UNIVERSITAT DE BARCELONA

Of flatworms and genes

Systematics, hybridization and evolutionary history of the Corso-Sardinian *Dugesia* (Platyhelminthes, Tricladida)

Daniel Dols Serrate

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

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

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

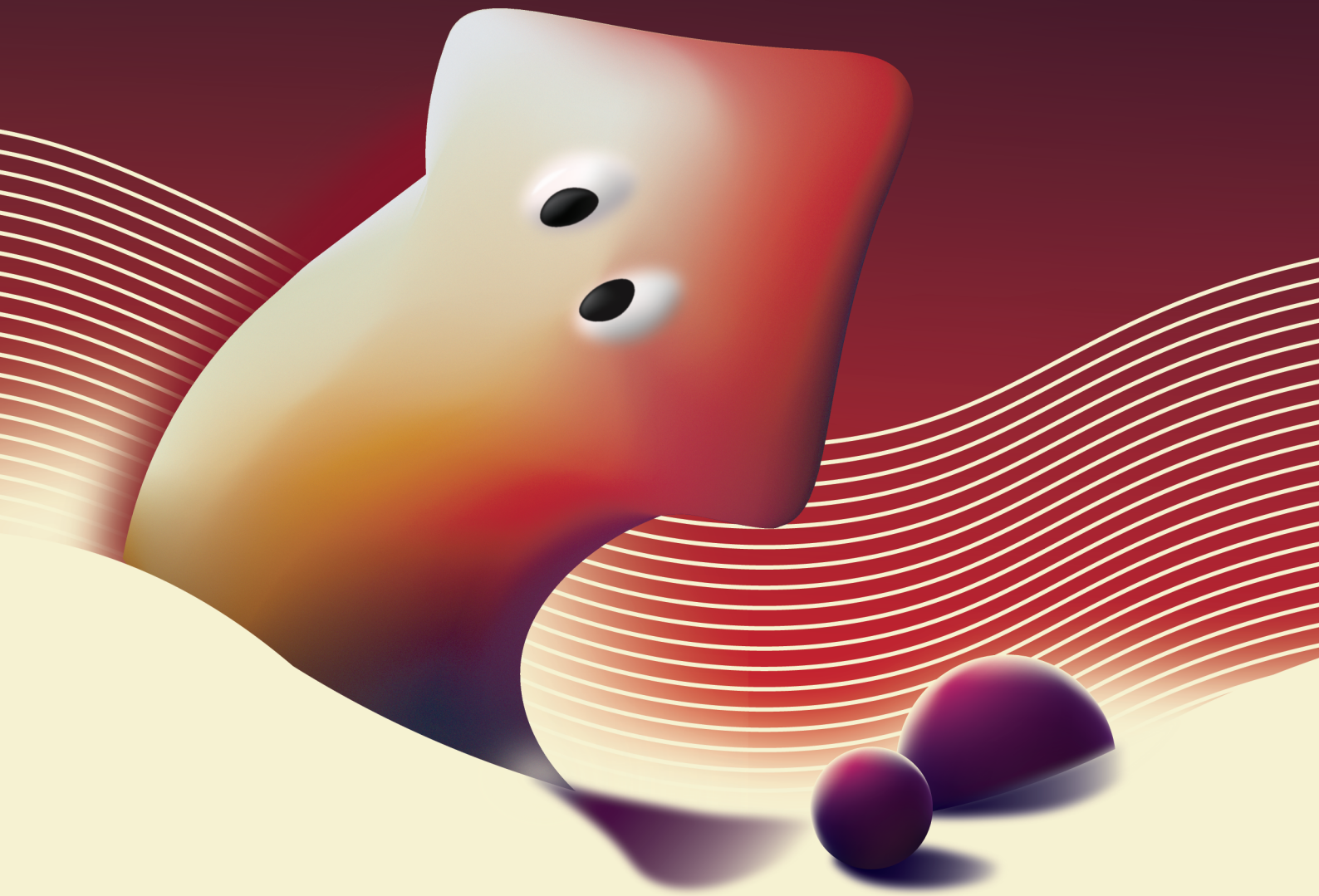
Doctoral Thesis

OF FLATWORMS AND GENES

Systematics, hybridization and evolutionary history of the Corso-Sardinian *Dugesia* (Platyhelminthes, Tricladida)

Daniel Dols Serrate

Barcelona, 2024



UNIVERSITAT DE BARCELONA

IRBio

Institut de Recerca de la Biodiversitat
UNIVERSITAT DE BARCELONA

Of flatworms and genes: Systematics, hybridization and evolutionary history of the Corso-Sardinian *Dugesia* (Platyhelminthes, Tricladida)

Memòria presentada per **Daniel Dols Serrate**
per optar al grau de Doctor per la Universitat de Barcelona

Facultat de Biologia

Departament de Genètica, Microbiologia i Estadística

Programa de Doctorat en Genètica

Memòria dirigida i tutoritzada per:
Dra. Marta Riutort León

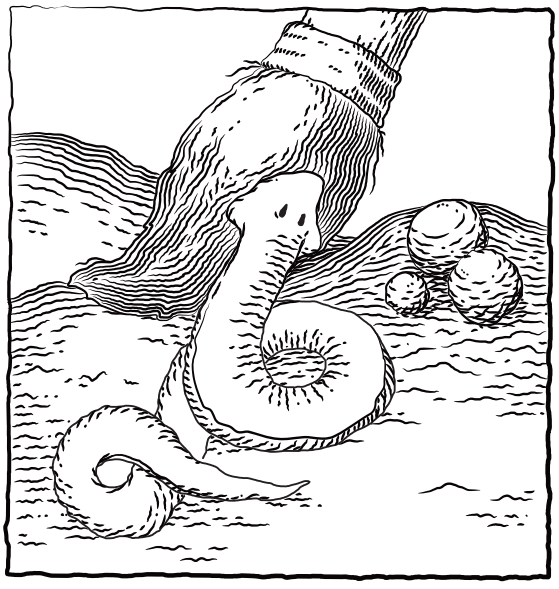
El doctorand:
Daniel Dols Serrate

Barcelona, Maig de 2024

Cover illustrations: Daniel Dols Serrate

ALS MEUS PARES

Per haver estat, en molts aspectes,
els meus mestres



The greatest adventure is what lies ahead.
Today and tomorrow are yet to be said.
The chances, the changes are all yours to make.
The mold of your life is in your hands to break.

J. R. R. Tolkien, *The Hobbit*

Contents

Agraïments/Acknowledgements I

Abstract VII

Sinopsi IX

Introduction 1

1. Evolutionary Biology 1

1.1. A historical perspective 1

1.2. Phylogenetics 2

1.2. Phylogenetic markers 2

2. Genomic sequencing technologies 5

2.1. Brief history of genomic sequencing 5

2.1.1. First-generation DNA sequencing 5

2.1.2. Second-generation DNA sequencing 6

2.1.3. Third-generation DNA sequencing 6

2.2. Chromosome-scale genomes 7

2.3. Markers in the NGS context 8

3. Phylogenetic inference methods 9

3.1. Brief accounts on distance-based methods 9

3.2. Character-based methods 10

4. Sources of gene tree discordance 13

5. Species delimitation in an integrative framework 17

6. The Tricladida Lang, 1884 19

6.1. The Bauplan 19

6.2. Regeneration 21

6.3. Ecology 22

6.4. Systematics 23

7. The family Dugesiidae Ball, 1974 27

8. The genus *Dugesia* Girard, 1850 29

9. The Corso-Sardinian clade 33

9.1. Geographical context 33

9.2. Historical context 34

9.3. An appealing case of study 36

Objectives 39

Publications 41

Supervisor's report 43

Chapter I 43

Dugesia hepta and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex? 49

Chapter II 91

Chapter II.1

Fantastic beasts and how to delimit them: an integrative approach using multispecies coalescent methods reveals two new, endemic *Dugesia* species (Platyhelminthes: Tricladida) from Corsica and Sardinia 93

Chapter II.2

New insights into the evolution and biogeography of freshwater planarians on islands in the Tyrrhenian Sea, Western Mediterranean Basin, with the integrative description of a new endemic species from Corsica (Platyhelminthes, Tricladida, *Dugesia*) 135

Chapter III 189

Mixed, not stirred: Genomic analyses of *Dugesia* (Platyhelminthes, Tricladida) shed light on the evolutionary history and diversification of the Corso-Sardinian clade and reveal the first case of interspecific hybridization in freshwater flatworms 191

Chapter IV 307

The first chromosome-scale assemblies of *Dugesia* shed light on the genome size evolution in flatworms 309

General discussion 339

1. On the use of molecular markers to identify species of the Corso-Sardinian clade 339

2. Hurdles of species delimitation in *Dugesia* 343

3. Novel genomic resources for the planarian research community 347

4. Karyological plasticity and its role in speciation in planarians 349

4.1. Ploidies in the Corso-Sardinian clade 349

4.2. On the aneuploidy and intraindividual karyotype variability 349

4.3. Remarks on ploidy and reproductive strategies: a key to success? 351

5. Chromosomal rearrangements and their role in *D. hepta*'s speciation 353

6. A new perspective on the putative hybrids 355

7. Accounts on the *leporii* × *benazzii* hybrids: a case of hybrid speciation? 357

Conclusions 361

References 367

Appendix: Other Publications 397

Appendix I

Amphibioplanidae: a new branch and family on the phylogenetic tree of the triclad flatworms (Platyhelminthes: Tricladida), represented by a species from Sardinian caves with a remarkable lifestyle 401

Appendix II

Island-specific evolution of a sex-primed autosome in a sexual planarian 439

Appendix III

Building phylogenies from transcriptomic data 461

Agraïments
Acknowledgements

Tota història té un inici. Si els motius que em van dur a realitzar aquesta tesi doctoral es poguessin remuntar, com si d'establir un llinatge d'inquietuds i decisions tractés la cosa, estic convençut que tots ells coalescerien en un mateix punt. Concretament, quatre anys abans que decidís embarcar-me en aquesta aventura.

Corria l'any 2015. Tres bons amics i jo havíem de triar un article científic sobre el qual fer un treball i presentació oral per a l'assignatura de Biologia d'Invertebrats Marins, del Grau de Biologia. L'atzar, però sobretot la dislèxia, van voler que escollíssim un article sobre espècies críptiques de poliquets marins tot pensant-nos que veuríem criatures vistoses, acolorides i exòtiques. No em cauen pas els anells si reconec que, en gran manera, l'instigador d'aquell error vaig ser jo, perquè d'alguna forma al meu cap els conceptes d'espècies críptiques i espècies aposemàtiques eren cosins germans. Així doncs, vaig insistir a fer el treball sobre aquell article. No cal dir que les expectatives van acabar distant molt de la realitat amb la qual vàrem topar, però com a mínim cripsi i aposematisme no es van tornar a confondre mai més. No obstant això, la tragicomèdia ja les té aquestes coses, i d'una situació subòptima en van sortir dues coses positives que, sense jo saber-ho, acabarien duent-me al món de les planàries d'aigua dolça. En primer lloc, aquell treball va despertar en mi l'interès per les filogènies moleculars i la sistemàtica. I, en segon lloc, l'aleshores professora de la Facultat de Biologia Dra. Ana Riesgo va ser qui va fer-nos de tutora. Fou ella qui em va encoratjar, amb arguments de pes com "*joder, pero si eso es muy chulo!*", a presentar-me a l'oferta de treball final de grau d'una tal Dra. Riutort malgrat les meves

inseguretats al respecte per no saber-ne prou ni de filogènies ni de planàries. Em va donar confiança quan la necessitava i gràcies a les seves paraules (no necessàriament les que he citat) he pogut gaudir d'uns anys meravellosos de ciència que m'han permès créixer com a persona i científic. Així doncs, el meu sincer agraïment.

Ara, però, que percebo amb més claredat com els llums d'aquesta etapa s'abaixen, no puc sinó pensar en tota aquella gent que d'una forma o altra m'han acompanyat durant la tesi. Un treball com aquest, per força, esdevé un esforç col·lectiu i, tanmateix, voldria agrair el bon humor, la generositat, l'aportació de coneixements i la paciència a tots i cadascun dels actors i figurants que han tingut un paper important en els darrers cinc anys de la meva vida i el decurs d'aquest doctorat. Gràcies a vosaltres les misèries han sigut més suportables i la joia més celebrada.

Voldria començar per mostrar el meu sentit agraïment a la persona que m'ha donat totes les oportunitats científiques de les quals he pogut gaudir durant tot aquest temps, na Marta Riutort. Marta, moltes gràcies per haver confiat en mi durant tots aquests anys. Gràcies per haver sabut animar-me quan ho necessitava i donar-me confiança en mi mateix i en la recerca que hem dut a terme quan la situació ho requeria. Crec que no et descobreixo res nou si et dic que ets una gran científica i, sobretot, una bona persona. Tens un tarannà i una manera de veure la ciència i la vida que contagia l'optimisme, el bon humor i el companyonatge. Estic convençut que aquesta visió la compartim tots aquells científics i científiques als que has format, motiu pel qual tots t'apreciem i veiem en tu tot un referent científic, laboral i humà. El cert és que treballar amb tu ha sigut una joia i em sento orgullós de formar part del teu llegat científic.

Així mateix, voldria agrair de tot cor a na Gema Blasco, la tècnic del nostre laboratori, per tots els anys que hi hem compartit junts. I en son molts! En tu sempre he trobat una companya excel·lent i una font inesgotable de xafarderies, riures, confidències i bon humor. Sempre has fet més lleugers els dies més pesats. Admiro el teu optimisme i el sentit de l'humor que gastes! Moltes gràcies, Gema, per haver cuidat tant i tan bé de les planàries i de mi.

D'altra banda, i sense moure'ns de l'àmbit del laboratori Riutort, la meva estabilitat mental i felicitat durant el doctorat no s'entendrien sense les amistats i els companys que hi he trobat pel camí. Amics, gràcies per tots els bons moments compartits. Laia, moltes gràcies per tot el que em vas ensenyar quan vaig arribar al laboratori. Gràcies per tots els consells, totes les bromes i totes les confidències compartides. Gràcies per alleugerir-me el pes del doctorat i per haver-me contagiat l'amor per les planàries i per aixecar còdols de riu. Guardo un molt bon record de totes les nostres aventures per Còrsega i Sardenya, així com de les deshores al laboratori xerrant de la vida! Lisy, t'agreixo moltíssim tot el que m'has ensenyat durant el doctorat. Sempre he admirat molt la teva resiliència i fortalesa. Ets tot un exemple de com afrontar la vida, malgrat que aquesta no sempre et posi les coses fàcils. Gràcies pel teu entusiasme per la ciència i per sempre confiar en mi i en la meva

recerca, fins i tot quan m'hi costava creure. Raquel, la teva energia i entusiasme han sigut tan contagiosos i necessaris! Sempre m'has escoltat quan ho he necessitat o animat quan el dia semblava un xic més gris. Tots hem passat per les penes d'un doctorat, però amb tu, com a mínim, es feien més suportables. Joan, tu encarnes tot allò que en el meu imaginari representa un bon científic. Ets reposat, afable i bona persona, alhora que treballador, professional, entregat i amb esperit crític. Ha estat tot un goig seure braç a braç al laboratori i poder aprendre de tu en converses disteses sobre ciència i les nostres recerques. Malgrat que no sempre m'ha resultat fàcil emmirallar-me en tu, sempre he mirat de tenir present el teu tarannà i manera de fer. Tots vosaltres, que heu anat per davant meu, heu sigut mentors, amics i una font de confiança en allò que feia, i les paraules se'm queden curtes per expressar el respecte i l'estima que us tinc.

De la mateixa manera, voldria agrair a la resta de companys de laboratori que m'han ajudat en incomptables ocasions i que m'han fet sentir com a casa en tot moment. Ona, Edu, Paula, Cristian, Ricard, Marina, us agraeixo sincerament totes les bones estones que m'heu fet passar amb vosaltres. Ja fos parlant de Magic l'Aplec, de política, de *salsejos* i *varietés*, del Barça, fent broma, o bé, senzillament, explicant-nos la vida, els neguits i allò que ens fa feliços. Moltes gràcies. En especial, però, voldria mostrar el meu agraïment a en Guillem i en Nacho. Guillem, veig en tu qualitats que envejo i una passió vers la ciència que fa goig. Es nota d'una hora lluny que malgrat que a vegades aquesta ciència et pugui atabalar, la gaudeixes, i això s'encomana. A tu et dec algunes de les discussions més interessants sobre les meves planàries i com afrontar segons quines problemàtiques associades. No tinc cap mena de dubte que te'n sortiràs la mar de bé en el teu doctorat, i ho voldré veure! Gràcies per tots els cops que m'has ajudat desinteressadament. Nacho, mi presidente, només puc dir-te que lamento que no hàgim coincidit abans en el meu doctorat. En tu he trobat un bon amic i un company que sap escoltar i compartir amb entusiasme aquelles coses que et fan gaudir. Has fet que la solitud que acompanya els darrers anys d'un doctorat fos lleugera i has fet que les hores al laboratori fossin més disteses a cop de mems, ossos rentadors, cartes Magic, Tuna Tuna o reportatges sobre les nits electorals a la FIT. Et desitjo tota la sort i encerts que puguis necessitar en la teva tesi i espero que l'amistat la puguem celebrar molts anys. Mantingues sempre aquest bon esperit que és tan teu. Gràcies i, ara i sempre, visca la Fitoplastika de Jaulín!

Bona part d'aquesta tesi no s'entendria sense l'ajut que m'ha brindat la bona gent del grup de recerca de *Evolutionary Genomics & Bioinformatics*. Començant pel patriarca, vull expressar la meva gratitud i agraïment a Julio Rozas. Julio, és ben cert que sense Hèrcules m'hagués estat del tot impossible acabar la tesi, però el meu agraïment va més enllà del que és pragmàtic o material. Tot i no formar part del teu grup, m'has acollit de la millor manera possible i també ajudat quan ho he necessitat. M'alegra poder haver anat de bars amb tu i haver gaudit de la teva hospitalitat a Blanes, així com dels teus consells en lo científic. Tu i la Marta sou dues rara avis i els vostres deixebles us respectem molt. A l'Álex i la Sara els vull agrair el bon humor i les inacabables anècdotes que sempre surten amb una bona copa de cervesa i un bon tiberi al davant. Sou molt divertits, però sobretot, en sabeu molt de lo

vostre i parlar amb vosaltres és sempre tan enriquidor! Així mateix, vull agrair a na Paula Escuer tots els consells i tots els cops que ens hem fet companyia i animat l'un a l'altre durant la tesi, que no han estat pocs. Els cafès amb tu sempre eren revitalitzadors i ben entretinguts, i que bé que ens ho vàrem passar gravant els vídeos de les tesis de la Raquel i la Lisy! Tanmateix, hi ha una persona en concret a qui dec no tan sols una gran amistat, sinó també haver pogut acabar en condicions aquesta tesi. Vadim Pisarenco, t'agraeixo de tot cor tota la generositat que has tingut vers a mi. Ets una magnífica persona que no entén la vida sense preocupar-se de les persones a les quals aprecia, i ho mostres cada dia. Sé que per a tu el doctorat, sovint, no ha estat senzill, i malgrat això sempre has tingut bons gestos per la gent del teu voltant. Mai tens un no per resposta quan es tracta d'ajudar a algú i malgrat que això pugui atraure a trepes, tu sempre tries confiar en la gent. Vadim, ets bondadós, treballador, sacrificat, bona persona i un amic excel·lent. Hem xalat molt fent cafès, parlant de Hans Zimmer, One Piece, Batman, la Juna, en Loki i la Sansa, visitant Canet de Mar o tastant les magnífiques postres que prepares amb la teva mare. Collons, què n'és de bo el шапка гугуцэ! Espero i desitjo que la vida et somrigui més sovint, i que la Marta i jo en puguem formar part per molt més temps. Gràcies per tot, Vadim. Alhora, vull agrair a n'Adrià, l'Hekate, la Sílvia i en Carlos haver fet més distesos aquests darrers mesos d'esforços i patiments. Sou collonuts i espero que les vostres respectives vides siguin plenes i felices.

Now, I believe this is a suitable paragraph to switch languages, as I am indeed indebted to several great, and international, collaborators. I would like to start off by expressing my gratitude to Giacinta Stocchino. My journeys to Sardinia would not have been the same without your guidance and great hosting. I still sometimes crave some of that wondrous pasta and lamb you and signora Pala prepared back in her apartment! I treasure my memories sampling with you and her. Most especially, when it came the time to take a break and share a loaf of bread by the river, while enjoying prosciutto, delicious slices of cheese, and oranges all while chatting about planarians, the weather or Sardinia in an often-trilingual conversation. I thoroughly enjoyed those days. Grazie di tutto, Giacinta! Likewise, I would like to thank Professor Maria Pala for her hospitality and contagious enthusiasm for freshwater planarians. I do not recall exactly how many planarians we sampled during my first visit to Sardinia, but I am convinced that signora Pala alone collected more than half of the specimens in that campaign, beating us all by a country mile! I will always cherish those dinners at Bakkus, carelessly conversating with you (or at least trying to) in Italian. Il mio soggiorno in Sardegna è stato davvero piacevole! Grazie di tutto e di quel vasetto di mermellata fatta in casa che così gentilmente mi hai regalato. I am very grateful to Ronald Sluys as well. I sincerely appreciate his thorough revisions and comments on each of the articles I had the pleasure to work on with him. Ronald, you may not know it, but your feedback has greatly influenced my way of conceiving a paper and communicating its key points. Hartelijk dank voor uw geduld en geweldige inzichten! (hopefully, I got that right).

Much in the same way, I would like to express my gratitude to Longhua Guo and Leonid Kruglyak, who both made possible my stay in Los Angeles, and who took great care

of me. Most especially, I am thankful to you, Long, for inviting me over Thanksgiving and sharing with me a lovely evening with your family and friends. Likewise, I would like to thank Tzitziki Lemus Vergara, Heriberto Marquez and the rest of the Kruglyak Lab for kindly taking me in.

Si hi ha res que sigui important en aquesta vida, és saber-se triar bé els amics. Durant els meus anys a la universitat he pogut tenir el privilegi de conèixer algunes de les millors persones que conec, i trobo que és obligat que hi tinguin cabuda en aquests agraïments. M'agradaria començar pel flamant doctor Cuesta Porta, a qui li tinc una estima immensa. Víctor, fa quasi mitja vida que et conec i et considero un dels meus millors amics. T'esforces i t'entregues en tot allò que fas, sobretot en les amistats. Sempre aportes alegria i un riure que es contagia amb molta facilitat i t'ho agraeixo de tot cor, doncs m'has fet tots els dies una mica millors a base de cafès, mems, xalant de One Piece o simplement explicant-nos la vida. T'agraeixo tots els anys que hem passat junts, i en els quals m'has ajudat, consolat, animat i entretingut un munt de vegades. T'admiro molt, Victor, com a científic i com a persona, i desitjo que pugui seguir-ho fent al teu costat per molts anys! Voldria agrair també a la seva eminència, el doctor Galià Camps, tots els anys d'amistat compartida. Carles, són molts anys, moltes hores de Dota, LoL i WoW (de fet, les enyoro), però també un munt de calçotades, barbacoes, excursions, cartes Magic, cafès i birres. Envejo la teva dedicació i passió per la ciència, el com t'has fet a tu mateix en un món de genomes i ordinadors, i el bé que te n'has sortit. Gràcies per haver compartit aquests anys de tesi amb mi. A l'excel·lentíssim doctor Benaiges li dec incomptables riallades i bones estones al seu costat. Robert, tens una qualitat innata que fa que la gent t'apreciï. Caus bé de forma natural, perquè ets de les persones més autèntiques i altruistes que conec. I, alhora, tens aquesta capacitat d'enutjar-te com una mona, cardar quatre renecs, i fer-me esclatar del riure. Ets una font inesgotable d'anècdotes, una persona formada i amb consciència de classe, un amic excel·lent i, per sobre de tot, una bona persona i un millor sotanero. Gràcies per totes les estones compartides al bar xerrant de les coses que ens agraden, per ser tan generós amb tots nosaltres, i per fer els partits del Barça més divertits de veure. I que duri! Així mateix, vull mostrar el meu profund agraïment a n'Oriol Castells i a na Mireia Pàmies. M'és impossible concebre l'un sense l'altre, perquè tots dos m'heu cuidat tant i tan bé durant tots aquests anys d'amistat compartida. Tots sou generosos, bondadosos i, en definitiva, molt bones persones. Passar temps amb vosaltres sempre m'alleugereix els mals. El cert és que, malgrat que no us ho digui sovint, us estimo.

Altrament, vull mostrar el meu sentit agraïment a en Pau Bonachela, l'Álvaro Míguez, l'Ariadna Vidal, l'Alba Enguídanos, la Paula Cabrera, en Sergi Albert, la Marta Company, la Núria Profitós, l'Adrià Chorro, l'Elena Prats, en Roger i Marçal Olivé i a tanta altra gent que m'ha fet sentir acollit en diferents moments d'aquesta tesi i els anys que la varen precedir. Perdoneu-me que no m'esplaï a gust, però no seria el més indicat tenir uns agraïments més llargs que la mateixa Introducció de la tesi. Us agreixo de veres tots i cadascun dels moments compartits i el bé que m'han fet.

Tanmateix, hi ha una persona en particular a qui dec moltíssim. Els darrers anys de la meva vida no s'entendrien sense la meva companya de vida, amiga i confident, na Marta Olivé. Marta, t'estic eternament agraït per tot el que has compartit amb mi. Em guareixes dels mals, em dones amor i joia, em recolzes sempre i junts les hem viscudes de tots colors. La vida amb tu és un regal. Admiro la teva bondat, generositat i bon humor, i, alhora, la perseverança i resiliència que tens i que tantes vegades t'ha costat de veure. Em fas millor i t'ho agraeixo de tot cor. Gràcies per construir una vida junts, t'estimo.

També voldria mostrar el meu sentit agraïment als sogres, la Pilar i l'Albert, que m'han fet sentir sempre com un més de la família. A tots dos, gràcies per cuidar-me tant i tan bé.

Per anar acabant, vull agrair al meu germà, en David, per tota una vida compartida al seu costat. Sempre ens tindrem l'un a l'altre i això em conforta.

I, finalment, dedicar unes últimes paraules als meus pares, en Joaquim i la Cristina, a qui els dec ser la persona que soc. Gràcies per l'educació i els valors que m'heu inculcat. Gràcies per tots els anys d'amor incondicional i suport que m'heu donat. Gràcies per tot el que heu hagut de sacrificar en algun moment o altre per a fer-nos feliços a en David i a mi. Sou la meva vida i la meva llar.

Papa, mama, ho he fet tot el bé que he sabut. Espero que us faci sentir orgullosos. Us estimo molt.

Abstract

Planarian flatworms (Platyhelminthes, Tricladida), commonly referred to as triclads, constitute a diverse taxonomic group encompassing a wealth of species exhibiting a broad spectrum of phenotypic diversity. Triclads are present in all biogeographical realms and have colonized terrestrial, marine, and freshwater environments distributed all across the globe, bearing testament to their evolutionary success.

Free-living freshwater planarians of the family Dugesiidae encompass some of the most well studied taxa in the group and stand out due to their outstanding regenerative capabilities and diversity of reproductive strategies. Among the 12 genera presently recognized in the family, *Dugesia* is possibly one of the most emblematic. This taxon constitutes the most speciose genus of Dugesiidae, probably even of Tricladida, and is a well-defined taxonomic group that has silently colonized freshwater environments spread across Africa, Europe, the Middle East and Australasia, in spite of the low vagility of its species. In recent years, numerous studies have placed European *Dugesia* in the limelight as rising model systems for phylogeography and, ultimately, evolutionary biology. Understanding the processes by which populations differentiate and evolve into new species is fundamental to the field of evolutionary biology and holds significant implications for the underpinning of novel conceptual frameworks and the generation of both local and global patterns of species-level biodiversity.

This thesis aims to contribute to the pursual of this knowledge by gaining insights into the processes that have shaped the genetic diversification, speciation and evolutionary

history of a seemingly small taxonomic group of the longstanding genus *Dugesia*, namely: the Corso-Sardinian clade. This particular group, which only comprises endemics from Corsica and Sardinia, caught my interest for several reasons: (i) no phylogenetic or phylogeographic studies dedicated to its taxa were ever carried out; (ii) its diversity had not been evaluated under the light of integrative taxonomy approaches; (iii) this group harbors the only species with a haploid chromosome complement of 7 chromosomes in the entire Western Palearctic region, posing a possible case of speciation by chromosomal rearrangement; (iv) the only putative case of interspecific hybridization in Dugesiidae (and Tricladida) was reported in this group.

