit, Okinawa Institute of Science and Technology Graduate University (OIST), Onna-son, Okinawa, 904-0495, Japan

***Corresponding author:** Cristian Cañestro* [1,2,] canestro@ub.edu

Keynotes: Tunicate evolution, appendicularian, *Oikopleura dioica*, FGF signaling, embryo development, gene loss, gene duplication, alternative splicing

Abstract

The impact of gene loss on the divergence of taxa and the generation of evolutionary innovations is a fundamental aspect of Evolutionary Biology that remains unclear. Here, using the evolution of the Fibroblast Growth Factors (FGFs) in appendicularians as a case study, we investigate how gene losses have influenced the evolution of chordates, especially the divergence among tunicates. Our work reveals an unprecedented case of massive losses of all *Fgf* gene subfamilies, except for the *Fgf9/16/20* and *Fgf11/12/13/14*, which in turn suffered two bursts of gene duplications. Phylogenetic inferences and genomic analyses of gene synteny conservation, gene architecture, alternative splicing and protein 3D-structure have allowed us to reconstruct the history of appendicularian *Fgf* genes in the context of chordate evolution, providing compelling evidence supporting the paracrine secreting functions and the intracellular functions of the *Fgf9/16/20* and *Fgf11/12/13/14* subfamilies, respectively. Exhaustive analysis of developmental *Fgf* expression in *Oikopleura dioica* as a model for appendicularians reveals a paradigmatic case of what could be referred as “less, but more”, providing a conceptual evolutionary framework characterized by four associated evolutionary patterns: conservation of ancestral *Fgf* expression domains; function shuffling between paralogs upon gene loss; innovation of new expression domains after the bursts of *Fgf* duplications; and the extinction of *Fgf* functions linked to gene losses. The findings of this work allow us to formulate novel hypotheses about the potential impact of losses and duplications of *Fgf* genes on the transition from an ancestral ascidian-like biphasic lifestyle to a fully free-living style of appendicularians. These hypotheses include the massive co-option of *Fgf* genes for the patterning of the oikoblast responsible of the house architecture, and for the development of the tail fin; the recruitment of *Fgf11/12/13/14* genes into the evolution of a new mouth, and their role modulating neuronal excitability; the evolutionary innovation of an “anterior tail” FGF signaling mesodermal source upon the loss of retinoic acid signaling; and the

potential link between the loss of *Fgf7/10/22* and *Fgf8/17/18* and the loss of drastic metamorphosis, mesenchymal cells and lack of tail absorption in appendicularians, in contrast to ascidians.

Introduction

The bloom of sequenced genomes has revealed that gene losses are pervasive and prevalent over gene gains throughout the tree of life, leaving little doubt of their great potential as a key evolutionary force that can generate adaptive phenotypic diversity (Krylov et al. 2003; Albalat and Cañestro 2016; Fernández and Gabaldón 2020; Guijarro-Clarke et al. 2020; Helsen et al. 2020; Xu and Guo 2020). There are many paradigmatic examples of gene losses that have been key for the evolution of adaptations in certain species, under what is known as the "less is more" hypothesis (Olson 1999). Examples of adaptive gene losses include positively selected null-mutations in certain receptors that provide resistance to malaria and HIV in humans (Novembre et al. 2005; Hodgson et al. 2014), the loss of gluconeogenic muscle enzyme that allowed the evolution of true hovering flight in hummingbirds (Osipova et al. 2023), and many gene losses that facilitated the reconquest of aquatic and air environments in mammals (Sharma et al. 2018). However, how evolutionary processes can drive the loss of essential genes without carrying an important detrimental load remains enigmatic. The loss of a gene often does not come as an isolated event, but it is accompanied by the co-elimination of other genes that are functionally linked to a distinctive pathway (Albalat and Cañestro 2016). Moreover, within a given gene family, the loss of some members is often accompanied by the duplication of others, what increases the robustness of the genetic system and may lead to processes of function shuffling among paralogs that facilitate the events of gene loss (McClintock et al. 2001; Cañestro et al. 2009). To understand the impact of the loss of essential genes, such as those governing embryo development, it is necessary to study cases in which events of gene co-elimination and duplication can be related to the loss or survival of ancestral traits still present in sister groups, or even to the origin of evolutionary adaptations.

During recent years, the appendicularian tunicate *Oikopleura dioica* has become an attractive animal model to study the impact of gene loss on the evolution of developmental mechanisms in our own phylum, the chordates (Ferrández-Roldán et al. 2019). For instance, the discovery of the deconstruction of the cardiopharyngeal gene regulatory network in *O. dioica* has been key to understanding the adaptive evolution of the heart, facilitating the transition from an

ancestral biphasic ascidian-like lifestyle to the complete free-living lifestyle in appendicularians (Ferrández-Roldán et al. 2021). *O. dioica* also appears to have suffered a large number of gene losses affecting important signaling pathways, such as the wingless (Wnt) and retinoic acid (RA) signaling pathways, that play fundamental roles in axial patterning and cell differentiation during embryo development in chordates (Martí-Solans et al. 2016; Martí-Solans et al. 2021). *O. dioica* has the minimal Wnt repertoire found among chordates, where only four out of the thirteen Wnt families are present. *O. dioica* also stands as the only chordate known to date that can be considered an evolutionary knockout model for RA signaling. The antagonistic action of the RA and the fibroblast growth factor (FGF) signaling pathways is well known in vertebrates during axial patterning, where it regulates for instance the expression of *Hox* genes (Diez del Corral and Storey 2004; Wilson et al. 2009). The fact that this antagonistic action has also been described in ascidians led to conclude that the RA-FGF antagonism was already present in the last common ancestor of olfactores (this is vertebrates + tunicates) (Pasini et al. 2012). Moreover, its absence in cephalochordates led to suggest that the evolutionary innovation of this RA-FGF antagonism could have facilitated the evolutionary origin and radiation of olfactores (Bertrand et al. 2015). The loss of RA signaling and the drastic reduction of Wnt families in *O. dioica* makes the evolution of FGF signaling in this species particularly intriguing. Its study can contribute to a better understanding of how living beings manage the loss of important genes while maintaining similar morphologies (i.e., the inverse paradox in Evo-Devo) (Cañestro et al. 2007), and to correlating gene loss events with evolutionary adaptations of the lifestyle of certain groups of organisms.

Fibroblast growth factors form a family of signaling proteins that emerged concomitantly with the origin of Eumetazoans and have been vastly conserved during animal evolution, regulating a plethora of important biological processes such as cell proliferation, migration, or differentiation during embryonic development and adult tissue homeostasis (Bertrand et al. 2014; Teven et al. 2014). In general, Fgfs are small proteins characterized by a conserved FGF core homology domain of 120-130 amino acids disposed in a β-trefoil topology (Plotnikov et al. 2001). The FGF domain contains essential motifs for binding both heparin and the extracellular region of the Fgf receptors on the cell surface, forming a dimeric ternary complex that can trigger canonical signal transduction cascades inside the cells (Schlessinger et al. 2000). Outside the FGF domain, sequences are not conserved among different subfamilies, and some *Fgf* genes have independently evolved extended N- or C-terminal regions that are not homologous (Popovici et al. 2005). Consequently, protein alignments between distant Fgf subfamilies are not possible

outside the FGF domain. Some of these extended regions of variable lengths often include signal peptides (SP) and nuclear localization signals (NLS) (Coulier et al. 1997). The SP is a short hydrophobic region that includes the first ~15-30 residues of a protein at its N-terminus and is important for the secretion of proteins outside the cell (Owji et al. 2018). In principle, Fgfs that lack an SP remain intracellular, although alternative secretion mechanisms that enable these Fgfs to perform paracrine signaling functions have also been described (Revest et al. 2000; Miyakawa and Imamura 2003; Schäfer et al. 2004; Kirov et al. 2012). Moreover, there is an increasing body of evidence for intracellular functions and interacting partners for non-secreted Fgfs (Goldfarb 2001; Schoorlemmer and Goldfarb 2002; Olsnes et al. 2003; Goldfarb 2005; Wu et al. 2012; Sluzalska et al. 2021). In addition, the presence of NLS allows some Fgfs to migrate into the nucleus and interact with other transcription factors to modulate the expression of target genes (Antoine et al. 2005; Bryant and Stow 2005; Sheng et al. 2005; Popovici et al. 2006).

The high sequence variability among *Fgfs* and the short length of their conserved core have hindered the phylogenetic classification of this family (Popovici et al. 2005). Recent evolutionary reconstructions suggested the existence of eight *Fgf* subfamilies in chordates (namely *Fgf1/2*, *Fgf3*, *Fgf4/5/6*, *Fgf7/10/22*, *Fgf8/17/18/24*, *Fgf9/16/20*, *Fgf11/12/13/14* and *Fgf19/21/23*) (Oulion et al. 2012). It has been proposed that the basally divergent chordate amphioxus might possess the full catalogue of chordate *Fgfs*, with one single member for each of the eight subfamilies. This contrasts with the large number of *Fgfs* within each subfamily in vertebrates due to the various rounds of genome duplication that occurred during the evolution of different lineages (e.g. 19 in sarcopterygians, 22 in mammals, 23 in chicken and 27 in zebrafish) (Dehal and Boore 2005; Bertrand et al. 2011; Oulion et al. 2012). In ascidian tunicates, seven *Fgf* genes have been found representing at least six of the eight chordate subfamilies, and suggesting that *Fgf1/2* and *Fgf3* were lost during the evolution of the lineage leading to ascidians, at the same time that a novel *Fgf* member, *Fgf-L*, was innovated probably as a duplicate of *Fgf7/10/22* (Dehal and Boore 2005; Popovici et al. 2005; Oulion et al. 2012).

According to their functions, *Fgf* subfamilies have been classically classified into: (I) canonical Fgfs (i.e. *Fgf1/2*, *Fgf3*, *Fgf4/5/6*, *Fgf7/10/22*, *Fgf8/17/18/24*, *Fgf9/16/20*), which act as paracrine and autocrine signaling factors binding and activating the Fgf-receptor tyrosine kinases; (II) endocrine Fgfs (i.e. *Fgf19/21/23*), which bind the cofactor Klotho instead of heparin and act as long-distance signaling molecules in vertebrates, and (III) intracellular non-secreted Fgfs (i.e.

Fgf11/12/13/14) with intracrine non-signaling functions that serve as cofactors to other proteins (reviewed in Ornitz and Itoh 2015; Ornitz and Itoh 2022).

In the present work, as a case study to better understand the impact of gene loss, we address the evolution of the FGF signaling pathway in appendicularian tunicates. The independent genome assemblies at the chromosome level of three different cryptic species of *O. dioica* from different parts of the globe (i.e. Barcelona, Osaka, and Okinawa) (Plessy et al. 2024) allow us to identify for the first time with high confidence the full catalogue of *Fgf* genes in an appendicularian species, and to describe its complete atlas of expression during embryonic and larval development. Our results reveal that only two out of the eight chordate *Fgf* subfamilies have survived in *O. dioica*, and that the presence of 10 *Fgfs* in this species is the result of several gene duplications and losses that occurred during the evolution of the appendicularian lineage. Our findings, moreover, allow us to discuss how the massive losses and burst of duplications affecting the *Fgf* family have impacted on the developmental mechanisms underlying the evolution of the appendicularian free-swimming lifestyle in this unique chordate, which functions as an evolutionary knockout for RA signaling. Finally, the results of this case study allow us to propose a “less, but more” scenario as an conceptual framework that facilitates a better understanding of the evolutionary impact of gene loss.

Material and methods

Laboratory culture of *Oikopleura dioica*

O. dioica specimens were acquired from animal colonies that have been maintained in our facility in the University of Barcelona for over five years. The founder individuals were originally obtained from the Mediterranean coast near Barcelona (Catalonia, Spain) and cultured as detailed in Martí-Solans et al. 2015 (Martí-Solans et al. 2015). This project did not raise any ethical concerns since the experimentation conducted on aquatic invertebrate animals does not fall under the regulations pertaining to animal experimentation, as stipulated in Real Decreto 223 14-3-1998 and Catalonia Ley 5/1995, DOGC2073,5172. Nonetheless, all experimental procedures adhered to the European Union (EU) guidelines for animal care and were formally approved by the Ethical Animal Experimentation Committee (CEEA-2009) of the University of Barcelona.

183 **Genome database searches, gene identification and phylogenetic analyses**

184 *Fgf* genes in *O. dioica* were first identified in the reference database of *O. dioica* relative to
185 the Norwegian population (http://oikoarrays.biology.uiowa.edu/Oiko/) (Danks et al. 2013) with
186 BLASTp and tBLASTn, using as queries the Fgf protein sequences from the vertebrate *Homo*
187 *sapiens* and the tunicate *Ciona robusta*, as well as several other chordates. The corresponding
188 orthologs were then identified in the telomere-to-telomere genomic assemblies of the Barcelona
189 and Okinawa *O. dioica* species, and in the superscaffolded version of the Osaka *O. dioica* cryptic
190 species using tBLASTn (Plessy et al. 2024). In each cryptic species, the newly identified *Fgf*
191 genes were used to search for further paralogs using tBLASTn to obtain the final *Fgf* catalogue.
192 The final catalogue in Barcelona was confirmed searching for proteins containing the FGF domain
193 (PF00167), retrieved from the Pfam database, with HMMscan against all the transcripts translated
194 in all reading frames from a Barcelona *O. dioica* transcriptome assembly (Plessy et al. 2024).

195 *Fgf* genes in other appendicularian species were identified using *O. dioica* and other
196 chordate Fgf proteins as queries for tBLASTn against publicly available genomes (i.e. *Oikopleura*
197 *albicans* SCLG01000000, *Oikopleura vanhoeffeni* SCLH01000000) (Naville et al. 2019). *Fgf*
198 genes in ascidian species other than *Ciona robusta* were identified using the Fgf protein
199 sequences from *C. robusta* and other chordates as queries in BLASTp and tBLASTn searches
200 against the gene models and whole genome assemblies of most species available in ANISEED
201 (i.e. *Ciona savignyi*, *Phallusia fumigata*, *Phallusia mammillata*, *Halocynthia roretzi*, *Halocynthia*
202 *aurantium*, *Botryllus schlosseri*, *Botryllus leachii*, *Molgula occulta*, *Molgula oculata* and *Molgula*
203 *occidentalis*) (Brozovic et al. 2018).

204 Protein alignments were generated with MUSCLE and MAFFT implemented in Aliview v1.28
205 (Larsson 2014) and reviewed by hand. Non-homologous independently extended N- and C-
206 terminus of different subfamilies were aligned in a non-overlapping manner to reduced
207 background noise among Fgf subfamilies in which no similarity was detected outside the FGF
208 core homology domain. Phylogenetic trees were based on Maximum Likelihood (ML) inferences
209 calculated with PhyML v3.0 (Guindon et al. 2010), as well as IQ-Tree (Nguyen et al. 2015). LG
210 was inferred as the best-fit substitution model according to Bayesian information criterion BIC to
211 Fgf data, with a gamma with 4 categories and a shape alpha of 2.4681 (Kalyaanamoorthy et al.
212 2017). Tree node support was inferred by fast likelihood-based methods aLRT SH-like, aLRT
213 chi2-based and aBayes, and by standard or ultrafast bootstraps (n=100) according to
214 computational capacity. Phylogenetic trees were inferred both in complete and trimmed protein

alignments, the later by removing all extended regions outside the FGF core homology domain, and in both cases producing the same tree topology regarding the *Fgf* subfamily homology of appendicularian genes. Gene names were assigned according to previous literature, and those that were described for the first time in this work were assigned according to the topology in the phylogenetic tree. Paralogs were named with letters in alphabetical order, reflecting orthology when the tree topology was conclusive, but not necessarily when additional species-specific duplications occurred, or the tree topology was not fully solved.

Protein structure analyses

The domain architecture and functional motifs of Fgf proteins were examined individually with InterProScan, a comprehensive software suite that combines the information stored in several protein databases (including InterPro, PFAM, SMART or PANTHER) to provide in silico functional characterization of queried protein sequences (Jones et al. 2014). Hydropathy plots were generated in ProtScale available in Expasy (Gasteiger et al. 2005), using the Kyte-Doolittle hydrophobic scale and an interval of 9 amino acids with a linear weight variation model in a normalized scale, following previous similar analysis reported on Fgf9 (Miyakawa et al. 1999). Sequence identity and similarity for every pair of Fgf sequences was obtained from the global pairwise sequence alignment with the Needleman–Wunsch algorithm implemented in EMBOSS needle (Madeira et al. 2024). Three-dimensional structures of *O. dioica* Fgf proteins were predicted de novo with AlphaFold2 (Jumper et al. 2021). For each Fgf protein, the top-ranked relaxed model was imported into USCF ChimeraX for its visualization, analysis, and image generation (Pettersen et al. 2021). Nuclear Localization Signals (NLS) were predicted with the NLStradammus software using the 4-state HMM static model and a posterior cutoff of 0.5 (Nguyen Ba et al. 2009), or manually identified in the case of classical NLS based on the consensus stated in (Lu et al. 2021). Signal peptide (SP) predictions were conducted using the SignalP 6.0 (Teufel et al. 2022) and Phobius (Käll et al. 2004) software.