To unravel the evolutionary history and processes that have shaped the diversification of this group, I employed an integrative approach, incorporating morphological, karyological and molecular data. This approach has enabled the re-evaluation of the taxonomic status of several *Dugesia* populations, resulting in the description of 3 new species for the group and the uncovering of multiple species candidates whose status and pertinent taxonomic decisions will be addressed in a forthcoming paper. Furthermore, divergence time estimation analyses indicated that most species within this group diverged during the Pleistocene, probably due to population fragmentation, contraction, and extinction events caused by the glacial phases. While it cannot be claimed that the phylogenetic relationships within the Corso-Sardinian clade have been fully resolved, this thesis has significantly contributed to elucidating its hidden diversity and phylogeography, thus laying the foundations for future research on this group.

Most notably, by using phylogenomic markers and combining phylogenetic, genetic clustering and introgression analyses, this thesis reports and characterizes the first case of interspecific hybridization in Dugesiidae, and likely Tricladida. Furthermore, I examine the potential origin of the hybrids, their status as a potential hybrid species, and their seemingly mixoploid chromosome set. In addition, I also debate on the established concepts of hybrid speciation and assess their suitability for the current case presented within this thesis.

Finally, this thesis has contributed to the planarian research community by presenting the first chromosome-scale genome assemblies of *Dugesia*. The results obtained from the analysis of these genomes have elucidated the origin of *Dugesia* hepta's unique karyotype, while shedding light on the evolution of genome size in the genus. These results highlighted a differential expansion of transposable elements (TEs) in *Dugesia* involving large DNA TEs as the likely cause for the increment of genome size in this genus with respect to *Schmidtea*, while discarding a whole-genome duplication event as the underlying cause. Most importantly, these new genome assemblies will benefit the planarian research community and contribute to improving the understanding of the structure, organization, and genome evolution in planarians.

Sinopsi

Les planàries (Platyhelminthes, Tricladida), sovint també anomenades tricl·àdides, constitueixen un grup taxonòmic divers que abasta una gran quantitat d'espècies que presenten un ampli espectre de diversitat fenotípica. Els tricl·àdides són presents a tots els reialmes biogeogràfics i han estat capaços de colonitzar entorns terrestres, marins i dolç-aqüícoles distribuïts per tot el món, fet pel qual el seu èxit evolutiu queda palès.

Les planàries d'aigua dolça i vida lliure de la família Dugesiidae comprenen alguns dels tàxons més ben estudiats d'aquest grup i destaquen gràcies a la seva excepcional capacitat regenerativa i la diversitat d'estratègies reproductives. D'entre els 12 gèneres actualment acceptats a la família, *Dugesia* és possiblement un dels més emblemàtics. Aquest tàxon constitueix el gènere més especiat de Dugesiidae, probablement fins i tot de Tricladida, i és un grup taxonòmic ben definit el qual ha colonitzat silenciosament ambients d'aigua dolça distribuïts per Àfrica, Europa, l'Orient Mitjà i Australàsia, malgrat la reduïda vagilitat de les seves espècies. En els darrers anys, nombrosos estudis han situat a les *Dugesia* europees en el punt de mira com a sistemes model per a la filogeografia i, en definitiva, la biologia evolutiva. Entendre els processos pels quals les poblacions es diferencien i evolucionen cap a noves espècies és fonamental per al camp de la biologia evolutiva i té implicacions significatives tant per a la sustentació de nous marcs conceptuals com per a comprendre la generació de patrons de biodiversitat a nivell d'espècie a escales locals o globals.

Aquesta tesi té com a objectiu contribuir a la recerca en aquest camp mitjançant l'obtenció de nous coneixements sobre els processos que han modelat la diversificació genèti-

ca, l'especiació i la història evolutiva d'un grup taxonòmic aparentment petit del gènere *Dugesia,* en concret: el clade corso-sard. Aquest grup en concret, el qual només inclou espècies endèmiques de les illes de Còrsega i Sardenya, va atraure el meu interès per diversos motius: (i) mai s'havien fet estudis filogenètics o filogeogràfics centrats en els seus tàxons; (ii) la seva diversitat no s'havia avaluat mai des d'una perspectiva de taxonomia integradora; (iii) aquest grup alberga l'única espècie en tota la regió del Paleàrtic Occidental que posseeix un complement cromosòmic haploide de 7 cromosomes, plantejant un possible cas d'especiació per reorganització cromosòmica; (iv) l'únic cas putatiu d'hibridació interespecífica en Dugesiidae (i Tricladida) es va documentar en aquest grup.

Per tal de desentranyar la història i els processos evolutius que han donat forma a la diversificació d'aquest grup, he fet ús d'una metodologia integradora, incorporant dades morfològiques, cariològiques i moleculars. Aquest mètode ha permès la reavaluació de l'estat taxonòmic de diverses poblacions de *Dugesia*, donant lloc a la descripció de 3 noves espècies dins del grup i al descobriment de múltiples espècies candidates. Tanmateix, l'estat taxonòmic d'aquestes darreres i les pertinents decisions taxonòmiques que en derivin s'abordaran en un pròxim article. A més, les anàlisis del temps de divergència indiquen que la majoria de les espècies d'aquest grup van divergir durant el Pleistocè, probablement degut a esdeveniments de fragmentació, contracció i extinció poblacionals causats per les fases glacials que van tenir lloc durant aquest període. Malgrat que no es pugui afirmar que les relacions filogenètiques dins del clade corso-sard hagin quedat completament resoltes, aquesta tesi ha contribuït significativament a aclarir la seva diversitat oculta i filogeografia, establint així les bases per a futures investigacions sobre aquest grup.

Cal remarcar que l'ús de marcadors filogenòmics en combinació amb anàlisis filogenètiques, de clusterització genètica i d'introgressió han permès documentar i caracteritzar en aquesta tesi el primer cas d'hibridació interespecífica en Dugesiidae, i probablement en Tricladida. A més, hi examino el potencial origen dels híbrids, el seu estatus com a potencial espècie híbrida i l'origen del seu set cromosòmic aparentment mixoploide. Alhora, també discuteixo sobre els conceptes d'especiació per hibridació establerts i n'avaluo la seva idoneïtat per al cas presentat dins d'aquesta tesi.

Finalment, aquesta tesi ha contribuït a la comunitat de recerca en planàries presentant els primers assemblatges genòmics a escala cromosòmica de *Dugesia*. Els resultats obtinguts de l'anàlisi d'aquests genomes han permès aclarir l'origen del cariotip únic de *Dugesia hepta*, alhora que també l'evolució de la mida del genoma en el gènere. Aquests resultats han evidenciat una expansió diferencial d'elements transposables (ET) a *Dugesia* que implicarien ET d'ADN d'una grandària considerable com a causa probable de l'augment de mida del genoma en aquest gènere respecte a *Schmidtea*. Alhora, les anàlisis descarten que un esdeveniment de duplicació del genoma sencer sigui la causa subjacent de les diferències en mida del genoma. El més important, però, és que aquests nous assemblatges beneficiaran a la comunitat de recerca en planàries i contribuiran a millorar la comprensió de l'estructura, l'organització i l'evolució del genoma en les planàries.

Introduction

1. Evolutionary Biology

1.1. A historical perspective

"*After having been twice driven back by heavy south-western gales, Her Majesty's ship «Beagle», a ten-gun brig, under the command of Captain Fitz Roy, R. N., sailed from Devonport on the 27th of December, 1831*" (Darwin, 1839, p. 1). Just like that, in a seemingly mundane manner, Darwin recounted the beginning of a journey that would become instrumental for the publication of his celebrated book *On the Origin of Species* (Darwin, 1859) and for underpinning the foundations of evolutionary biology. Evidently, much has been added to the evolutionary theory since Darwin's original work and the concept we have today is not the same, not any more than today's physics is "Newtonian" in nature (Pigliucci, 2009; Pigliucci & Müller, 2010). And yet, reconstructing the Tree of Life is still at the core of evolutionary biology almost two centuries later.

Influenced by Dobzhansky's evolutionary genetics framework, systematists, paleontologists and geneticists realized that there was no fundamental conflict between their respective discipline views, and their practices could be bind into an evolutionary network grounded in genetics and selection theory (see Smocovitis, 1992). Following pressures to unify the "fractured" biological sciences under a cooperative organization, the *Society for the Study of Evolution* was founded in 1946 as per Ernst Mayr's initiative, with the objective to promote the use of new experimental practices rather than just descriptive or taxonomic practices of evolution.

Thus, evolutionary biology was constituted as a discipline, and thrived thanks to the development of comparative methods that were based on explicit evolutionary and statistical models, allowing to perform hypothesis tests on evolution (Penny, 1992). Most notably, Felsenstein's contributions (1985) were instrumental to revolutionize the field of evolutionary biology, as it was the first to develop an explicit method to address the problem of statistical dependence of comparative data due to the hierarchical ancestor-descendant structure of phylogenies (Carvalho et al., 2005).

Concurrently, the continuous development of sequencing technologies and novel theoretical frameworks (e.g., evo-devo, genomics, and epigenetics) has elevated evolutionary biology to a new era of bioinformatics and high-throughput sequencing.

1.2. Phylogenetics

Phylogenies have been of great importance to tackle the reconstruction of the Tree of Life. They represent the evolutionary relationships between taxa, following the phylogenetic hypothesis that all organisms share a common ancestor (Huelsenbeck and Rannala, 1997). A phylogenetic tree is composed of nodes that are connected by branches. If a tree represents the relationships among a group of species, then nodes represent the speciation events.

Phylogenies can be inferred based on phenotypic and/or molecular data, and the use of phylogenetic comparative methods in cross-species studies has achieved a paradigmatic status. While the former type of data was thoroughly used during the first half of the 20th century, molecular data only started to be available mainly after the 1980s. However, it led to the establishment of a new discipline, the molecular phylogenetics, whose objective is the correct reconstruction of genealogical relationships between biological entities, the estimation of the time of divergence between them, and the sequence of events throughout an evolutionary lineage (Graur and Li, 2000).

Initially, phylogenetic trees were almost exclusively used for inferring the evolutionary relationships among species. However, nowadays their use extends beyond that, and phylogenies have become customary for addressing various biological questions such as the relationships of paralogues in gene families, infer the evolutionary history of populations, model epidemiological dynamics of pathogens, describe the genealogical relationship of somatic cells during differentiation and cancer development, and even describing the evolution of language (Yang and Rannala, 2012; and references therein).

1.3. Phylogenetic markers

Most of what we know about the reconstruction of the Tree of Life stems from classifications grounded on morphological data, yet paradoxically its use for phylogeny inference has dwindled significantly over the last decades in favor of molecular data.

Such paradigmatic shift is owed to differences in the number of potentially unambiguous characters available, their ease and speed of discovery, and their suitability for analysis using transformational models (Scotland et al., 2003). Although morphological and molecular data are similar in that criteria of topological correspondence used to assess primary homology, morphological character definitions tend to be problematic and prone to subjectivity, as they depend on the expertise and perception of the researcher (Scotland et al., 2003; and references therein). However, morphology remains a relevant and often necessary source of evidence for testing molecular clades in integrative systematics (Dayrat, 2005), and still is the primary means for time-scaling phylogenies through fossil phenotypes (Lee and Palci, 2015).

Much like morphology, molecular phylogenetics also relies on homology. Two genes, or molecular markers, will be homologous if they derived from a common ancestral gene. Yet, homology may come from two different processes. When duplication is the rooting cause, we refer to such markers as *paralogs*, whereas when speciation is the cause, we refer to them as *orthologs*. Orthologous genes reflect the relationships among species. Consequently, phylogenetics rely in a great part on well-curated data sets comprising this type of genes. However, more recently the potential of paralogs for reconstructing true phylogenetic relationships has also been explored (Hellmuth et al., 2015). During the early years of molecular phylogenetics, most studies relied on a small set of universal orthologous genes in which ribosomal RNAs (Woese and Fox, 1977) and mitochondrial genes (Zardoya and Meyer, 1996) were commonly used provided their differential rates of evolution. This property allowed researchers to resolve biological questions on a phylogenetic scale of interest. Furthermore, these markers were comparatively easy to amplify with universal primers, their orthology was well established, and databases of these markers grew rapidly, most notably thanks to the introduction of the polymerase chain reaction (PCR) in the 1980s (Saiki et al., 1988), which revolutionized the field of molecular biology. However, it would not take too long before researchers started to realize that, despite the remarkable improvements in phylogenetic resolution, some of the most contentious branches (e.g., rapid radiations, pervasiveness of horizontal gene transfer, hybridization, incomplete lineage sorting) of the Tree of Life could not be easily resolved with those genes (Edwards et al., 1991; Farias et al., 2001).

2. Genomic sequencing technologies

2.1. Brief history of genomic sequencing

2.1.1. First-generation DNA sequencing

Major improvements in biology have often gone hand in hand with equivalently relevant breakthroughs in sequencing technologies. Much have changed since Friedrich Mietscher successfully performed the very first DNA isolation in 1869, and methodological developments of DNA sequencing can be summarized in three phases, popularly acknowledged as generations (reviewed in Heather and Chain, 2016).

The three-dimensional structure of DNA was solved in 1953 thanks to the crystallographic data produced by Rosalind Franklin and Maurice Wilkins (Watson and Crick, 1953; Zallen, 2003), which contributed to a conceptual framework for both DNA replication and encoding proteins in nucleic acids. However, it would not be until decades later that the first techniques for *reading* DNA sequences appeared, that is, the first-generation DNA sequencing. The founding DNA sequencing methods were the Sanger dideoxy synthesis (Sanger and Coulson, 1975; Sanger et al., 1977) and the Maxam-Gilbert chemical cleavage (Maxam and Gilbert, 1977). While both techniques worked on the same principle (that of producing all possible incremental lengths sequences and labelling the ultimate nucleotide), Sanger's sequencing would become the most commonly used technology for decades due to its accuracy, robustness and ease of use (Heather and Chain, 2016). Furthermore, the development of the PCR (Saiki et al., 1988) and other improvements, such as the fluorometric based detection through capillary based electrophoresis or the devel-

opment of automated DNA sequencing machines (Slatko et al., 2018), helped establish the Sanger sequencing-by-synthesis (SBS) method as the sequencing standard, still to this day, for most evolutionary studies.

2.1.2. Second-generation DNA sequencing

The onset of the second-generation DNA sequencing (also known as next-generation sequencing, NGS) coincide with the appearance of a novel sequencing approach: the pyrosequencing. This approach was based on a luminescent method for measuring the production of pyrophosphate, a byproduct of nucleotide incorporation (Nyrén and Lundin, 1985; Ronaghi et al., 1996). This technique, pioneered by Pål Nyrén and collaborators, had the advantage of not requiring modified nucleotides among other beneficial features (Nyrén, 1987). Before being discontinued in 2013, pyrosequencing was licensed to 454 Life Sciences (later purchased by Roche), whose sequencing machines became a paradigm shift which allowed the mass parallelization of sequencing reactions, achieving high throughput sequencing of relatively long reads (600 to 800 nucleotides), which facilitated genome assembling (Slatko et al., 2018).

Concurrently, other sequencing techniques sprung up following the success of 454 sequencing, but the most relevant was that of Solexa (later acquired by Illumina). This technique was based on using *bridge amplification* (see https://www.youtube.com/watch?v=fCd6B5HRaZ8), which allows paired-end (PE) sequencing, meaning that both ends of a DNA fragment are amplified. Having PE data is useful when mapping reads to reference sequences, as the accuracy improves, particularly in repetitive regions, and improves our capability to detect insertion-deletion (indel) variants, spliced exons, gene fusions or DNA rearrangements, and differential transcripts (Freedman et al., 2020; Roach et al., 1995; Treangen and Salzberg, 2011). At present, Illumina sequencing still is the powerhouse of the second-generation sequencing and supports a variety of protocols including whole-genome sequencing (WGS), exome and targeted sequencing, metagenomics, RNA sequencing, methylome sequencing, single-cell profiling and chromatin analysis (details on the sequencing machines and their respective capabilities can be found in https://www.illumina.com/systems/sequencing-platforms.html).

2.1.3. Third-generation DNA sequencing

First-generation DNA sequencing brought about the foundations of reliable DNA sequencing. Second-generation sequencing's hallmark feature was its massive throughput at a modest cost and its remarkable variety of protocols. It is acknowledged that we are now at the dawning of a new generation, a third one, which stands out for a reason: the capability to sequence long (very long) DNA and RNA molecules (Schadt et al., 2010). At present, the most widely used third-generation technology is probably the single molecular real time (SMRT) sequencing platform from Pacific Biosciences (PacBio) (van Dijk

et al., 2014), which allows sequencing of non-amplified DNA, thus avoiding all associated biases and errors. Sequencing long reads (>10 kb) is a matter of hours with SMRT and since it occurs at the rate of the polymerase, it produces kinetic data that can be used to detect modified bases (Slatko et al., 2018; and references therein). Although this type of technology was characterized, at first, by reduced read accuracy, the optimization of circular consensus sequencing (CCS) has greatly improved the accuracy of SMRT to a level comparable to that of Illumina sequencing (PacBio HiFi; Wenger et al., 2019). Needless to say, SMRT has had a significant impact on assembling de novo genomes and, subsequently, on evolutionary biology studies.

However, perhaps the most anticipated area of development within this new generation is the promise of reliable nanopore sequencing. Nanopore-based sequencing was first proposed in the 1990s, when researchers demonstrated that single-stranded RNA and DNA molecules could be driven across a lipid bilayer through alpha hemolysin ion channels and *Mycobacterium smegmatis* porin A (MspA) by electrophoresis (Deamer et al., 2016), but only recently commercialization has been achieved by Oxford Nanopore Technologies (ONT). Sequencers use protein nanopores embedded in an electrically resistant polymer membrane through which current changes occur as each nucleotide passes through the pore, allowing a detector to interpret a given sequence. Although nanopore has been regarded as a promising yet prone to errors sequencing technology, in 2022, ONT developed a novel sequencing chemistry (i.e., R10), which promises accuracies on par with that of Illumina sequencing (i.e., 99.9%). Recent studies have shown the improved performance of the R10 chemistry over previous iterations (Sanderson et al., 2023; Wagner et al., 2023), which in turn has increased the popularity of nanopore-based sequencing as the method of choice for many studies dealing with pathogens or focused on clinical investigations (Bogaerts et al., 2024; and references therein). Although nanopore sequencing technology was developed for DNA, it is likely that other applications will emerge to provide information on other types of polymers (i.e., RNA and proteins) (Bonini et al., 2024; Dorey and Howorka, 2024).

2.2. Chromosome-scale genomes

The organization of the chromatin in the nucleus is nonrandom and extremely relevant to many biological functions including chromosome morphogenesis, gene regulation, transcription, and replication. Chromosome conformation capture methods, such as Hi-C (Lieberman-Aiden et al., 2009; Belton et al., 2012), combine chemical crosslinking of chromatin with fragmentation, DNA ligation, and high-throughput DNA sequencing to detect neighboring loci genome-wide, which will be sequenced in pairs (Eagen, 2018). In consequence, the number of read pairs between intrachromosomal regions is a slowly decreasing function of the genomic distance between them. Burton and collaborators (2013) developed a computational method that was capable of exploiting the signal of genomic proximity in Hi-C data to cluster and organize contigs (or scaffolds) on the same chromosome based on the amount of Hi-C links. Contigs on the same chromosome tend to have

more Hi-C links between them than to contigs on different chromosomes and, thus, it was possible to assemble the first individual chromosomes. At present, one of the most prevalent pipelines for chromosome scaffolding using Hi-C data is the HiRise pipeline (Putnam et al., 2016).

2.3. Markers in the NGS context

The advent of high throughput sequencing technologies certainly has fueled evolutionary studies, uncovering hundreds or thousands of polymorphic loci across the genome at a modest cost. Consequently, disciplines such as molecular phylogenetics have been able to spring forward into the omics era. One of the most significant developments within the last two decades has been, without a doubt, the emergence of reduced-representation genomic sequencing approaches, among which the restriction site-associated DNA sequencing (RADseq) (Miller et al., 2007).

This approach relies on restriction enzymes that target a subset of the genome, generating a high density of genome-wide markers. Although the term RADseq was originally used to describe one particular method (Baird et al., 2008), nowadays it has been adopted to describe a range of related techniques that rely on restriction enzymes (Andrews et al., 2016). Unlike other methods for generating genome-wide data, RADseq does not require any prior genomic knowledge for the taxa being studied (e.g., genome size, ploidy), which has boost its popularity among evolutionary studies focused on non-model systems that often lack a reference genome. RADseq has been primarily used for conducting population genomic analyses (Hohenlohe et al., 2010). However, it has also been successfully used for phylogeographic, interspecific phylogenomic, and species delimitation studies (Eaton and Ree, 2013; Newton et al., 2020; Rancilhac et al., 2019; Rubin et al., 2012; Stobie et al., 2018, 2019). The number of orthologous loci recovered using RADseq approaches is dependent on the divergence times among the lineages being studied (Rubin et al., 2012; Cariou et al., 2013; Harvey et al., 2016). Thus, RADseq approaches are often preferred for shallow phylogenomics, wherein the yield of polymorphic loci can be of thousands or tens of thousands of loci. However, under proper conditions for taxon sampling and sequencing coverage, RADseq approaches can be useful for large-scale phylogenomics as well (Eaton et al., 2017).

3. Phylogenetic inference methods

3.1. Brief accounts on distance-based methods

The progress made towards elucidating the evolutionary history of species and, ultimately, resolving the Tree of Life cannot be comprehended without the appearance and refinement of phylogenetic inference methods. Phylogeny reconstruction methods are either distance-based or character-based.

Distance-based methods were the first ones to be developed out of the interest of using molecular data for inferring phylogenies. The genetic distance between each pair of sequences is computed and integrated in a matrix of distances that will be used to infer the tree. Distance methods have the advantage of being computationally efficient, fast and able to handle large numbers of sequences (Roch, 2010), but their performance is often erroneous when rates of substitution vary greatly or gaps are present among examined lineages (Yang and Rannala, 2012; and references therein). However, in recent years, some distance-based methods for high-resolution phylogeny inference have been developed, claiming robustness in front of elevated sequence divergence (Bhardwaj et al., 2012; Xia, 2016). Interestingly, other developed methods based on distances, such as phylogenetic networks, have remained as useful approaches to detect patterns of reticulation (Huson and Bryant, 2006).

3.2. Character-based methods

Character-based methods simultaneously compare all sequences in the alignment, considering one site (character) at a time in the alignment to calculate a score for each tree considered. Maximum parsimony (MP) was originally developed for use in analyzing discrete morphological characters, but soon became a customary method for inferring phylogenies based on molecular data during the late 1970s. This method minimizes the number of changes on a phylogenetic tree by assigning character states to interior nodes on the tree. Thus, the character length becomes the minimum number of changes required for that site and the sum of all lengths becomes the score of the tree. Hence, the MP tree is the tree with the lowest tree score. Although MP is computationally efficient, it is highly prone to long-branch attraction artifacts, as it fails to correct for multiple substitutions at a same given site. Furthermore, it is statistically inconsistent under more complex, and often realistic, evolutionary models (Huelsenbeck, 1998; Yang, 1996).

In contrast with distance-based methods or MP, character-based methods often are probabilistic, and two approaches are widely used: maximum likelihood (ML) and Bayesian inference (BI). In this case, the tree scoring calculation is dependent on evolutionary models that take into account the substitution rates and nucleotide frequencies (Jukes and Cantor, 1969; Kimura, 1980; Rodríguez et al., 1990). Moreover, these models can also include extra parameters, such as invariant sites or a discrete gamma distribution, enabling the modeling of variation in the rate of evolution across sites (Yang, 1994). ML approaches rely on the likelihood function, which is defined as the probability of the data given the parameters of the model. From a statistical point of view, the tree's topology is the model while the length of its branches and the substitution parameters are the parameters of the model (Felsenstein, 1981). The maximum likelihood estimates (MLEs) of parameters are the parameter values that maximize the likelihood of observing these data. Consequently, maximum likelihood tree inference is equivalent to compare many statistical models, each with the same number of parameters, and search for the model (topology + branch lengths + substitution parameters) that maximizes the likelihood (best MLE) to establish it as the estimate of phylogeny. Thus, the maximum likelihood tree is not necessarily the 'true' species tree, but the one that explains best the evolution of the observed data (the sequences). BI differs from ML in that parameters in the model are considered to be random variables with statistical distributions rather than unknown fixed constants (Rannala and Yang, 1996). Prior distributions are assigned to the parameters and combined with the data to generate the posterior distribution. All inferences concerning the parameters are then based on the posterior distribution, from which Markov chain Monte Carlo algorithms (MCMC) will generate samples because these parameters cannot be directly calculated (Yang and Rannala, 2012).

There is an ongoing debate when analyzing multilocus data, particularly relevant in the context of the current phylogenomic era, over whether concatenation or coalescent methods are more appropriate (Gatesy and Springer, 2014; Liu et al., 2015). In the concat-

enation or supermatrix approach, sequences from multiple genes are concatenated into a supermatrix that is then used to infer a single tree using ML or BI approaches, thus assuming that the same phylogeny underlies all genes. However, if differences in evolutionary dynamics among genes are not accounted the phylogenetic inference can be inaccurate (Buckley et al., 2001; Lu et al., 2018). Thus, datasets can be partitioned to accommodate a fitting evolutionary model for each partition (Lanfear et al., 2017).

On the other hand, coalescent methods account for the heterogeneity in gene trees, which has been acknowledged as a significant issue in phylogenetics (Maddison, 1997). These methods treat gene trees as conditionally independent random variables given the species tree and parameters such as species divergence times and population sizes under the multispecies coalescent (MSC) model (Kingman 1982; Rannala and Yang 2003). MSC based methods can combine the information of gene trees to infer the species tree in two ways (Fig. 1). The summary or two-step methods use ML or BI to infer separate gene trees and then use summary statistics or pseudolikelihood functions to infer the species tree (e.g., Liu et al., 2010; Mirarab and Warnow, 2015). These methods are computationally efficient and can analyses thousands of gene trees in a short time. Concurrently, single-step methods estimate both gene trees and species tress at the same time according to multilocus sequence data, priors, and a MSC model (e.g., Bouckaert et al., 2019; Flouri et al., 2018). These methods can be more robust than two-step methods but are computationally intensive.

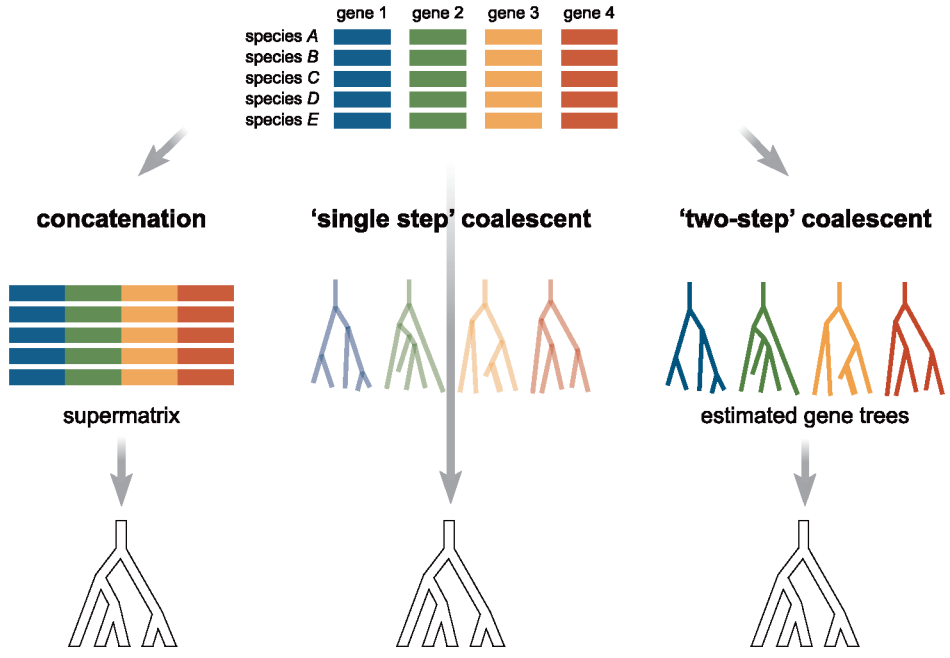


Fig. 1. Schematic of the most commonly used phylogenetic approaches for the analysis of phylogenomic datasets. On the left, the concatenation or supermatrix approach which relies on the concatenation of gene alignments and posterior inference of a phylogeny using ML or BI approaches. In the centre, the full coalescent approach in which both gene trees and the species trees are simultaneaously estimated according to multilocus sequence data, priors and the MSC. On the right, the two-step species tree methods or summary methods, in which gene trees are estimated using ML or BI approaches and then used as input data to estimate a species tree. Extracted from Liu et al. (2015).

4. Sources of gene tree discordance

While it holds true that the emergence of genome-scale data has provided unprecedented opportunities to resolve some contentious relationships in the Tree of Life (Jarvis et al., 2014; Misof et al., 2014; Rokas et al., 2003), it also has made clearer that gene tree heterogeneity is pervasive across the Tree of Life (Degnan and Rosenberg, 2006; Nichols, 2001). The conflict between gene trees and species tree may be due to multiple biological processes such as horizontal gene transfer (HGT), incomplete lineage sorting (ILS), hybridization and introgression, or gene duplication to name a few (Maddison, 1997).

HGT is a well-known and common cause of discordance, especially in the microbial world, so much so that it has been question whether a coherent Tree of Life exists for microbes (e.g., Doolittle and Bapteste, 2007). It concerns the transfer of genetic material between distinct evolutionary lineages and often plays an important role in microbial evolution, leading to the gain of novel functions, virulence, or resistance to antibiotics. In many prokaryotes, HGT has contributed to 10-20% of the genes (Koonin et al., 2001) However, although much less frequently, HGT also occurs in eukaryotes and might have been common early in eukaryotic evolution (Keeling and Palmer, 2008; Richardson and Palmer, 2007).

ILS (or deep coalescence) is considered one of the most widespread of the aforementioned biological causes of gene tree heterogeneity (Edwards, 2009). It describes a phenomenon in which interspecific orthologous genes coalesce into their pertinent ancestral copy before (or way before) the speciation event (Fig. 2). It is an intrinsic property of populations and as such, ILS may occur in any taxonomic group or gene. Consequently, genes may not reflect the species phylogeny and yield different tree topologies. ILS will be more prevalent when the rate of genetic drift is low (large populations) compared to the lengths of internodes in the species tree (Pamilo and Nei, 1988). Thus, it is more likely to occur when ancestral species had large populations that diverged in a short period of time (short internodes).

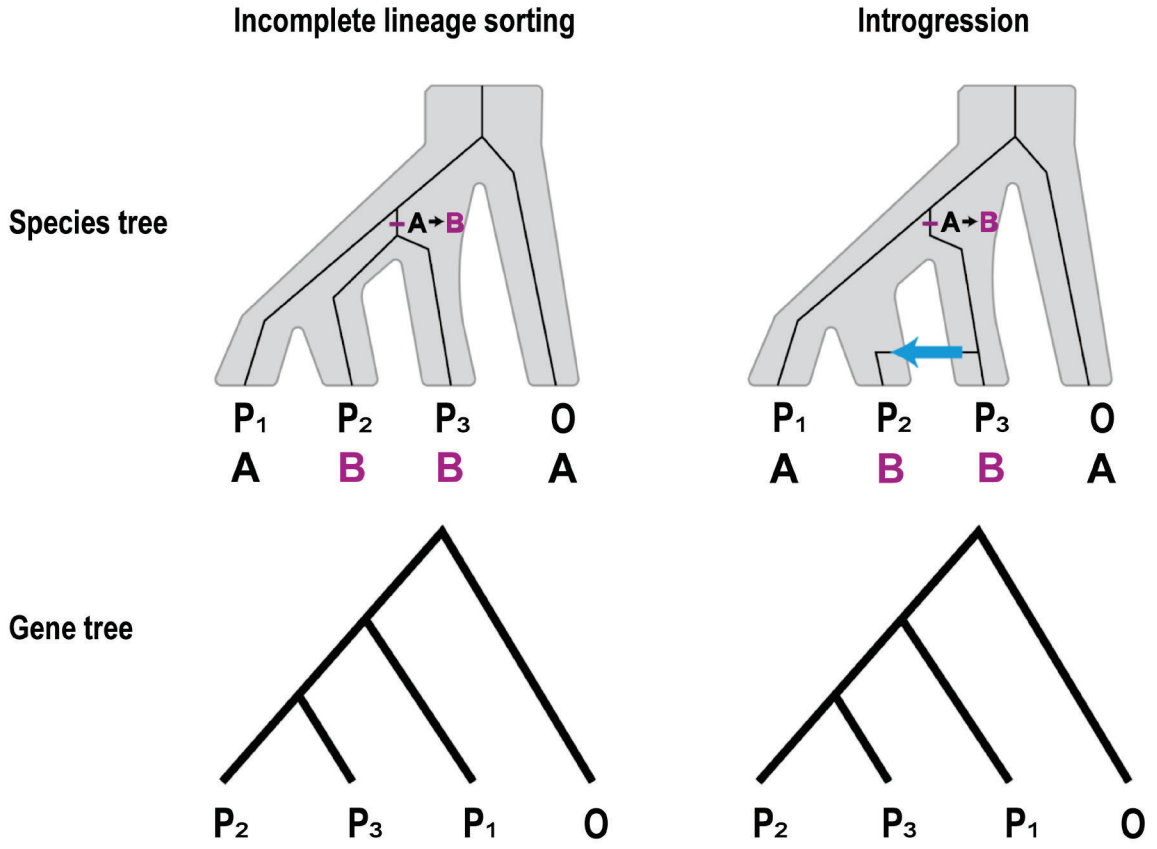


Fig. 2. Schematic depicting how ILS and introgression, two of the most common sources of gene tree heterogeneity, can result in gene tree and species tree discordance. Top diagrams show a species tree in which two sister species or populations (P_1, P_2), a third external species or population (P_3), and an outgroup (O) are represented. The black lines represent the gene tree for a given locus. Purple dash indicates a mutation from the ancestral allele A to a derived allele B. Diagramans on the bottom depict the resulting topology of the respective gene trees, both of them in conflict with the true species tree. Both ILS and introgression may lead to the same conflicting topologies. However, if ILS is the true underlying cause of the discordance, the frequencies of ABBA and BABA patterns are expected to be equal. Alternatively, if introgression is the cause of discordance then one of the two patterns is expected to occur at a higher frequency, which will depend on who P_3 introgressed with. Modified from Ferrer Obiol (2020).

Hybridization and introgression are an important source of interspecific incorporation of novel alleles. Both processes are widespread across the Tree of Life (Abbot et al., 2016; Mallet, 2007) and the outcome of evolutionary consequences can influence either positively or negatively the formation and persistence of independently evolving lineages (Mallet, 2007). Gene flow may lead to adaptive introgression across species boundaries via hybridization and backcrossing (Hedrick, 2013) or the formation of new species (hybrid speciation) through the emergence of novel genotypes and reproductive isolation between parental and hybrid lineages (e.g., Hermansen et al., 2011; Lamichhaney et al., 2018; Noguerales and Ortego, 2022). Either of these processes will result in discordant gene trees, reminiscent of the ILS. However, while ILS is more stochastic, hybridization and introgression are often directional and/or asymmetrical. Meaning that there is a possibility to distinguish between ILS and hybridization/introgression based on the distributions of conflicting phylogenetic patterns in the genome (Durand et al., 2011).

Most studies make use of the *D*-statistic (Durand et al., 2011; Patterson et al., 2012), also referred to as ABBA-BABA test, to detect signatures of introgression or hybridization. This method works on the assumption that the ancestral populations were randomly mating, is sensitive to asymmetry under alternative demographic scenarios, and it is based on a four-taxon test. Suppose that three populations (viz. P_1, P_2 and P_3) and an outgroup population, denoted O, are arranged in a pectinate tree topology as follows: (((P$_1$, P$_2$), P$_3$), O). Ancestral (or outgroup) alleles are denoted as A and derived alleles as B. For a given site, there are three possible allele configurations, viz. BBAA, ABBA or BABA. If gene heterogeneity is caused by ILS, the expected frequencies of ABBA and BABA configurations should be equal. Alternatively, if hybridization or introgression are the underlying causes of gene tree discordance, one of the patters is expected to occur with a higher frequency (Durand et al., 2011).

5. Species delimitation in an integrative framework

Darwin (1859, p. 44) wrote: "*No one definition [of species] has as yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species*".

We treat species as tangible entities, more so than other subspecific groups that are often used as a placeholder category for populations with fuzzy boundaries (Burbrink et al., 2022; Zink and McKitrick, 1995), yet determining when a population should conceptually and operationally be considered a species remains the thorniest of problems. With over 20 different species concepts in circulation (Agapow, 2005), systematists require statistical methods for testing the alternative models that may describe the number of species in a clade in order to delimit species with accuracy (Fujita et al., 2012). Taxonomical conflict derived from the use of different species definitions stems from the nature of evolution itself, that acts as a continuous process that not always follows a bifurcating path (Mallet et al., 2016). Thus, uncertainty is to be expected especially when species find themselves in the "gray zone" of speciation (de Queiroz, 1998, 2007), in which incomplete genealogical exclusivity is prevalent. Therefore, alternative contemporary species concepts come into conflict, for they adopt different properties as their respective cutoffs, e.g. two lineages being phenetically distinguisable, or reproductively isolated, or ecologically distinct and so on (de Queiroz, 2007).

The general lineage species concept, which treats species as independently evolving metapopulation lineages (de Queiroz, 1998), has provided a good framework for statistical species delimitation. Most notably, the combination of high-resolution genomic data with the use of multispecies coalescent methods (MSC) (Grummer et al., 2014; Yang and Rannala, 2010; Zhang et al., 2014) has proven a powerful resource to test multiple species hypotheses (Knowles and Carstens, 2007; Yang and Rannala, 2010), increasing the robustness of species delimitations and, progressively, moving taxonomist away from exclusively using qualitative or quantitative phenotypic traits. However, this gain in resolving power entails the risk of mistaking population structure for species boundaries (Chambers and Hillis, 2020; Frankham et al., 2012; Sukumaran and Knowles, 2017), which can lead to taxonomic inflation. This may have consequences on subsequent biological studies, conservation and management policies (Carstens et al., 2013; Mace, 2004; Sukumaran and Knowles, 2017; Thomson et al., 2021). A way to overcome artefactual over-splitting is to combine appropriate geographical sampling with multiple lines of evidence (e.g., morphological, karyological or genetic data) to provide a robust framework for species delimitation in what is known as an integrative species delimitation approach (Dayrat, 2005). Many studies have benefitted from this approach (e.g., Dejaco et al., 2016; Frajman et al., 2019; Spriggs et al., 2019; Steiner et al., 2018; Tonzo et al., 2019), which has the potential to solve long-standing taxonomic problems, or unearth new ones.

6. The Tricladida Lang, 1884

6.1. The Bauplan

Many may be familiar with a few examples of triclad flatworms (Fig. 3), often seen by the general public as flattened worm-like animals with googly eyes and a gift for regeneration, unbeknownst to the widespread and rich diversity they encompass. Triclads, popularly known as planarians, are members of an order of free-living Platyhelminthes known as the Tricladida, whose representatives are present in all biogeographical regions (Schockaert et al., 2008). Approximately, Tricladida harbors 1800 species (Tyler et al., 2006-2024) which may greatly differ in body size, coloration, reproductive structures, life cycle, reproductive strategies, and regenerative abilities (Sluys and Riutort, 2018).

Triclads owe their name to the three-branched intestine they possess, a blind sac located in a pouch approximately in the middle of the body that extends anteriorly (one branch) and posteriorly (two branches), ramifying into numerous diverticula (Fig. 4). Further, planarians are acoelomate bilaterians in which the space between organs is filled by connective tissue known as parenchyma or mesenchyme. Concurrently, they are also characterized by presenting a muscular pharynx (to which the intestine is connected), serial arrangement of the nephridiopores, cerebral position of the female gonads, and a marginal adhesive zone (Sluys, 1989).

Fig. 3. External appearance of an unmoved specimen of *Dugesia xeropotamica*, endemic from Corsica. Discovered and described in this thesis. Photography curtesy of Eduardo Mateos

Planarians lack specialized respiratory or circulatory systems. In consequence, oxygen or substances from digested food can only reach their internal tissues via diffusion, be it across the entire body (oxygen) or through the intestine to the surrounding tissues (nutrients). However, they do present an excretory system in which two rows of dorsoventrally arranged protonephridia excrete cellular waste through the nephridiopores. A pair of ventral nerve cords, which are connected by a diffuse nerve plexus, dominates the nervous system. These cords thicken in the head area, where the eyes and auricular grooves are housed, to form a bilobed cerebral ganglion. In turn, the eyes and auricular grooves are an important part of the sensory system, which is constituted by photoreceptors (i.e., the eyes) and chemoreceptors (e.g., auricular grooves).

Triclads are simultaneous hermaphroditic animals that generally exhibit cross-fertilization during copulation, with the exception of two dioic marine species, viz. *Sabussowia dioica* (Claparède, 1863) and *Cercyra teissieri* Steinmann, 1930. Numerous follicular testes are situated among the diverticula of the gut branches and a pair of ovaries is usually located near the cerebral ganglion, always in the prepharyngeal region of the body. Contrastingly, the copulatory apparatus, in which both male and female genitalia are comprised, is always situated in the postpharyngeal part of the body and holds great taxonomic value for classification and identification. However, planarians, most notably freshwater ones, are also capable of reproducing asexually either by fission or parthenogenesis (e.g., Beukeboom et al., 1998; Beukeboom and Vrijenhoek, 1998; Stocchino and Manconi, 2013).

Regarding their musculature, triclads present two types of muscle: cutaneous and parenchymatic. The former type forms the epidermal body wall, which is comprised of circular, longitudinal and diagonal fibres that may be arranged differently depending on

the taxonomic group. Concurrently, the parenchymatic fibres can also be longitudinal and diagonal, but transverse and dorsoventral as well.

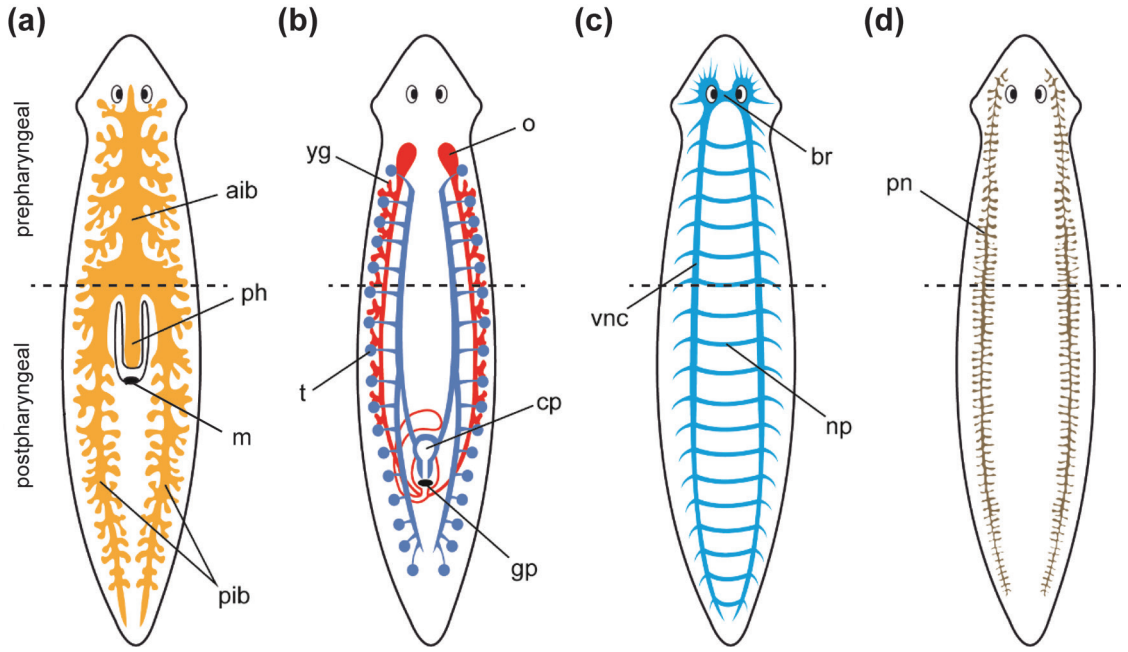


Fig. 4. Schematic of the different organ systems of a triclad flatworm. Digestive system with its numerous diverticula (a), the reproductive system (b), the nervous system (c), and the excretory system (d). Abbreviations: aib, anterior intestinal branch; br, brain; cp, copulatory apparatus; gp, gonopore; m, mouth; np, nerve plexus; o, ovary; ph, pharynx; pib, posterior intestinal branch; pn, protonefridia; t, testis; vnc, ventral nerve cord; yg, yolk glands. Modified from Leria (2018), based on Solà (2014).

Finally, a layer of mucus covers the entire body of planarians, which is involved in protective, adhesive and predatory functions (Pedersen, 1963; Prasniski and Leal-Zanchet, 2009). Secreted by diverse subepidermal glands, this mucus layer also provides planarians with a non-abrasive substrate over which they will glide using the combined action of their muscles and ventral cilia to move around their surroundings.

6.2. Regeneration

Planarians, with their remarkable regenerative abilities, would not make for the most thrilling adversaries in a sword fight, but they certainly provide a most interesting subject for studying regeneration and stem cell dynamics. This capability is due to the proliferative activity of an abundant type of adult stem cells known as the neoblasts (Rink, 2013; and references therein). This cell type can be found throughout the body, especially in the parenchyma which acts as a reservoir of neoblasts, but not in the pharynx and the tip of the head (Reddien and Alvarado, 2004).

Stem cells are generally defined as cells that have both the capacity of long-term self-renewal and to produce at least one type of differentiated cells (Watt and Hogan, 2000). How stem cells are classified depends on their potential to differentiate into distinct cell

types, being pluripotency (capacity to make all somatic cell types) and totipotency (capacity to make all somatic and extraembryonic tissues) only limited to the earliest stages of embryonic development in most organisms. Neoblasts are likely totipotent adult stem cells, as demonstrated by Wagner and collaborators (2011) who showed that a single neoblast transplanted into a stem cell depleted specimen was able to completely rescue the host. However, neoblasts are often referred to as pluripotent in a cautionary acknowledgement of the fact that the transplanted stem cell in the above-mentioned experiment was extracted from an asexual strain donor, thus lacking the reproductive system and, perhaps, the capacity to differentiate extraembryonic tissues.

Instead of presenting multipotent (capacity to make multiple cell types of a specific tissue) and tissue-specific stem cells, like vertebrate systems do, planarians only harbor neoblasts in quantity (20-35%, González-Estévez and Saló, 2010; Plass et al., 2018). These cells are the only cells that can divide mitotically (Morita and Best, 1974; Newmark and Alvarado, 2000) and maintain a continuous somatic cell turnover (Baguñà, 2012) to preserve tissue homeostasis, cell number and body size. Triclad flatworms are neither unique in possessing pluripotent stem cells, nor are they the only animals that undergo continuous tissue turnover. However, in the continuum of regenerative systems across the animal kingdom they are an extreme. So much so that recent studies have shown that at least some freshwater planarians are able to accumulate numerous mutations, including non-synonymous ones, which render them like authentic living mosaics (Leria et al., 2019; Nishimura et al., 2015).

6.3. Ecology

Triclad flatworms have colonized a great variety of freshwater, marine and land environments across the globe. However, their survival and dispersal are dependent on the availability of contiguous water bodies (freshwater and marine planarians) or high environmental humidity (land planarians). Thus, planarians are highly sensitive to dryness and desiccation. Interestingly, resistant forms have only been reported in planariids (family Planariidae), viz. *Atrioplana notadena* de Beauchamp, 1937 (Ginet and Puglisi, 1964), *Dendrocoelum* sp. (Manenti et al., 2022), *Hymanella retenuova* Castle, 1941, *Phagocata velata* (Stinger, 1909) (Ball et al., 1981), and *Polycelis nigra* (Müller, 1774) (Vila-Farré et al., 2011). Therefore, the reason why planarians are often referred to as low vagility organisms is mainly due to their intrinsic dependency on moisture (Ball, 1974).

While many may find it challenging to envision planarians as predators, the reality is that triclad flatworms often have a sit atop their local trophic chains (Vila-Farré and Rink, 2018 and references therein), with a varied diet that consists, for the most part, of small invertebrates (Cuevas-Caballé et al., 2019; Kreuzinger-Janik et al., 2018). In some cases, they may prey on other planarian species, or even on members of the same species (cannibalism) (Hay and Ball, 1979), although this behavior is mostly observed under lab

conditions. However, planarians may prey or scavenge on other items of animal origin (e.g., recently deceased animals). Some species may show preference for indicated preys, even displaying specific predatory behaviors (Boll and Leal-Zanchet, 2016; Reynoldson and Davies, 1970), yet generally planarians are generalists that may even share the same dietary resources, leading to potential niche overlaps (Lock and Reynoldson, 1976; Reynoldson and Davies, 1970).

Contrastingly, their role as appetizers for other species is largely unexplored, although there are studies that have shown that ciliates (Janovy, 1963) and salamanders (Gillespie, 2013) may prey on freshwater planarians. In fact, flatworms of the genus *Dugesia* constitute the most relevant prey item for the endangered salamander species *Eurycea sosorum* Chippindale, Prince & Harris, 1993 and it has been shown that planarians harbor a higher energy and protein content than amphipods and chironomids. Further, larval *Salamandra salamandra* (Linnaeus, 1758) and *Salamandrina perspicillata* (Savi, 1821) have been observed feeding on planarians (*Dendrocoelum* sp. and *Dugesia* sp.) in cave streams in Italy (extracted from Gillespie, 2013).

Interestingly, planarians have a remarkable ability to survive long periods of starvation, ranging from several months to well over a year, depending on the species. However, prolonged starvation results in organ degeneration (resorption of the reproductive apparatus), and reduction of size (loss of body mass) in a process known as degrowth (Berninger, 1911; Hyman, 1951). For example, the land planarian *Arthurdendyus triangulatus* (Dendy, 1894) can reduce its body mass up to 82-95% after 11 months of starvation (Baird et al., 2005; Boag et al., 2006). Degrowth is a rather unique and interesting phenomenon that may act as a regulator of the total biomass of populations to avoid substantial loss of individuals in situations when the habitat capacity is at its limit and competition is most relevant (Reynoldson, 1964, 1983). Most interestingly, starved individuals resemble juvenile planarians, regenerate faster and show higher oxygen consumption, which has led some researchers to consider that starved planarians actually rejuvenate (González-Estévez, 2008; and references therein).

6.4. Systematics

The Order Tricladida is one of numerous orders within Platyhelminthes (see Riutort et al., 2012 for a detailed recount). Along with the orders Fecampiida and Prolecitophora, it forms a clade known as Adiaphanida (Norén and Jondelius, 2002), for which there are no morphological synapomorphies, and molecular data is the only evidence that supports it (Laumer and Giribet, 2017).

Initially, Hallez (1894) recognized three taxonomic groups that reflected the ecological habitats that triclads occupied: Paludicola (freshwater planarians), Maricola (marine planarians), and Terricola (land planarians). Sluys (1990) proposed a fourth clade, Caverni-

cola, based on morphological features related to the reproductive apparatus that involved species that had been previously assigned to Maricola and Paludicola. Furthermore, said species presented a discontinuous distribution, and they were often found in hypogean environments as well. However, the internal relationships of the Tricladida would radically change when a study by Carranza and collaborators (1998), using 18S ribosomal sequences, showed a phylogenetic tree that proved instrumental to establish the current classification. In said tree, Terricola were shown to be the sister taxon to one of Paludicola's families: Dugesiidae. Consequently, Paludicola was rendered paraphyletic. Thereafter, this hypothesis would be validated by subsequent molecular studies (Baguñà et al., 2001; Álvarez-Presas et al., 2008). Thus, a new taxon emerged which replaced the obsolete taxa Terricola and Paludicola: Continenticola.

Presently, triclad flatworms can be classified into the three subsequent suborders, viz. Maricola Hallez, 1892, Cavernicola Sluys, 1990, and Continenticola Carranza *et al.*, 1998 (Fig. 5). The relationships between these suborders remain unclear, yet recent molecular studies have shown that Continenticola is likely the sister taxon to Cavernicola and Maricola (Harrath et al., 2016; Laumer and Giribet, 2014). However, the relationships within each suborder have received uneven attention, being Continenticola the most studied. Extensive molecular analyses have been performed on this taxon, while only a very preliminary study has been conducted on Maricola (Charbagi-Barbirou et al., 2011) and the internal relationships of Cavernicola have only been recently examined from a molecular point of view (Benítez-Álvarez et al., 2020; Stocchino et al., 2021).

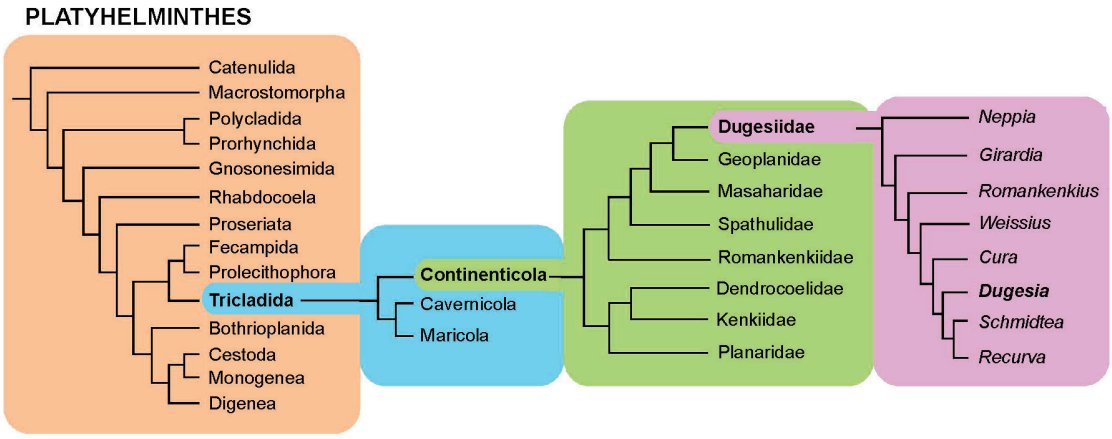


Fig. 5. Schematic phylogeny of the Tricladida focused on the family Dugesiidae. Different taxonomic ranks are color-coded as follows: orders (orange), suborders (cyan), families (green), and genera (lilac). Modified from Leria (2018). Phylogenetic relationships are based on results reported in Laumer et al. (2015), Harrath et al. (2016), and Grant (2016).

At present, Continenticola houses to superfamilies: Planarioidea Stimpson, 1857 and Geoplanoidea Stimpson, 1857. The former superfamily comprises three families of freshwater planarians, viz. Dendrocoelidae Hallez, 1892, Kenkiidae Hyman, 1937, and Planariidae Stimpson, 1857, being kenkiid flatworms the only group within Continenticola for which there is not even a single molecular sequence published. The latter superfamily, Geoplanoidea, houses two families, viz. Dugesiidae Ball, 1974 (freshwater planarians) and Geoplanidae Stimpson, 1857, formerly known as Terricola (land planarians) (Sluys and Riutort, 2018). While this view of Continenticola is formally accepted, it is worth mentioning that recent work on freshwater triclads from Australia (Grant, 2016; unpublished data) has challenged the notion of Dugesiidae as a monophyletic group. In light of its likely polyphyletic status, significant revisions within Dugesiidae have been considered, proposing the creation of three new families to accommodate the diversity of Australian freshwater flatworms, viz. Masaharidae, Romankenkiidae and Spathulidae (Fig. 5). However, as of present date, this proposal has not been formalized in a published paper.

7. The family Dugesiidae Ball, 1974

Members of the family Dugesiidae, occasionally referred as dugesiids, can be distinguished due to the entrance of the oviducts into the bursal canal, above the zone of the shell glands, and the bilayered arrangement of the inner pharyngeal muscles (Ball, 1974). Dugesiidae representatives are distributed worldwide; however, Antarctica, Greenland, Iceland, and some oceanic islands are exceptions to this distribution. It is acknowledged that based on said distribution pattern dugesiids are, most certainly, an ancient family. Biogeographical hypotheses on the origin of this family have been lengthily discussed (e.g., Ball, 1974, 1975; Kawakatsu, 1968; Kawakatsu et al., 1976), yet nowadays the prevailing hypothesis is that of Sluys and collaborators (1998), in which it is proposed that Dugesiidae must have originated and diversified already on Pangea. However, this hypothesis remains to be tested on the basis of molecular data which, considering the large distribution and the lack of molecular resources on the family, will not be an easy task.

As of present day, 12 genera are established within this family: *Bopsula* Marcus, 1946; *Cura* Strand, 1942; *Dugesia* Girard, 1850; *Eviella* Ball, 1977; *Girardia* Ball, 1974; *Neppia* Ball, 1974; *Reynoldsonia* Ball, 1974; *Romankenkius* Ball, 1974; *Schmidtea* Ball, 1974; *Spathula* Nurse, 1950; *Weissius* Sluys, 2007 and *Recurva* Sluys, 2013 (Sluys and Riutort, 2018). Further, ongoing morphological and molecular studies on Australian dugesiids show that there is yet a broader diversity to be unraveled within this family, possibly even in

the form of new general (Grant, 2016; unpublished data). While it is true that Dugesiidae has received much more attention from a systematic point of view than other families of triclads, little is known about the phylogenetic relationships among its genera. There are only a few studies that have molecularly explored the dugesiids as a whole (Álvarez-Presas et al., 2008, 2014), and the obtained results suggest that probably its monophyly should be carefully and thoroughly re-evaluated.