Cloning and expression analyses

O. dioica Fgf genes were PCR amplified from cDNA or gDNA obtained from individuals from the Barcelona population as previously described in Martí-Solans et al. 2016. The PCR products were cloned using the Topo TA Cloning Kit (K4530-20, Invitrogen), and the resulting plasmid was digested with the adequate restriction enzyme to synthesize antisense digoxigenin (DIG) riboprobes for whole-mount in situ hybridization (WMISH) (Bassham and Postlethwait 2000;

Cañestro and Postlethwait 2007; Martí-Solans et al. 2016). Primers used for the cloning of *O. dioica Fgf* genes, as well as the DNA used as template and the length of the clone and probe are indicated in **Supplementary Table 1**.

Results

Extensive loss of *Fgf* subfamilies in appendicularian tunicates

The gold-standard chromosome arm level genome assembly of three cryptic species of the appendicularian *O. dioica* (BAR, OSA and OKI) (Plessy et al. 2024) has allowed us to identify the full *Fgf* catalogue made up of 10 genes in each species (**Supplementary Table 2**). Phylogenetic analysis showed a clear one-to-one orthology between each *Fgf* gene in the three *O. dioica* cryptic species, providing strong evidence that we had identified the full catalogue of *Fgf* genes in *O. dioica*. Importantly, the use of three independently assembled genomes in our gene survey minimized the possibility of undetected unassembled regions containing additional *Fgf* genes, providing further evidence that the full *Fgf* catalogue in *O. dioica* was made of 10 genes. Phylogenetic analyses provided strong evidence with high node support values indicating that the 10 *Fgf* genes were paralogs originated by two bursts of appendicularian-specific duplications (red solid circled nodes in **Figure 1**). The phylogenetic tree topology also indicated with high support values that all *O.dioica Fgf* genes belonged to only two subfamilies: *Fgf11/12/13/14* and *Fgf9/16/20* (black solid circled nodes in **Figure 1**). Analyses on gene structure, protein domains and protein sequence motifs further supported this phylogenetic classification (see below). In contrast to *O. dioica*, our survey of 11 ascidian species' genomes revealed that most of them had 7 *Fgf* genes orthologous to those previously described in *C. robusta*, which are representatives of all chordate *Fgf* subfamilies except *Fgf1/2* and *Fgf3* (Satou et al. 2002; Popovici et al. 2005; Oulion et al. 2012). The exceptions were *Ciona savingyi*, in which we could not identify an ortholog for *Fgf4/5/6*, and *Halocynthia roretzi*, *Halocynthia auriantum*, *Botrylloides leachii* and *Botryllus spp*., in which we could not identify orthologs for *Fgf-NA1*. We also surveyed the genomes of two additional appendicularians species whose genome assemblies were not too fragmented (i.e. *Oikopleura albicans* and *Oikopleura vanhoeffeni*) (Naville et al. 2019). All *Fgf* genes identified in these species belonged to the same two subfamilies found in *O. dioica* (**Figure 1** and **Supplementary table 3**). These findings across ascidian and appendicularians tunicates suggested that the loss of *Fgf1/2* and *Fgf3* likely occurred in the last common ancestor of all tunicates, before the divergence of these two groups. Interestingly, while ascidians have not

systematically lost any further *Fgf* subfamily, the appendicularian ancestor lost four additional subfamilies (i.e. *Fgf4/5/6*, *Fgf7/10/22*, *Fgf8/17/18/24*, *Fgf19/21/23*) before the radiation of this clade (**Figure 1**). Regarding the evolutionary origin of *Fgf11/12/13/14*, which until now has remained unclear (Oulion et al. 2012), our phylogenetic inferences suggested that none of the cephalochordate *Fgf* genes belonged to this subfamily. This subfamily appeared to be restricted to tunicates and vertebrates (**Figure 1**)**,** supporting previous work suggesting that *Fgf11/12/13/14* was not present in amphioxus (Bertrand et al. 2011), nor in ambulacrarian genomes, including hemichordates (Oulion et al. 2012; Fan and Su 2015) and echinoderms (Lapraz et al. 2006; Röttinger et al. 2008; Czarkwiani et al. 2021). Our results, therefore, suggested that the *Fgf11/12/13/14* subfamily was an innovation of olfactores,

Appendicularian expansion of the surviving *Fgf11/12/13/14* and *Fgf9/16/20* subfamilies

The phylogenetic analysis showed that the massive loss of *Fgf* subfamilies during the evolution of appendicularians was accompanied by a burst of duplications of the two surviving subfamilies, resulting in four *Fgf11/12/13/14a-d* paralogs and six *Fgf9/16/20a-f* paralogs (**Figure 1**). The fact that all *Fgf* genes identified in the other two analyzed appendicularians species (namely *O. vannhoeffeni* and *O. albicans*) also appeared as paralogs within these two subfamilies indicated that the expansions might predate the radiation of the clade. Moreover, the tree topology also suggested that further independent lineage-specific gene duplications might have occurred within each *Fgf* subfamily (**Figure 1**).

Analysis of microsynteny revealed a strong conservation of neighboring genes of each *Fgf* ortholog across all *O. dioica* cryptic species (i.e. Barcelona, Osaka, and Okinawa), providing strong support to the assigned homologies (**Figure 2A** and **Supplementary Figure 1**). Despite this overall microsyntenic conservation, we also observed occasional small inversions and translocations of neighboring genes in most *Fgf* gene neighborhoods. Additionally, we noted positional changes of the *Fgf* orthologs within the same chromosomal arm among the three cryptic species, consistent with the characteristic genome scrambling described in *O. dioica* (**Figure 2B** and **Supplementary Table 2**) (Plessy et al. 2024). We did not detect synteny conservation among *Fgf* paralogs within each subfamily, reinforcing the idea that most of the *Fgf* gene duplications that expanded each subfamily were not recent but likely ancestral, occurring before the radiation of the appendicularian clade.

Comparative analysis of gene structure among appendicularian *Fgf*s and those from cephalochordates, ascidians, and vertebrates provided further support to the conclusion that all *Fgf* genes in *O. dioica* belonged to only two subfamilies. The observation that all *Fgf* genes in amphioxus retain two conserved introns in the core FGF domain (i.e. internal core intron 1 and 2: ici1 and ici2) suggested that the ancestral *Fgf* gene also had these two introns. In the *Fgf11/12/13/14* subfamily, all vertebrate members had retained these two ancestral introns, while in appendicularian and ascidian tunicates, ici1 had been lost, and only ici2 had been preserved in some of their members (**Figure 3**). The presence of an additional internal core intron (ici3) exclusively in all *O. dioica Fgf11/12/13/14* paralogs, and its identification in some *Fgf11/12/13/14* genes from other appendicularian species, further supported the notion that they were paralogs resulting appendicularians-specific duplications and suggested that ici3 could be a synapomorphic feature of the *Fgf11/12/13/14* subfamily in this lineage. Moreover, the fact that we identified two core-flanking introns (i.e. cfi1 and cfi2) in all vertebrate and tunicate *Fgf11/12/13/14* genes that were absent in all *B. floridae* and ambulacrarian *Fgf* genes, suggested that these two core-flanking introns could be a conserved synapomorphy of the *Fgf11/12/13/14* subfamily innovated in the clade olfactores (**Figure 3**).

Mapping of RNAseq and EST data in the genomes of the three *O. dioica* cryptic species revealed at least four alternative first exons (namely, distal, proximal, and middle 1, middle 2, etc.) that could give rise to different isoforms due to alternative splicing and transcription start site usage in most *Fgf11/12/13/14* paralogs (**Figure 3**). The presence of alternative isoforms differing in their N-terminus in members of the *Fgf11/12/13/14* subfamily has been also described in vertebrates (Munoz-Sanjuan et al. 2000; Pablo and Pitt 2016). Our genome database surveys revealed evidence of similar alternative splice variants of the first exon in *Ciona robusta* (GeneID:445758 in NW_004190431.2) and other ascidian species, as we have also found in *O. dioica,* thus suggesting that this feature might be an ancestral characteristic of this *Fgf* subfamily in olfactores. Although many of these alternative first exons were rich in positively charged residues (i.e. Lysines and Arginines) and displayed similar hydrophobicity profiles among paralogs, sequence conservation was poor, supporting the hypothesis that the duplications that originated them were ancient in the evolution of appendicularians (**Figure 3**). The presence of small differences in the alternative splice variants among the three *O. dioica* cryptic species (e.g. the first intron of the *Fgf11/12/13/14d* has been incorporated in the open reading frame in BAR, but not in OSA or OKI), together with the presence of cryptic species-specific introns (i.e.

Fgf11/12/13/14c showed a unique intron in the FGF domain exclusively in OKI) illustrated the rapid evolution of the *Fgf* genes in appendicularians, and provided an example of genetic variation among the cryptic species (**Figure 3**).

Analysis of gene structure of *Fgf9/16/20* paralogs in *O. dioica* revealed that while *Fgf9/16/20a* had retained the 2 ancestral introns in the FGF domain (e.g. ici1 and ici2), which were also conserved in human and ascidian *Fgf9/16/20* orthologs, all the other *O. dioica Fgf9/16/20* paralogs (i.e. *Fgf9/16/20b-f*) showed an intronless structure (**Figure 3**). This intronless structure suggested that an ancestral *Fgf9/16/20a*-like gene, most similar in sequence to *Fgf9/16/20* orthologs in other chordate species, could have been duplicated by the integration of a retrotranscribed form followed by further gene duplications. The absence of introns in most *Fgf9/16/20* genes found in other appendicularian species further reinforced the idea that this subfamily was expanded ancestrally in this clade, and that the six *Fgf9/16/20* paralogs in *O. dioica* belonged to the same subfamily. Further evidence supporting that all *O. dioica Fgf9/16/20* paralogs belonged to this subfamily was the conservation of two cysteine residues that were present in all tunicate *Fgf9/16/20* orthologs and absent in all other *Fgf* subfamilies (**Supplementary Figure 3**).

Canonical Fgf9/16/20 paracrine and Fgf11/12/13/14 intracellular functions

Analysis of protein sequence similarity among Fgfs revealed that during the expansion of the two surviving subfamilies in appendicularians, one or two of their members (namely, Fgf9/16/20a and Fgf11/12/13/14a-b) conserved high similarity with their co-orthologs in other chordate species, while the other paralogs suffered a remarkable sequence divergence, especially within the Fgf9/16/20 subfamily. Thus, for instance, while sequence identity among paralogs of the Fgf9/16/20 subfamily in humans ranges from 62%-69.6% throughout the entire protein (80.6%-87.8% throughout the FGF core), in *O. dioica* sequence similarity among some Fgf9/16/20 paralogs was as low as 18.7% (21.2% in the FGF core) (**Supplementary table 4**). To understand how this sequence divergence might have impacted the function of the paralogs within each subfamily, we examined their conserved protein domains with HMMscan, as well as the presence of potential signal peptides (SP), nuclear localizations signals (NLS), or other conserved motifs known to interact with other cofactors and proteins.

In the Fgf9/16/20 subfamily, despite the variability of the HMMscan e-values of the FGF domains, ranging from 1^{-35} of the Fgf9/16/20a to 1^{-7}-1^{-13} of the Fgf9/16/20b-f, the AlphaFold2

371 software predicted that most *O. dioica* Fgf9/16/20 paralogs displayed the characteristic twelve β-
372 sheets conforming the typical β-trefoil fold structure of vertebrate Fgf9/16/20 proteins (**Figure 3**
373 and **Supplementary Figures 2 and 3**). Moreover, the presence of regions enriched in positively
374 charged residues (e.g. Arginine and Lysine) in all *O. dioica* Fgf9/16/20 paralogs, especially near
375 the end of the FGF domain where heparin binding sites (HBS) have been identified in vertebrates
376 (**Figure 3** and **Supplementary Figure 3**) (Xu et al. 2012), suggested that these regions could
377 bind heparin or heparan sulfate proteoglycans. The presence of a β-trefoil fold and HBS,
378 therefore, suggested that all *O. dioica* Fgf9/16/20 paralogs could potentially interact with Fgf
379 receptors on the cell surface to function through the canonical FGF signaling. To investigate the
380 secretion potential of *O. dioica* Fgf9/16/20 paralogs, we examined the presence of signal peptides
381 (SP) with signalP and Phobius softwares. The results predicted the presence of SP
382 (likelihood>0.5) in the N-terminus of four of the Fgf9/16/20 paralogs (namely, Fgf9/16/20abc and
383 f), but not in the other two (d and e). The presence of SPs in four of the Fgf9/16/20 paralogs was
384 comparable to the SP described in the N-terminus of Fgf9/16/20 of *C. robusta* (Satou et al. 2002),
385 and it can explain the degeneration of the non-canonical signal peptide EFISIA motif within the
386 FGF core, which is required for extracellular secretion via the endoplasmic reticulum in other
387 organisms (Popovici et al. 2004). Interestingly, the absence of an SP in Fgf9/16/20d-e correlated
388 with a certain conservation of the EFISIA motif (i.e. TFIQIA), producing a peak of hydrophobicity
389 like those observed in the EFISIA motif of Fgf9/16/20 in other species (**Figure 3** and
390 **Supplementary Figure 4**) (Miyakawa et al. 1999; Popovici et al. 2004). These observations,
391 therefore, further supported the paracrine nature of the Fgf9/16/20 subfamily in appendicularians
392 and suggested that different paralogs might have evolved different secretion mechanisms.
393 In the Fgf11/12/13/14 subfamily, its members have been traditionally associated with
394 intracellular functions through interacting with various proteins. such as regulators of voltage-
395 gated channels (i.e. Na_vs or Ca_vs), as regulators of transcription factors (i.e. islet brain-2 or
396 NEMO), or as players of neuronal cytoskeleton architecture and cell morphology (Pablo and Pitt
397 2016). Like their vertebrate counterparts, all appendicularian Fgf11/12/13/14 paralogs lacked a
398 signal peptide (**Figure 3**), providing the first clue of a conserved intracellular function. The
399 conservation in all *O. dioica* Fgf11/12/13/14 paralogs of a Leucine and an Arginine in positions
400 that have been described to be critical for the interaction with Na_vs and islet brain-2 (Olsen et al.
401 2003; Pablo and Pitt 2016), and that were conserved in all vertebrate and ascidian members of
402 the Fgf11/12/13/14 subfamily, but not in other Fgf subfamilies, reinforced the idea that this

subfamily also played an intracellular function in appendicularians (**Supplementary Figure 3**). The presence of multiple alternative splice variants with different first exons found in most *O. dioica Fgf11/12/13/14* paralogs (**Figure 3**) was also consistent with the same feature described in vertebrates for *Fgf11/12/13/14* genes with intracellular functions related to the modulation of voltage-gated channels regulating neural excitability (Munoz-Sanjuan et al. 2000; Laezza et al. 2009).

Upon the growing evidence that Fgf proteins may also have intranuclear functions (Popovici et al. 2006), we conducted NLS predictions with NLStradamus and searched for KRVR motifs known to provide NLS in *O. dioica* (Clarke et al. 2007). Our analysis revealed that most *O. dioica* Fgf11/12/13/14 paralogs possess an NLS at the end of the FGF core, similarly to vertebrate and ascidian Fgf11/12/13/14 proteins. Interestingly, we also found NLS in some of the alternative first exons that generate different isoforms diverging at the N-terminus, suggesting that these different isoforms not only might have different promoter usage, but also different intracellular localizations (**Figure 3**). We also found NLS in three out of the six *O. dioica* Fgf9/16/20 paralogs, including the two that lack an SP, implying that these paralogs may have also evolved non-secreted functions.

Overall, our findings from the structural analysis were consistent with paracrine functions for the *Fgf9/16/20* subfamily and intracellular functions for the *Fgf11/12/13/14* subfamily in appendicularians. The high sequence divergence, variation in the presence of putative SP and NLS, and the formation of different isoforms due to differential splice variants in the N-terminus raised the possibility that multiple functions might have also evolved among the different paralogs duplicated during the expansion of these two surviving families in appendicularians.