8. The genus *Dugesia* Girard, 1850

The longstanding genus *Dugesia* is, contrary to other dugesiid genera, a well-defined and homogenous taxon and, most certainly, the most speciose of the 12 genera presently recognized in the family. It encompasses more than 110 nominal species, all of which are characterized by the presence of a diaphragm between the seminal vesicle and the ejaculatory duct, and the emergence of the oviducts from the dorsal surface of the ovaries (Sluys, 2001; de Vries and Sluys, 1991).

While it holds true that planarians are generally regarded as low vagility species, *Dugesia*'s large distribution range challenges this view and have supposed an intriguing aspect of this taxon, despite decades of biogeographic speculation on its origin and dispersal (Ball, 1974; Kawakatsu, 1968; Sluys et al. 1998). Species of the genus can be found inhabiting freshwater bodies in Africa, Europe, the Middle East, and Australasia (Fig. 6). Such a widespread and disjointed distribution could be attributable to dispersal and vicariant events. Solà et al. (2022), in what can only be considered to be as the most ambitious biogeographic study on *Dugesia* to date, proposed a hypothesis, based on extensive molecular data, in which both vicariance and dispersion, occasionally even transoceanic dispersion, were the main drivers of *Dugesia*'s current distribution and diversification, shaping its evolutionary history. A history, on the other hand, that started somewhere in the eastern part of Gondwana about 160-130 Mya, at the Jurassic-Cretaceous boundary, when

the first *Dugesia* lineages diverged in a region that would give rise to the Afro-Malagasy landmasses, which as suggested by their name, would become the present-day Africa and Madagascar.

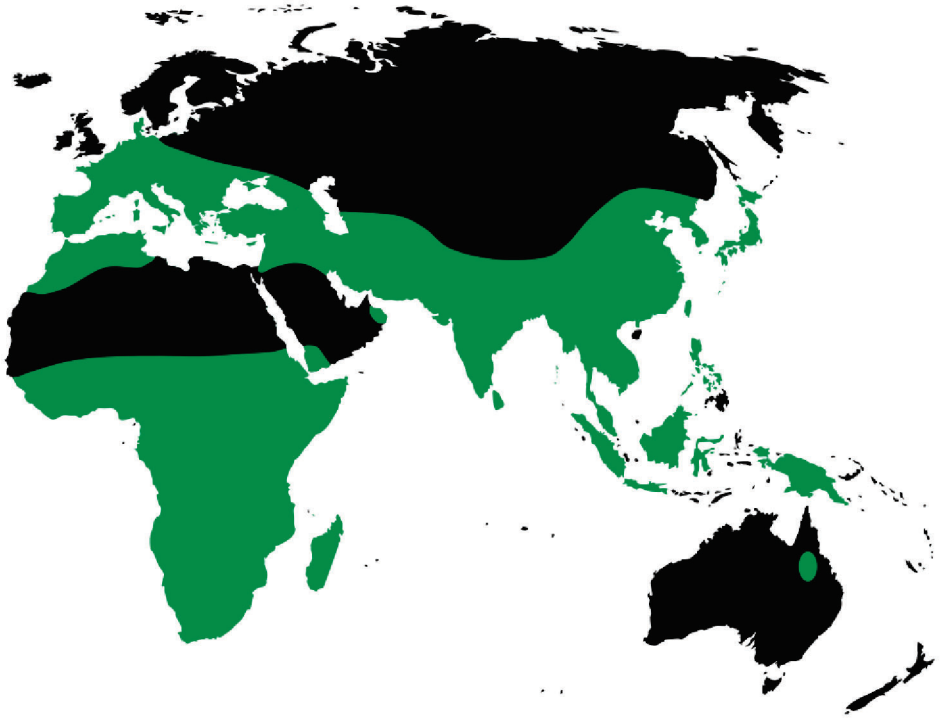


Fig. 6. Distribution range of the genus *Dugesia*, highlighted in green. Extracted from Solà (2014).

As of present date, seven major clades of *Dugesia* can be distinguished (Solà et al., 2022). Briefly recounting, an initial split separated the first Malagasy clade (160-130 Mya), followed by a second split, likely caused by the separation of Africa and Madagascar during the Mesozoic era (c. 130 Mya). As a result, a second Malagasy clade and a first African clade diverged. It is unclear how diversification might have occurred within Africa, yet a second African clade emerged and this one is considered to be the sister group to the Eurasian species. Within the Eurasian group, three clades can be discerned: an Asian clade and two European clades, divided into the Eastern and Western European clades, often referred to as the Eastern and Western Mediterranean clades.

The Mediterranean is considered to be one of the most biodiverse areas in the Palearctic region, and many researchers have turned their inquiring eyes to its taxa to address a wealth of biological questions (e.g., Barbato et al., 2018; Bisconti et al., 2013; Dapporto et al., 2007; Guo et al., 2022; Salvi et al., 2017; Solà et al., 2013). European *Dugesia*, most notably the Mediterranean populations, have been subject of numerous studies focused on delving into the evolutionary history of the aforementioned clades (e.g., Benítez-Álvarez et al., 2023a; Lázaro et al., 2009; Leria, 2019; Solà et al., 2013). However, the interest in these species extends beyond solely elucidating their phylogenetic relationships or understanding how they came to reach most of their present-day distribution ranges. More recently, some species, viz. *Dugesia subtentaculata* (Draparnaud, 1801) and *Dugesia etrusca*

Benazzi, 1946, have also been used as model systems for studying the effects of long-term asexuality and compare the genetic footprints of different reproductive strategies (i.e., fissiparity versus sexuality) (Benítez-Álvarez et al., *in prep.*; Leria et al., 2019).

While the Eastern Mediterranean diversity and biogeographic history have been lengthily studied (Sluys et al., 2013; Solà et al., 2013; de Vries, 1984, 1985, 1988), only in recent years the focus of evolutionary studies has started to veer into the Western Mediterranean (Benítez-Álvarez et al., 2023a; Lázaro et al., 2009; Leria, 2019). Therein, like the small village of Gauls conceived by René Goscinny, lies the subject of study of this thesis: a small taxonomic group known as the Corso-Sardinian clade.

9. The Corso-Sardinian clade

9.1. Geographical context

As suggested by its name, the Corso-Sardinian clade comprises autochthonous species, or endemics, of Corsica (France) and Sardinia (Italy). Presently located amidst the waters of the Ligurian and Tyrrhenian Seas, these two islands share a rich geological history that started approximately 25 Mya, during the Oligocene-Miocene transition, when a landmass comprising what would later become the Balearic Islands, Corsica and Sardinia, among other geographical regions, detached from southwestern Europe and started a slow-paced drift southwards (Alvarez et al., 1974; Rosebaum et al., 2002). During the course of its journey, said landmass fragmented several times, leading to the formation of smaller tectonic units known as microplates. The Sardo-Corsican microplate migrated eastwards following a counter clockwise rotation (Alvarez, 1972; Carmignani et al., 2004) and, concurrently, Corsica and Sardinia slowly split apart in a process that started c. 20-15 Mya and was completed 9 Mya. Further, microplate tectonics of the Western Mediterranean has been shown to be an important driver for the dispersal and diversification of many species (e.g., Bidegaray-Batista and Arnedo, 2011; Mendes et al., 2017; Opatova et al., 2016; Pérez-Losada et al., 2011; Pfenninger et al., 2010), and freshwater planarians are not an exception (Lázaro et al., 2011; Leria, 2019). Nowadays, Corsica and Sardinia are separated by the Bonifacio Strait, a narrow (11 km) and shallow (60-100 m deep) water channel. As a consequence of the above mentioned, and the assumed low vagility of planarians, it is agreed that the Corso-Sardinian clade, and most likely the lineage from

which it diverged, has been isolated from other continental and insular dugesiid species ever since the Sardo-Corsican microplate started to drift apart (c. 20-15 Mya).

9.2. Historical context

The scientific history of this clade can be separated in two periods, to which I will refer as pre- (1951-1999) and post-molecular (1999-present) based on the lack or availability of genetic sequences of this group. The wealth of pre-molecular knowledge on the triclad fauna of these islands is broad and rich, especially for Sardinian populations, and the first records are due to the planariologist Mario Benazzi who studied freshwater planarians of northern and southwestern Sardinia during his stay as professor at the University of Sassari (1935-1936). Benazzi documented the occurrence of three different species on Sardinia (Benazzi, 1936, 1938a, b), viz. *Euplanaria gonocephala* (presently *Dugesia* spp.), *Euplanaria lugubris* s.l. (presently *Schmidtea* spp.), and *Fonticula* sp. (presently *Atrioplanaria* sp.). Thereafter, numerous morphological, taxonomic, cytogenetic, karyological, and molecular studies would follow from his initial reports, leading to a better understanding of the taxonomic status of the triclad inhabitants on Corsica and Sardinia, the number of species and endemics, the variety of karyotypes within and between species, their reproductive biology, and their geographic distribution (e.g., Benazzi, 1950; Benazzi-Lentati, 1979; Benazzi and Deri, 1980; Canovai, 1989; Deri, 1971, 1981; Lázaro et al., 2009, 2011; Lepori, 1948, 1951; Lepori and Pala, 1982; Pala et al., 1981, 1982, 2000; Stocchino, 2018; Stocchino et al., 2005; Stocchino and Manconi, 2013; Vacca et al., 1988, 1993).

However, only one of the three initially reported species by Benazzi has meaningful relevance to this thesis: *Euplanaria gonocephala*. Shortly after Benazzi's initial reports, *Euplanaria* was synonymized to *Dugesia* (Hyman, 1939), thus *E. gonocephala* became *Dugesia gonocephala* (Dugès, 1830). Initially, this species was considered a "superspecies" comprised of numerous "microspecies" distributed across the Palearctic region (de Vries and Ball, 1980), yet subsequent studies challenged this view. Lepori, a colleague and assistant teacher to Benazzi who had been studying Corsican and Sardinian populations of *D. gonocephala*, erected a new endemic species based on morphological evidence, viz. nominal *Dugesia benazzii* Lepori, 1951. Thus, unbeknownst to its discoverer, the first member of the Corso-Sardinian clade was welcomed to the planarian research community. However, almost seventy years would pass before the first molecular data supported the existence of this taxonomic group (Lázaro et al., 2009).

Dugesia benazzii was thought to be a single species distributed across Corsica, Sardinia, and Capraia (Benazzi, 1955), wherein various karyological biotypes were identified (Benazzi, 1945; Benazzi-Lentati and Piegaia, 1953), all sharing the same haploid chromosomal complement, n = 8. This notion of a single species would be challenged for the first time by Pala and collaborators (1981), who discovered a new biotype (n = 7) (Fig. 7) and elevated it to the species status, thus describing *Dugesia hepta* Pala, Casu & Vacca,

1981, a Sardinian endemic with a chromosomal complement unique among the *Dugesia* species of the Western Palearctic region. Thereafter, a new species from Sardinia would be described based on morphological and karyological evidence, viz., *Dugesia leporii* Pala, Stocchino, Corso & Casu, 2000 (Pala et al., 2000). This species also presented a haploid chromosomal complement of n = 8 and would become the last endemic *Dugesia* to be described in the Sardo-Corsican region. In addition, it is worth mentioning that after its formal description, all subsequent efforts to collect wild specimens of *D. leporii* have been in vain.

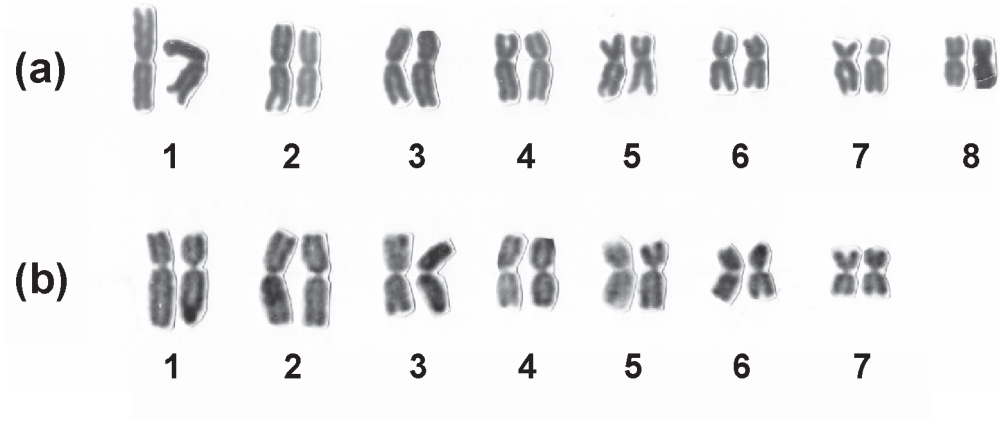


Fig. 7. Chromosome complements of *D. benazzii* (a) and *D. hepta* (b). Extracted from Dols-Serrate et al. (2020).

As previously mentioned, during the pre-molecular period (1951-2009) extensive research was conducted on the endemic *Dugesia* populations of Corsica and Sardinia, covering a wide variety of biological disciplines. However, the evolutionary history and phylogenetic relationships of these species were largely understudied in comparison. Sluys and collaborators (1998) were the first ones to establish a phylogenetic hypothesis based on morphological data that included *D. benazzii* and *D. hepta*. Evidently, both species were grouped with other Western Palearctic *Dugesia* species, but the resolution within such clade was lacking.

With the advent of molecular techniques, the prospects of exploring the Western Mediterranean *Dugesia* under a renewed perspective were enticing, and a study by Baguñà and collaborators (1999) would be the starting point of the post-molecular period (1999-present) of the Corso-Sardinian clade. Based on the internal transcribed spacer-1 (*ITS-1*) marker, it was possible to infer the first molecular phylogenetic tree of *Dugesia* in the Mediterranean. Therein, a clade comprising *D. benazzii* and *D. hepta* specimens was reported, although the number of representatives was meagre (two *D. benazzii* and one *D. hepta*), and the clade included an asexual specimen that was posteriorly identified as *D. etrusca*. Ten years later, a molecular phylogenetic study by Lázaro and collaborators (2009) on Mediterranean *Dugesia*, among other groups, using *ITS-1* and cytochrome c oxidase subunit I (*COI*) would report anew a clade, this time, comprised exclusively of *D. benazzii* and *D. hepta* individuals, which later on would be baptized as the Corsica-Sar-

dinia (or Corso-Sardinian) clade by Benítez-Álvarez and collaborators (2023a). However, although subsequent phylogeographic studies on the Mediterranean *Dugesia* fauna have included representatives of both species (Leria, 2019; Solà et al., 2013, 2022), their presence has always been testimonial and there are no published studies dedicated to the Corso-Sardinian clade.

9.3. An appealing case of study

Beyond the simple appeal of unravelling the unexplored, the Corso-Sardinian clade poses an interesting case of study for reasons I will subsequently elaborate.

Dugesia species commonly share a haploid chromosomal complement of 8 chromosomes (n = 8), in which length and centromeric position may vary between species. Nonetheless, some species may show different complements, such as n = 7 (e.g., *Dugesia ryukyuensis* Kawakatsu, Oki, Tamura & Sugino, 1976) or n = 9 (e.g., *Dugesia aethiopica* Stocchino, Corso & Manconi, 2002, *Dugesia maghrebiana* Stocchino, Manconi, Corso, Sluys, Casu & Pala 2009 or *Dugesia sicula* Lepori, 1948). Euploidies (i.e. changes involving the gain or loss of entire chromosome sets) are frequent in *Dugesia* and often associated to the reproductive strategy. Furthermore, supernumerary chromosomes (also known as B-chromosomes), chromosomic translocations and pericentric inversions have also been shown to occur in this genus (Ribas, 1990; Tamura et al., 1991, 1998; de Vries, 1986).

Dugesia hepta supposes the only known case of a haploid chromosomal complement of n = 7 among congenerics of the Western Palearctic region. It is assumed that a chromosomal rearrangement was the underlying cause of its peculiar karyotype, yet such hypothesis has never been examined. Furthermore, it is an iteroparous univoltine (breeds repeatedly throughout its life, but once per season) that only presents exclusively sexual natural populations (Stocchino and Manconi, 2013), all of which have a stable diploid karyotype (2n = 14).

In contrast, its closest known relative, viz. *D. benazzii*, presents the common haploid chromosomal complement n = 8. Furthermore, it houses natural iteroparous sexual and mixed (Fig. 8) (coexistence of sexual and fissiparous individuals) populations, and different chromosomic portraits, also referred to as biotypes, have been described for this species. Most notably, there is a particular biotype that had been object of debate in the past, known as the biotype *G* (Benazzi, 1945). This biotype was characterized by presenting an aneuploid set that showed great intraindividual variability, being 32 the most frequent number of chromosomes (Benazzi-Lentati and Piegaia, 1953). Pala and collaborators (1982) hypothesized that this biotype was in fact a *natural stabilized hybrid* between *D. hepta* and *D. benazzii*. However, a subsequent karyological study by Benazzi-Lentati and Benazzi (1985) questioned said hypothesis.

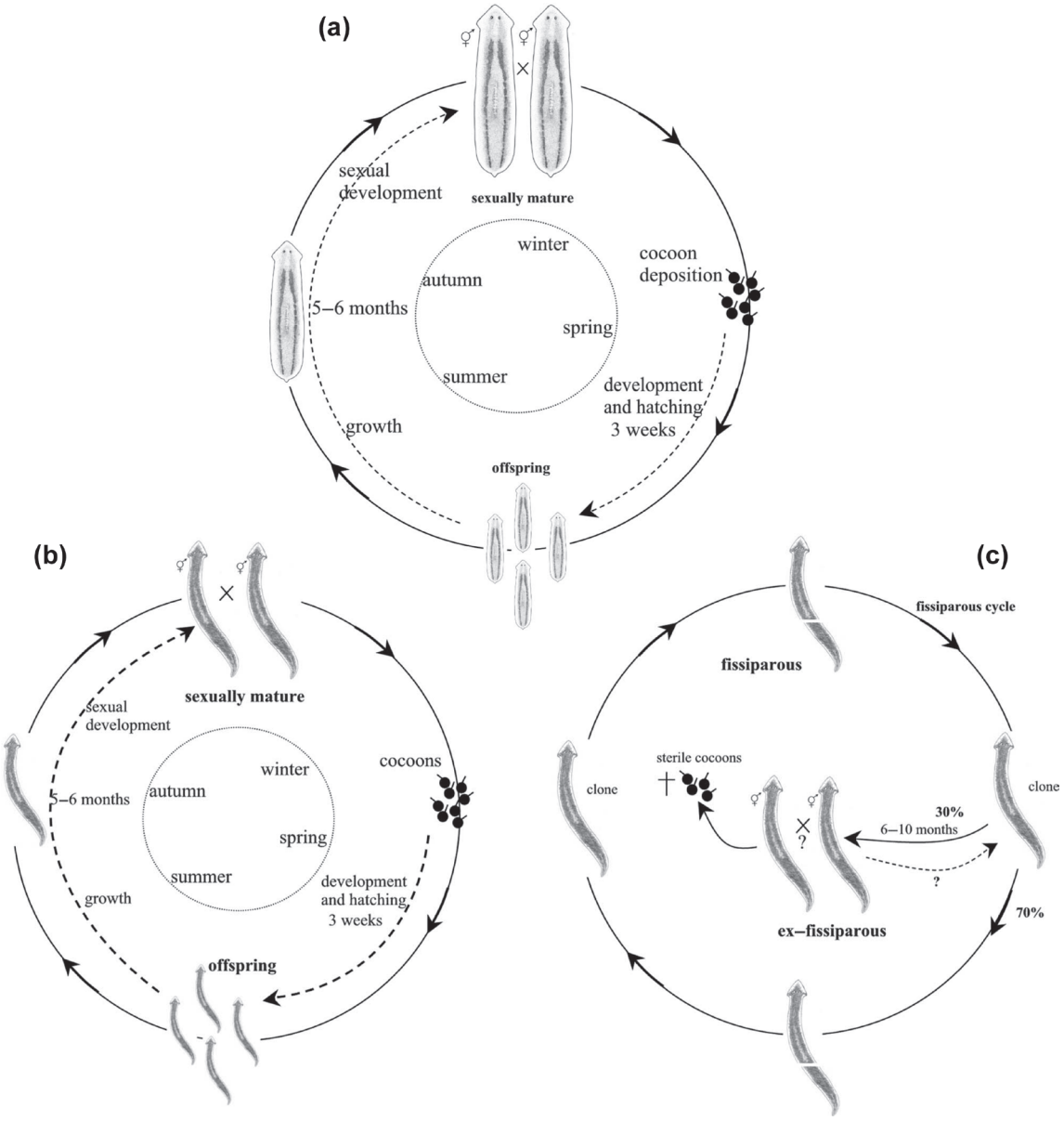


Fig. 8. Schematic overview of the life cycles and seasonal reproductive patterns of *D. hepta* (a) and *D. benazzii* (b, c). (a, b) Sexual cycles with their respective seasonal rhythms in both sexual populations and sexual individuals from mixed populations. (c) Life cycle of fissiparous *D. benazzii* individuals from mixed populations. Extracted and modified from Stocchino and Manconi (2013).

Thus, all things considered, the Corso-Sardinian clade offers an opportunity to not only study chromosomal rearrangements as speciation drivers, but also to study the only case of interspecific hybridization ever reported in triclads.

Objectives

All that I have poured, scientifically, into this thesis is but a small contribution to the understanding of the evolutionary history of Tricladida and the processes that have deftly shaped the diversification of this wondrous taxonomic group. My goal, the ultimate objective of this thesis, has been to pursue the understanding of how genetic diversity is generated, changed and lost throughout the scales of space and time using as model systems the freshwater planarians of the Corso-Sardinian clade. To attain such objective, I made use of multiple sources of data, integrating morphological, karyological and molecular data to shed light on the patterns and processes that have contributed to the evolution of this taxon.

The specific objectives of this thesis are:

• Obtain the first molecular phylogeny of the Corso-Sardinian clade and ascertain the reciprocal molecular monophyly of *D. benazzii* and *D. hepta*.

• Evaluate and update the taxonomy of the Corso-Sardinian clade using an integrative taxonomical approach combining molecular data obtained from legacy Sanger and RADseq loci, morphological and karyological data.

• Resolve the phylogenetic relationships of this taxonomic group.

• Validate the hypothesis of interspecific hybridization between *D. benazzii* and *D. hepta* and characterize said putative hybrids from a molecular and karyological perspective.

• Explore the historical processes that have shaped the biogeographical history of this taxonomic group.

• Sequence, annotate, and analyze the first chromosome-scale genome assemblies of *Dugesia* species of the Western Mediterranean, *D. hepta* included among them, to establish the bases for understanding the role of chromosomal rearrangements as drivers for speciation and provide new genomic resources for the planarian research community.

Publications

Supervisor’s report

UNIVERSITAT DE BARCELONA

Dr. Marta Riutort
Catedràtic de Genètica

Departament de Genètica, Microbiologia i Estadística
Diagonal 643
Edifici Prevosti
08028 Barcelona
Tel. +34 93 4035432
mriutort@ub.edu

La Dra. Marta Riutort León, directora de la Tesi Doctoral elaborada pel Sr. Daniel Dols Serrate, amb el títol "**Of flatworms and genes: Systematics, hybridization and evolutionary history of the Corso-Sardinian *Dugesia* (Platyhelminthes, Tricladida)**"

INFORMA que la tesi doctoral està elaborada com a compendi de 5 publicacions amb dades originals (articles 1-2-3-4-5 en el cos central de la tesi):

1. **Dols-Serrate, D**., Leria, L., Aguilar, J.P., Stocchino, G.A., Riutort, M. (2020). *Dugesia hepta* and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex? *Organisms Diversity and Evolution* 20, 369–386. https://doi.org/10.1093/zoolinnean/zlad135
   Factor d'impacte 3 (2020). Ocupa la posició **15** de 174 **(D1**) dins la categoria de *Zoologia*. Cites: 6

2. **Dols-Serrate, D.**, Stocchino, G.A., Sluys, R., Riutort, M. (2023). Fantastic beasts and how to delimit them: an integrative approach using multispecies coalescent methods reveals two new, endemic *Dugesia* species (Platyhelminthes: Tricladida) from Corsica and Sardinia. *Zoological Journal of the Linnean Society*, zlad135. https://doi-org.sire.ub.edu/10.1093/zoolinnean/zlad135
   Factor d'impacte 2,8 (2022). Ocupa la posició **12** de 177 **(D1**) dins la categoria de *Zoologia*.

3. **Dols-Serrate, D.**, Stocchino, G.A., Nuin-Villabona, P., Sluys, R., Riutort, M. (acceptat). New insights into the evolution and biogeography of freshwater planarians on islands in the Tyrrhenian Sea, Western Mediterranean Basin, with the integrative description of a new endemic species from Corsica (Platyhelminthes, Tricladida, *Dugesia*). *Zoological Journal of the Linnean Society*,
   Factor d'impacte 2,8 (2022). Ocupa la posició **12** de 177 **(D1**) dins la categoria de *Zoologia*.

4. **Dols-Serrate, D.**, Guo, L., Kruglyak, L., Riutort, M. (enviat). Mixed, not stirred: Genomic analyses of *Dugesia* (Platyhelminthes, Tricladida) shed light on the evolutionary history and diversification of the Corso-Sardinian clade and reveal the first case of interspecific hybridization in freshwater flatworms. Enviat a *Molecular Phylogenetics and Evolution*

5. **Dols-Serrate, D.**, Tenaguillo, I., Pisarenco, V.A., Olivé-Muñiz, M., Rozas, J., Riutort, M. (en preparació). The first chromosome-scale assemblies of *Dugesia* shed light on the genome size evolution in flatworms.

Aportació del doctorand als articles. El doctorand ha participat en la majoria dels mostrejos realitzats a Sardenya i Còrsega per a totes les publicacions (alguns d'ells realitzats en període de pandèmia COVID). S'ha encarregat de les extraccions d'àcids nucleics i de l'obtenció dels cariotips. Ha obtingut les dades moleculars tant per PCR i seqüenciació Sanger com les de RADseq. Ha realitzat les anàlisis filogenètiques, biogeogràfiques, d'estructuració genètica i de delimitació d'espècies dels articles 1 a 4. Es va fer càrrec de la revisió de la qualitat i neteja del genoma de *D. hepta* així com de la seva anotació estructural i funcional i de les anàlisis

UNIVERSITAT DE BARCELONA

Dr. Marta Riutort
Catedràtic de Genètica

Departament de Genètica, Microbiologia i Estadística
Diagonal 643
Edifici Prevosti
08028 Barcelona
Tel. +34 93 4035432
mriutort@ub.edu

estadístiques de la comparació de genomes amb altres espècies de *Dugesia* i *S. mediterranea* de l'article 5. Va escriure la primera versió dels manuscrits i ha participat en tot el procés d'elaboració fins a la versió final de tots els articles. Ha actuat com a autor de correspondència dels articles 2, 3 i 4. En cap cas s'ha utilitzat, implícitament o explícitament, els treballs corresponent als articles 1 a 4 presentats en aquesta tesi per a l'elaboració d'una altra tesi doctoral prèvia. L'article 5 en la seva forma final formarà part d'una altra tesi doctoral, ja que inclou dos genomes més que han estat analitzats per un altre doctorand.

A més s'inclouen en l'apèndix 3 publicacions més realitzades en paral·lel amb els treballs de la tesi. Aquests han resultat de col·laboracions científiques on el doctorand va contribuir a les anàlisis moleculars de laboratori humit i/o fent servir les eines bioinformàtiques o analítiques que també ha emprat en la seva tesi doctoral.

6. Benítez-Álvarez L., Leria L., **Dols-Serrate D.**, Riutort M. (2023). Building phylogenies from transcriptomic data. Book chapter in "The *Schmidtea mediterranea* toolbox" (Luca Gentile, ed.) within Springer Nature series Methods in Molecular Biology. https://link.springer.com/book/10.1007/978-1-0716-3275-8

7. Guo, L., Bloom, J., **Dols-Serrate**, **D.**, Ben David, E., Schubert, O.T., Kazuma, K., Ho, K., Wei, Y., Leighton, D., Boocock, J., Vergara, T.L., Riutort, M., Sánchez Alvarado, A., Kruglyak, L. (2022). Island-specific evolution of a sex-primed autosome in the planarian *Schmidtea mediterranea. Nature,* 606: 329–334 https://www.nature.com/articles/s41586-022-04757-3
Factor d'impacte 64 (2022). Ocupa la posició **1** de 73 **(D1)** dins la categoria de *Multidisciplinary Sciences*. Cites: 13

8. Stocchino, G.A., **Dols-Serrate, D.**, Sluys, R., Riutort, M., Onnis, C., Manconi, R. (2021) Amphibioplanidae: a new branch and Family on the phylogenetic tree of the triclad flatworms (Platyhelminthes, Tricladida), represented by a species from Sardinian caves with a remarkable lifestyle. *Zoological Journal of the Linnean Society*, 193(4): 1364–1391 https://doi.org/10.1093/zoolinnean/zlaa183
Factor d'impacte 3,83 (2021). Ocupa la posició **8** de 176 **(D1)** dins la categoria de *Zoologia*. Cites: 4

Barcelona, 2 de maig de 2024

Dr. Marta Riutort León
Catedràtica de Genètica
Universitat de Barcelona

Chapter I

Dugesia hepta and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex?

Dols-Serrate, D., Leria, L., Aguilar, J.P., Stocchino, G.A., Riutort, M.

2020. *Organisms Diversity and Evolution*, 20, 369-386.
doi: 10.1007/s13127-020-00438-z

Organisms Diversity & Evolution (2020) 20:369–386
https://doi.org/10.1007/s13127-020-00438-z

Organisms Diversity & Evolution

ORIGINAL ARTICLE

Dugesia hepta and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex?

Daniel Dols-Serrate[1] · Laia Leria[1] · Juan Pablo Aguilar[1] · Giacinta Angela Stocchino[2] · Marta Riutort[1]

Received: 19 December 2019 / Accepted: 17 April 2020 / Published online: 29 May 2020
© Gesellschaft für Biologische Systematik 2020

Abstract
Dugesia hepta Pala, Casu & Vacca, 1981 and *Dugesia benazzii* Lepori, 1951 are two freshwater planarian species from the islands of Corsica and Sardinia. *Dugesia hepta* is endemic of Sardinia and distributed in four northern hydrographic basins where it co-occurs with *D. benazzii*, which has a wider Tyrrhenian distribution. Although these species have been broadly studied—especially *D. benazzii*—as regards to their variety of reproductive patterns as well as for their karyological diversity, little is known about them from a molecular phylogenetic perspective. For the first time, we present a molecular phylogenetic tree of the two species and their populations based on two molecular markers—one mitochondrial, *Cox1*, and one nuclear, *Dunuc12*. Our results not only confirm that both species are molecularly distinct but also show that *D. benazzii*'s Corsican and Sardinian populations could belong to separate species. Furthermore, we present the first demonstration of a natural hybridization between different species in the genus *Dugesia* on the basis of molecular data.

Keywords *Dugesia* · Hybridization · Hybrid speciation · Molecular phylogeny · Haplotype network

Introduction

Tricladida (Lang 1884)—most commonly known as planarians—have been the subject to an in-depth research in a wide-ranging spectrum of scientific fields, such as regeneration, pattern formation, genomics, and transcriptomics (Abril et al. 2010; Newmark and Alvarado 2002; Reddien and Alvarado 2004; Robb et al. 2007, 2015) as well as on diversity and phylogeographical analyses (Álvarez-Presas and Riutort 2014; Leria et al. 2018; Leria et al. 2020; Sluys et al. 2013; Solà et al. 2013). However, historically, the Tricladida have been a challenging group for taxonomists and systematists due to the unsuspected complexity for classifying their specimens, which relies, ironically, in their morphological simplicity. Within any genus, most planarian species share a common external morphology; hence, the diagnostic characters are mainly found in their most complex anatomical system, the copulatory apparatus. Nonetheless, the existence of fully fissiparous species (Stocchino and Manconi 2013) as well as the fact that freshwater planarians are known for being able to resorb their own reproductive organs during starvation periods (Berninger 1911; Newmark and Alvarado 2002; Schultz 1904) can sometimes render the morphological approach useless.

A good example of this situation is the longstanding freshwater flatworm genus *Dugesia*. This genus inhabits the Afrotropical, Palearctic, Oriental, and Australasian biogeographic regions and comprises an approximate number of 90 described species (Harrath et al. 2019; Leria et al. 2020; Stocchino et al. 2017). In this genus, three main types of life cycle related to their reproductive strategy can be found: sexual, asexual (fissiparous), and facultative, which seasonally alternate between the former modes (Stocchino and Manconi 2013). Asexual reproduction in this genus occurs by transverse fission of the architomic type (i.e. cellular differentiation, organ formation, and body reorganization occur in the new individual after the fission process). Due to the existence

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13127-020-00438-z) contains supplementary material, which is available to authorized users.

✉ Marta Riutort
mriutort@ub.edu

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Catalonia, Spain

[2] Dipartimento di Medicina Veterinaria, University of Sassari, Via Vienna 2, I-07100 Sassari, Italy

of the aforementioned fissiparous-reproducing populations that do not develop the copulatory organs, it has not been unusual to group together several different species into one unique species or species group. For instance, *Dugesia gonocephala* s.l. (Dugès 1830) was a hodgepodge where all the sexual, asexual, and nonconclusive forms of European *Dugesia* were put together (Benazzi 1955; Benazzi and Banchetti 1972; Benazzi and Benazzi-Lentati 1976; Benazzi and Deri 1980; De Vries 1984, 1986; Sluys and De Jong 1984). The incorporation of molecular data not only helped to identify fissiparous individuals but also helped to delimit new species and to clarify the phylogenetic relationships within the genus (Álvarez-Presas and Riutort 2014; Lázaro et al. 2009; Leria et al. 2019; Leria et al. 2020; Riutort et al. 2012; Sluys et al. 2013).

Two species once belonging to the *D. gonocephala* s.l. complex, *Dugesia benazzii* Lepori 1951 and *Dugesia hepta* Pala et al. 1981, are the main focus of the present study. The former inhabits the islands of Corsica, Sardinia, and Capraia (Benazzi and Benazzi-Lentati 1976; De Vries 1985), whereas the latter is endemic of Sardinia restricted to four fluvial basins in the northern region, where both species occur in sympatry. *Dugesia hepta* differs from other Sardinian *Dugesia* species in its haploid chromosomal number ($n = 7$), which is unique in the Western Palearctic region (Stocchino et al. 2005), while in contrast, *D. benazzii* presents the most common haploid chromosomal number among *Dugesia* species, $n = 8$ (Fig. 1). Moreover, *D. benazzii* is known for comprising diploid ($2n = 16$), triploid ($2n = 3x = 24$), tetraploid ($2n = 4x = 32$), hexaploid ($2n = 6x = 48$), and aneuploid (with a mean chromosomal number of 32) populations (Lepori 1951; Pala et al. 1982, 1999; Stocchino 2018). Concurrently, *D. benazzii* characteristically presents sexual and mixed (coexistence of sexual and fissiparous individuals) natural populations, whereas for *D. hepta*, there are no reports on asexual and/or polyploid individuals (Stocchino and Manconi 2013). At first, both species were considered to be identical and indistinguishable except for their karyotypes (Pala et al. 1981), yet in a later study (Stocchino et al. 2005), differential morphological features were described in the copulatory apparatus and in the external morphology. Molecularly, *D. benazzii* and *D. hepta* have had a meager presence in the current molecular phylogenetic era, being represented in all cases only by a few individuals (Lázaro et al. 2009; Solà et al. 2013). From those studies, we learnt that they are sister groups and closely related to the *Dugesia* species from the Western Palearctic Region.

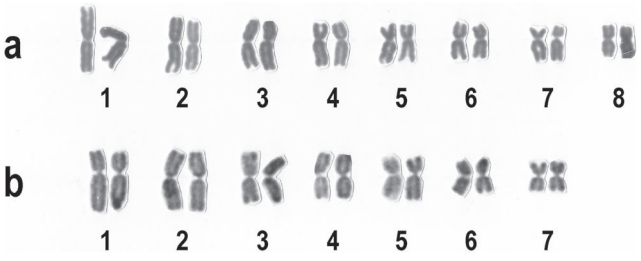


Fig. 1 Karyogram with chromosome complements arranged in pairs of *Dugesia benazzii* (**a**) and *Dugesia hepta* (**b**)

These two sister species externally nearly identical and biogeographically sympatric pose an interesting case study. Since no thorough molecular study with a broad taxon sampling centered on these species have been carried out before, and adding to the fact that *D. hepta* is restricted to only four fluvial basins where co-occurs with *D. benazzii* individuals, one may wonder whether *D. hepta* could have had multiple origins or if it is really a monophyletic species. In the first scenario, *D. hepta* would be the result of a recurrent chromosomal disorder from *D. benazzii* specimens and, thus, should not be considered as a species per se. In the second scenario, we could face a possible case of biogeographical sympatric speciation due to a chromosomal rearrangement, with a concomitant parallel dispersion over the same fluvial basins. Speciation due to chromosomal rearrangements has been proposed to take place in other planarian genera (Benazzi 1982; Leria et al. 2018) as well as in several turbellarian groups (Curini-Galletti et al. 1985; Galleni and Puccinelli 1986). On the other hand, regardless of the taxonomical status of *D. hepta*, given its external morphological similarity to *D. benazzii* and their spatial distribution, one could wonder whether they are able to intercross and if they do it naturally. In fact, there are reports on an aneuploid ($2n = ?x = 32$) population of *D. benazzii*—reproducing by gynogenesis and referred to as the "biotype G" (Benazzi 1949)—located in Rio Bunnari (where the two species co-occur) that was considered to be a stabilized natural hybrid population (Pala et al. 1982) based on karyological data, yet further research rejected that hypothesis (Benazzi-Lentati and Benazzi 1985).

Hence, *D. hepta* and *D. benazzii* pose a case in which first, sympatric speciation may have occurred as the consequence of a chromosomal reorganization, and in the second place, posterior hybridization may have occurred. In the case of plants, approximately 25% of flowering species are considered to be involved in natural hybridization and introgression (Mallet 2007), and the prevalence of this processes has been demonstrated to facilitate speciation and adaptive radiation (Mallet 2007; Pennisi 2016). In animals, hybridization is less frequent, affecting only 10% of species (Arnold et al. 2008; Mallet et al. 2016). Moreover, the existence of this reticulate evolutionary process leads to difficulties in phylogenetic inferences and species delimitation studies, challenging for instance the concept of species grounded on the evolutionary independence of lineages. Hence, studying the patterns and processes of reticulate evolution is not only important to understand evolutionary processes generating new species but may also help to resolve the relationship among closely related taxa affected by them.

In the present study, we aim to answer the following questions using molecular tools: (i) is *D. hepta* a monophyletic distinct species from *D. benazzii* or the result of a recurrent chromosomal disorder from the last species? (ii) Do these two species (or taxonomic entities) hybridize? To accomplish our goals, we have performed a broad sampling of both species from Sardinia; we make use of karyological data to identify some individuals and molecular data to infer the phylogeny of the Sardinian populations of *D. hepta* and *D. benazzii* as well as to reconstruct a haplotype network to assess the existence of hybridism. The results obtained show that *D. hepta* and *D. benazzii* are indeed two different species. Interestingly, we also uncover a complex genetic scenario for their relationships compatible with the existence of ancestral and recent hybridization events between them.

Material and methods

Sampling

Dugesia benazzii and *Dugesia hepta* specimens were sampled from 32 localities distributed on the islands of Corsica and Sardinia (Fig. 2; Table S1) between 1997 and 2010. For each locality, some specimens were fixed and preserved in absolute ethanol for molecular analysis. Others were kept alive and taken to the laboratory to obtain karyotypes.

Assignment of individuals to species

Diagnostic differences between the two species rely on their karyotypes and differences in the copulatory apparatus. Aguilar (2011) found the external morphological differences proposed by Stocchino et al. (2005) to be misleading due to the fact that, in some populations, these differences were not always clearly recognizable but, in contrast, it was envisaged that *D. hepta* and *D. benazzii* differed in 6 divergent sites in their *Internal Transcribed Spacer-1* (*ITS-1*) sequences of the ribosomal cluster (Aguilar 2011) (Table 1). A situation that may allow the assignment of individuals from sympatric populations without a tedious work of karyotyping and/or obtaining histological sections from all the individuals. In the present work in the first place, we checked whether this correlation was univocally true by karyotyping 31 individuals from different populations to assign them to a species based on their chromosomal number. Once the karyological assignation was established, *ITS-1* sequences were obtained. Our newly obtained data corroborated Aguilar's (2011) hypothesis. Thereafter, *ITS-1* sequences were obtained to assign individuals to species.

Karyotyping

Karyotypes were obtained for 31 animals from four populations where *D. hepta* and *D. benazzii* were known to coexist (Table S1) in order to assign them to species and corroborate the *ITS-1* criterion. Chromosome metaphasic plates were obtained by the squashing method. Regenerative blastemas of caudal fragments were treated with a solution of colchicine (0.3%) for 4 h then transferred onto glass slides and treated with a solution of acetic acid (5%) for 5 min. Subsequently, they were stained with acetic orcein for 2 h and squashed using a small coverslip (cf. Stocchino et al. 2019).

The criterion to assign the individuals to species based on karyological data was the following: sexual specimens were assigned to *D. hepta* when the karyotype was $2n = 14$ and any other case with a chromosomal number equal or higher than $2n = 16$ was considered a *D. benazzii* individual. This criterion was used because *D. hepta* is only described from diploid sexual populations, while *D. benazzii* is known to present diploid, triploid, tetraploid, hexaploid, and also different types of aneuploids biotypes (Lepori 1951; Pala et al. 1982). We compared our karyological results with our molecular phylogenetic results to ascertain the usefulness of the *ITS-1* criterion as well as the correct species assignation.

DNA extraction, quantification, and sequence amplification

Total genomic DNA extraction was performed for 161 individuals using DNAzol Reagent (Molecular Research Center Inc., Cincinnati, OH) and Wizard Genomics DNA Purification Kit (Promega Corporation) following the manufacturer's instructions. DNA quantification was performed for each sample using a spectrophotometer NanoDrop™ 1000 (Thermo Fisher Scientific Inc.) using 2 μl per sample.

Specific primers were used to amplify a fragment of the mitochondrial gene *Cytochrome c oxidase I* (*Cox1*), the *ITS-1*, and the *transmembrane p24 trafficking protein 9* (*TMED9*; referred in the present study as *Dunuc12*). Sequences and annealing temperatures of each pair of primers are given in Table 2. Final PCR reaction volume for all markers was 25 μl, consisting of (1) 5 μl of Promega 5× Green GoTaq Flexi Buffer, (2) 2 μl of $MgCl_2$ (25 mM), (3) 1 μl of dNTP (0.5 mM), (4) 0.5 μl of each primer (either 10 or 25 μM), (5) 0.15 μl of *Taq* polymerase (5 u/μl) (GoTaq Flexi DNA Polymerase of Promega), and (6) 1 μl of genomic DNA sample (50 ng/μl). Autoclaved Milli-Q water was added to obtain the final PCR volume. It was necessary in many cases to vary the annealing temperatures or the amount of $MgCl_2$ and/or DNA in order to achieve sequence amplification. The resulting PCR products were visualized in a 1% agarose gel in order to verify the correct amplification of the different molecular markers.

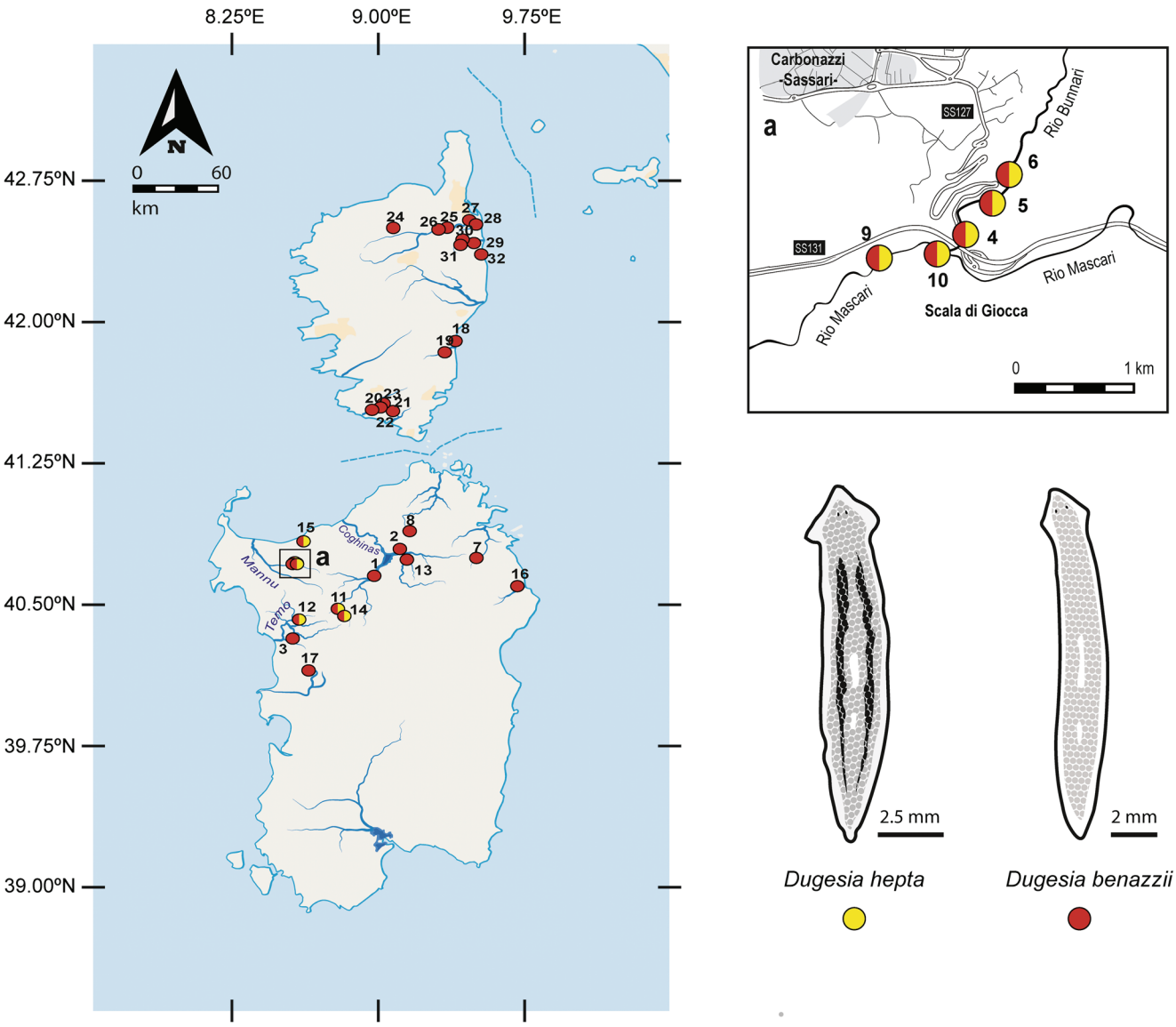


Fig. 2 Sampling localities used for this study and geographical distribution of *D. hepta* and *D. benazzii*. Localities are numbered according to Supporting information Table S1. Red-colored circles indicate the presence of *D. benazzii* species, whereas yellow-colored circles indicate *D. hepta*'s presence. **a** Zoom in of Scala di Giocca

Viable-checked PCR products were purified before sequencing using a vacuum system (MultiScreen™ HTS Vacuum Manifold of Millipore) or a digestion with exonuclease I (0.2 u/µl; Tebu-Bio) and shrimp alkaline phosphatase (0.2 u/µl; SAP, Sigma-Aldrich) (1:2 ExoSAP per PCR product at 37 °C for 15 min plus an additional phase at 80 °C for 15 min). Sequencing reactions were performed using Big Dye (3.1, Applied Biosystems) and ran in an automated sequencer ABI Prism 3730 (Unitat de Genòmica dels Serveis Científico-Tècnics de la Universitat de Barcelona) or at Macrogen Inc. (Amsterdam). The same primers used to amplify were used for sequencing both strands. Chromatograms were visually checked for quality with Geneious R8 (Biomatters, http://www.geneious.com/ last visited June 2019) and then contig and consensus sequences were obtained.

Since some individuals presented double-band patterns in their *Dunuc12* chromatograms, we decided to clone their PCR products to inspect their origin; we also cloned some animals

Table 1 Polymorphic diagnostic site positions of *D. benazzii* and *D. hepta* in the *ITS-1* alignment and GenBank accession numbers for the type haplotypes

Species	Alignment position						
	12	24	49	223	372	458	Acc. number
D. benazzii	C	–	G	C	A	C	MN162692
D. hepta	T	T	A	T	G	T	MN162693

Table 2 Sets of primers used in this study

Gene	Primers	Sequence (5′-3′)	Ann. Temp.[a]	Source
Dunuc12	Jon_12F	GATTACGAAAGCTATTTATAATT	52	Present study
	Snow_12R	CATGCACAAGATTACAAAG	52	Present study
	Elo_12F	AAGCTATTTATAATTCAGCG	54	Present study
	Hell_12R	AAAGAAATTGCTGCTAAAG	54	Present study
	Dunuc12_1F	CTCGTATCTCTGAATCTAGCCTC	55	Leria et al. 2019
	Dunuc12_1R	GTTCATACAACTCATTCTTC	55	Leria et al. 2019
Cox1	SamCF	GCTAATAATTTGAGTTTTTG	51	Present study
	TarlyCR	CATTTTAAAACAACATTACC	51	Present study
	COIF	CCNGGDTTTGGDATDRTWTCWCA	49	Lázaro et al. 2009
	COIR	CCWGTYARMCCHCCWAYAGTAAA	49	Lázaro et al. 2009
ITS-1	ITS-9F	GTAGGTGAACCTGCGGAAGG	45	Baguñà et al. 1999
	ITSR	TGCGTTCAAATTGTCAATGATC	45	Baguñà et al. 1999

[a] Annealing temperature (C°)

that presented a mito-nuclear discordance or other peculiar molecular features (see results), and animals not presenting such features as a control. In total, 17 individuals (4 controls and the rest presenting double bands, mito-nuclear discordances, or other peculiarities) were cloned using a TOPO® TA Cloning Kit (Thermo Fisher Scientific Inc.) following the manufacturer's instructions. Approximately fifteen to thirty colonies from each individual were sequenced using the T3 and T7 primers (included in the kit).

Phylogenetic analysis

Alignments of the sequences were performed with the online software MAFFT v.7 (Katoh and Standley 2013) and posteriorly revised in BioEdit v.7.2.5 (Hall 1999). Prior to analyses, *Cox1* and *Dunuc12* sequences were translated into amino acids to verify that there were no stop codons within coding regions. Three alignments were obtained: (1) *ITS-1*, used to perform the species assignment; (2) *Cox1* (dataset I) and (3) *Dunuc12* (dataset II), which were used to perform the subsequent phylogenetic analyses. *Cox1* and *Dunuc12* sequences of four specimens belonging to three other *Dugesia* species were used as outgroup for the phylogenetic inferences (Supporting information Table S1).

Levels of sequence saturation were assessed by means of the Xia et al. (2003) test implemented in the software DAMBE (Xia and Xie 2001). The best substitution model was selected with jModelTest (Posada 2008) based on the Akaike information criterion (AIC). In both cases the best fitting model resulted in a HKY + G. Phylogenetic analyses were performed using two inference methods: Maximum Likelihood (ML) and Bayesian inference (BI). ML analyses were performed with RaxML 7.0.3 (Stamatakis 2006), applying a GTR + G substitution model—owing to the absence of the HKY model in the aforementioned RaxML version. A total of 5000 replicates were calculated to obtain bootstrap supports (bs) conducting a rapid bootstrap analysis, and the ML search was performed starting from a random tree. Furthermore, we applied an optimization of both branches and model parameters on bootstrapped trees. BI analyses were conducted using MrBayes v.3.2 (Ronquist et al. 2012) and applying a HKY + G substitution model. Prior to run the analyses, nexus files were generated with MEGA6 (Tamura et al. 2013). We ran one cold and three heated chains for two parallel runs. Both topological and model parametrization convergence were surveyed by checking that the standard deviation of the split frequencies reached a value below 0.01. A total of 10,000,000 generations were performed for each gene, saving a tree every 5000 generations. We applied the default burn-in—set at 25%—in order to avoid the inclusion of trees obtained before likelihood values had stabilized to infer the topology and the posterior probabilities (pp). Obtained trees were visualized with FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/last visited June 2019).

Haplotype networks

In order to construct the haplotype networks, we generated two additional alignments only comprising specimens from Sardinia: (1) comprising haplotypes from the Cox1 sequences (dataset III) and (2) consisting of cloned Dunuc12 sequences from 17 individuals (dataset IV). In the latter, we included Sardinian individuals that presented chromatograms either with double-bands or patterns indicating possible heterozygosity for indels, which had aroused our interest (individuals MR0092-05, MR0092-11, MR0025-02, MR0091-01, MR0092-04, and MR0030-06). We also included individuals belonging to the COI mixed clade (see results) that did not

present polymorphic bands (MR0088-02 and MR0092-03) as well as two individuals whose karyotype was known yet formed part of the mixed clade (MR0353-01 and MR0352-01). Furthermore, individuals MR0022-04, MR0022-05, and MR0172-01 were included as they showed signs of having mito-nuclear discordance. In addition, MR0022-04 and MR0172-01 had double-band patterns too. Finally, four individuals previously identified as *D. hepta* (MR0354-01 and MR0355-01) and *D. benazzii* (MR0368-01 and MR0370-01) based on karyological data were also analyzed as controls.

As for the *Cox1* haplotype network (dataset III), all individuals with polymorphic positions in dataset I were excluded from the alignment. Sequence ends were trimmed using BioEdit v.7.2.5 to avoid overestimating the number of underlying haplotypes due to the terminal missing data. DnaSP v.5 (Librado and Rozas 2009) was used to determine and assign the haplotypes for each individual analyzed and Network v.5.0.0.0 (Fluxus Technology Ltd.) was used to construct the haplotype networks. We used a default *epsilon* value set at zero.

Genetic diversity

Levels of nucleotide and haplotype diversity were calculated using DnaSP v.5. for each Cox1 haplogroup. We also calculated a Cox1 distance matrix using Kimura 2P evolutionary model for all the individuals included in the study.

Figures edition

The maps in figures were created with Q-GIS v.3.2.2 (https://qgis.org/es/site/ last visited June 2019). Figures 2, 3, 4, 5, and 6 were edited with Illustrator CC v.22.0.1 (https://www.adobe.com/products/illustrator.html last visited June 2019).

Results

ITS-1 as diagnostic marker

The karyological analysis of the 31 individuals from localities 4, 5, 7, 9, and 14 revealed that nine of them matched to *D. benazzii* and 22 of them to *D. hepta*. *ITS-1* sequences could be recovered from only 23 of the aforementioned individuals. In 19 cases, the six species-specific sites proposed by Aguilar (2011) revealed the individuals to belong to *D. hepta* and four cases to *D. benazzii*, coinciding with the expected for the karyologically identified species. These results confirm the *ITS-1*'s validity as a marker for molecular diagnosis of the two species. Henceforth, we refer to it as the *ITS-1* criterion in the present manuscript.

We attained an overall number of 146 *ITS-1* sequences; none presented double peaks in the chromatograms. The resulting alignment had a length of 496 bp—base pairs. The six species-specific sites were highlighted as key elements for the species diagnose. An overall number of 91 individuals were identified as *D. benazzii*, 55 as *D. hepta*, and 15 remained unidentified since our attempts to amplify and sequence their respective *ITS-1* were unsuccessful (*Dugesia* sp. in the Supporting information Table S1).