Fgf expression atlas during the development of *Oikopleura dioica*.

To better understand the functional consequences of gene loss and gene expansion on the *Fgf* subfamilies in *O. dioica*, we performed an exhaustive expression analysis of all the *Fgf9/16/20* and *Fgf11/12/13/14* paralogs by whole mount in situ hybridization (WISH) throughout development, from eggs to late-hatchling stages (**Figure 4 A-J**). In general, we found that the level of expression signal of most *Fgf* genes was low, and long periods of staining (e.g., between 1 to 14 days) were required to visualize some of the tissue-specific expression domains. Our description here will mostly focus on tissue-specific *Fgf* expression domains repeatedly observed in different embryos over background levels, but we cannot discard that in addition to those specific domains some of the *Fgf* genes also had a generalized basal expression scattered in

other parts of the embryo, as it has been also described in other animals including *C. robusta* and amphioxus (Imai et al. 2004; Bertrand et al. 2011).

In oocytes, some *Fgf* genes (i.e. *Fgf9/16/20a,d,e*, and *Fgf11/12/13/14a,b,d*) showed a weak staining signal in the cytoplasm that was difficult to distinguish over the staining background. This suggested that, in general, *Fgf* transcripts were not a major component of the maternal contribution (**Figure 4 A-J1**). In the case of *Fgf9/16/20d* and *Fgf9/16/20e*, however, we observed a small but intense staining spot in the cortical area of many unfertilized eggs (n=11/13 and n=9/11, for *Fgf9/16/20d* and *Fgf9/16/20e*, respectively), but not in all. This suggested that the formation of this *Fgf* transcripts spot might be transient and difficult to capture, or perhaps not present in all individuals (**Figure 4 E-F1, black arrowheads**). In the early stages of development, from 8- to 64-cell, all *Fgf9/16/20* paralogs started showing staining signals, suggesting that their expression onset took place concomitant with the activation of zygotic transcription (Wang et al. 2015) (**Figure 4 A-F2 & A-F3**). At the 8-cell stage, many *Fgf9/16/20* paralogs (i.e. *Fgf9/16/20c*, *Fgf9/16/20d*, *Fgf9/16/20e* and *Fgf9/16/20f*) showed staining signal restricted to the smaller pair of blastomeres in the vegetal pole (**Figure 4 C-F2, black double arrowheads**). These cells corresponded to the A/A4.1 blastomere pair in Delsman/Conklin nomenclature, which gives rise to most of the nervous system, the notochord, and other endomesodermal derivatives (Nishida 2008; Stach et al. 2008). At the gastrula stage (32-64 cells), we found that all *Fgf9/16/20* paralogs were expressed in the precursor blastomeres of either mesodermal or ectodermal derivatives. *Fgf9/16/20a*, *Fgf9/16/20b* and *Fgf9/16/20c* staining was detected in the endomesodermal blastomeres (**Figure 4 A-C3, orange arrowheads**); *Fgf9/16/20b*, *Fgf9/16/20d* and *Fgf9/16/20e* staining was detected in the neural plate (**Figure 4 B3 & E-F3, cyan arrowheads**); *Fgf9/16/20f* was detected in notochord precursor cells (**Figure 4 D3, yellow arrowheads**); and *Fgf9/16/20d* staining was detected in muscle precursor blastomeres (**Figure 4 E3, red arrowheads**). Consistent with these observations in early developmental stages, the majority of tissue-specific *Fgf* expression domains observed during later embryogenesis were also predominantly associated with ectodermal derivatives (e.g., nervous system and epidermis) or mesodermal derivatives (e.g., notochord and muscle).

Among ectodermal derivatives, the staining signal of *Fgf9/16/20* paralogs observed in the neural plate of 64-cells stage embryos persisted in cells of the developing nervous system up to the hatchling stage. These signals were observed in precursors of the brain, the caudal ganglion, and the spinal cord (**Figure 4 A-F, different tones of blue arrowheads**). The fact that some of

these neural expression domains disappeared at specific developmental stages suggested that some *Fgf9/16/20* paralogs might be precisely regulated in specific subsets of neural populations along the anteroposterior axis during the formation of the central nervous system. For instance, *Fgf9/16/20d* expression domains were clearly observed in the neural plate at 64-cell stage and in the posterior part of the brain in incipient tailbud (ITB) embryos (**Figure 4 E3-4, cyan arrowheads**), but it was absent in subsequent stages until the just-hatchling stage, when it reappeared in the developing brain (**Figure 4 E8, light blue arrowhead**). We also observed staining signals for *Fgf11/12/13/14* paralogs in neural domains. However, in contrast to *Fgf9/16/20* genes, which were predominantly expressed before hatching, *Fgf11/12/13/14* genes were mainly expressed during the hatchling stages (**Figure 4 G-J, different tones of blue arrowheads**). Neural expression domains of *Fgf11/12/13/14* genes were detected in various locations within the nervous system, including specific cells of the brain dorsal to the sensory vesicle, the ventral region of the ciliary funnel (**Figure 4 G10-11, light blue arrowheads**), groups of cells in the caudal ganglion (**Figure 4 G9-11, H10-11, I10-11 & J11, dark blue arrowheads**), and isolated cells at different positions along the nerve cord in the tail (**Figure 4 G10-11, cyan arrowheads**). Overall, these results revealed a complex pattern of *Fgf* expression in the developing neural system, suggesting that despite the extensive loss of *Fgf* subfamilies, the expansion of the surviving *Fgf* genes has allowed the preservation of neural functions during appendicularian development, similar to other chordates.

Among ectodermal derivatives, we also observed specific expression domains for various *Fgf* genes in the epidermis, both in the tail and in the trunk (**Figure 4, different tones of green arrowheads and dashed lines**). In the tail, we observed a dynamic expression pattern with different *Fgf* genes expressed at different levels of the anteroposterior axis, mainly in two areas: the developing fin and the tip of the tail. In the precursor cells of the fin, *Fgf11/12/13/14b* expression was first detected in a bilateral pair of epidermal cells located in the middle region of the tail (**Figure 4 H7, green dashed lines**). This expression later spread to the first third of the tail at the just-hatch stage and eventually extended, along with other *Fgf* paralogs (namely, *Fgf11/12/13/14c*, *Fgf9/16/20a*, *Fgf9/16/20b* and *Fgf9/16/20c*) to the posterior half of the tail in mid- and late-hatchlings (**Figure 4 A10-11, B-C9-11, H7-11 & I10-11, green dashed lines**). In the tip of the tail, a pair of epidermal cells started showing strong staining for *Fgf9/16/20d* and *Fgf9/16/20e* at the ITB stage. While *Fgf9/16/20e* expression was downregulated by the mid-tailbud stage, *Fgf9/16/20d* signal persisted until the mid-hatchling stage (**Figure 4 E4-10 & F4-5,**

dark green arrowheads). In the trunk, bilateral groups of epidermal cells expressed *Fgf9/16/20d* at different levels of the anteroposterior axis (**Figure 4 E4-10, light green arrowheads**). The most posterior group included the primordia of the Langerhans receptors, which also expressed *Fgf11/12/13/14c* (**Figure 4 E8-9 & I9, light green double arrowheads**). In the most rostral region of the trunk epidermis, *Fgf11/12/13/14a* showed an expression domain in a group of subepidermal cells in the area of the mouth at the just-hatchling stage. This expression domain was later expanded to the epidermal surface by the mid-hatched stage, coinciding with the opening of the mouth (**Figure 4 G8-10, magenta arrowheads**). Interestingly, *Fgf11/12/13/14a* expression was also observed in the pharyngeal slits in the mid- and late-hatchling larvae (**Figure 4 G10-11, magenta double arrowheads**).The expression of *Fgf11/12/13/14a* in the primordia of organs in which ciliated sensory cells developed (i.e. the stomodeum, the ciliary funnel, and the ciliary rings), together with the expression of *Fgf9/16/20d* and *Fgf11/12/13/14c* in the primordia of the Langerhans receptors, suggested that FGF signaling might be involved in the development of placodial derivatives in appendicularians, as well as in structures in which epithelial perforation and fusions occur, as described in other chordates (Bassham and Postlethwait 2005; Kourakis and Smith 2007; Bassham et al. 2008). In late hatchling stages, nearly all *Fgf* genes were strongly expressed in different parts of the oikoblast, the organ responsible for the architecture and secretion of the house (**Figure 4 A-J11**). Some showed generalized patterns, while others were restricted to or excluded from specific regions. For example, *Fgf9/16/20f* was restricted to cells adjacent to the anterior cells of the field of FoI, the anterior rosette, the field of Martini, the posterior rosette, and its adjacent lateral bands (**Figure 4 D11**). In contrast, *Fgf9/16/20a-b* were expressed throughout the entire oikoblast but excluded from the ring of the mouth and the Giant cells (**Figure 4 A-B11**). This finding suggested that FGF signaling has been recruited for the development of this innovative organ responsible for the formation of the house, as described for many other developmental genes in appendicularians (Mikhaleva et al. 2018).

Among endomesodermal derivatives, the notochord exhibited an *Fgf9/16/20a* expression domain restricted to the first and third cells from the early tailbud (ETB) to late tailbud (LTB) stages (**Figure 4 A5-7, yellow arrowheads**)**.** In these stages, *Fgf9/16/20a* staining was also observed in few internal cells bilaterally located in the anterior half of the trunk, whose positions were compatible with endomesodermal progenitors of the pharynx, endostyle or buccal glands (**Figure 4 A5-8, purple arrowheads**). At the LTB stage, a new mesodermal expression domain of *Fgf9/16/20a* appeared restricted to the first and eighth pairs of muscle cells**,** and it was maintained

until the early hatchling stage (**Figure 4 A7-9, red arrowheads**). Other *Fgf* paralogs with broad, ubiquitous expression patterns also showed stronger expression in muscle cells compared to other parts of the embryo (i.e. *Fgf9/16/20b*, *Fgf9/16/20c*, *Fgf11/12/13/14b* and *Fgf11/12/13/14c*; **Figure 4, red arrowheads**). From ITB to LTB stages, two cells located on the right side of the anterior part of the notochord, which later at the early hatchling stage were located anteriorly in a rostral position to the notochord, expressed a very distinct expression of *Fgf11/12/13/14c* (**Figure 4 I6-8, purple arrowheads**). We could not determine the identity of these endomesodermal cells but, considering their position, they could be related to the development of endodermal progenitors of the digestive system or the gonad (Olsen et al. 2018).

Discussion

Less, but more: massive gene losses accompanied by bursts of duplications of the surviving paralogs

The study of FGF signaling is central for understanding many fundamental functions of cell biology, including proliferation, differentiation, migration, apoptosis, and survival of cells, from embryo development to adult tissue homeostasis, as well as for understanding how its malfunctioning can cause several diseases (reviewed in Dorey and Amaya 2010; Xie et al. 2020). FGF signaling has an ancient evolutionary origin, at least already present in the ancestral eumetazoan (Bertrand et al. 2014). Some of the conserved core functions of the FGF signaling also have an ancestral origin, such as mesodermal induction which predates the Cambrian explosion and the origins of Bilateria (Matus et al. 2007). Different taxa, however, have innovated a great variety of other FGF functions, in many cases associated to gene duplications, and often accompanied by gene losses (Popovici et al. 2005; Oulion et al. 2012). During the evolution of chordates, for instance, the expansion of eight single-gene subfamilies up to 27 *Fgf* genes early in the evolution of vertebrates due to the two rounds of genome duplication has been linked to the innovation and sophistication of many of the characteristic vertebrate traits, including axial patterning, somitogenesis, limb bud formation, visceral and skeletal development (Thisse and Thisse 2005), and even the invention of a "new head" (Bertrand et al. 2011). Studies of FGF signaling in ascidian tunicates have contributed to reveal that some of the traits that characterize vertebrates, indeed, were not vertebrate innovations, but were already present in the last common ancestor of olfactores, such as the role of *Fgf8/17/18* in the organizer activity and the compartmentalization of the central nervous system and in placodal derivatives (Kourakis and

561 Smith 2007; Imai et al. 2009; Wagner and Levine 2012; Stolfi et al. 2015; Horie et al. 2018). Here,
562 our results in *O. dioica* unveil an unprecedented case among chordates, in which massive gene
563 losses have erased all *Fgf* subfamilies but two, *Fgf9/16/20* and *Fgf11/12/13/14*. Interestingly, the
564 massive gene losses have been accompanied by two bursts of duplications that have given rise
565 to several *Fgf* paralogous genes within each subfamily describing an evolutionary scenario of
566 "less, but more" (**Figure 5A**).

567 Our analysis of gene phylogenies, synteny conservation, gene architecture and structural
568 protein motifs across various appendicularians species provide solid evidence that all *Fgf* genes
569 in appendicularians belong to only two subfamilies. This indicates that the losses likely occurred
570 at the base of this clade, and that different species might have independently expanded their *Fgf*
571 catalogue in a very dynamic fashion. This dynamic evolution of *Fgf* genes is particularly evident
572 when comparing different cryptic species of *O. dioica*, where divergences are observed in the
573 presence of different isoforms due to alternative splicing, significant sequence divergence outside
574 the FGF domain, presence of novel introns and microsyntenic rearrangements. The vast
575 conservation of the FGF domain with their typical β-trefoil topology supports the idea that Fgf
576 paralogs in *O. dioica*, despite their sequence divergence, can function as the typical Fgf ligands
577 of other animals. The common presence of secretion motifs in *Fgf9/16/20* paralogs suggests that
578 members of this subfamily can act extracellularly through the canonical signaling pathway typical
579 of this subfamily (Itoh and Ornitz 2011; Ornitz and Itoh 2015). On the other hand, the extensive
580 presence of nuclear localization signals in Fgf11/12/13/14 paralogs suggests that members of
581 this subfamily might have intracellular functions, as it has been described for members of this *Fgf*
582 subfamily in other chordates (Smallwood et al. 1996; Pablo and Pitt 2016).

583 **Evolutionary patterns associated to the "less, but more" scenario of *Fgf* evolution**
584 **in tunicates**

585 Our comparative *Fgf* expression analysis between appendicularians, ascidians and other
586 chordates allow us to identify cases that can be categorized into four different evolutionary
587 patterns associated to the "less, but more" scenario (**Figure 5A**):

588 **(1) conservation of ancestral expression domains**

589 In the first category, we found that the early expression domains of the single ascidian co-
590 ortholog *Fgf9/16/20* in the A4.1 derived vegetal blastomeres of *C. intestinalis* have been
591 preserved in four of the *Fgf9/16/20* paralogs of *O. dioica* (i.e. *Fgf9/16/20c*, *d*, *e* and *f*) in the same

equivalent vegetal blastomeres (Kaoru S. Imai et al. 2002; Bertrand et al. 2003; Hudson et al. 2016; Satou 2020). This expression has been related to an ancestral function of the FGF signaling conserved among most bilaterians in initiating mesodermal and endodermal gene regulatory networks (Technau and Scholz 2003). In ascidians, *Fgf9/16/20* is sufficient for the first phase of initiation of mesodermal induction, although *Fgf8/17/18* is also required for a second phase of fate maintenance (Yasuo and Hudson 2007). This second phase might have been modified in *O. dioica* upon the loss of the Fgf8/17/18 subfamily. Moreover, the early expression of *Fgf9/16/20* paralogs at the eight-cell stage in *O. dioica* is compatible with recent findings in ascidians in which FGF acts as a timer for zygotic genome activation whose responsiveness sharply starts between the 8- and 16-cell stage (Treen et al. 2023).

A second example of conservation of ancestral expression domains is shown by *Fgf9/16/20d-e* in the developing central nervous system from the 64-cell to tailbud stages, comparable to the expression of the ascidian *Fgf9/16/20* co-ortholog in equivalent positions in the central nervous system dorsally at the level of the anterior tip of the notochord. This similarity suggests conserved *Fgf9/16/20* ancestral functions in the anteroposterior patterning of the central nervous system among tunicates (Karou S. Imai et al. 2002; Miyazaki et al. 2007) (**Figure 5B**).

For the *Fgf11/12/13/14* subfamily, our data show that many of the expression domains of its paralogs in *O. dioica* are associated to the nervous system (i.e. dorsal and anterior ventral part of the brain, subset of cells of the caudal ganglion, as well as in isolated neurons located at different levels) (**Figure 5B**). This neural *Fgf* expression is a shared characteristic with their vertebrate homologs, many of which are expressed in neurons where they perform Fgf receptor-independent intracellular functions. These functions involve interactions with voltage-gated sodium channels, affecting neuronal excitability, and have been implicated in human neuronal diseases (Wang et al. 2011). Supporting this neuronal function, our findings show that most *O. dioica Fgf11/12/13/14* paralogs produce different isoforms corresponding to alternative splice variants of the first exons, a feature also shared with vertebrate *Fgf11/12/13/14* genes that interact with voltage-gated sodium channels (Laezza et al. 2009). In ascidians, although *Fgf11/12/13/14* has no detectable expression during development (Satou et al. 2002; Treen et al. 2014), our finding of alternative splicing in the first exon also suggests a conserved neural function. This is consistent with recent single cell RNA-seq data in ascidians showing abundant *Fgf11/12/13/14* transcripts in specific neurons of the tail, including bipolar tail neurons, which have been proposed to share properties with neural-crest-derived dorsal root ganglia (Stolfi et al. 2015; Horie et al.

2018). Future expression analyses of the ascidian *Fgf11/12/13/14* in post-metamorphic animals to uncover more types of neurons that express this intracrine Fgf ligand, as well as functional characterization of its neural role in modulating neuronal excitability, will be of great interest to develop new animal models to better understand the molecular basis of related human neuronal disorders. Moreover, considering that our results reinforce the hypothesis that amphioxus and ambulacrarians lack *Fgf11/12/13/14* (Lapraz et al. 2006; Röttinger et al. 2008; Bertrand et al. 2011; Fan and Su 2015; Czarkwiani et al. 2021), and suggest that its origin could be an innovative synapomorphy of olfactores (**Figure 5A**), further studies of this gene subfamily in appendicularians and other tunicates could shed light on the role of intracrine Fgf ligands in the evolution of the nervous system within this clade following its divergence from cephalochordates.

(2) function shuffling among surviving paralogs upon the loss of genes

In this second category, our study reveals two paradigmatic examples of function shuffling between the ascidian *Fgf8/17/18* or *Fgf7/10/22* and the *O. dioica Fgf9/16/20d*, all of which belong to the group of paracrine/autocrine secreted Fgf ligands. First, the *Fgf8/17/18* epidermal expression domain in the tip of the tail of ascidians, which acts as a secreted posterior tail FGF source (PTFS) (Pasini et al. 2012; Kim et al. 2020), is comparable to the equivalent domain of *Fgf9/16/20d* in the posterior tip of the tail in *O. dioica* (**Figure 5B**). Second, *Fgf8/17/18* and *Fgf7/10/22* in ascidians are also expressed during the development of the atrial siphon, which has been related to the evolution of otic placode homologs (Kourakis &Smith, 2007). In *O. dioica*, in the absence of these two genes, it is *Fgf9/16/20d* and *Fgf11/12/13/14c* the paralogs that show equivalent expression domains in the Langerhans receptor primordia, which have been proposed to be homologous placodial structures in appendicularians (Bassham and Postlethwait 2005) (**Figure 5B**).

(3) innovation of novel expression domains in novel paralogs

In the third category, among the novel expression domains innovated by the duplicated paralogs in *O. dioica*, we find at least three examples. First, in early hatchling stages *Fgf11/12/13/14a* is specifically expressed in the stomodeum, coinciding with the time when the mouth is opening (**Figure 5B**). No comparable expression has been observed for the ascidian *Fgf11/12/13/14* gene, which shows no detectable expression during development (Satou et al. 2002; Treen et al. 2014). Interestingly, the stomodeum and mouth opening derive from the anterior neuropore in ascidians (Veeman et al. 2010), while in *O. dioica* these structures develop in the most rostral part of the trunk directly connecting to the pharynx. In *O. dioica*, the recruitment

of *Pax2/5/8a* expression in the primordium of the stomodeum has been suggested to be related to cellular functions of perforation, adhesion and fusion of epithelial openings, including the mouth (Bassham et al. 2008). Therefore, despite the expression of placodial markers such as *Pitx* suggests deep genetic homology among mouths and adenohypophysis-like organs –i.e. the ciliary funnels in ascidians and appendicularians, and the Hatschek's pit in amphioxus– (Bassham and Postlethwait 2005) we cannot discard the possibility that the mouth of ascidians and appendicularians have independent evolutionary origins recruiting a common cassette of placodial genes as has been suggested for other placodial-derived structures (Bassham and Postlethwait 2005). The involvement of different Fgf ligands in the late development of the mouth and pharyngeal slits in various animals across diverse taxa suggests that the FGF signaling might have been repeatedly recruited during the evolution of perforated structures (Crump et al. 2004; Röttinger et al. 2008; Bertrand et al. 2011; Fan et al. 2018; Rees et al. 2024). Moreover, the high expression of *Fgf11/12/13/14a* in the stomodeum, pharyngeal slits, and in the rostro-ventral part of the brain related to the ciliary funnel, regions also expressing *Pax2/5/8a*, suggests that *O. dioica* may have innovated the recruitment of *Fgf11/12/13/14a* in the evolution of mechanisms related to ciliary cells and perforated structures. To our knowledge, this would be the first case in which an intracrine Fgf ligand has been related with the development of the mouth.

The second example is related to *Fgf* expression in mesodermal derivatives, such as the notochord or muscle cells, in the anterior region of the tail during tailbud and hatchling stages. In ascidians, no *Fgf* expression has been detected in cells of the notochord or tail muscles in tailbud/hatchling stages, except for *Fgf9/16/20* being expressed in the most posterior pair of muscle cells near the tip of the tail (Imai et al. 2004; Pasini et al. 2012). In contrast, *O. dioica* exhibits strong expression of *Fgf9/16/20a* in the most posterior muscle cells of the tail and the most anterior pair, as well as in the first and third cells of the notochord. Additionally, *Fgf9/16/20b-d* expression is observed above background levels in the three most anterior and central muscle cells at tailbud stages. These results reveal that the anterior part of the tail in appendicularians could act as an anterior tail-mesoderm FGF source (Anterior Tail FGF Source, ATFS) of secreted Fgf9/16/20 ligands (**Figure 5B**). This may drastically differ with ascidians in which the Fgf9/16/20 source is restricted to the most posterior part of the tail (Posterior Tail FGF Source, PTFS), often associated with tail elongation and posterior cell identity differentiation or survival, similar to vertebrates (Diez del Corral and Storey 2004; Imai et al. 2004; Olivera-Martinez et al. 2012; Pasini et al. 2012). Interestingly, considering that the anterior region of the tail in ascidians serves as a

source of RA signaling (Anterior Tail Retinoic Acid Source, ATRAS, **Figure 5B**) from the most anterior *Aldh1a*-positive cells of the tail muscle (Nagatomo and Fujiwara 2003), and given the antagonistic action between FGF and RA signaling conserved in ascidians and vertebrates, where down-regulation of RA signaling often leads to increased FGF production (Diez del Corral and Storey 2004; Olivera-Martinez et al. 2012; Pasini et al. 2012; Paschaki et al. 2013), it is tempting to speculate that the evolutionary innovation of this ATFS in appendicularians might be related to their loss of RA signaling (Cañestro and Postlethwait 2007; Martí-Solans et al. 2016) (**Figure 5**). This loss would have reduced selective constrains, allowing some *Fgf9/16/20* genes to gain novel expression domains in the anterior part of the tail. This drastic difference could represent a major shift in developmental signaling sources between appendicularians and ascidians, potentially driving the divergent evolution of developmental processes associated with the distinct body plans and lifestyles that characterize these two groups of tunicates.

The last examples of novel *Fgf* expression domains in *O. dioica* are related to the patterning of the epidermis in late hatchling stages. In the tail, the lateral wings that will develop into the tail fin express three *Fgf9/16/20* and two *Fgf11/12/13/14* paralogs. In the trunk, the oikoblast, which is the epidermal organ responsible for building the house, has recruited the expression of nearly all *Fgf9/16/20* and *Fgf11/12/13/14* paralogs, most of them in a broad fashion, although some appear to be restricted or excluded from certain fields within this complex organ (**Figure 4A11-J11**). The fact that no similar *Fgf* expression domains have been observed in the tail or the trunk of ascidian larvae (Satou et al. 2002; Imai et al. 2004) suggests that the recruitment of FGF for epidermis patterning could be an innovation of the appendicularian lineage, linked to the evolution of the house building organ and the tail movements that characterize its fully free-living lifestyle.

(4) extinction of ancestral expression domains linked to gene losses

In ascidians, *Fgf7/10/22* (which previously had been also referred as *Fgf3*) is strongly expressed throughout the ventral row of cells of the neural tube at tadpole stages, and its signaling function has been described to be crucial for the convergent extension of the notochord that underlies just underneath the neural tube (Shi et al. 2009). In *O. dioica*, the loss of *Fgf7/10/22* predicts that the convergent extension of the notochord might have become independent of the FGF signaling from the ventral neural tube, a prediction that can be tested in future experiments interfering with FGF signaling. Moreover, in ascidians *Fgf7/10/22* knockout makes tail absorption to be arrested during metamorphosis, which has led to suggest that *Fgf7/10/22* might play an inductive cue for the metamorphosis in ascidians (Treen et al. 2014). In this context, it is tempting

to speculate that the loss of *Fgf7/10/22* in appendicularians could be related with the loss of a drastic metamorphosis and the lack of absorption of the tail, as it has been suggested in the ascidian-like biphasic tunicate ancestor (**Figure 5A**). Additionally, ascidian metamorphosis involves mesenchymal tissue, which consists of mesodermal cells that remain in a pluripotent state during development until postmetamorphic differentiation into adult tissues and structures. The ascidian *Fgf8/17/18* ortholog plays a role in the early differentiation of mesenchymal cells, with its expression maintained throughout embryonic development (Imai et al., 2004; Satou, 2020). Therefore, the loss of *Fgf8/17/18* appendicularians could be related to the loss of mesenchymal tissue and the lack of a drastic metamorphic process **(Figure 5)**.

Altogether, our work highlights the evolution of the *Fgf* family in appendicularians as a paradigmatic example of what could be referred as “less, but more”, where massive gene losses, but also extensive duplications, result both in an overall conservation of *Fgf* expression domains, in many cases due to function shuffling among paralogs, and in the innovation of new expression domains. Interestingly, many of these innovations can be related to the transition from an ancestral ascidian-like biphasic lifestyle to the fully free-living lifestyle that characterizes appendicularians. Future functional analyses will be crucial to investigate several key aspects: the role of *Fgf11/12/13/14* gene on the innovation of a new mouth opening and the modulation of neural excitability; the involvement of multiple *Fgf* genes in the patterning of the oikoblast and the innovation of the house; the role of FGF signaling in notochord convergence and in the development of the fin; the impact of the emergence of an ATFS and its potential connection to the fact that these organisms are evolutionary knockouts of RA signaling; and finally, how the loss of *Fgf7/10/22* and *Fgf8/17/18* might be related to the loss of tail absorption and the absence of metamorphosis, both of which occurred during the evolutionary innovation of a fully free-living lifestyle of appendicularians.

Bibliography

Albalat R, Cañestro C. 2016. Evolution by gene loss. *Nat Rev Genet* 17:379–391.

Antoine M, Reimers K, Wirz W, Gressner AM, Müller R, Kiefer P. 2005. Fibroblast growth factor 3, a protein with a dual subcellular fate, is interacting with human ribosomal protein S2. *Biochem Biophys Res Commun* 338:1248–1255.

749 Bassham S, Cañestro C, Postlethwait JH. 2008. Evolution of developmental roles of Pax2/5/8
750 paralogs after independent duplication in urochordate and vertebrate lineages. *BMC Biol*
751 6:35.
752 Bassham S, Postlethwait J. 2000. Brachyury (T) expression in embryos of a larvacean
753 urochordate, Oikopleura dioica, and the ancestral role of T. *Dev Biol* 220:322-32.
754 Bassham S, Postlethwait JH. 2005. The evolutionary history of placodes: a molecular genetic
755 investigation of the larvacean urochordate Oikopleura dioica. *Development* 132:4259–4272.
756 Bertrand S, Aldea D, Oulion S, Subirana L, de Lera AR, Somorjai I, Escriva H. 2015. Evolution of
757 the Role of RA and FGF Signals in the Control of Somitogenesis in Chordates. *PLoS One*
758 10:e0136587.
759 Bertrand S, Camasses A, Somorjai I, Belgacem MR, Chabrol O, Escande ML, Pontarotti P, Escriva
760 H. 2011. Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits.
761 *Proc Natl Acad Sci U S A* 108:9160–9165.
762 Bertrand S, Iwema T, Escriva H. 2014. FGF signaling emerged concomitantly with the origin of
763 eumetazoans. *Mol Biol Evol* 31:310–318.
764 Bertrand V, Hudson C, Caillol D, Popovici C, Lemaire P. 2003. Neural tissue in ascidian embryos
765 is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription
766 factors. *Cell* 115:615–627.
767 Brozovic M, Dantec C, Dardaillon J, Dauga D, Faure E, Gineste M, Louis A, Naville M, Nitta KR,
768 Piette J, et al. 2018. ANISEED 2017: Extending the integrated ascidian database to the
769 exploration and evolutionary comparison of genome-scale datasets. *Nucleic Acids Res*
770 46:D718–D725.
771 Bryant DM, Stow JL. 2005. Nuclear Translocation of Cell-Surface Receptors: Lessons from
772 Fibroblast Growth Factor. *Traffic* 6:947–953.
773 Cañestro C, Catchen JM, Rodríguez-Marí A, Yokoi H, Postlethwait JH. 2009. Consequences of
774 lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic
775 acid signaling in vertebrate genomes. *PLoS Genet* 5:e1000496.
776 Cañestro C, Postlethwait JH. 2007. Development of a chordate anterior-posterior axis without
777 classical retinoic acid signaling. *Dev Biol* 305:522–538.
778 Cañestro C, Yokoi H, Postlethwait JH. 2007. Evolutionary developmental biology and genomics.
779 *Nat Rev Genet* 8:932–942.

780 Clarke T, Bouquet J-M, Fu X, Kallesøe T, Schmid M, Thompson EM. 2007. Rapidly evolving
781 lamins in a chordate, Oikopleura dioica, with unusual nuclear architecture. *Gene* 396:159–
782 169.
783 Coulier F, Pontarotti P, Roubin R, Hartung H, Goldfarb M, Birnbaum D. 1997. Of worms and men:
784 an evolutionary perspective on the fibroblast growth factor (FGF) and FGF receptor families.
785 *J Mol Evol* 44:43–56.
786 Crump JG, Maves L, Lawson ND, Weinstein BM, Kimmel CB. 2004. An essential role for Fgfs in
787 endodermal pouch formation influences later craniofacial skeletal patterning. *Development*
788 131:5703–5716.
789 Czarkwiani A, Dylus D V., Carballo L, Oliveri P. 2021. FGF signalling plays similar roles in
790 development and regeneration of the skeleton in the brittle star Amphiura filiformis.
791 *Development* 148.
792 Danks G, Campsteijn C, Parida M, Butcher S, Doddapaneni H, Fu B, Petrin R, Metpally R, Lenhard
793 B, Wincker P, et al. 2013. OikoBase: A genomics and developmental transcriptomics
794 resource for the urochordate Oikopleura dioica. *Nucleic Acids Res* 41:1–9.
795 Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate.
796 *PLoS Biol* 3:e314.
797 Diez del Corral R, Storey KG. 2004. Opposing FGF and retinoid pathways: a signalling switch that
798 controls differentiation and patterning onset in the extending vertebrate body axis. *Bioessays*
799 26:857–869.
800 Dorey K, Amaya E. 2010. FGF signalling: diverse roles during early vertebrate embryogenesis.
801 *Development* 137:3731–3742.
802 Fan T-P, Su Y-H. 2015. FGF signaling repertoire of the indirect developing hemichordate
803 Ptychodera flava. *Mar Genomics* 24:167–175.
804 Fan T-P, Ting H-C, Yu J-K, Su Y-H. 2018. Reiterative use of FGF signaling in mesoderm
805 development during embryogenesis and metamorphosis in the hemichordate Ptychodera
806 flava. *BMC Evol Biol* 18:120.
807 Fernández R, Gabaldón T. 2020. Gene gain and loss across the metazoan tree of life. *Nat Ecol*
808 *Evol* 4:524–533.
809 Ferrández-Roldán A, Fabregà-Torrus M, Sánchez-Serna G, Duran-Bello E, Joaquín-Lluís M,
810 Bujosa P, Plana-Carmona M, Garcia-Fernàndez J, Albalat R, Cañestro C. 2021.
811 Cardiopharyngeal deconstruction and ancestral tunicate sessility. *Nature* 599:431–435.