Fig. 3 Bayesian Inference tree of dataset I (*Cox1*). Node values are displayed qualitatively using squares for posterior probability (pp) and circles for bootstrap support (bs) values. Used colors indicate fully supported (black), significantly supported (gray), and nonsupported (white) nodes. Locality numbers from Table S1 are highlighted in black hexagons. Sampling localities are displayed in the map as follows: in Sardinia—grouped according to hydrographical distribution, and in Corsica—grouped into northern and southern geographical regions as showed in the tree

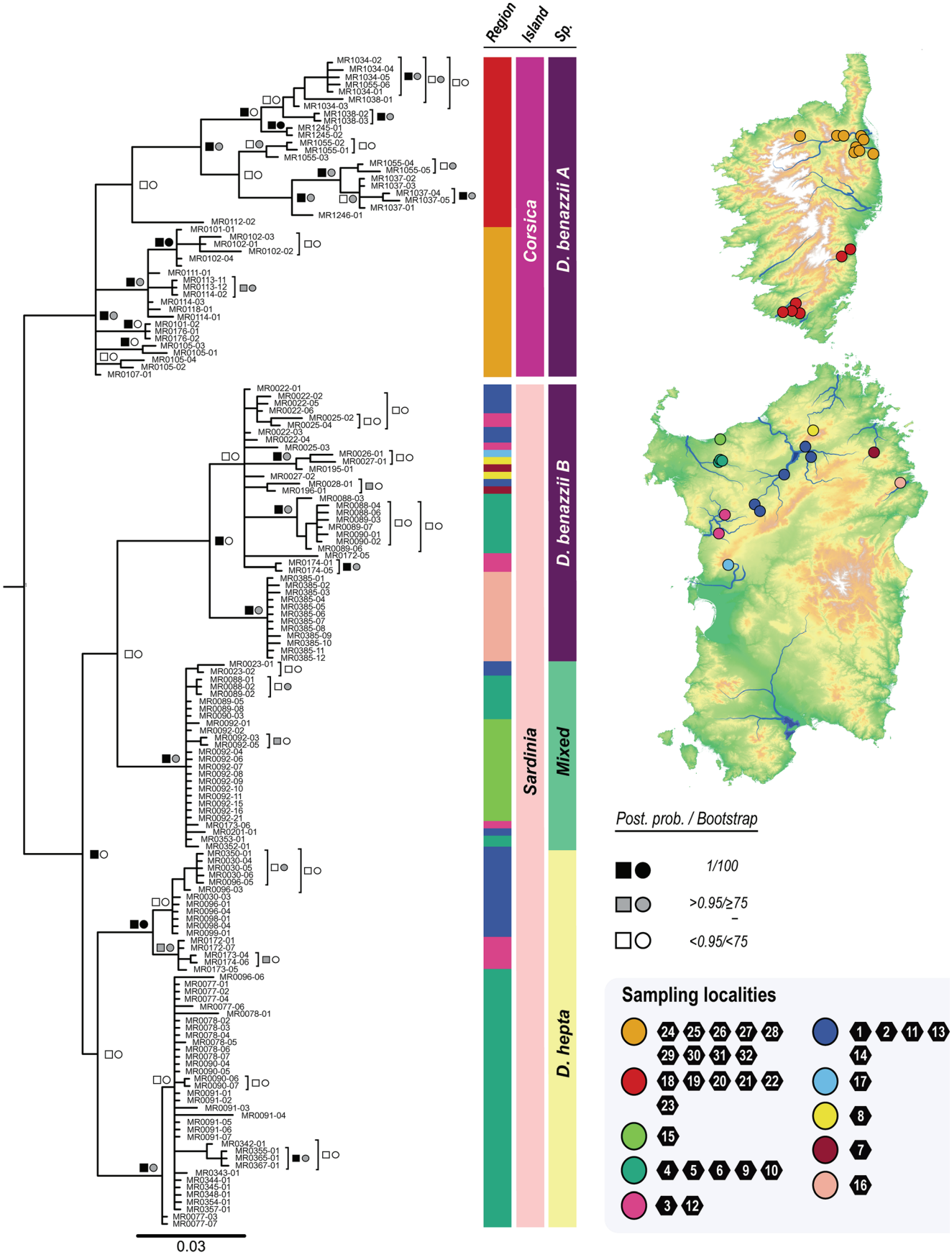
Dataset characteristics

We set four alignments to be analyzed either to estimate phylogenies (dataset I and II) or to construct the haplotype networks (dataset III and IV). Dataset I was comprised of 163 *Cox1* sequences (706 bp, 44 from Corsica, 115 from Sardinia, and 4 outgroup sequences) and dataset II was comprised of 102 *Dunuc12* sequences (642 bp total, 81 exonic bp, 2 from Corsica, 96 from Sardinia, and 4 outgroup sequences). Dataset III was constructed with nonpolymorphic *Cox1* sequences (685 bp, 53 sequences representing the 43.1% of the original Sardinian alignment)—from which 19 haplotypes were identified (Supporting information Table S2). Lastly, dataset IV consisted of cloned *Dunuc12* sequences (530 bp, 55 exonic bp, 259 sequences) from 17 individuals, from which a total of 176 haplotypes were recovered (Supporting information Table S3) with a mean of 10.0 ± 4.83 different haplotypes per individual. All new sequences have been deposited in GenBank (accession numbers available in Supporting information Tables). The tests revealed no significant saturation signals from any of the alignments.

Phylogenetic analyses

The phylogenetic inferences carried out by means of both Bayesian Inference (BI) and maximum likelihood (ML) yielded no topological incongruities.

Based on the mitochondrial marker (dataset I), the BI tree showed four major clades (Fig. 3): (1) *D. benazzii* specimens from Corsica (*D. benazzii* A), (2) Sardinian *D. benazzii* specimens (*D. benazzii* B), (3) *D. hepta* individuals belonging to Sardinia (*D. hepta*), and (4) an unexpected, apparently mixed clade comprised of 13 individuals identified via *ITS-1* as *D. benazzii*, 5 individuals identified as *D. hepta*, and 8 *Dugesia* sp. individuals. Most of the clades were well-


Region
Island
Sp.
Corsica
Sardinia
D. benazzii A
D. benazzii B
Mixed
D. hepta
0.03
Post. prob. / Bootstrap
1/100
>0.95/≥75
<0.95/<75
Sampling localities
24 25 26 27 28 29 30 31 32
18 19 20 21 22 23
15
4 5 6 9 10
3 12
1 2 11 13 14
17
8
7
16

supported as indicated by posterior probability (pp ≥ 0.95) and *bootstrap* (bs ≥ 75) values—with the exception of the *D. hepta* clade, yet the resolution within clades was scarce. *Dugesia benazzii* did not result in a monophyletic group but a paraphyletic one, being the Corsican clade (A), the sister group to a clade constituted by *D. benazzii* from Sardinia (B), *D. hepta*, and the mixed clade. However, the phylogenetic relationship among these last three clades was unclear due to low support values.

The topology yielded by the nuclear marker (dataset II; Fig. 4) differed from the one obtained with *Cox1*. First of all, the populations of *D. benazzii* from Corsica and Sardinia clustered as a monophyletic clade. Likewise, the main clades—namely one for *D. benazzii* and one for *D. hepta*—were unambiguously supported as indicated by pp and bs values, albeit the resolution within the groups was again poor. Secondly, there were no traces of the aforementioned mixed clade. Instead, the individuals belonging to the mitochondrial mixed clade were integrated within either the *D. hepta* or the *D. benazzii* group, in concordance with the *ITS-1* criterion. However, two individuals (MR0022-04 and MR0022-05) showed signs of mito-nuclear discordance given that their *Dunuc12* sequences belonged to the *D. hepta* clade—contrary to the *ITS-1* identification—while the

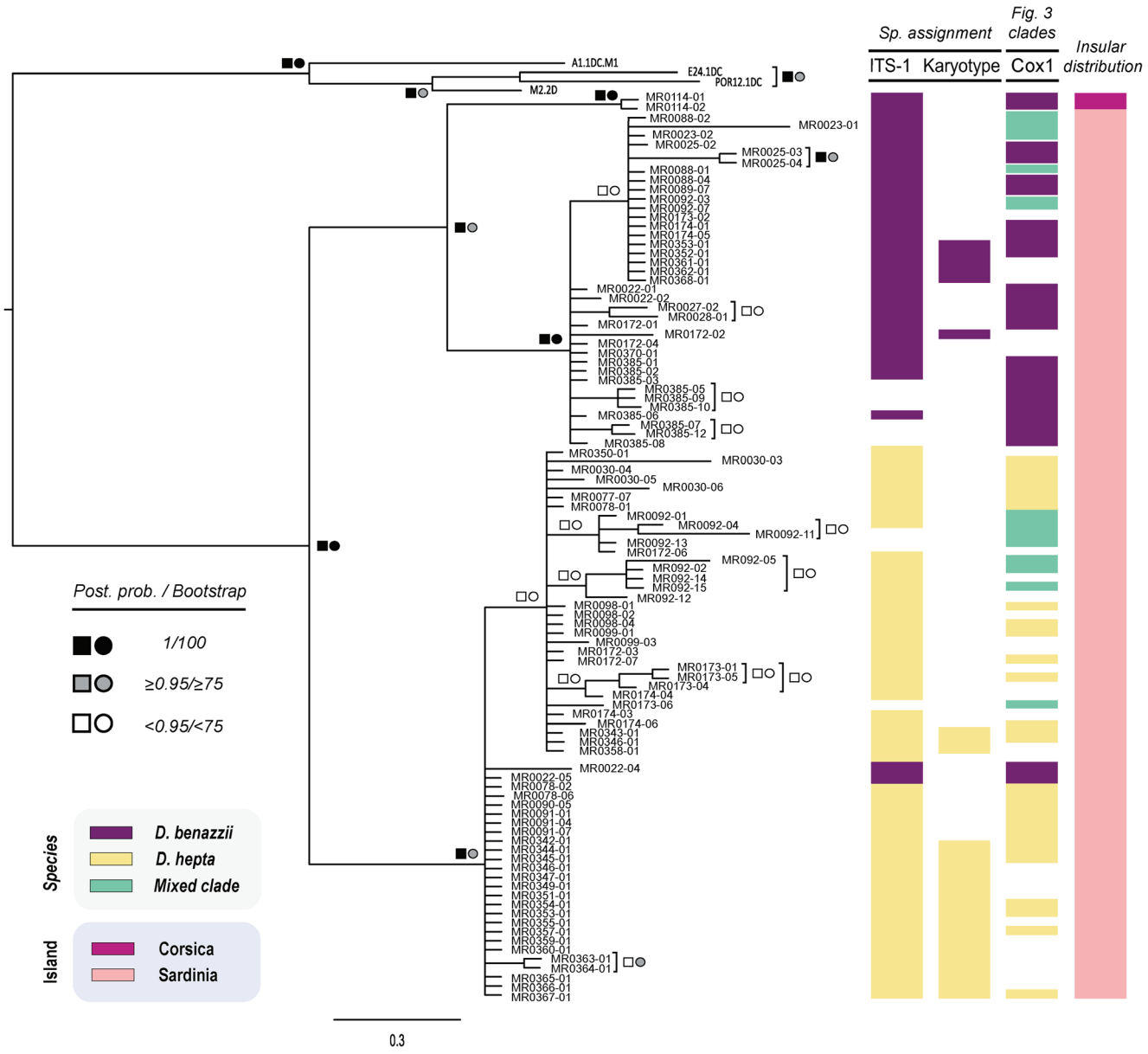


Fig. 4 Bayesian inference tree of dataset II (*Dunuc12*). Node support values are displayed as in Fig. 3. Bar diagrams indicate the species assignation of samples based on the *ITS-1* criterion and karyology. It is also showed the phylogenetic position in dataset I's tree and the insular distribution (fuchsia—Corsica and light pink—Sardinia)

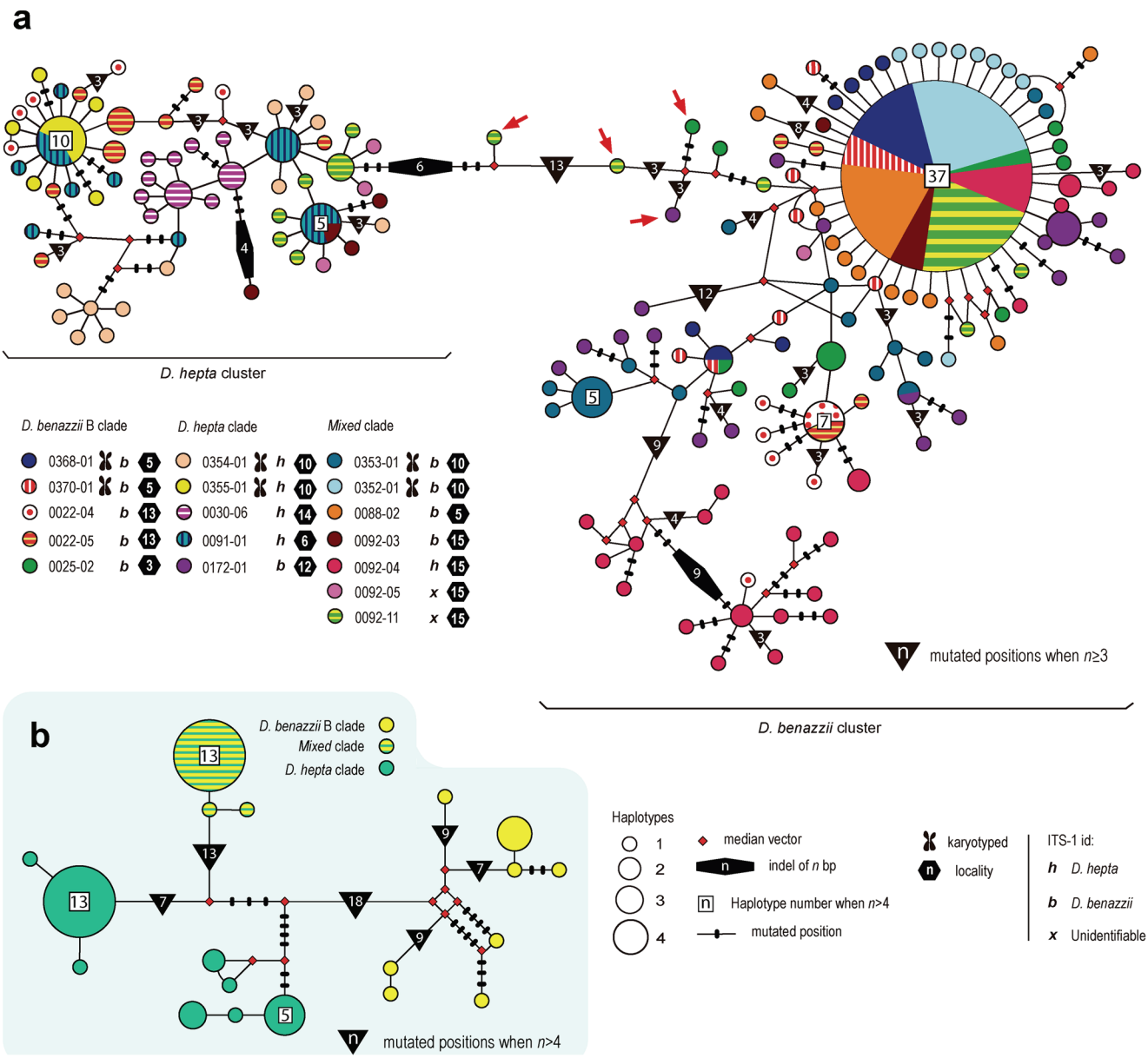


Fig. 5 Haplotype networks for datasets IV (**a** *Dunuc12*) and III (**b** *Cox1*). Haplotypes are depicted as individual circles which are proportional to their abundancy (number of sequences), highlighted in a white square. Mutations are either depicted with black bars or black triangles when the number of mutations between linked haplotypes is equal to or exceeds a certain threshold number. Insertions and deletions are represented with an elongated hexagon indicating numerically the length of the *indel*. **a** For each cloned individual, information regarding *ITS-1* species identification, sampling locality, and availability of karyotype is given. Recombinant haplotypes are highlighted with red arrows

mitochondrial sequences placed them unequivocally within the *D. benazzii* B clade.

Haplotype networks

As regards to the *Dunuc12* (dataset IV) cloned haplotype network (Fig. 5a), we found two distinct haplotype clusters separated by at least 24 substitutions and a 6-nucleotide indel. The individuals karyologically identified as *D. hepta* (MR0354-01 and MR0355-01) had all their haplotypes assigned to one cluster, while those karyologically identified as *D. benazzii* (MR0368-01 and MR0370-01—from *D. benazzii* B clade) had their haplotypes in the other cluster, thus indicating that each cluster could be matched to a different species. Although some of the haplotypes sequenced may have been artificially generated by errors of the polymerase, the high differentiation between the two species' clusters ensures that their differentiation is real so that we can assign haplotypes from individuals as belonging to one species or the other depending on the cluster they belong to. Out of all the individuals assigned to the species *D. hepta* based solely on the *ITS-1* criterion, MR0030-06 and MR0091-01 had all

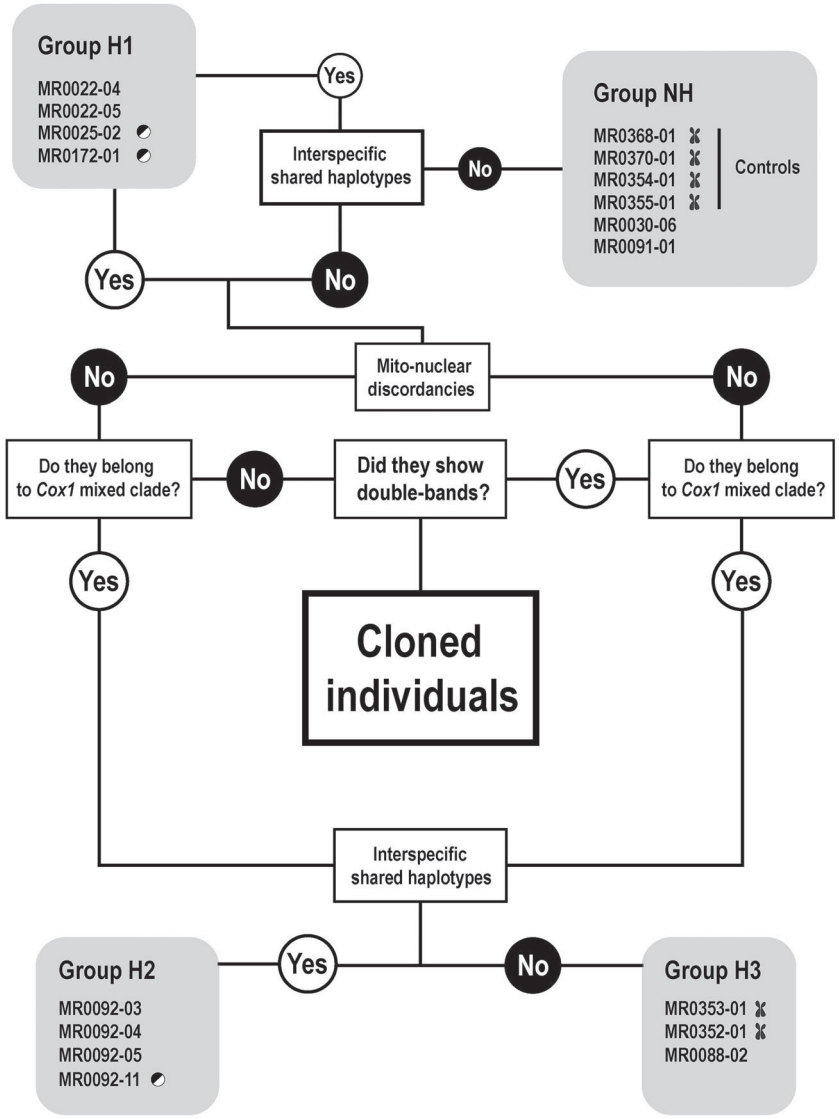


Fig. 6 Summarization of dataset IV groups and their characteristics. Individuals whose karyotype is stablished are highlighted with bivalent chromosome symbol. Black and white split-up circles indicate individuals with recombinant sequences

their haplotypes unequivocally assigned to the *D. hepta* cluster, in accordance also with their mitochondrial data. In contrast, MR0092-04 had all its haplotypes associated to the *D. benazzii* cluster despite being part of the mitochondrial mixed clade. On the other hand, the individuals assigned to *D. benazzii* based on the *ITS-1* criterion showed more disparate results. Out of the specimens from mitochondrial *D. benazzii* B clade, MR0022-04 and MR0022-05 presented haplotypes in both clusters, while MR0025-02 had all its haplotypes in the *D. benazzii* cluster. Individual MR0172-01 from mitochondrial *D. hepta* clade presented only *benazzii* haplotypes, coinciding with its *ITS-1* assignment. As regards to the individuals belonging to the mitochondrial mixed clade, individual MR0088-02 had all its haplotypes associated to the *benazzii* cluster, while MR0092-03 showed haplotypes in both clusters. At the same time, karyotyped individuals MR0353-01 and MR0352-01 showed only *benazzii* haplotypes. Lastly, individuals MR0092-05 and MR0092-11 that were not possible to identify based on *ITS-1* had haplotypes assigned to both clusters. It should be pointed out that intercluster recombinant haplotype variants were found in three individuals—MR0025-02, MR0092-11, and MR0172-01. In summary (Fig. 6), the clonal analyses of the nuclear gene of different "anomalous" individuals have resulted in the finding of four groups of individuals: some that are "pure" *D. hepta* or *D. benazzii* and probably presented double bands in their sequences due to their heterozygosity (Fig. 6, group NH); a second group that presents mitochondrial haplotypes either from *hepta* or *benazzii*, but in whose nucleus we can find haplotypes from both species or some recombinant (group H1); a third group presenting the mixed clade mitochondrial haplotypes and also presenting in the nucleus haplotypes from both species (group H2); and finally a group presenting the mixed clade mitochondrial haplotype but only *benazzii* nuclear haplotypes (group H3).

As for the *Cox1* haplotype network (Fig. 5b), the same three Sardinian clades observed in the tree (Fig. 3) were recovered. Curiously enough, in this case, the mixed clade derives from within the *hepta* cluster, separated by 13 substitutions, instead of being closer to the *benazzii* cluster as shown in the phylogenetic tree (a relationship that receives a low support, Fig. 3). The *benazzii* clade is separated from the *hepta* clade by 18 substitutions. Hence, the three groups are well separated, but in fact, the internal differentiation within the *hepta* and the *benazzii* clades is also quite high.

Genetic diversity

The *Cox1* distances estimated with Kimura 2P are shown in Table 3.

We also calculated nucleotide (π) and haplotype diversity (H_D) within each of the three *Cox1* haplogroups (Table 4). The values of H_D were high for the *hepta* and *benazzii* haplogroups in contrast to a low value for the mixed group. For π again, *hepta* and *benazzii* presented higher values than the mixed clade; however its values were also moderately low.

Table 4 Estimations of nucleotide (π) and haplotype diversity (H_D) for the haplotype groups in dataset III (*Cox1*)

Group	Sequences	Haplotype num.	H_D	π
Hepta	$n = 27$	8	0.738 ± 0.005	0.0131
Benazzii	$n = 11$	8	0.891 ± 0.008	0.0178
Mixed	$n = 15$	3	0.257 ± 0.020	0.0006

Discussion

Outstanding intraindividual haplotype diversity: artifact or commonality?

We cloned the PCR products of the nuclear gene *Dunuc12* to unveil the haplotypes of the putative heterozygous individuals of *D. hepta* and *D. benazzii*. Surprisingly, we found a mean of 10 different haplotypes per individual. Some of these haplotypes may have resulted from polymerase errors during the PCR step before the cloning procedure; however, our results exceed whatever expectations of errors caused by the malfunction of the polymerase. Even though the recovered number of haplotypes was surprisingly high, similar results have been previously found in *D. japonica* (Nishimura et al. 2015) and in a closely related group of species—*D. subtentaculata*, *D. aurea*, *D. corbata*, and *D. vilafarrei* (Leria et al. 2019; . Leria and coworkers demonstrated in said study the presence of mosaicism due to the accumulation of mutations in planarian stem cells during homeostatic and fission processes, resulting in multiple closely related haplotypes in sexual animals (star-like patterns) and in multiple distantly related haplotypes in fissiparous animals (divergent pattern). Thus, we have reasons to believe that this might be characteristic of *Dugesia* species—perhaps even of other dugesid genera—especially when asexuality is involved as a reproductive strategy in the species.

Dugesia benazzii species status

The clues provided by the phylogenetic *Cox1* tree obtained in this study (Fig. 3) point out that *D. benazzii* could constitute more than one species. We base this suggestion on the paraphyletic arrangement of the *D. benazzii* clades in the tree and especially on the high genetic differentiation among them. The genetic differentiation between the populations from Corsica and Sardinia for COI has a mean value of 5.8 ± 0.8%, far superior to the ones found between populations within each of the islands (Table 3). Moreover, the values found between islands concur with some of the interspecific genetic distance values for *Dugesia* species from the Western and Eastern Mediterranean (Lázaro et al. 2009; Solà et al. 2013) that vary between 2.8% for closely related species in the Aegean region and 11% for some species on the Western region. On the other hand, the phylogenetic tree obtained from the nuclear marker (*Dunuc12*, Fig. 4) showed a monophyletic *D. benazzii* clade, yet the two individuals from Corsica appeared again to be highly differentiated from the Sardinian populations. Nonetheless, the fact that only two sequences for the aforementioned nuclear marker of Corsican individuals were used in this study gives us little information regarding genetic diversity of the populations of the Corsican *D. benazzii*. On the other hand, populations of *D. benazzii* from both islands are identical regarding the *ITS-1*, which will

Table 3 Distance matrix of mean K2P distances (%) with standard deviations for the *Cox1* marker. Number of sequences is given inside the parentheses

	D. benazzii A	*D. benazzii B*	*Mixed clade*	*D. hepta*
D. benazzii A (44)	2.9 ± 1.4			
D. benazzii B (39)	5.8 ± 0.8	1.6 ± 0.9		
Mixed clade (26)	5.2 ± 0.8	4.2 ± 0.4	0.1 ± 0.1	
D. hepta (52)	4.9 ± 0.7	4.0 ± 0.5	3.2 ± 0.1	1.3 ± 1.2

support the monophyly of *D. benazzii* but not its division in more than one species, unless this is a very recent event and ribosomal clusters are still being kept similar by concerted evolution. As for the morphology, the original description of the species given by Lepori (1951) did not establish any remarkable differences between Corsican and Sardinian *D. benazzii* populations regarding the copulatory apparatus, but it did point out some minor dissimilarities that he deemed insufficient neither to consider the Corsican and Sardinian populations as distinct geographical subspecies. However, considering that modern descriptions take into account more characters than in the past, the possibility to find valid differences supporting a specific differentiation cannot be ruled out. To this purpose, a new detailed morphological study has been already undertaken on populations of *D. benazzii* from the two islands that will be the subject of a companion paper.

Hence, there is incongruence between the mitochondrial history and the nuclear and morphological accounts. In some cases, a potentially high degree of genetic variation may only be reflected by recondite morphological traits (according to Kucera and Darling 2002) that are not evident at first sight. Sibling species often have minor morphological differences that are only noticed once species are recognized for other reasons—such as karyological data or molecular evidences. The species that fit this profile are known as pseudo-cryptic (Knowlton 1993), and this may be the case for Corsican and Sardinian *D. benazzii* populations. Nonetheless, speciation is a continuum. Theoretically, the further we stray from the starting point, the clearer and more evident should be the differences between descendant lineages, but in the first stages of speciation, there can be divisiveness among sources of evidence—i.e. genetic data versus morphological data, nuclear versus mitochondrial DNA—because changes do not accumulate uniformly and at a fixed rate. This interval of speciation is known as the "gray zone" (De Queiroz 2007) and could explain why we find differences regarding the *Cox1* sequences between Corsican and Sardinian *D. benazzii* populations but not in the *ITS-1* or in their morphology.

These results point to the need for a revision of the taxonomic status of *D. benazzii*, based on more data ranging from an increase of the number of nuclear markers and the use of molecular methods for species delimitation to a morphological and karyological revision of the individuals. A similar situation has been resolved in a close relative, *D. subtentaculata*, by the concurrent use of all these lines of evidence in an integrative way, resulting in the description of three new species that are morphologically cryptic with *D. subtentaculata* (Leria et al. 2020).

Species status and origin of *D. hepta*

Dugesia hepta is a monophyletic species beyond questioning as ascertained by both phylogenetic trees. We sustained a reasonable doubt regarding its taxonomical status owing to (1) the atypical chromosomal number ($n = 7$) and (2) their geographical distribution—restricted to four fluvial basins and in co-occurrence with *D. benazzii* individuals. However, our results rule out the possibility that *D. hepta* could be an aberrant chromosomal form of *D. benazzii* in which case we might expect to find *D. hepta* as a polyphyletic ensemble appearing in the phylogenetic trees independently from different *D. benazzii* clades. Our phylogenetic trees do not support that hypothetical scenario but rather show that *D. hepta* and *D. benazzii* are two different species that shared a common ancestor.