812 Ferrández-Roldán A, Martí-Solans J, Cañestro C, Albalat R. 2019. Oikopleura dioica: An
813 Emergent Chordate Model to Study the Impact of Gene Loss on the Evolution of the
814 Mechanisms of Development. In: Tworzydlo W, Bilinski S, editors. Evo-Devo: Non-model
815 Species in Cell and Developmental Biology. Vol. 68. Springer, Cham. p. 63–105.
816 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. 2005. Protein
817 Identification and Analysis Tools on the ExPASy Server. In: The Proteomics Protocols
818 Handbook. Totowa, NJ: Humana Press. p. 571–607.
819 Goldfarb M. 2001. Signaling By Fibroblast Growth Factors: The Inside Story. *Science's STKE*
820 2001.
821 Goldfarb M. 2005. Fibroblast growth factor homologous factors: Evolution, structure, and function.
822 *Cytokine Growth Factor Rev* 16:215–220.
823 Guijarro-Clarke C, Holland PWH, Paps J. 2020. Widespread patterns of gene loss in the evolution
824 of the animal kingdom. *Nat Ecol Evol* 4:519–523.
825 Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms
826 and methods to estimate maximum-likelihood phylogenies: assessing the performance of
827 PhyML 3.0. *Syst Biol* 59:307–321.
828 Helsen J, Voordeckers K, Vanderwaeren L, Santermans T, Tsontaki M, Verstrepen KJ, Jelier R.
829 2020. Gene loss predictably drives evolutionary adaptation. *Mol Biol Evol* 37:2989–3002.
830 Hodgson JA, Pickrell JK, Pearson LN, Quillen EE, Prista A, Rocha J, Soodyall H, Shriver MD,
831 Perry GH. 2014. Natural selection for the Duffy-null allele in the recently admixed people of
832 Madagascar. *Proc Biol Sci* 281:20140930.
833 Horie R, Hazbun A, Chen K, Cao C, Levine M, Horie T. 2018. Shared evolutionary origin of
834 vertebrate neural crest and cranial placodes. *Nature* 560:228–232.
835 Hudson C, Sirour C, Yasuo H. 2016. Co-expression of Foxa.a, Foxd and Fgf9/16/20 defines a
836 transient mesendoderm regulatory state in ascidian embryos. *Elife* 5.
837 Imai KS, Hino K, Yagi K, Satoh N, Satou Y. 2004. Gene expression profiles of transcription factors
838 and signaling molecules in the ascidian embryo: towards a comprehensive understanding of
839 gene networks. *Development* 131:4047–4058.
840 Imai Kaoru S., Satoh N, Satou Y. 2002. Early embryonic expression of FGF4/6/9 gene and its role
841 in the induction of mesenchyme and notochord in Ciona savignyi embryos. *Development*
842 129:1729–1738.

Imai Karou S., Satoh N, Satou Y. 2002. Region specific gene expressions in the central nervous system of the ascidian embryo. *Mech Dev* 119 Suppl 1:S275–S277.

Imai KS, Stolfi A, Levine M, Satou Y. 2009. Gene regulatory networks underlying the compartmentalization of the Ciona central nervous system. *Development* 136:285–293.

Itoh N, Ornitz DM. 2011. Fibroblast growth factors: from molecular evolution to roles in development, metabolism and disease. *J Biochem* 149:121–130.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589.

Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 338:1027–1036.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589.

Kim K, Gibboney S, Razy-Krajka F, Lowe EK, Wang W, Stolfi A. 2020. Regulation of Neurogenesis by FGF Signaling and Neurogenin in the Invertebrate Chordate Ciona. *Front Cell Dev Biol* 8:477.

Kirov A, Al-Hashimi H, Solomon P, Mazur C, Thorpe PE, Sims PJ, Tarantini F, Kumar TKS, Prudovsky I. 2012. Phosphatidylserine externalization and membrane blebbing are involved in the nonclassical export of FGF1. *J Cell Biochem* 113:956–966.

Kourakis MJ, Smith WC. 2007. A conserved role for FGF signaling in chordate otic/atrial placode formation. *Dev Biol* 312:245–257.

Krylov DM, Wolf YI, Rogozin IB, Koonin E V. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res* 13:2229–2235.

Laezza F, Lampert A, Kozel MA, Gerber BR, Rush AM, Nerbonne JM, Waxman SG, Dib-Hajj SD, Ornitz DM. 2009. FGF14 N-terminal splice variants differentially modulate Nav1.2 and Nav1.6-encoded sodium channels. *Mol Cell Neurosci* 42:90–101.

873 Lapraz F, Rottinger E, Duboc V, Range R, Duloquin L, Walton K, Wu SY, Bradham C, Loza MA,
874 Hibino T, et al. 2006. RTK and TGF-beta signaling pathways genes in the sea urchin
875 genome. *Dev Biol* 300:132–152.
876 Larsson A. 2014. AliView: A fast and lightweight alignment viewer and editor for large datasets.
877 *Bioinformatics* 30:3276–3278.
878 Lu J, Wu T, Zhang B, Liu S, Song W, Qiao J, Ruan H. 2021. Types of nuclear localization signals
879 and mechanisms of protein import into the nucleus. *Cell Communication and Signaling*
880 19:60.
881 Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, Lopez R, Butcher S.
882 2024. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic*
883 *Acids Res*.
884 Martí-Solans J, Belyaeva O V., Torres-Aguila NP, Kedishvili NY, Albalat R, Cañestro C. 2016.
885 Coelimination and Survival in Gene Network Evolution: Dismantling the RA-Signaling in a
886 Chordate. *Mol Biol Evol* 33:2401–2416.
887 Martí-Solans J, Ferrández-Roldán A, Godoy-Marín H, Badia-Ramentol J, Torres-Aguila NP,
888 Rodríguez-Marí A, Bouquet JM, Chourrout D, Thompson EM, Albalat R, et al. 2015.
889 Oikopleura dioica culturing made easy: a low-cost facility for an emerging animal model in
890 EvoDevo. *Genesis* 53:183–193.
891 Martí-Solans J, Godoy-Marín H, Diaz-Gracia M, Onuma TA, Nishida H, Albalat R, Cañestro C.
892 2021. Massive Gene Loss and Function Shuffling in Appendicularians Stretch the
893 Boundaries of Chordate Wnt Family Evolution. *Front Cell Dev Biol* 9:700827.
894 Matus DQ, Thomsen GH, Martindale MQ. 2007. FGF signaling in gastrulation and neural
895 development in Nematostella vectensis, an anthozoan cnidarian. *Dev Genes Evol* 217:137–
896 148.
897 McClintock JM, Carlson R, Mann DM, Prince VE. 2001. Consequences of Hox gene duplication
898 in the vertebrates: an investigation of the zebrafish Hox paralogue group 1 genes.
899 *Development* 128:2471–2484.
900 Mikhaleva Y, Skinnes R, Sumic S, Thompson EM, Chourrout D. 2018. Development of the house
901 secreting epithelium, a major innovation of tunicate larvaceans, involves multiple
902 homeodomain transcription factors. *Dev Biol* 443:117–126.

903 Miyakawa K, Hatsuzawa K, Kurokawa T, Asada M, Kuroiwa T, Imamura T. 1999. A hydrophobic
904 region locating at the center of fibroblast growth factor-9 is crucial for its secretion. *J Biol*
905 *Chem* 274:29352–29357.
906 Miyakawa K, Imamura T. 2003. Secretion of FGF-16 requires an uncleaved bipartite signal
907 sequence. *Journal of Biological Chemistry* 278:35718–35724.
908 Miyazaki Y, Nishida H, Kumano G. 2007. Brain induction in ascidian embryos is dependent on
909 juxtaposition of FGF9 / 16 / 20-producing and -receiving cells.
910 Munoz-Sanjuan I, Smallwood PM, Nathans J. 2000. Isoform diversity among fibroblast growth
911 factor homologous factors is generated by alternative promoter usage and differential
912 splicing. *J Biol Chem* 275:2589–2597.
913 Nagatomo K, Fujiwara S. 2003. Expression of Raldh2, Cyp26 and Hox-1 in normal and retinoic
914 acid-treated Ciona intestinalis embryos. *Gene Expr Patterns* 3:273–277.
915 Naville M, Henriet S, Warren I, Sumic S, Reeve M, Volff JN, Chourrout D. 2019. Massive Changes
916 of Genome Size Driven by Expansions of Non-autonomous Transposable Elements. *Current*
917 *Biology* 29:1161–1168.
918 Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: a simple Hidden Markov
919 Model for nuclear localization signal prediction. *BMC Bioinformatics* 10:202.
920 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective
921 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 32:268–
922 274.
923 Nishida H. 2008. Development of the appendicularian Oikopleura dioica: culture, genome, and cell
924 lineages. *Dev Growth Differ* 50 Suppl 1:S239–S256.
925 Novembre J, Galvani AP, Slatkin M. 2005. The geographic spread of the CCR5 Delta32 HIV-
926 resistance allele. *PLoS Biol* 3:e339.
927 Olivera-Martinez I, Harada H, Halley PA, Storey KG. 2012. Loss of FGF-Dependent Mesoderm
928 Identity and Rise of Endogenous Retinoid Signalling Determine Cessation of Body Axis
929 Elongation. *PLoS Biol* 10:e1001415.
930 Olsen LC, Kourtesis I, Busengdal H, Jensen MF, Hausen H, Chourrout D. 2018. Evidence for a
931 centrosome-attracting body like structure in germ-soma segregation during early
932 development, in the urochordate Oikopleura dioica. *BMC Dev Biol* 18:4.

933 Olsen SK, Garbi M, Zampieri N, Eliseenkova A V., Ornitz DM, Goldfarb M, Mohammadi M. 2003.
934 Fibroblast Growth Factor (FGF) Homologous Factors Share Structural but Not Functional
935 Homology with FGFs. *Journal of Biological Chemistry* 278:34226–34236.
936 Olsnes S, Klingenberg O, Wie˛dłocha A. 2003. Transport of Exogenous Growth Factors and
937 Cytokines to the Cytosol and to the Nucleus. *Physiol Rev* 83:163–182.
938 Olson M V. 1999. When less is more: gene loss as an engine of evolutionary change. *Am J Hum*
939 *Genet* 64:18–23.
940 Ornitz DM, Itoh N. 2015. The fibroblast growth factor signaling pathway. *Wiley Interdiscip Rev Dev*
941 *Biol* 4:215–266.
942 Ornitz DM, Itoh N. 2022. New developments in the biology of fibroblast growth factors. *WIREs*
943 *mechanisms of disease* 14:e1549.
944 Osipova E, Barsacchi R, Brown T, Sadanandan K, Gaede AH, Monte A, Jarrells J, Moebius C,
945 Pippel M, Altshuler DL, et al. 2023. Loss of a gluconeogenic muscle enzyme contributed to
946 adaptive metabolic traits in hummingbirds. *Science* 379:185–190.
947 Oulion S, Bertrand S, Escriva H. 2012. Evolution of the FGF Gene Family. *Int J Evol Biol* 2012:1–
948 12.
949 Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. 2018. A comprehensive review
950 of signal peptides: Structure, roles, and applications. *Eur J Cell Biol* 97:422–441.
951 Pablo JL, Pitt GS. 2016. Fibroblast Growth Factor Homologous Factors: New Roles in Neuronal
952 Health and Disease. *Neuroscientist* 22:19–25.
953 Paschaki M, Schneider C, Rhinn M, Thibault-Carpentier C, Dembélé D, Niederreither K, Dollé P.
954 2013. Transcriptomic analysis of murine embryos lacking endogenous retinoic acid signaling.
955 *PLoS One* 8:e62274.
956 Pasini A, Manenti R, Rothbacher U, Lemaire P. 2012. Antagonizing retinoic acid and FGF/MAPK
957 pathways control posterior body patterning in the invertebrate chordate Ciona intestinalis.
958 *PLoS One* 7:e46193.
959 Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. 2021.
960 UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein*
961 *Sci* 30:70–82.
962 Plessy C, Mansfield MJ, Bliznina A, Masunaga A, West C, Tan Y, Liu AW, Grašič J, del Río Pisula
963 MS, Sánchez-Serna G, et al. 2024. Extreme genome scrambling in marine planktonic
964 *Oikopleura dioica* cryptic species. *Genome Res*.

Plotnikov AN, Eliseenkova A V., Ibrahimi OA, Shriver Z, Sasisekharan R, Lemmon MA, Mohammadi M. 2001. Crystal Structure of Fibroblast Growth Factor 9 Reveals Regions Implicated in Dimerization and Autoinhibition. *Journal of Biological Chemistry* 276:4322–4329.

Popovici C, Conchonaud F, Birnbaum D, Roubin R. 2004. Functional Phylogeny Relates LET-756 to Fibroblast Growth Factor 9. *Journal of Biological Chemistry* 279:40146–40152.

Popovici C, Fallet M, Marguet D, Birnbaum D, Roubin R. 2006. Intracellular trafficking of LET-756, a fibroblast growth factor of C. elegans, is controlled by a balance of export and nuclear signals. *Exp Cell Res* 312:1484–1495.

Popovici C, Roubin R, Coulier F, Birnbaum D. 2005. An evolutionary history of the FGF superfamily. *Bioessays* 27:849–857.

Rees JM, Palmer MA, Gillis JA. 2024. Fgf signalling is required for gill slit formation in the skate, Leucoraja erinacea. *Dev Biol* 506:85–94.

Revest JM, DeMoerlooze L, Dickson C. 2000. Fibroblast growth factor 9 secretion is mediated by a non-cleaved amino- terminal signal sequence. *Journal of Biological Chemistry* 275:8083–8090.

Röttinger E, Saudemont A, Duboc V, Besnardeau L, McClay D, Lepage T. 2008. FGF signals guide migration of mesenchymal cells, control skeletal morphogenesis [corrected] and regulate gastrulation during sea urchin development. *Development* 135:353–365.

Satou Y. 2020. A gene regulatory network for cell fate specification in Ciona embryos. *Curr Top Dev Biol* 139:1–33.

Satou Y, Imai KS, Satoh N. 2002. Fgf genes in the basal chordate Ciona intestinalis. *Dev Genes Evol* 212:432-8.

Schäfer T, Zentgraf H, Zehe C, Brügger B, Bernhagen J, Nickel W. 2004. Unconventional Secretion of Fibroblast Growth Factor 2 Is Mediated by Direct Translocation across the Plasma Membrane of Mammalian Cells. *Journal of Biological Chemistry* 279:6244–6251.

Schlessinger J, Plotnikov AN, Ibrahimi OA, Eliseenkova A V., Yeh BK, Yayon A, Linhardt RJ, Mohammadi M. 2000. Crystal Structure of a Ternary FGF-FGFR-Heparin Complex Reveals a Dual Role for Heparin in FGFR Binding and Dimerization. *Mol Cell* 6:743–750.

Schoorlemmer J, Goldfarb M. 2002. Fibroblast Growth Factor Homologous Factors and the Islet Brain-2 Scaffold Protein Regulate Activation of a Stress-activated Protein Kinase. *Journal of Biological Chemistry* 277:49111–49119.