It is not clear from our results whether *D. hepta* is sister to only the Sardinian *D. benazzii* or to an older lineage that gave rise to the *D. benazzii* group from Sardinia (group B) and Corsica (group A). Nuclear data seem to indicate that *D. hepta* could be in fact the sister group of *D. benazzii* A and B (Fig. 4), yet such relationship is questioned when the mitochondrial data is considered. However, discordances in nodal support depending on the inference method used to build the tree based on *Cox1* data (Fig. 3) let open the possibility that *D. hepta* could be the sister group of *D. benazzii* A and B. It is worth noticing that although both species present similar values of diversity for their *Cox1* sequences (Tables 3 and 4), they differ in how this genetic variation is geographically distributed. The Sardinian populations of *D. benazzii* from the sampled localities showed no remarkable signs of geospatial structure or isolation, with the exception of the samples from Monte Albo (Table S1; Fig. 3 locality 16) that are appreciably genetically isolated from the rest of *D. benazzii* B populations. In contrast, the populations of *D. hepta* appeared to be more structured with no apparent admixture of individuals from different fluvial basins (Fig. 3). A plausible explanation for the differences in mitochondrial structuration degree could reside within the reproductive strategy of each species. *Dugesia hepta* is exclusively sexual, while *D. benazzii* is strategically more flexible having both sexual and sexual-fissiparous mixed populations, which could be advantageous towards rapidly colonizing new fluvial basins, as it has been shown for other *Dugesia* species (Lázaro and Riutort 2013; Leria et al. 2020). This also could explain why *D. benazzii* has a broader geographic distribution. A speculative scenario will be that, *D. hepta*'s ancestral populations may had undergone through a constraint in numbers due to direct competition with other species—possibly *D. benazzii* itself who could have colonized Sardinia from Corsica and displaced *D. hepta*. However, similar results could be expected if recent bottlenecks caused by abiotic phenomena—such as the desiccation of the brooks and springs where they can usually be found—affect *D. hepta* population's survival more than those of *D. benazzii* due to its exclusive sexual way of reproduction. There was no apparent correlation between the genetic lineages and their geographic distribution for the

nuclear gene, for any of the species. This can be a result of the gene analyzed being highly conserved and hence lacking information for recent dispersal events.

Whether *D. hepta* is sister to *D. benazzii* from Sardinia or to the lineage that gave rise to *D. benazzii* group from Sardinia and Corsica, the speciation event may have been related to a chromosomal rearrangement. Bearing in mind that *D. hepta*'s chromosomal number ($n = 7$) is uncommon within the whole Western Palearctic Region and that most species of *Dugesia* commonly share the $n = 8$ chromosomal number, we suggest that the ancestor that gave rise to the lineages leading to the current *D. hepta* and *D. benazzii* species might have shared the same chromosomal number, $n = 8$. Therefore, *D. hepta* could pose a case of speciation due to a chromosomal rearrangement. In most animal and plant groups, there are differences regarding the chromosomal number among closely related species (King 1993). Nonetheless, not all of the changes that may operate on the chromosomes are implicated in speciation phenomena (King 1987) but only those that have potential to diminish the biological efficiency of the hybrids, which are known as negative heterotic (Forsdyke 2004; King 1987, 1992; Rieseberg 2001), or that even impede their viability. Both cases, at shorter or longer term, give rise to reproductive isolation among populations and therefore are likely to cause speciation (White 1978). Chromosomal rearrangement speciation cases have gained presence over time (Coates and Shaw 1984; Kawakami et al. 2011; Talavera et al. 2013). There are other cases in freshwater planarians where a chromosomal rearrangement is suspicious of being the speciation cause, within the dugesiid genus *Schmidtea*. *Schmidtea nova* and *S. lugubris* are two sibling species with haploid chromosomal numbers of $n = 3$ and $n = 4$, respectively. Within the genus, $n = 4$ is the plesiomorphic karyological state. *Schmidtea nova* would have originated from a common ancestor through a Robertsonian translocation plus a pericentric inversion resulting in its three basic chromosomes that would have rapidly isolated reproductively the descendent lineages (Benazzi and Puccinelli 1973; Leria et al. 2018). In *D. hepta*, we also have a reduction of the chromosomal number as well as changes in the chromosomal structure—being the most remarkable a large submetacentric chromosome 1—within a predominantly metacentric set—that could be the by-product of a nonreciprocal translocation that led to the loss or the assimilation of the eighth chromosome. We can conclude that these two species exhibit a great karyological plasticity regarding ploidy and chromosomal composition as it has been previously proposed for other planarian groups (Leria et al. 2018; Leria et al. 2020; Ribas 1990) in comparison with their conservative morphology, and this plasticity may in some cases be related to speciation events. However, does the chromosomal difference between *D. hepta* and *D. benazzii* really impede their intercrossing?

Dangerous liaisons: a complex relationship between *D. hepta* and *D. benazzii*

Pala et al. (1982) had proposed that *D. hepta* and *D. benazzii* might be able to intercross to explain the presence of individuals reproducing gynogenetically bearing a variable number of chromosomes (being the most frequent number 32 but never eutetraploid) in "Scala di Giocca" locality (Rio Bunnari), although it was posteriorly refuted by Benazzi-Lentati and Benazzi (1985) based on karyometric analyses. To try to elucidate whether hybrids exist, we have planned our analyses to detect a classical basic case of hybridization, the detection of nuclear haplotypes from both parent species in the putative hybrids, together with the mitochondrial haplotypes from only one of them. Our results, however, show a much more complex and interesting situation that should be deeply looked into in further researches. We have found evidence that prove the existence of at least three types of "anomalous" individuals collected in the four hydrographic basins in which the two species co-occur (H1 to H3, Fig. 6) that most probably could be the result of hybridizations. Even though it would be tempting to jump into hasty conclusions, we cannot unerringly relate Pala's et al. individuals to our hybrids. The situation is certainly much more complex than we could expect.

H1 individuals could be the result of a recent hybridization. Those individuals have *Cox1* sequences either belonging to the Sardinian *D. benazzii* group—as well as their *ITS-1*—or to *D. hepta* group yet the *Dunuc12* nuclear marker presents haplotypes from both species or presumptive recombinant alleles. These individuals could be the result of a recent hybridization in which *D. benazzii* or *D. hepta* will have acted as a mother so that the hybrids have one or the other mitochondrial DNA. In the nucleus, we will in this case expect to find haplotypes from both parents, which is the case for individuals MR0022-04 and M0022-05, while in the other two individuals, we only find benazzii haplotypes but some presumptive recombinants. This latter case could be a consequence of the hepta *Dunuc12* haplotypes not having been PCR-amplified as efficiently as benazzii *Dunuc12* haplotypes (so a methodological artifact) or else that hybrid individuals have been able to backcross with *D. benazzii* parental species resulting in the loss of the hepta nuclear haplotypes.

For the *ITS-1* sequences (showing *D. benazzii* origin in the four H1 individuals), either a similar situation is found (lack of amplification or backcross to parental species) or else the concerted evolution processes that regularly homogenize the multiple copies of the ribosomal clusters (Dover 1982; Hillis and Dixon 1991) may have resulted in the original *D. benazzii* cluster having overruled the *D. hepta* cluster. Subsequently, the validity of the *ITS-1* criterion as a highly reliable method to identify the species that Aguilar (2011) originally proposed is questioned. It will work for the parental species, but it will

certainly fail to determine hybrids unless it is cloned. Since a hybrid will have the genomes of both parental species, theoretically one could be amplifying the *ITS-1* of any of the two parental genomes.

The H2 group individuals also present nuclear *Dunuc12* haplotypes from both species (or a recombinant), hence likely being of hybrid origin. However, what makes these hybrids special is that they bear the mixed clade *Cox1* haplotypes that appear as a monophyletic clade in the *Cox1* tree, completely independent from the *D. hepta* and *D. benazzii* clades. Therefore, we have individuals bearing in their nucleus haplotypes coming from both species, while their mitochondrial genome seems to have differentiated from both parentals, showing a closer relationship to the hepta haplogroup from which most probably derived (Fig. 5b). Moreover, to make the picture more complex, within this mixed clade we also find the group H3 presenting the mixed clade *Cox1* haplotypes but only benazzii nuclear sequences.

Many of the members in the mixed clade (H2-H3 individuals) belong to the same river where Pala et al. (1982) found and described the putative stabilized 32-aneuploid hybrids (Bunnari). Yet the 32-aneuploid can also be found in Rio Silis (locality 15) where a karyological study carried out on the *D. gonocephala* s.l. planarian populations by Vacca et al. (1988) discovered another anomalous karyotype of constantly 22 chromosomes with low frequency (described in 10 individuals out of the 95 studied). Vacca et al. (1988) were unable to neither reconstruct the idiogram nor to establish a solid ploidy due to the differences in size and shape of chromosomes of the same plates and also between different chromosome plates. Furthermore, all individuals had copulatory apparatus and, under laboratory conditions, they were able to lay cocoons, but these were always sterile.

This 22 chromosomal number could result from the sum of 14 + 8 chromosomes, which would be possible if a diploid *D. hepta* gamete (most likely an oocyte) and a haploid *D. benazzii* gamete (most probably a sperm) joined, which could point to these animals to be the putative H2 or H3 hybrids. The fact that in the *Cox1* network (Fig. 5b), the mixed clade *Cox1* haplotypes derive from the *D. hepta* haplotypes would give further support to this latter possibility. Thus, we find two karyotypes in Rio Silis that could be a match to our H2-H3 hybrids. Unfortunately, we only have karyological information from two individuals of the mixed clade, and they were found not to bear a diploid set of chromosomes (neither 14 nor 16 chromosomes), but the exact number was not registered.

Molecular evidence could point to a hybrid lineage that originated through the cross of a *D. hepta* oocyte with a *D. benazzii* sperm and that now seems to be stabilized with its individuals reproducing by their own (H2 individuals) so that no mitochondrial genomes from any of both parental species is newly introduced in this lineage. A possibility would be that the hybrid populations could carry on reproducing by fission, a reproductive strategy frequently used in *Dugesia* when they become triploids (Stocchino and Manconi 2013) and, on time, evolve their own mitochondrial lineage by accumulating changes, as we observe in this case. That they use this type of reproduction will be supported by the low nucleotide and haplotype diversity found within this group, especially as compared with those found for the hepta and benazzii haplogroups (Table 4), expected for clonal individuals. This hypothesis would explain why the hybrids bear nuclear haplotypes of the two parental haplogroups but would not explain the existence of a recombinant haplotype (individual MR0092-11), neither the individuals of the H3 group. Howbeit, if the hybrids are able to intercross these latter cases would be explained, but this hypothesis has the problem of how the two different karyotype compositions can combine to produce viable gametes in the hybrids. A possibility would be that these animals use a similar strategy to that observed in triploid ex-fissiparous lineages of *D. ryukyuensis* (Chinone et al. 2014). In a lineage from this species before spermatogenesis begins, the spermatogonia eliminate a whole set of chromosomes, hence in a triploid hybrid between *D. hepta* and *D. benazzii*, two sets of the same species could remain in a certain proportion of cases and pass a regular meiosis. In the female oogenesis in *D. ryukyuensis*, the three sets of chromosomes are retained until the metaphase I occurs. During the meiosis, two chromosome sets pair and the third remains alone. Thus, there is a certain probability that either the two sets of chromosomes from the same species pair or that those homologous chromosomes from the two parental species pair. This process would provide some haploid genetically equilibrated (bearing one copy of each gene) oocytes and also some diploid oocytes, either bearing two sets of chromosomes coming both from one species or even recombinants between the two species' chromosomes. This situation would clearly render these animals mostly sterile (explaining for instance the observations of Vacca et al. 1988) since the probability of getting two gametes with an equilibrated set of chromosomes each to mate and give offspring would be low. Nonetheless, even if this happened with a low frequency, it would be enough to explain the presence of a recombinant haplotype among so many sequenced and especially that some individuals may have only nuclear haplotypes from one of the parent species. Thus, in this hypothesis, the hybrids may mostly reproduce by fission but could be able to mate and produce some fertile offspring from time to time. In any case, both hypotheses point to a probable case of speciation by hybridization since the hybrids would have stabilized and have stablished populations reproducing on their own.

There is also a third possibility that these mixed clade hybrids can cross with the parental species. Almost all the individuals in the mixed clade belong to localities where both species coexist (localities 5, 6, 10, 12, 14 and 15; Supporting

information Table S1; Fig. 2), thus giving the hybrids that have produced some genetically viable gamete the opportunity to backcross. When the hybrids act as females, the mitochondrial lineage is retained to evolve independently, while the nuclear genome is continuously being introgressed by parental species chromosomes. A situation of this type is known from the vertebrate genus *Rana* (Blankenhorn, 1977). We wonder if such reproductive strategy may explain the *D. hepta*, *D. benazzii*, and putative hybrids conundrum and if environmental and/or competition factors may explain the existence of the hybrids and the co-occurrence of the three lineages.

However, similar results can be expected under incomplete lineage sorting (ILS) when a radiation takes place as it has been demonstrated in other cases (Suh et al. 2015). What we consider to be the *hepta* and *benazzii* exclusive haplotypes for *Dunuc12* and *ITS-1* would have been population alleles in a polymorphic ancestor. Stochastically, the *benazzii* variant could have gone lost in the *D. hepta* lineage and preserved as a polymorphic state in *D. benazzii* populations. Thus, the mitochondrial mixed clade would be a distinct *D. benazzii* clade—C—that would have diverged from the other Sardinian *D. benazzii*. We could expect to find in said clade homozygous *Dunuc12* individuals for the *D. benazzii* variant—even for the *D. hepta* variant, though we found none—as well as heterozygous individuals that would be our "hybrids." This could also justify why there are individuals whose karyotype is *benazzii*-like and homozygous for the *benazzii* haplotypes within the mixed clade. However, this hypothesis fails to explain why there are individuals bearing anomalous karyotypes and showing infertility as found in previous studies. On the other hand, from an ecological point of view, three differentiated lineages, probably depending on the same resources, co-occurring in the same localities seem a more difficult scenario to explain than two separated lineages co-occurring with their hybrids.

Conclusions

We present for the first time molecular evidence of the species status for *D. hepta* as a sister group and not derived from *D. benazzii*. In addition, we have found that *D. benazzii* individuals from Corsica may in fact be a different species. At the same time, we have uncovered an unexpected and complex situation in those rivers from Sardinia where the two species, *D. benazzii* and *D. hepta*, co-occur. *Dugesia benazzii* was thought to be a complex species presenting different ploidies and even aneuploids and with sexual and asexual reproduction, while *D. hepta* is exclusively diploid and sexual. Our results show that some aneuploids may in fact be the result of crossings between both species, which represents the first demonstration of planarian hybridism in natural conditions on the base of molecular data, and what is more relevant is that they may even have become a new species. But the complexity of the mitochondrial and nuclear haplotype combinations found makes present information not enough to solve the riddle on how these hybrids may have originated and how they reproduce (if they do) and point to the need of a thorough study. An extensive sampling in the rivers where they co-occur, followed by a study at the genomic level of karyotyped individuals, so that the reproductive behavior, morphology, karyotype, and genomic information is known from each individual as well as the performance of interbreeding tests might render an interesting view on how this complex situation has been generated and is evolving.

Acknowledgments We kindly thank Prof. Maria Pala and Dr. Eva Lázaro for providing specimens of *D. hepta* and *D. benazzii*. We are also grateful to Dr. Marta Álvarez-Presas for her always-welcomed advice and guiding throughout the development of the analytical part of this study, and to all of our laboratory members for their help.

Data availability statement The sequences generated and analyzed during the current study are available in the GenBank repository (https://www.ncbi.nlm.nih.gov/genbank/), under accession numbers MN162692, MN162693(ITS-1 haplotypes), MN442429 to MN442516 (Dunuc12), MN442517 to MN442535 (COI cloned haplotypes), MN525628 to MN525771 (COI), and MN565083 to MN565258 (Dunuc12 cloned haplotypes).

Funding information This research was supported by the Ministerio de Economía y Competitividad, Spain (project CGL2015-63527-P).

References

Abril, J. F., Cebrià, F., Rodríguez-Esteban, G., Horn, T., Fraguas, S., Calvo, B., Bartscherer, K., & Saló, E. (2010). Smed454 dataset: unravelling the transcriptome of *Schmidtea mediterranea*. *BMC Genomics, 11*(1), 731. https://doi.org/10.1186/1471-2164-11-731.

Aguilar, J.P. (2011) Genética Evolutiva de *Dugesia benazzii* y *Dugesia hepta* de las islas de Córcega y Cerdeña. Master thesis, Universitat de Barcelona. Research Gate. https://doi.org/10.13140/RG.2.2.23010.27849. .

Álvarez-Presas, M., & Riutort, M. (2014). Planarian (Platyhelminthes, Tricladida) diversity and molecular markers: a new view of an old group. *Diversity, 6*, 323–338. https://doi.org/10.3390/d6020323.

Arnold, M. L., Sapir, Y., & Martin, N. H. (2008). Review. Genetic exchange and the origin of adaptations: prokaryotes to primates. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1505), 2813–2820. https://doi.org/10.1098/rstb.2008.0021.

Baguñà, J., Carranza, S., Pala, M., Ribera, C., Giribet, G., Arnedo, M. A., Ribas M., Riutort M. (1999). From morphology and karyology to molecules. New methods for taxonomical identification of asexual populations of freshwater planarians. A tribute to Professor Mario

Benazzi. *Italian Journal of Zoology*, 66, 207–214. https://doi.org/10.1080/11250009909356258.

Benazzi, M. (1949). Problemi di differenziazione specifica e razziale nei Tricladi. *Bollettino della Società Italiana di Biologia Sperimentale, 25*, 666.

Benazzi, M. (1955). Appunti sulla distribuzione dei Tricladi in Italia. *Bollettino di Zoologia, 22*(2), 149–164. https://doi.org/10.1080/11250005509439195 .

Benazzi, M. (1982). Speciation events as evidenced in Turbellaria. In C. Barigozzi (Ed.), *Mechanisms of speciation* (pp. 307–344). New York: Alan R. Liss.

Benazzi, M., & Banchetti, R. (1972). Descrizione di *Dugesia biblica*, nuova microspecie del "gruppo *Dugesia gonocephala*" trovata nel fiume Giordano (Israele). Atti della Società Toscana di Scienze Naturali., 79, 83–91. http://www.stsn.it/AttiB1972/Benazzi_Banchetti_M.pdf. Accessed 20 February 2020.

Benazzi, M., & Benazzi-Lentati, G. (1976). Platyhelminthes. In B. John (Ed.), *Animal cytogenetics* (pp. 1–182). Berlin-Suttgart: Gebründer Borntraeger.

Benazzi, M., & Deri, P. (1980). Histo-cytological study of fissiparous planarian testicles (Tricladida, Paludicola). *Italian Journal of Zoology, 14*(3), 151–163. https://doi.org/10.1080/00269786.1980.10736353.

Benazzi, M., & Puccinelli, I. (1973). A Robertsonian translocation in the fresh-water triclad *Dugesia lugubris*: karyometric analysis and evolutionary inferences. Chromosoma, 198, 193–198. https://fdocuments.in/document/a-robertsonian-translocation-in-the-fresh-water-triclad-dugesia-lugubris.html. Accessed 20 February 2020.

Benazzi-Lentati, G., & Benazzi, M. (1985). On the origin and the true nature of the tetraploid asynaptic planarian called Benazzi's G biotype. *Caryologia, 38*(3–4), 269–279. https://doi.org/10.1080/00087114.1985.10797750.

Berninger, J. (1911). Über die Einwirkung des Hungers auf Planarien. *Zoologische Jahrbücher, Abteilung für allgemeine Zoologie und Physiologie der Tiere, 30*, 181–216.

Chinone, A., Nodono, H., & Matsumoto, M. (2014). Triploid planarian reproduces truly bisexually with euploid gametes produced through a different meiotic system between sex. *Chromosoma, 123*(3), 265–272. https://doi.org/10.1007/s00412-013-0449-2.

Coates, D. J., & Shaw, D. D. (1984). The chromosomal component of reproductive isolation in the grasshoper *Caledia captiva* III. Chiasma distribution patterns in a new chromosomal taxon. *Heridity, 53*, 85–100. https://doi.org/. https://doi.org/10.1038/hdy.1984.65.

Curini-Galletti, M. C., Puccinelli, I., & Martens, P. M. (1985). Meccanismi di evoluzione del cariotipo nei monocelididi (Turbellaria, Proseriata). Atti della Società Toscana di Scienze Naturali (Serie B)., 92, 291–297. http://www.vliz.be/imisdocs/publications/276721.pdf. Accessed 20 February 2020.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology, 56*(6), 879–886. https://doi.org/10.1080/10635150701701083.

De Vries, E. J. (1984). On the species of the *Dugesia gonocephala* group (Platyhelminthes, Turbellaria, Tricladida) from Greece. *Bijdragen tot de Dierkunde, 54*(1), 101–126 http://cat.inist.fr/?aModele=afficheN&cpsidt=9712729. .

De Vries, E. J. (1985). The biogeography of the genus *Dugesia* (Turbellaria, Tricladida, Paludicola) in the Mediterranean region. *Journal of Biogeography, 12*(6), 509–518. https://doi.org/10.2307/2844906.

De Vries, E. J. (1986). On the karyology of *Dugesia gonocephala* s.l. (Turbellaria, Tricladida) from Montpellier, France. *Hydrobiologia, 132*(1), 251–256. https://doi.org/10.1007/BF00046257.

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. *Nature, 299*(5879), 111–117. https://doi.org/10.1038/299111a0.

Dugès, A. (1830). Aperçu de quelques observations nouvelles sur les Planaires et plusienrs genres voisins. Annales des Sciences Naturelles, 21, 72–91. http://gallica.bnf.fr/ark:/12148/bpt6k55070078/f4.image.r. Accessed 20 February 2020.

Forsdyke, D. R. (2004). Chromosomal speciation: a reply. *Journal of Theoretical Biology, 230*(2), 189–196. https://doi.org/10.1016/j.jtbi.2004.04.020.

Galleni, L., & Puccinelli, I. (1986). Chromosomal evolution in marine triclads and polyclads (Turbellaria). *Hydrobiologia, 132*, 239–242. https://doi.org/10.1007/BF00046255.

Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. https://doi.org/citeulike-article-id:691774.

Harrath, A. H., Sluys, R., Mansour, L., Lekeufack Folefack, G., Aldahmash, W., Alwasel, S., Solà, E., & Riutort, M. (2019). Molecular and morphological identification of two new African species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from Cameroon. *Journal of Natural History, 53*(5–6), 253–271. https://doi.org/10.1080/00222933.2019.1577508.

Hillis, D. M., & Dixon, M. T. (1991). Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology, 66*(4), 411–453. https://doi.org/10.1086/417338.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution, 30*(4), 772–780. https://doi.org/10.1093/molbev/mst010.

Kawakami, T., Butlin, R. K., & Cooper, S. J. B. (2011). Chromosomal speciation revisited: Modes of diversification in Australian morabine grasshoppers (*Vandiemenella viatica* species group). *Insects, 2*(1), 49–61. https://doi.org/10.3390/insects2010049.

King, M. (1987). Chromosomal rearrangements, speciation and the theoretical approach. *Heredity, 59 (Pt 1)*(1), 1–6. https://doi.org/10.1038/hdy.1987.90.

King, M. (1992). A dual level model for speciation by multiple pericentric inversions. *Most, 68*(May 1991), 437–440. https://doi.org/10.1038/hdy.1992.63.

King, M. (1993). *Species evolution: the role of chromosome change*. Cambridge: Cambridge University Press.

Knowlton, N. (1993). Sibling species in the sea. *Annual Review of Ecology and Systematics, 24*, 189–216. https://doi.org/10.2307/2097177.

Kucera, M., & Darling, K. F. (2002). Cryptic species of planktonic foraminifera: their effect on palaeoceanographic reconstructions. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 360, 695–718. https://doi.org/10.1098/rsta.2001.0962 .

Lang, A. (1884). *Die polycladen (seeplanarien) des colfes von Neapel und der angrenzenden meeresabschnitte*. Leipzig: W. Engelmann. http://www.biodiversitylibrary.org/bibliography/10545#/summary . Accessed 20 February 2020.

Lázaro, E., & Riutort, M. (2013). *Dugesia sicula* (Platyhelminthes, Tricladida): the colonizing success of an asexual Planarian. *BMC Evolutionary Biology, 13*(1), 268. https://doi.org/10.1186/1471-2148-13-268.

Lázaro, E., Sluys, R., Pala, M., Stocchino, G. A., Baguñà, J., & Riutort, M. (2009). Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution, 52*(3), 835–845. https://doi.org/10.1016/j.ympev.2009.04.022.

Lepori, N. G. (1951). Sulle caratteristiche morfologiche e sulla posizione sistematica della planaria di Sardegna e Corsica già ascritta a *Dugesia* (*Euplanaria*) *gonocephala* (Dugès). *Atti della Società Toscana di Scienze Naturali, 58*, 28–47.

Leria, L., Sluys, R., & Riutort, M. (2018). Diversification and biogeographic history of the Western Palearctic freshwater flatworm genus

Schmidtea (Tricladida: Dugesiidae), with a redescription of *Schmidtea nova*. *Journal of Zoological Systematics and Evolutionary Research, 56*(3), 335–351. https://doi.org/10.1111/jzs.12214.

Leria, L., Vila-Farré, M., Solà, E., & Riutort, M. (2019). Outstanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. *BMC Evolutionary Biology, 19*, 1–19. https://doi.org/10.1186/s12862-019-1440-1.

Leria, L., Vila-Farré, M., Álvarez-Presas, M., Sánchez-Gracia, A., Rozas, J., Sluys, R., & Riutort, M. (2020). Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhelminthes, Tricladida): Extreme intraindividual genetic diversity, morphological stasis, and karyological variability. *Molecular Phylogenetics and Evolution, 143*(May), 106496. https://doi.org/10.1016/j.ympev.2019.05.010.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25*(11), 1451–1452. https://doi.org/10.1093/bioinformatics/btp187.

Mallet, J. (2007). Hybrid speciation. *Nature, 446*, 279–283. https://doi.org/10.1038/nature05706.

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays, 38*(2), 140–149. https://doi.org/10.1002/bies.201500149.

Newmark, P. A., & Alvarado, A. S. (2002). Not your father's planarian: a classic model enters the era of functioning genomics. *Nature Reviews Genetics, 3*(3), 210–219. https://doi.org/10.1038/nrg759.

Nishimura, O., Hosoda, K., Kawaguchi, E., Yazawa, S., Hayashi, T., Inoue, T., Umesono, Y., & Agata, K. (2015). Unusually large number of mutations in asexually reproducing clonal planarian *Dugesia japonica*. *PLoS One, 10*(11), 1–23. https://doi.org/10.1371/journal.pone.0143525.

Pala, M., Casu, S., & Vacca, R. A. (1981). *Dugesia hepta*, nuova specie di planaria di acqua dolce di Sardegna appartenente alla superspecie *Dugesia gonocephala* (Dugès). Bollettino della Società Sarda di Scienze Naturali, 20, 97–107. https://core.ac.uk/reader/11689490. Accessed 20 February 2020.

Pala, M., Casu, S., & Lepori, N. G. (1982). Stabilized natural interspecific hybrid population of the fresh water planarians *Dugesia gonocephala* s. l. (Turbellaria, Tricladida). *Caryologia, 35*(2), 247–256. https://doi.org/10.1080/00087114.1982.10796929.

Pala, M., Casu, S., & Stocchino, G. A. (1999). Karyology and karyotype analysis of diploid freshwater planarian populations of the *Dugesia gonocephala* group (Platyhelminthes, Tricladida) found in Sardinia. *Hydrobiologia, 392*(2), 113–119. https://doi.org/10.1023/A:1003534507632.

Pennisi, E. (2016). Shaking up the tree of life. *Science, 354*(6314), 817–821. https://doi.org/10.7748/ns.16.21.24.s37.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution, 25*(7), 1253–1256. https://doi.org/10.1093/molbev/msn083.

Reddien, P. W., & Alvarado, A. S. (2004). Fundamentals of planarian regeneration. *Annual Review of Cell and Developmental Biology, 20*, 725–757. https://doi.org/10.1146/annurev.cellbio.20.010403.095114.

Ribas, M. (1990). *Cariologia*. Sistemàtica i biogeografia de les planàries d'aigües dolces als Països Catalans: Universitat de Barcelona.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution, 16*(7), 351–358. https://doi.org/10.1016/S0169-5347(01)02187-5.

Riutort, M., Álvarez-Presas, M., Lázaro, E., Solà, E., & Paps, J. (2012). Evolutionary history of the Tricladida and the Platyhelminthes: an up-to-date phylogenetic and systematic account. *International Journal of Developmental Biology, 56*(1–3), 5–17. https://doi.org/10.1387/ijdb.113441mr.

Robb, S. M. C., Ross, E., & Alvarado, A. S. (2007). SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Research, 36*, 599–606. https://doi.org/10.1093/nar/gkm684.

Robb, S. M. C., Gotting, K., Ross, E., & Alvarado, A. S. (2015). SmedGD 2.0: the *Schmidtea mediterranea* genome database. *Genesis, 53*(8), 535–546. https://doi.org/10.1002/dvg.22872.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology, 61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029.

Schultz, E. (1904). Über Reduktionen. *I.* Über Hungerserscheinungen bei *Planaria lactea*. *Arch. Entwm. Org., 18*, 555–557.

Sluys, R., & De Jong, H. (1984). Chromosome morphological studies of *Dugesia gonocephala* s.l. (Platyhelminthes, Tricladida). *Caryologia, 37*(1–2), 9–20. https://doi.org/10.1080/00087114.1984.10797683.