997 Sharma V, Hecker N, Roscito JG, Foerster L, Langer BE, Hiller M. 2018. A genomics approach
998 reveals insights into the importance of gene losses for mammalian adaptations. *Nat Commun*
999 9:1215.
1000 Sheng Z, Liang Y, Lin C-Y, Comai L, Chirico WJ. 2005. Direct Regulation of rRNA Transcription
1001 by Fibroblast Growth Factor 2. *Mol Cell Biol* 25:9419–9426.
1002 Shi W, Peyrot SM, Munro E, Levine M. 2009. FGF3 in the floor plate directs notochord convergent
1003 extension in the Ciona tadpole. *Development* 136:23–28.
1004 Sluzalska KD, Slawski J, Sochacka M, Lampart A, Otlewski J, Zakrzewska M. 2021. Intracellular
1005 partners of fibroblast growth factors 1 and 2 - implications for functions. *Cytokine Growth*
1006 *Factor Rev* 57:93–111.
1007 Smallwood PM, Munoz-Sanjuan I, Tong P, Macke JP, Hendry SH, Gilbert DJ, Copeland NG,
1008 Jenkins NA, Nathans J. 1996. Fibroblast growth factor (FGF) homologous factors: new
1009 members of the FGF family implicated in nervous system development. *Proc Natl Acad Sci*
1010 *U S A* 93:9850–9857.
1011 Stach T, Winter J, Bouquet J-MM, Chourrout D, Schnabel R. 2008. Embryology of a planktonic
1012 tunicate reveals traces of sessility. *Proc Natl Acad Sci U S A* 105:7229–7234.
1013 Stolfi A, Ryan K, Meinertzhagen IA, Christiaen L. 2015. Migratory neuronal progenitors arise from
1014 the neural plate borders in tunicates. *Nature* 527:371–374.
1015 Technau U, Scholz CB. 2003. Origin and evolution of endoderm and mesoderm. *Int J Dev Biol*
1016 47:531–539.
1017 Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O,
1018 Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides
1019 using protein language models. *Nat Biotechnol* 40:1023–1025.
1020 Teven CM, Farina EM, Rivas J, Reid RR. 2014. Fibroblast growth factor (FGF) signaling in
1021 development and skeletal diseases. *Genes Dis* 1:199–213.
1022 Thisse B, Thisse C. 2005. Functions and regulations of fibroblast growth factor signaling during
1023 embryonic development. *Dev Biol* 287:390–402.
1024 Treen N, Chavarria E, Weaver CJ, Brangwynne CP, Levine M. 2023. An FGF timer for zygotic
1025 genome activation. *Genes Dev* 37:80–85.
1026 Treen N, Yoshida K, Sakuma T, Sasaki H, Kawai N, Yamamoto T. 2014. Tissue-specific and
1027 ubiquitous gene knockouts by TALEN electroporation provide new approaches to
1028 investigating gene function in Ciona. :481–487.

1029 Veeman MT, Newman-Smith E, El-Nachef D, Smith WC. 2010. The ascidian mouth opening is
1030 derived from the anterior neuropore: reassessing the mouth/neural tube relationship in
1031 chordate evolution. *Dev Biol* 344:138–149.
1032 Wagner E, Levine M. 2012. FGF signaling establishes the anterior border of the Ciona neural tube.
1033 *Development* 139:2351–2359.
1034 Wang Chaojian, Wang Chuan, Hoch EG, Pitt GS. 2011. Identification of novel interaction sites that
1035 determine specificity between fibroblast growth factor homologous factors and voltage-gated
1036 sodium channels. *J Biol Chem* 286:24253–24263.
1037 Wang K, Omotezako T, Kishi K, Nishida H, Onuma TA. 2015. Maternal and zygotic transcriptomes
1038 in the appendicularian, Oikopleura dioica: novel protein-encoding genes, intra-species
1039 sequence variations, and trans-spliced RNA leader. *Dev Genes Evol* 225:149–159.
1040 Wilson V, Olivera-Martinez I, Storey KG. 2009. Stem cells, signals and vertebrate body axis
1041 extension. *Development* 136:1591–1604.
1042 Wu Q-F, Yang L, Li S, Wang Q, Yuan X-B, Gao X, Bao L, Zhang X. 2012. Fibroblast Growth Factor
1043 13 Is a Microtubule-Stabilizing Protein Regulating Neuronal Polarization and Migration. *Cell*
1044 149:1549–1564.
1045 Xie Y, Su N, Yang J, Tan Q, Huang S, Jin M, Ni Z, Zhang B, Zhang D, Luo F, et al. 2020.
1046 FGF/FGFR signaling in health and disease. *Signal Transduct Target Ther* 5:181.
1047 Xu R, Ori A, Rudd TR, Uniewicz KA, Ahmed YA, Guimond SE, Skidmore MA, Siligardi G, Yates
1048 EA, Fernig DG. 2012. Diversification of the structural determinants of fibroblast growth factor-
1049 heparin interactions: implications for binding specificity. *J Biol Chem* 287:40061–40073.
1050 Xu Y-C, Guo Y-L. 2020. Less Is More, Natural Loss-of-Function Mutation Is a Strategy for
1051 Adaptation. *Plant Commun* 1:100103.
1052 Yasuo H, Hudson C. 2007. FGF8/17/18 functions together with FGF9/16/20 during formation of
1053 the notochord in Ciona embryos. *Dev Biol* 302:92–103.

1054 Acknowledgements

1055 We thank present and past team members on CC's laboratory for assistance and fruitful
1056 discussions on FGF signaling, gene loss, and evolution, specially to Sebastian Artime Paoletti for
1057 running the Oikopleura facility in the University of Barcelona. We thank to Centres Científics i
1058 Tecnològics de la UB for sea water supply and sequencing services.

Author contributions

Conceptualization: CC, GS; formal Analysis: GS; funding acquisition: CC; investigation: GS, PB, JBR, AF, AFR, MF, NTA, JNW; methodology: GS, PB, JBR, AF, AFR, MF, NTA, RA; project administration: CC; resources: JNW, MJM, CP, NL; software: NTA, JNW, MJM, CP; CP supervision: CC; validation: GS, CC; visualization: GS, CC; writing (original draft): GS, CC; Writing (review & editing): GS, CC.

Funding

CC was funded by PID2019-110562GB-I00 and PID2022-141627NB-I00 from the Spanish Ministerio de Ciencia e Innovación and by ICREA Acadèmia Ac2215698 and 2021-SGR00372 AGAUR, Generalitat de Catalunya; VR by 2017BP00139 AGAUR, Generalitat de Catalunya and 2019IRBio001 from IRBio, Universitat de Barcelona; GSS by FPU18/02414 fellowship from Ministerio de Educación y cultura M.P.-C. by colaboración-2015/16, MFT by a PREDOC2020/58 fellowship from Universitat de Barcelona; AFR by MS12 Margarita Salas from Ministerio de Universidades (Spain).

Figure Legends

Figure 1. Evolutionary tree of the *Fgf* subfamilies in chordates. ML phylogenetic tree of the *Fgf* family in chordates reveals that the 10 *Fgf* genes found *Oikopleura dioica*, together with all other genes found in other appendicularians species (in red) group in two clusters with high support values (nodes with red solid circles). The tree topology indicates that the two clusters belong with high support (nodes with black solid circles) to two subfamilies, *Fgf9/16/20* (red background) and *Fgf11/12/13/14* (blue background). The presence of *Fgfs* from ascidians (in blue), vertebrates (in black) and cephalochordates (in green) allowed to infer that appendicularians has lost the subfamilies *Fgf8/17/18*, *Fgf19/21/23*, *Fgf7/10/22* and *Fgf4/5/6*. The absence of *Fgf11/12/13/14* in cephalochordates suggests that this subfamily might be a synapomorphy of the olfactores. Well supported nodes of other Fgf subfamilies (eBayes=1) with members of more than one subphylum are indicated with grey solid circles. Node support values correspond to likelihood-based method aLRT-SH-like/aBayes/uf-boostrap. The scale bar indicates amino-acid substitutions. Species abbreviations: Vertebrates (in black): *Danio rerio* (Dre), *Gallus gallus* (Gga), *Homo sapiens* (Hsa), *Latimeria chalumnae* (Lch), *Mus musculus*

(Mmu), *Xenopus tropicalis* (Xtr); Ascidian tunicates (in blue): *Botrylloides leachii* (Ble), *Botrylloides schlosseri* (Bsc), *Ciona robusta* (Cro), *Ciona savignyi* (Csa), *Halocynthia aurantium* (Hau), *Halocynthia roretzi* (Hro), *Molgula occidentalis* (Mocci), *Molgula occulta* (Moccu), *Molgula oculata* (Mocul), *Phallusia fumigata* (Pfu), *Phallusia mammillata* (Pma); Appendicularian tunicates (in red): *Oikopleura albicans* (Oal), *Oikopleura dioica* (Odi), *Oikopleura vanhoeffeni* (Ova); Cephalochordates (in green): *Branchiostoma belcheri* (Bbe), *Branchiostoma floridae* (Bfl), *Branchiostoma lanceolatum* (Bla)

Figure 2. Comparative synteny analysis of *Fgf* genes among the three *Oikopleura dioica* cryptic species from Barcelona (BAR), Osaka (OSA), and Okinawa (OKI). (A) Comparison of microsynteny conservation between the genomic neighborhoods of *Fgf* genes (black arrow, and ten adjacent genes on each side). The BAR genome was used as the reference. Here, we show two illustrative examples, in which the *Fgf9/16/20e* neighborhood represents a case of high level of microsynteny conservation, and the *Fgf9/16/20a* neighborhood represents a case of low level of microsinteny conservation, especially when compared with OKI. The microsynteny comparison of the full *O. dioica Fgf* catalogue is provided in **Supplementary Figure 1**. (B) Macrosynteny analysis comparing the position of *Fgf* genes at arm chromosome level. Each *Fgf* ortholog is labeled with a distinctive color.

Figure 3. Comparative gene structures of *O. dioica* and other chordate *Fgf* genes. Exon intron organization supports phylogenetic classification of *Fgf9/16/20* and *Fgf11/12/13/14* paralogs of *O. dioica*. "cfi" denotes conserved FGF domain flanking introns, and "ici" denotes conserved internal FGF domain introns. Bfl_Fgfs represent the common structure of cephalochordate *Fgf* genes, featuring two internal core introns (ici) within the FGF domain coding sequence. Gene-specific introns are depicted as arrowheads and dashed lines at their respective locations. Predicted functional motifs are indicated as described in the legend. Black underlines highlight the presence and location of β-sheets as predicted by AlphaFold2. Orange dashed underlines highlight the presence and location of β-sheets that have been empirically determined, even though the AlphaFold2 software does not predict them (Goetz et al., 2009; Olsen et al., 2003; Plotnikov et al., 2001). For comparative purposes, genes and motifs are not drawn to scale. Dashed lines indicate alternative splicing variants of *Fgf11/12/13/14* (Dis = distal, Med = medial, Pro = proximal),

and black dashed lines boxes indicate exon length differences between *O. dioica* cryptic species.

Figure 4. Developmental expression atlas of *O. dioica Fgf* genes. Whole-mount in situ hybridization images of *O. dioica* at various developmental stages: eggs (A-J1), 8-cell embryos (A-J2), 64-cell embryos (A-J3), incipient tailbud (ITB) embryos (A-J4), early tailbud (ETB) embryos (A-J5), mid tailbud (MTB) embryos (A-J6), late tailbud (LTB) embryos (A-J7), just hatchlings (A-J8), early hatchling larvae (A-J9), mid hatchling larvae (A-J10), and late hatchling larvae (A-J11). Central images in each panel are left lateral views, oriented anterior to the left and dorsal to the top. Upper-right image insets (') are dorsal views of optical cross-sections at the levels indicated by black dashed lines. Black arrowheads label an stained cortical spot in unfertilized eggs; black double arrowheads label the A pair blastomeres in 8-cell embryos; orange arrowheads mark ingressing vegetal blastomeres in the 64-cell embryos; blue arrowheads label neural derivatives (cyan-blue labels the neural plate in 64-cell and ITB embryos, and the nerve cord in later stages; dark-blue labels the caudal ganglion, and pale-blue labels the anterior brain); light green arrowheads label epidermal domains in the trunk and light green double arrowheads label the area of the Langerhans receptors primordia; dark green arrowheads label epidermal domains in the tailbud tip; green dashed lines mark the lateral epithelium of the tail and the fins; purple arrowheads label undetermined endomesodermal domains in the trunk; magenta arrowheads mark the mouth primordium; magenta double arrowheads mark the pharyngeal slits; yellow arrowheads label notochord cells; red arrowheads label muscle precursor cells and muscle cells in the tail.

Figure 5. Evolutionary scenario of *Fgf* subfamilies in chordates. (A) Evolutionary tree of chordate subphyla indicating main events of losses (L), gains (G), duplications (D) or expansions by burst of duplications (E), as well as main associated patterns of conservation, innovation, function shuffling and extinction of *Fgf* expression domains. 2R-WGD: two rounds of whole genome duplication. (B) Comparative schematic representation of the main expression domains of *Fgf* subfamily members between ascidians and appendicularians (tailbud stage in the left, and hatchling in the right). Distinct colors are assigned to each *Fgf* as indicated in the figure legend. *FgfNA1* expression has not been described in ascidians, and *Fgf4/5/6** has been detected maternally and widely throughout development with no

obvious tissue-specific domains (Imai et al., 2004). In appendicularians, expression of *Fgf9/16/20* paralogs is the most abundant in tailbud stages, and expression of *Fgf11/12/13/14* paralogs is more obvious at hatchling stages. Asterisk (as in br*) denotes that expression was found in a slightly earlier stage than the one represented in the figure. Ascidians show a characteristic PTFS (Posterior-Tail FGF Source) and ATRAS (Anterior-Tail RA Source), while in appendicularians the loss of RA signaling (Cañestro & Postlethwait, 2007; Martí-Solans et al., 2016) might be related to the innovation of an ATFS (Anterior-Tail FGF Source), considering that the conserved RA-FGF antagonistic action emerged at the base of olfactores (Pasini et al. 2012; Bertrand et al. 2015). Abbreviations: an: anterior notochord, am: anterior muscle, br: brain, cf: ciliary funnel, cns: central nervous system, gs: gill slit, m: mouth, o: oikoblastic epithilium, pl: placode, pm: posterior muscle, te: tail epidermis, tte: terminal tail epidermis.

Supplementary files

Figure S1. Comparative microsynteny conservation in the *Fgf* genes among the three *O. dioica* cryptic species. Comparison of microsynteny conservation between the genomic neighborhoods of Fgf genes (black arrow), and ten adjacent genes on each side. The BAR genome was used as the reference in comparisons with OSA and OKI.

Figure S2. Three-dimensional models for *O. dioica* Fgf proteins. Models are colored according to their predicted local distance difference test (pLDDT) score. Predicted β-sheets are highlighted in light green in the models as well as in the protein sequences. Predicted α-helices are highlighted in light yellow in the protein sequences. All Fgf11/12/13/14 paralogs models and sequences correspond to the proximal isoform.

Figure S3. Protein alignment of *O. dioica* Fgf orthologs. The alignment includes the FGF domain of each ortholog. Sequence conservation is depicted according to the Clustal X default coloring. Black solid line boxes denote β-sheets as predicted by the AlphaFold2 software. Black dashed line boxes in Hsa_FGF9 and Hsa_FGF12 indicate β-sheets empirically confirmed but not predicted by AlphaFold2 (Goetz et al. 2009; Plotnikov et al. 2001). Red arrowheads and boxes in the *C. robusta* and *O. dioica* Fgf9/16/20 sequences highlight the positions of distinctive and conserved cysteines found in all tunicate Fgf9/16/20

paralogs. Yellow arrowheads and boxes in Fgf11/12/13/14 sequences denote the positions of the Leucine-Arginine pair, characteristic of the intracellular Fgf11/12/13/14 orthologs. Magenta shadings and dashed lines indicate the regions involved in binding heparin (Xu et al. 2012). Abbreviations: *Branchiostoma floridae* (Bfl), *Homo sapiens* (Hsa), *Ciona robusta* (Cro), *Oikopleura dioica* (Odi)

Figure S4. Hydropathy profile of chordate Fgf9/16/20 proteins. The graphs show the Kyte-Doolittle hydropathy score (y-axis) along the length of each protein sequence for *Homo sapiens* (Hsa), *Branchiostoma floridae* (Bfl), *Ciona robusta* (Cro), and *Oikopleura dioica* (Odi) Fgf9/16/20 orthologs. Scores are scaled to 1 for comparison purposes. Green squares and shaded areas indicate the presence and position of a cleavable signal peptide (SP) as predicted by Phobius. Red squares and shaded areas mark the position of well-conserved EFISIA motifs. Gray squares and shaded areas represent the positions of degraded EFISIA motifs. The amino acid letters in the EFISIA motif of each protein are colored red if they are conserved with the vertebrate counterpart or if they have been substituted by an amino acid with similar physicochemical properties.

Table S1. Primers used for the cloning of *O. dioica Fgf* genes.

Table S2. Genomic loci of *Fgf* genes in the four sequenced genomes of *O. dioica.*

Table S3. List of sequences used for phylogenetic analyses.

Table S4. Fgf protein identity and similarity.

Supplementary file 1. Fgf protein alignment

Supplementary file 2. *Fgf* phylogenetic tree

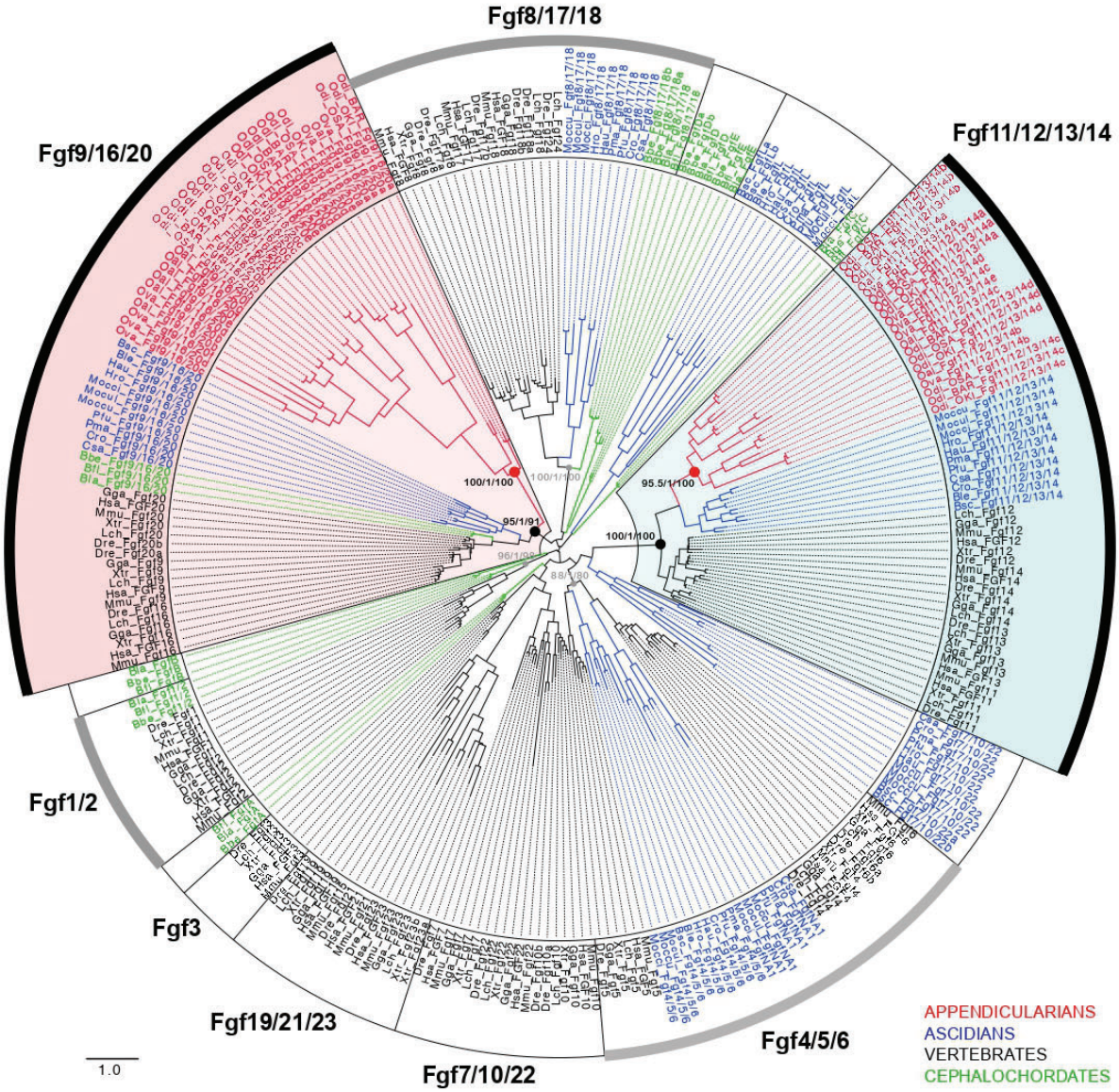

Fgf8/17/18
Fgf11/12/13/14
Fgf9/16/20
Fgf1/2
Fgf3
Fgf19/21/23
Fgf7/10/22
Fgf4/5/6
100/1/100
95/1/91
95.5/1/100
100/1/100
1.0
APPENDICULARIANS
ASCIDIANS
VERTEBRATES
CEPHALOCHORDATES

Figure 1.

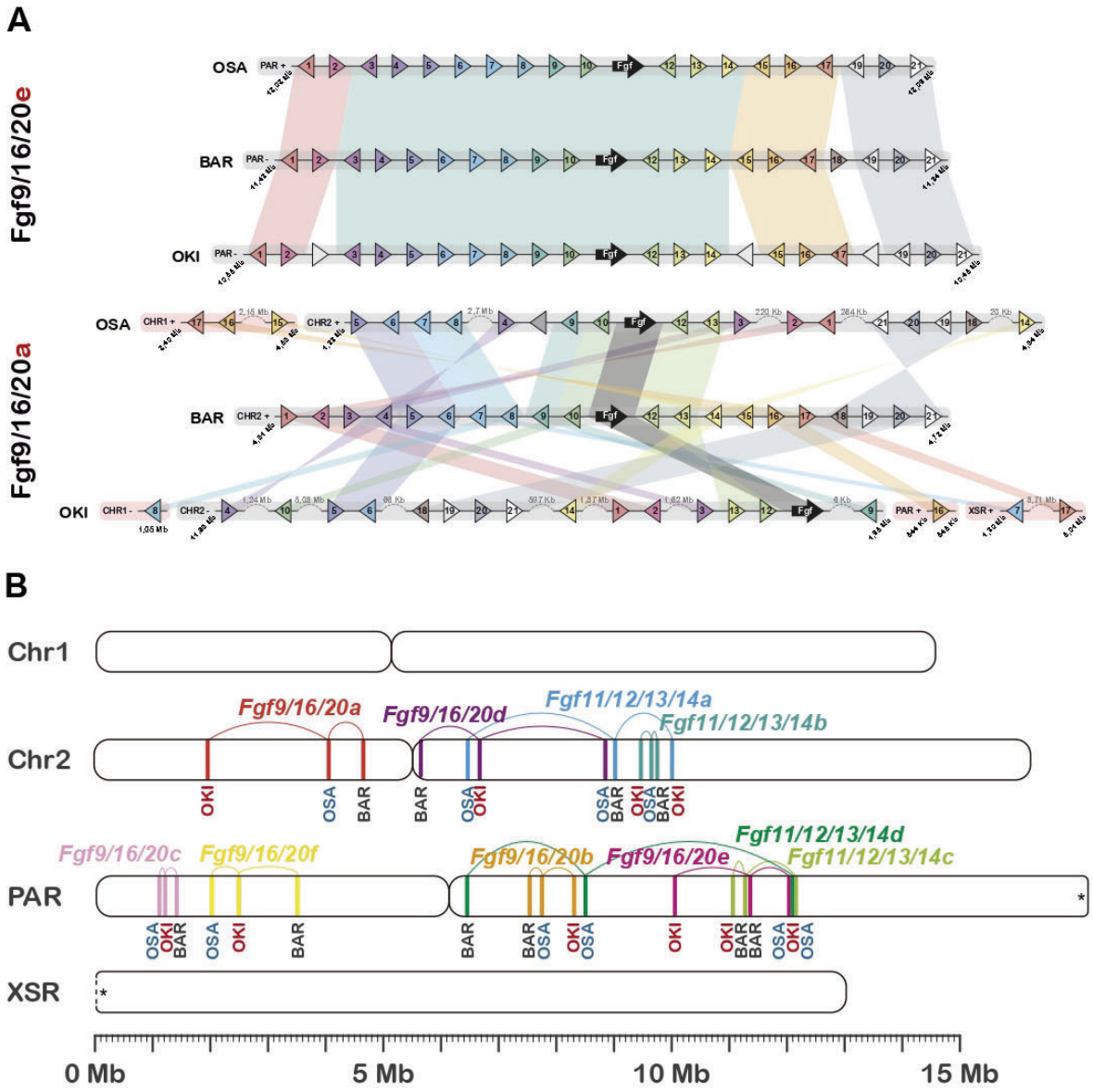

A
Fgf9/16/20e
OSA
BAR
OKI
Fgf9/16/20a
OSA
BAR
OKI
B
Chr1
Chr2
Fgf9/16/20a
Fgf9/16/20d
Fgf11/12/13/14a
Fgf11/12/13/14b
PAR
Fgf9/16/20c
Fgf9/16/20f
Fgf9/16/20b
Fgf9/16/20e
Fgf11/12/13/14d
Fgf11/12/13/14c
XSR
0 Mb
5 Mb
10 Mb
15 Mb

Figure 2.

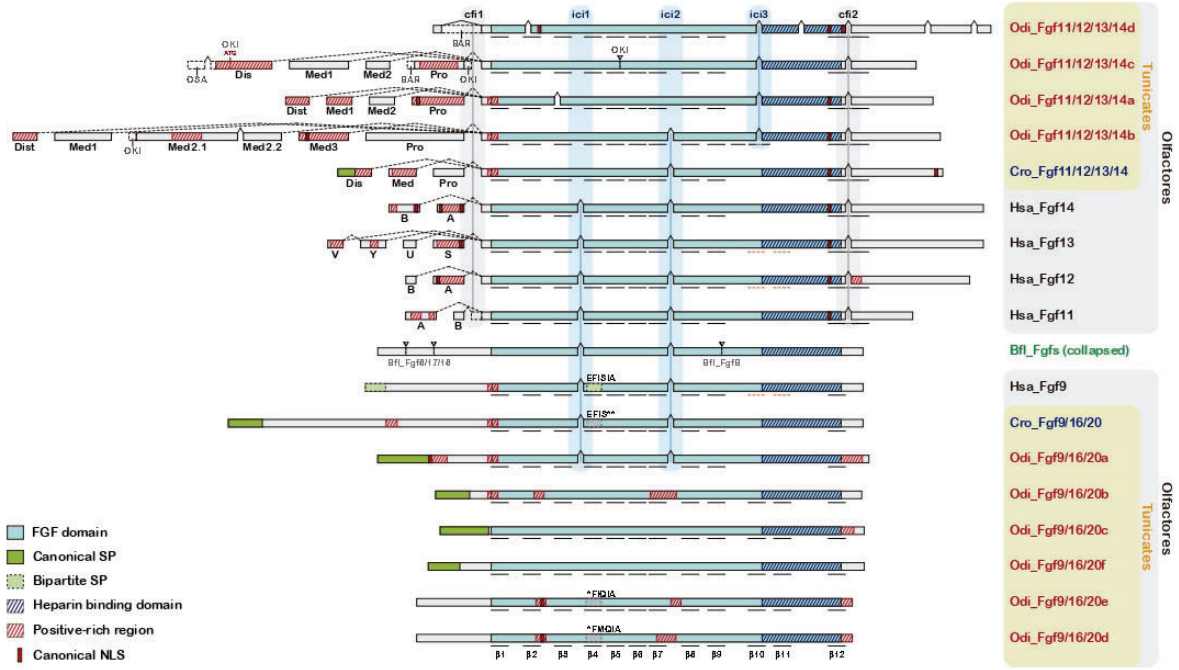

cfi1
ici1
ici2
ici3
cfi2
Odi_Fgf11/12/13/14d
Odi_Fgf11/12/13/14c
Odi_Fgf11/12/13/14a
Odi_Fgf11/12/13/14b
Cro_Fgf11/12/13/14
Hsa_Fgf14
Hsa_Fgf13
Hsa_Fgf12
Hsa_Fgf11
Bfl_Fgfs (collapsed)
Hsa_Fgf9
Cro_Fgf9/16/20
Odi_Fgf9/16/20a
Odi_Fgf9/16/20b
Odi_Fgf9/16/20c
Odi_Fgf9/16/20f
Odi_Fgf9/16/20e
Odi_Fgf9/16/20d
Tunicates
Olfactores
Olfactores
Tunicates
FGF domain
Canonical SP
Bipartite SP
Heparin binding domain
Positive-rich region
Canonical NLS

Figure 3.

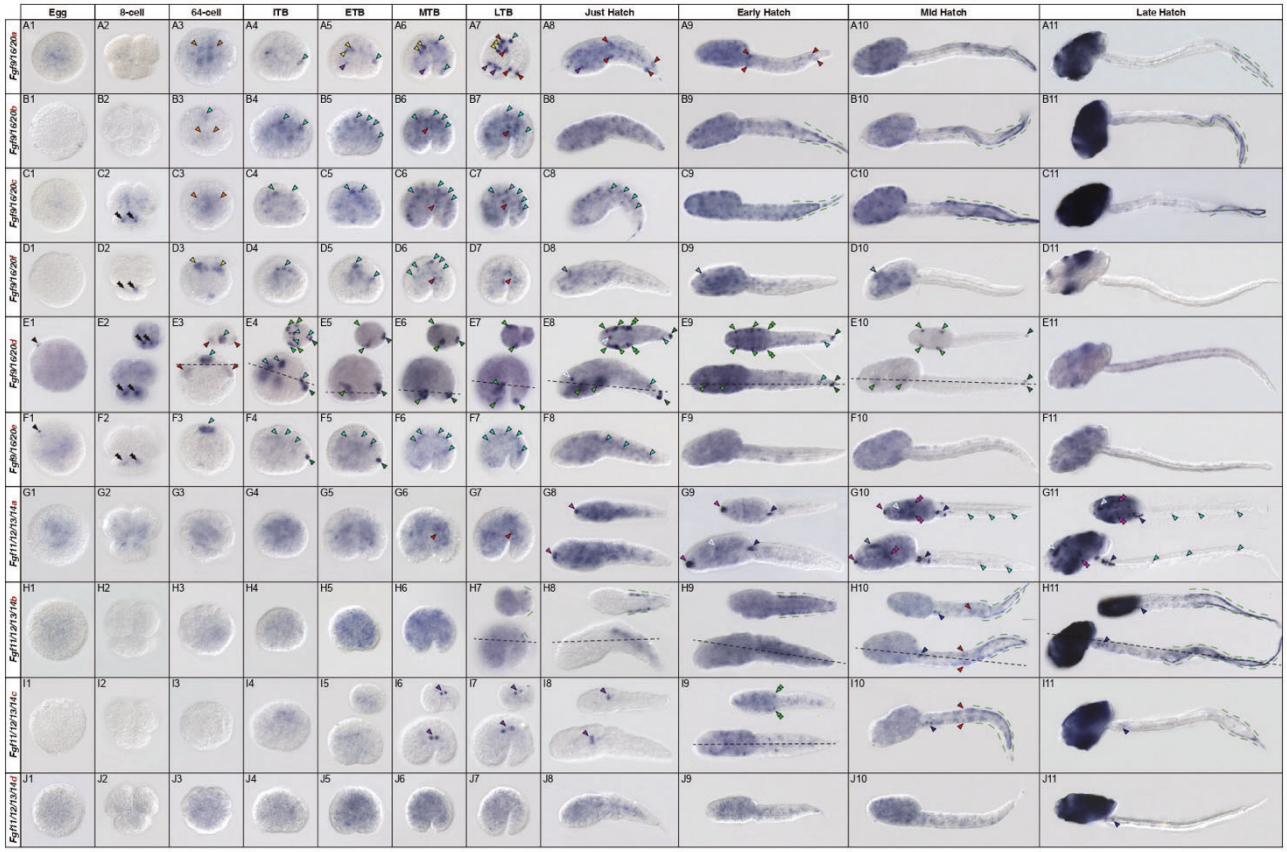

Egg
8-cell
64-cell
ITB
ETB
MTB
LTB
Just Hatch
Early Hatch
Mid Hatch
Late Hatch
Fgf9/16/20a
Fgf9/16/20b
Fgf9/16/20c
Fgf9/16/20f
Fgf9/16/20d
Fgf9/16/20e
Fgf11/12/13/14a
Fgf11/12/13/14b
Fgf11/12/13/14c
Fgf11/12/13/14d

Figure 4.

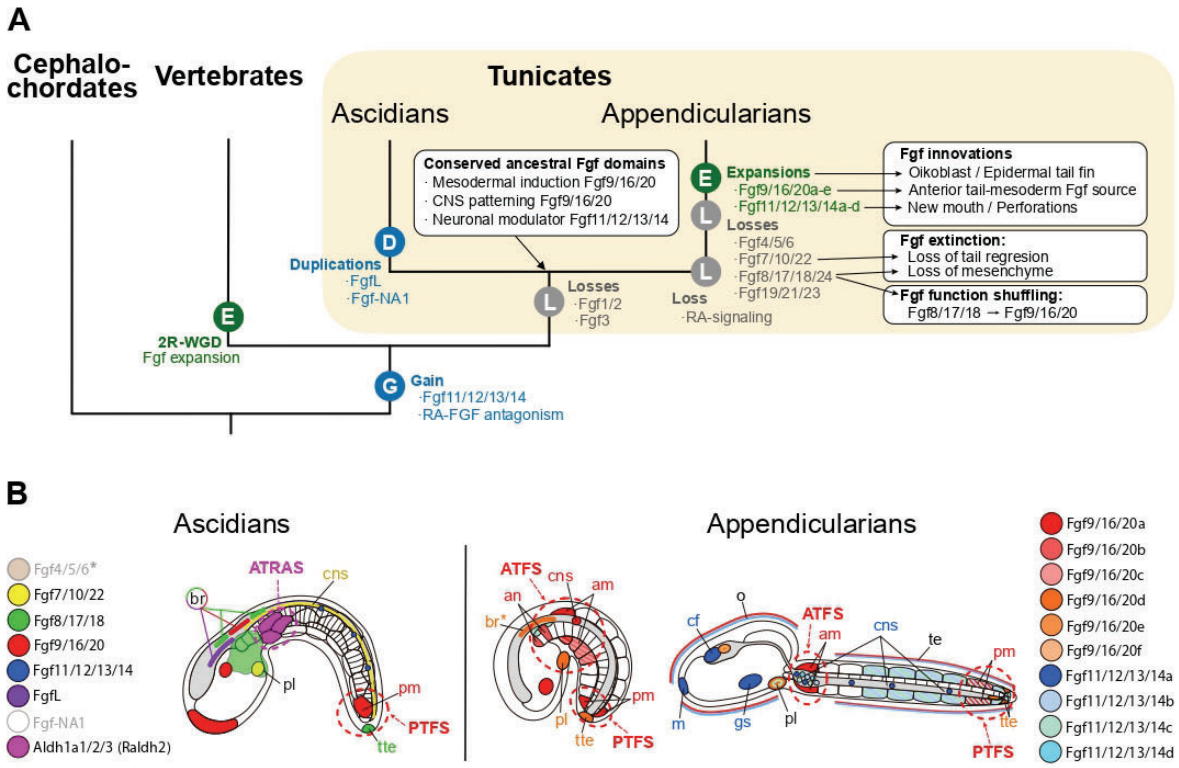
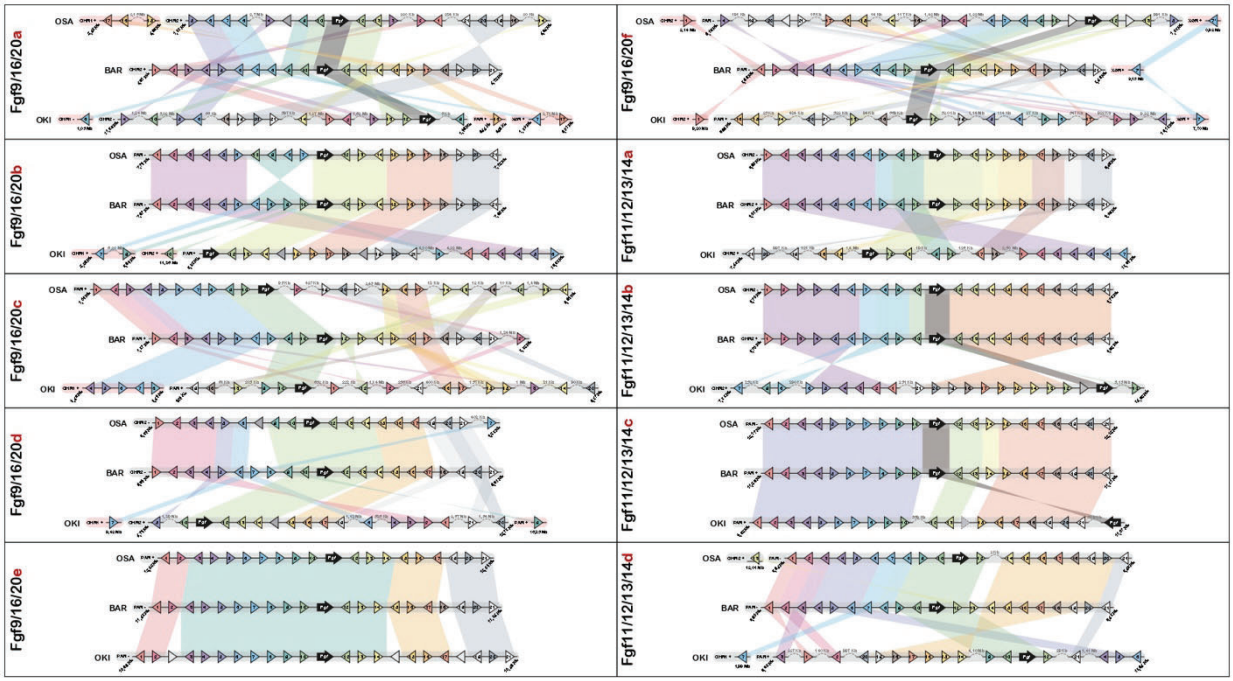
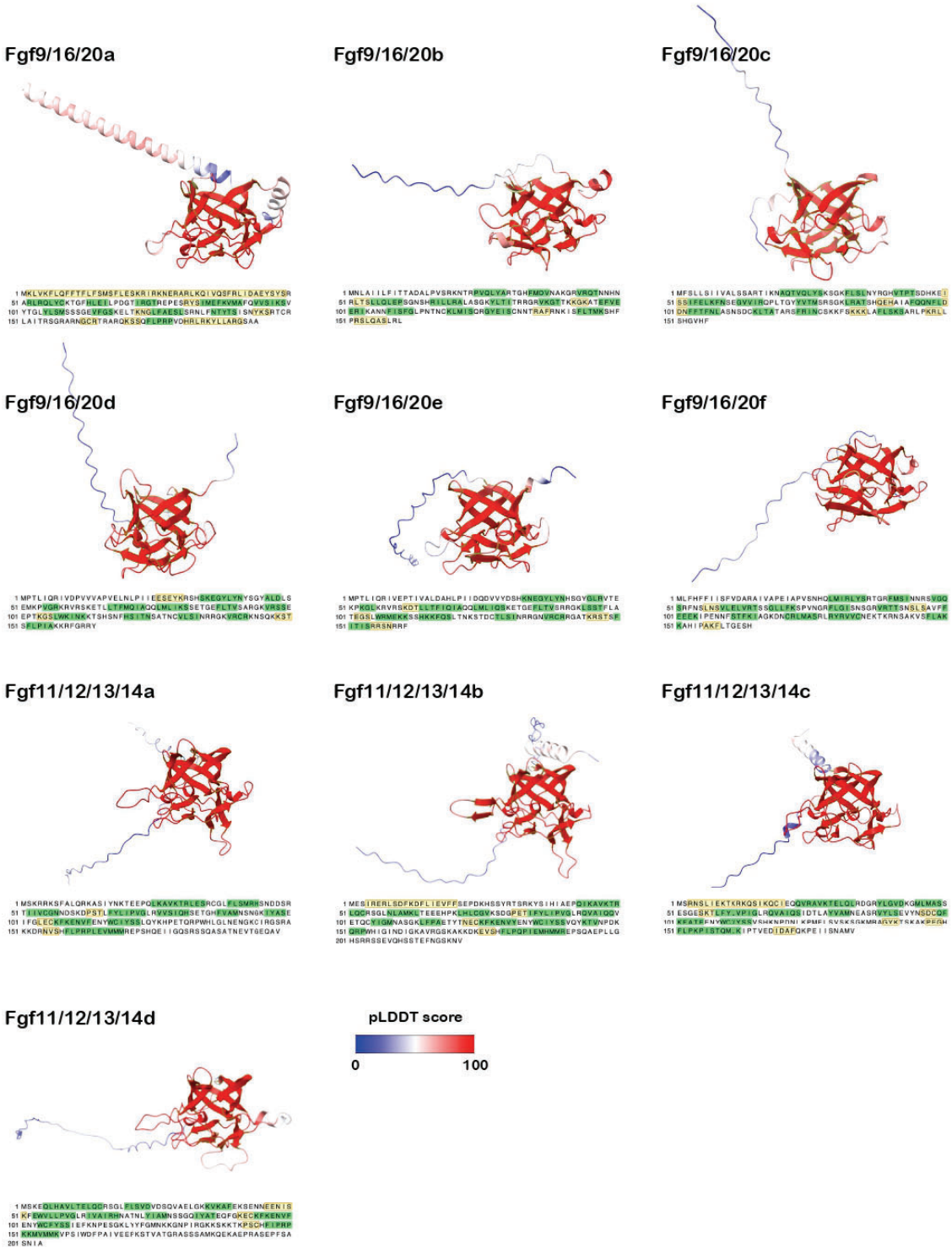

A
Cephalo-chordates
Vertebrates
Tunicates
Ascidians
Appendicularians
Conserved ancestral Fgf domains
· Mesodermal induction Fgf9/16/20
· CNS patterning Fgf9/16/20
· Neuronal modulator Fgf11/12/13/14
E
Expansions
·Fgf9/16/20a-e
·Fgf11/12/13/14a-d
L
Losses
·Fgf4/5/6
·Fgf7/10/22
·Fgf8/17/18/24
·Fgf19/21/23
Loss
·RA-signaling
Fgf innovations
Oikoblast / Epidermal tail fin
Anterior tail-mesoderm Fgf source
New mouth / Perforations
Fgf extinction:
Loss of tail regresion
Loss of mesenchyme
Fgf function shuffling:
Fgf8/17/18 → Fgf9/16/20
D
Duplications
·FgfL
·Fgf-NA1
L
Losses
·Fgf1/2
·Fgf3
E
2R-WGD
Fgf expansion
G
Gain
·Fgf11/12/13/14
·RA-FGF antagonism
B
Ascidians
Appendicularians
Fgf4/5/6*
Fgf7/10/22
Fgf8/17/18
Fgf9/16/20
Fgf11/12/13/14
FgfL
Fgf-NA1
Aldh1a1/2/3 (Raldh2)
ATRAS
cns
br
pl
pm
PTFS
tte
ATFS
an
am
br*
o
cf
te
m
gs
Fgf9/16/20a
Fgf9/16/20b
Fgf9/16/20c
Fgf9/16/20d
Fgf9/16/20e
Fgf9/16/20f
Fgf11/12/13/14a
Fgf11/12/13/14b
Fgf11/12/13/14c
Fgf11/12/13/14d

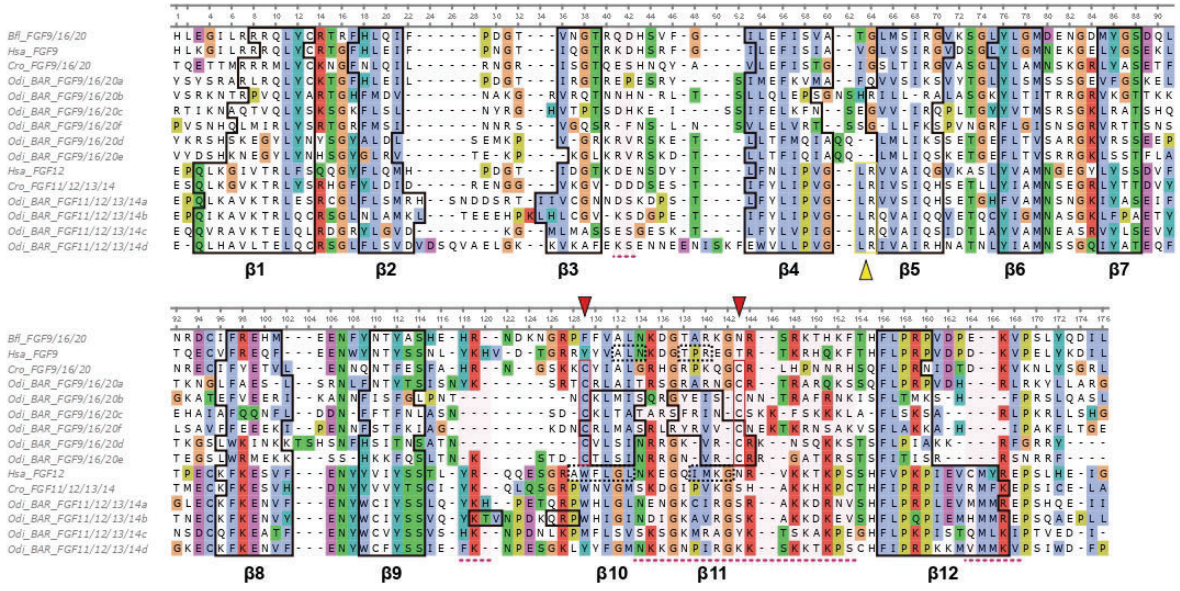
Figure 5.



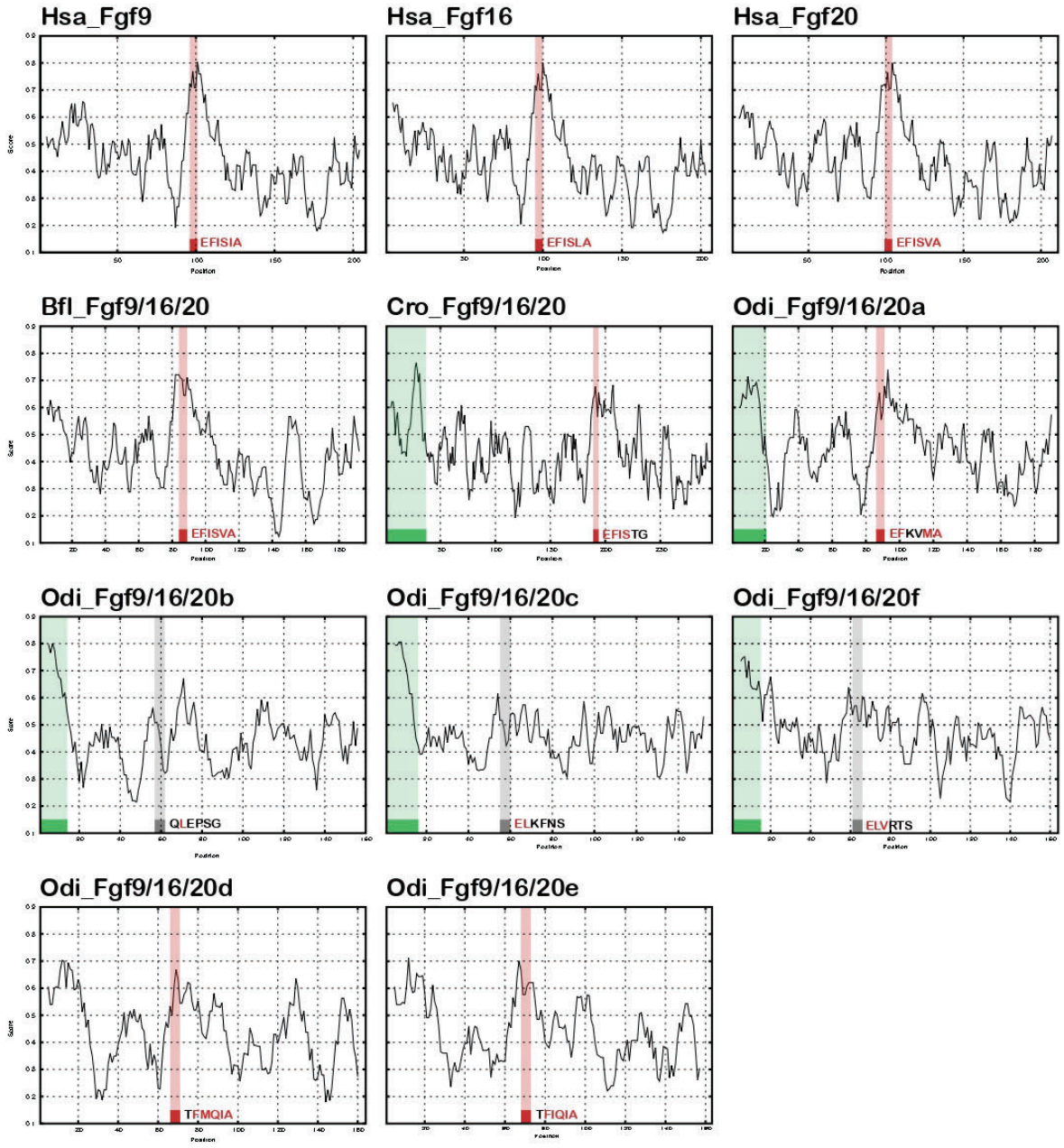
Supplementary Figure 1.



Supplementary Figure 2.



Supplementary Figure 3.



Supplementary Figure 4.