Sluys, R., Solà, E., Gritzalis, K., Vila-Farré, M., Mateos, E., & Riutort, M. (2013). Integrative delineation of species of Mediterranean freshwater planarians (Platyhelminthes: Tricladida: Dugesiidae). *Zoological Journal of the Linnean Society, 169*(3), 523–547. https://doi.org/10.1111/zoj.12077.

Solà, E., Sluys, R., Gritzalis, K., & Riutort, M. (2013). Fluvial basin history in the northeastern Mediterranean region underlies dispersal and speciation patterns in the genus *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution, 66*(3), 877–888. https://doi.org/10.1016/j.ympev.2012.11.010.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics, 22*(21), 2688–2690. https://doi.org/10.1093/bioinformatics/btl446.

Stocchino, G. A. (2018). 80 years of research on planarians (Platyhelminthes, Tricladida) from Sardinia, Italy: an annotated checklist. *Zootaxa, 4532*(4), 539–552. https://doi.org/10.11646/zootaxa.4532.4.5.

Stocchino, G. A., & Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology, 803*(3), 319–328. https://doi.org/10.1080/11250003.2013.822025.

Stocchino, G. A., Corso, G., Manconi, R., Casu, S., & Pala, M. (2005). Endemic freshwater planarians of Sardinia: redescription of *Dugesia hepta* (Platyhelminthes, Tricladida) with a comparison of the Mediterranean species of the genus. *Journal of Natural History, 39*(22), 1947–1960. https://doi.org/10.1080/00222930500060025.

Stocchino, G. A., Sluys, R., Riutort, M., Solà, E., & Manconi, R. (2017). Freshwater planarian diversity (Platyhelminthes: Tricladida: Dugesiidae) in Madagascar: new species, cryptic species, with a redefinition of character states. *Zoological Journal of the Linnean Society, 181*, 727–756. https://doi.org/10.1093/zoolinnean/zlx017.

Stocchino, G. A., Sluys, R., Harrath, A. H., Mansour, L., & Manconi, R. (2019). The invasive alien freshwater flatworm *Girardia tigrina* (Girard, 1850) (Platyhelminthes, Tricladida) in Western Europe: new insights into its morphology, karyology and reproductive biology. *Contributions to Zoology, 88*, 1–21. https://doi.org/10.1163/18759866-20191406.

Suh, A., Smeds, L., & Ellegren, H. (2015). The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology, 13*(8), 1–18. https://doi.org/10.1371/journal.pbio.1002224.

Talavera, G., Lukhtanov, V. A., Rieppel, L., Pierce, N. E., & Vila, R. (2013). In the shadow of phylogenetic uncertainty: the recent diversification of *Lysandra* butterflies through chromosomal change. *Molecular Phylogenetics and Evolution, 69*(3), 469–478. https://doi.org/10.1016/j.ympev.2013.08.004 .

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0.

Molecular Biology and Evolution, 30(12), 2725–2729. https://doi.org/10.1093/molbev/mst197.

Vacca, R. A., Casu, S., & Pala, M. (1988). Popolamento planariologico dei fiumi del nord Sardegna: 1. I cariotipi delle planarie d'acqua dolce del gruppo *Dugesia gonocephala* (Turbellaria, Tricladida) presenti nel fiume Silis (Sassari). Bollettino della Società Sarda di Scienze Naturali, 26, 131–147. https://www.researchgate.net/publication/277857627. Accessed 20 February 2020.

White, M. J. D. (1978). *Modes of speciation*. San Francisco: W. H. Freeman and Co..

Xia, X., & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity, 92*(4), 371–373. https://doi.org/10.1093/jhered/92.4.371.

Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution, 26*(1), 1–7. https://doi.org/10.1016/S1055-7903(02)00326-3.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information:

Dugesia hepta and *Dugesia benazzii* (Tricladida, Platyhelminthes): two sympatric species with occasional sex?

Daniel Dols-Serrate[1], Laia Leria[1], Juan Pablo Aguilar[1], Giacinta Stocchino[2], Marta Riutort[1]

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2]Dipartimento di Scienze della Natura e del Territorio, Via Muroni 25, University of Sassari, Sassari I-07100, Italy

Corresponding author: M. Riutort mriutort@ub.edu

Table S1. List of specimens used for the present study with information on species identification based on *ITS-1* criterion, karyology and molecular markers used. Gen Bank accession numbers are provided for all sequences.

Code name	Species (ITS-1)	Karyotype	Locality (n)	Coordinates	Accession number - GenBank		Collectors	Date of collection
					Cox1	*Dunuc12*		
MR0023-01	*Dugesia benazzii*	-	Rio s'Adde (1), Sardinia	40°37'42"N, 8°58'53"E	MN525634	MN442432	P	17/07/1997
MR0023-02	*Dugesia benazzii*	-	"	"	MN525635	MN442433	P	17/07/1997
MR0028-01	*Dugesia benazzii*	-	Rio Badde Manna (2), Sardinia	40°46'52"N, 9°06'59"E	MN525640	MN442438	P	17/07/1997
MR0025-02	*Dugesia benazzii*	-	Rio Badu de Mola (3), Sardinia	40°18'11"N, 8°34'57"E	MN525636	MN442434	P	17/07/1997
MR0025-03	*Dugesia* sp.	-	"	"	MN525637	MN442435	P	17/07/1997
MR0025-04	*Dugesia benazzii*	-	"	"	MN525638	MN442436	P	17/07/1997
MR0358-01	*Dugesia hepta*	n = 7	Rio Bunnari-1 (4), Sardinia	40°42'05"N, 8°35'30"E	-	MN442497	S	01/01/2013
MR0359-01	*Dugesia hepta*	n = 7	"	"	-	MN442498	S	01/01/2013
MR0360-01	*Dugesia hepta*	n = 7	"	"	-	MN442499	S	01/01/2013
MR0361-01	*Dugesia benazzii*	n = 8	"	"	-	MN4424500	S	01/01/2013
MR0362-01	*Dugesia benazzii*	n = 8	"	"	-	MN442501	S	01/01/2013
MR0088-01	*Dugesia benazzii*	-	Rio Bunnari-2 (5), Sardinia	40°42'05"N, 8°35'30"E	MN525657	MN442446	P, S, L	05/05/2010
MR0088-02	*Dugesia benazzii*	-	"	"	MN525658	MN442447	P, S, L	05/05/2010

MR0088-03	*Dugesia benazzii*	-	“	“	MN525659	-	P, S, L	05/05/2010
MR0088-04	*Dugesia benazzii*	-	“	“	MK712640	MK713164	P, S, L	05/05/2010
MR0088-05	*Dugesia benazzii*	-	“	“	-	-	P, S, L	05/05/2010
MR0088-06	*Dugesia benazzii*	-	“	“	MN525660	-	P, S, L	05/05/2010
MR0089-02	*Dugesia benazzii*	-	“	“	MN525661	-	S	01/01/2013
MR0089-03	*Dugesia benazzii*	-	“	“	MN525662	-	S	01/01/2013
MR0089-04	*Dugesia benazzii*	-	“	“	-	-	P, S, L	05/05/2010
MR0089-05	*Dugesia benazzii*	-	“	“	MN525663	-	P, S, L	05/05/2010
MR0089-06	*Dugesia benazzii*	-	“	“	MN525664	-	P, S, L	05/05/2010
MR0089-07	*Dugesia benazzii*	-	“	“	MN525665	MN442448	P, S, L	05/05/2010
MR0089-08	*Dugesia benazzii*	-	“	“	MN525666	-	P, S, L	05/05/2010
MR0363-01	*Dugesia hepta*	n = 7	“	“	-	MN442502	S	01/01/2013
MR0364-01	*Dugesia hepta*	n = 7	“	“	-	MN442503	S	01/01/2013
MR0365-01	*Dugesia hepta*	n = 7	“	“	MN525738	MN442504	S	01/01/2013
MR0366-01	*Dugesia hepta*	n = 7	“	“	-	MN442505	S	01/01/2013
MR0367-01	*Dugesia hepta*	n = 7	“	“	MN525739	MN442506	S	01/01/2013
MR0368-01	*Dugesia benazzii*	n = 8	“	“	-	MN442507	S	01/01/2013
MR0370-01	*Dugesia benazzii*	n = 8	“	“	-	MN442508	S	01/01/2013
MR0090-01	*Dugesia benazzii*	-	Rio Bunnari-3 (6), Sardinia	40°42'17"N, 8°35'46"E	MN525667	-	P, S, L	06/05/2010
MR0090-02	*Dugesia benazzii*	-	“	“	MN525668	-	P, S, L	06/05/2010
MR0090-03	*Dugesia benazzii*	-	“	“	MN525669	-	P, S, L	06/05/2010

MR0090-04	*Dugesia hepta*	-	“	“	MN525670	-	P, S, L	06/05/2010
MR0090-05	*Dugesia hepta*	-	“	“	MN525671	MN442449	P, S, L	06/05/2010
MR0090-06	*Dugesia hepta*	-	“	“	MN525672	-	P, S, L	06/05/2010
MR0090-07	*Dugesia hepta*	-	“	“	MN525673	-	P, S, L	06/05/2010
MR0091-01	*Dugesia hepta*	-	“	“	MN525674	MN442450	P, S, L	06/05/2010
MR0091-02	*Dugesia hepta*	-	“	“	MN525675	-	P, S, L	06/05/2010
MR0091-03	*Dugesia hepta*	-	“	“	MN525676	-	P, S, L	06/05/2010
MR0091-04	*Dugesia hepta*	-	“	“	MN525677	MN442451	P, S, L	06/05/2010
MR0091-05	*Dugesia hepta*	-	“	“	MN525678	-	P, S, L	06/05/2010
MR0091-06	*Dugesia hepta*	-	“	“	MN525679	-	P, S, L	06/05/2010
MR0091-07	*Dugesia hepta*	-	“	“	MN525680	MN442452	P, S, L	06/05/2010
MR0195-01	*Dugesia benazzii*	n = 8	Rio Lernu (7), Sardinia	40°44'27"N, 9°30'42"E	FJ646933+FJ646977	-	P	17/07/1997
MR0196-01	*Dugesia benazzii*	n = 8	“	“	FJ646934+FJ646978	-	P	17/07/1997
MR0027-01	*Dugesia benazzii*	-	Rio di Limbara (8), Sardinia	40°52'09"N, 9°09'26"E	FJ646937+FJ646982	-	P	17/07/1997
MR0027-02	*Dugesia benazzii*	-	“	“	MN525639	MN442437	P	17/07/1997
MR0077-01	*Dugesia hepta*	-	Rio Mascari-1 (9), Sardinia	40°41'57"N, 8°35'03"E	MN525644	-	P, S, L	06/05/2010
MR0077-02	*Dugesia hepta*	-	“	“	MN525645	-	P, S, L	06/05/2010
MR0077-03	*Dugesia hepta*	-	“	“	MN525646	-	P, S, L	06/05/2010
MR0077-04	*Dugesia hepta*	-	“	“	MN525647	-	P, S, L	06/05/2010

MR0077-05	*Dugesia hepta*	-	“	“	-	-	P, S, L	06/05/2010
MR0077-06	*Dugesia hepta*	-	“	“	MN525648	-	P, S, L	06/05/2010
MR0077-07	*Dugesia hepta*	-	“	“	MN525649	MN442442	P, S, L	06/05/2010
MR0078-01	*Dugesia hepta*	-	“	“	MN525650	MN442443	P, S, L	06/05/2010
MR0078-02	*Dugesia hepta*	-	“	“	MN525651	MN442444	P, S, L	06/05/2010
MR0078-03	*Dugesia hepta*	-	“	“	MN525652	-	P, S, L	06/05/2010
MR0078-04	*Dugesia hepta*	-	“	“	MN525653	-	P, S, L	06/05/2010
MR0078-05	*Dugesia hepta*	-	“	“	MN525654	-	P, S, L	06/05/2010
MR0078-06	*Dugesia hepta*	-	“	“	MN525655	MN442445	P, S, L	06/05/2010
MR0078-07	*Dugesia hepta*	-	“	“	MN525656	-	P, S, L	06/05/2010
MR0342-01	*Dugesia hepta*	n = 7	“	“	MK385923	MK713158	S	01/01/2013
MR0343-01	*Dugesia hepta*	n = 7	“	“	MN525730	MN442483	S	01/01/2013
MR0344-01	*Dugesia hepta*	n = 7	“	“	MK385924	MK713159	S	01/01/2013
MR0345-01	*Dugesia hepta*	n = 7	“	“	MN525731	MN442484	S	01/01/2013
MR0346-01	*Dugesia hepta*	n = 7	“	“	-	MN442485	S	01/01/2013
MR0347-01	*Dugesia hepta*	n = 7	“	“	-	MN442486	S	01/01/2013
MR0348-01	*Dugesia hepta*	n = 7	“	“	MN525732	MN442487	S	01/01/2013
MR0349-01	*Dugesia hepta*	n = 7	“	“	-	MN442488	S	01/01/2013
MR0350-01	*Dugesia hepta*	n = 7	“	“	-	MN442489	S	01/01/2013
MR0351-01	*Dugesia hepta*	n = 7	“	“	-	MN442490	S	01/01/2013

MR0352-01	*Dugesia benazzii*	n = 8	Rio Mascari-2 (10), Sardinia	40°41'60"N, 8°35'23"E	MN525733	MN442491	S	01/01/2013
MR0353-01	*Dugesia benazzii*	n = 8	“	“	MN525734	MN442492	S	01/01/2013
MR0354-01	*Dugesia hepta*	n = 7	“	“	MN525735	MN442493	S	01/01/2013
MR0355-01	*Dugesia hepta*	n = 7	“	“	MN525736	MN442494	S	01/01/2013
MR0356-01	*Dugesia hepta*	n = 7	“	“	-	MN442495	S	01/01/2013
MR0357-01	*Dugesia hepta*	n = 7	“	“	MN525737	MN442496	S	01/01/2013
MR0096-01	*Dugesia hepta*	-	Ponte Valenti (11), Sardinia	40°28'09"N, 8°48'21"E	MN525694	-	P, S, L	10/05/2010
MR0096-03	*Dugesia hepta*	-	“	“	MN525695	-	P, S, L	10/05/2010
MR0096-04	*Dugesia hepta*	-	“	“	MN525696	-	P, S, L	10/05/2010
MR0096-05	*Dugesia hepta*	-	“	“	MN525697	-	P, S, L	10/05/2010
MR0096-06	*Dugesia hepta*	-	“	“	MN525698	-	P, S, L	10/05/2010
MR0098-01	*Dugesia hepta*	-	“	“	MN525699	MN442463	P, S, L	10/05/2010
MR0098-02	*Dugesia hepta*	-	“	“	-	MN442464	P, S, L	10/05/2010
MR0098-04	*Dugesia hepta*	-	“	“	MN525700	MN442465	P, S, L	10/05/2010
MR0099-01	*Dugesia hepta*	-	“	“	MN525701	MN442466	P, S, L	10/05/2010
MR0099-03	*Dugesia hepta*	-	“	“	-	MN442467	P, S, L	10/05/2010
MR0172-01	*Dugesia benazzii*	-	Rio Mulinu (12), Sardinia	40°24'25"N, 8°37'35"E	MN525720	MN442468	P, S, L	13/05/2010
MR0172-02	*Dugesia benazzii*	-	“	“	-	MN442469	P, S, L	13/05/2010
MR0172-03	*Dugesia hepta*	-	“	“	-	MN442470	P, S, L	13/05/2010

MR0172-04	*Dugesia benazzii*	-	"	"	-	MN442471	P, S, L	13/05/2010
MR0172-05	*Dugesia benazzii*	-	"	"	MN525721	MN442472	P, S, L	13/05/2010
MR0172-06	*Dugesia* sp.	-	"	"	-	-	P, S, L	13/05/2010
MR0172-07	*Dugesia hepta*	-	"	"	MK712639	MK713161	P, S, L	13/05/2010
MR0173-01	*Dugesia hepta*	-	"	"	-	MN442473	P, S, L	13/05/2010
MR0173-02	*Dugesia benazzii*	-	"	"	-	MN442474	P, S, L	13/05/2010
MR0173-04	*Dugesia hepta*	-	"	"	MN525722	MN442475	P, S, L	13/05/2010
MR0173-05	*Dugesia hepta*	-	"	"	MN525723	MN442476	P, S, L	13/05/2010
MR0173-06	*Dugesia* sp.	-	"	"	MN525724	MN442477	P, S, L	13/05/2010
MR0174-01	*Dugesia benazzii*	-	"	"	MN525725	MN442478	P, S, L	13/05/2010
MR0174-03	*Dugesia hepta*	-	"	"	-	MN442479	P, S, L	13/05/2010
MR0174-04	*Dugesia hepta*	-	"	"	-	MN442480	P, S, L	13/05/2010
MR0174-05	*Dugesia benazzii*	-	"	"	MN252726	MN442481	P, S, L	13/05/2010
MR0174-06	*Dugesia hepta*	-	"	"	MN525727	MN442482	P, S, L	13/05/2010
MR0022-01	*Dugesia benazzii*	-	Rio di Otti (13), Sardinia	40°43'55"N, 9°08'36"E	MN525628	MN442429	P	28/04/2005
MR0022-02	*Dugesia benazzii*	-	"	"	MN525629	MN442430	P	28/04/2005
MR0022-03	*Dugesia benazzii*	-	"	"	MN525630	-	P	28/04/2005
MR0022-04	*Dugesia benazzii*	-	"	"	MN525631	-	P	28/04/2005
MR0022-05	*Dugesia benazzii*	-	"	"	MN525632	MN442431	P	28/04/2005
MR0022-06	*Dugesia benazzii*	-	"	"	MN525633	-	P	28/04/2005

MR0030-03	*Dugesia hepta*	-	Rio Santa Lucia (14), Sardinia	40°25'53"N, 8°50'09"E	MN525641	MN442439	P	01/02/2006
MR0030-04	*Dugesia hepta*	-	"	"	MN525642	MN442440	P	01/02/2006
MR0030-05	*Dugesia hepta*	-	"	"	MK712638	MK713160	P	01/02/2006
MR0030-06	*Dugesia hepta*	-	"	"	MN525643	MN442441	P	01/02/2006
MR0201-01	*Dugesia benazzii*	n = 8	"	"	FJ646938+FJ64983	-	P	01/02/2006
MR0092-01	*Dugesia hepta*	-	Rio Fiume Silis (15), Sardinia	40°49'47"N, 8°35'39"E	MN525681	MN442453	P, S, L	08/05/2010
MR0092-02	*Dugesia* sp.	-	"	"	MN525682	MN442454	P, S, L	08/05/2010
MR0092-03	*Dugesia benazzii*	-	"	"	MK712643	MN442455	P, S, L	08/05/2010
MR0092-04	*Dugesia hepta*	-	"	"	MN525683	MN442456	P, S, L	08/05/2010
MR0092-05	*Dugesia* sp.	-	"	"	MN525684	MN442457	P, S, L	08/05/2010
MR0092-06	*Dugesia hepta*	-	"	"	MN525685	MN442458	P, S, L	08/05/2010
MR0092-07	*Dugesia benazzii*	-	"	"	MN525686	MK713165	P, S, L	08/05/2010
MR0092-08	*Dugesia hepta*	-	"	"	MN525687	-	P, S, L	08/05/2010
MR0092-09	*Dugesia benazzii*	-	"	"	MN525688	-	P, S, L	08/05/2010
MR0092-10	*Dugesia hepta*	-	"	"	MN525689	-	P, S, L	08/05/2010
MR0092-11	*Dugesia* sp.	-	"	"	MN525690	MN442459	P, S, L	08/05/2010
MR0092-12	*Dugesia hepta*	-	"	"	-	MN442460	P, S, L	08/05/2010
MR0092-13	*Dugesia* sp.	-	"	"	-	MN442461	P, S, L	08/05/2010
MR0092-14	*Dugesia hepta*	-	"	"	-	MN442462	P, S, L	08/05/2010
MR0092-15	*Dugesia hepta*	-	"	"	MN525691	-	P, S, L	08/05/2010

MR0092-16	*Dugesia* sp.	-	"	"	MN525692	-	P, S, L	08/05/2010
MR0092-21	*Dugesia* sp.	-	"	"	MN525693	-	P, S, L	08/05/2010
MR0385-01	*Dugesia benazzii*	-	Monte Albo (16), Sardinia	40°34'30"N, 9°40'31"E	MK385925	MK713162	S	01/05/2012
MR0385-02	*Dugesia benazzii*	-	"	"	MK385926	MK713163	S	01/05/2012
MR0385-03	*Dugesia benazzii*	-	"	"	MN525740	MN442509	S	01/05/2012
MR0385-04	*Dugesia benazzii*	-	"	"	MN525741	-	S	01/05/2012
MR0385-05	*Dugesia benazzii*	-	"	"	MN525742	MN442510	S	01/05/2012
MR0385-06	*Dugesia benazzii*	-	"	"	MN525743	MN442511	S	01/05/2012
MR0385-07	*Dugesia* sp.	-	"	"	MN525744	MN442512	S	01/05/2012
MR0385-08	*Dugesia* sp.	-	"	"	MN525745	MN442513	S	01/05/2012
MR0385-09	*Dugesia* sp.	-	"	"	MN525746	MN442514	S	01/05/2012
MR0385-10	*Dugesia* sp.	-	"	"	MN525747	MN442515	S	01/05/2012
MR0385-11	*Dugesia* sp.	-	"	"	MN525748	-	S	01/05/2012
MR0385-12	*Dugesia* sp.	-	"	"	MN525749	MN442516	S	01/05/2012
MR0026-01	*Dugesia benazzii*	-	Rio Sos Molinos (17), Sardinia	40°07'53"N, 8°38'31"E	FJ646936+FJ646981	-	P	17/07/1997
MR1245-01	*Dugesia benazzii*	-	Chiola (18), Corsica	41°88'22.12"N, 9°39'35.41"E	MN525769	-	P, S, L	2009
MR1245-02	*Dugesia benazzii*	-	"	"	MN525770	-	P, S, L	2009
MR1034-01	*Dugesia benazzii*	-	La Solenzara (19), Corsica	41°50'39.6"N, 9°20'34"E	MN525750	-	P, S, L	2010
MR1034-02	*Dugesia benazzii*	-	"	"	MN525751	-	P, S, L	2010

MR1034-03	*Dugesia benazzii*	-	"	"	MN525752	-	P, S, L	2010
MR1034-04	*Dugesia benazzii*	-	"	"	MN525753	-	P, S, L	2010
MR1034-05	*Dugesia benazzii*	-	"	"	MN525754	-	P, S, L	2010
MR1055-01	*Dugesia benazzii*	-	Cioccia (20), Corsica	41°32'41.7"N, 9°01'24.1"E	MN525763	-	P, S, L	2009
MR1055-02	*Dugesia benazzii*	-	"	"	MN525764	-	P, S, L	2009
MR1055-03	*Dugesia benazzii*	-	"	"	MN525765	-	P, S, L	2009
MR1055-04	*Dugesia benazzii*	-	"	"	MN525766	-	P, S, L	2009
MR1055-05	*Dugesia benazzii*	-	"	"	MN525767	-	P, S, L	2009
MR1055-06	*Dugesia benazzii*	-	"	"	MN525768	-	P, S, L	2009
MR1037-01	*Dugesia benazzii*	-	Pero Longo (21), Corsica	41°32'27.6"N, 8°58'07.2"E	MN525755	-	P, S, L	2009
MR1037-02	*Dugesia benazzii*	-	"	"	MN525756	-	P, S, L	2009
MR1037-03	*Dugesia benazzii*	-	"	"	MN525757	-	P, S, L	2009
MR1037-04	*Dugesia benazzii*	-	"	"	MN525758	-	P, S, L	2009
MR1037-05	*Dugesia benazzii*	-	"	"	MN525759	-	P, S, L	2009
MR1246-01	*Dugesia benazzii*	-	Rio Spartano (22), Corsica	41°30'20.1"N, 9°01'43.1"E	MN525771	-	P, S, L	2009
MR1038-01	*Dugesia benazzii*	-	San Gavino (23), Corsica	41°31'34.3"N, 9°04'22.4"E	MN525760	-	P, S, L	2009
MR1038-02	*Dugesia benazzii*	-	"	"	MN525761	-	P, S, L	2009
MR1038-03	*Dugesia benazzii*	-	"	"	MN525762	-	P, S, L	2009

MR0101-01	*Dugesia benazzii*	-	Barchetta (24), Corsica	42°30'26.73"N, 9°04'22.4"E	MN525702	-	P, S, L	2010
MR0101-02	*Dugesia benazzii*	-	"	"	MN525703	-	P, S, L	2010
MR0176-01	*Dugesia benazzii*	-	"	"	MN525728	-	P, S, L	2010
MR0176-02	*Dugesia benazzii*	-	"	"	MN525729	-	P, S, L	2010
MR0102-01	*Dugesia benazzii*	-	Ponti strada (25), Corsica	42°30'34.41"N, 9°21'16.15"E	MN525704	-	P, S, L	2010
MR0102-02	*Dugesia benazzii*	-	"	"	MN525705	-	P, S, L	2010
MR0102-03	*Dugesia benazzii*	-	"	"	MN252706	-	P, S, L	2010
MR0102-04	*Dugesia benazzii*	-	"	"	MN525707	-	P, S, L	2010
MR0105-01	*Dugesia benazzii*	-	Rist. Accendi Pipa (26), Corsica	42°30'21.99"N, 9°18'41.24"E	MN525708	-	P, S, L	2010
MR0105-02	*Dugesia benazzii*	-	"	"	MN525709	-	P, S, L	2010
MR0105-03	*Dugesia benazzii*	-	"	"	MN525710	-	P, S, L	2010
MR0105-04	*Dugesia benazzii*	-	"	"	MN525711	-	P, S, L	2010
MR0107-01	*Dugesia benazzii*	-	N198 Loc. Arena (27), Corsica	42°29'57.21"N, 9°28'1.12"E	MN525712	-	P, S, L	2010
MR0111-01	*Dugesia benazzii*	-	Ocagnano (28), Corsica	42°28'43.54"N, 9°28'53.78"E	MN525713	-	P, S, L	2010
MR0112-01	*Dugesia benazzii*	-	Fium Alto (29), Corsica	42°25.648'N, 9°26.286'E	MN525714	-	P, S, L	2010
MR0112-02	*Dugesia benazzii*	-	"	"	MN525715	-	P, S, L	2010
MR0113-11	*Dugesia benazzii*	-	Ecole Communale (30), Corsica	42°26.648'N, 9°26.386'E	MN525716	-	P, S, L	2010

MR0113-12	*Dugesia benazzii*	-	“	“	MN525717	-	P, S, L	2010
MR0114-01	*Dugesia benazzii*	-	Strada per San Gavino (31), Corsica	42°25.52' N, 9°25.972'E	MK712641	MK713166	P, S, L	2010
MR0114-02	*Dugesia benazzii*	-	“	“	MK712642	MK713167	P, S, L	2010
MR0114-03	*Dugesia benazzii*	-	“	“	MN525718	-	P, S, L	2010
MR0118-01	*Dugesia benazzii*	-	N198 (32), Corsica	42°23.673' N, 9°31,827'E	MN525719	-	P, S, L	2010
A2.1DC.M1	*Dugesia vilafarrei*	-	Andalucia, Spain	36°45'40"N, 5°30'20.9"E	MK712533	MK712926	A, R, Le	17/06/2013
E24.1DC	*Dugesia subtentaculata*	-	Ribeira de Nissa, Portugal	39°19'23.7"N 7°24'32.8"W	MK712611	MK713134	A, R, Le	04/06/2013
M2.2D	*Dugesia aurea*	-	Soller, Mallorca, Spain	39°47'36.0"N 2°40'59.1"E	MK712632	MK713147	V	01/06/2013
POR12.1DC	*Dugesia subtentaculata*	-	Alte, Portugal	37°14'17.8"N 8°10'19.2"W	MK712619	MK713142	A, R, Le	06/10/2013

P: M. Pala S: G. A. Stocchino L: E. Lázaro A: M. Álvarez-Presas Le: L. Leria V: M. Vila-Farré R: M. Riutort

Table S2. List of *Cox1* haplotypes from data set III.

Haplotype	*Sample*	*Species*	*Locality*	*Karyotype*	*Accession num.*
HCOI_1	MR0030-03	*D. hepta*	14	-	MN442520
	MR0096-01	*D. hepta*	11	-	=
	MR0096-04	*D. hepta*	11	-	=
	MR0098-04	*D. hepta*	11	-	=
	MR0099-01	*D. hepta*	11	-	=
HCOI_2	MR0030-04	*D. hepta*	14	-	MN442521
	MR0030-05	*D. hepta*	11	-	=
	MR0096-05	*D. hepta*	11	-	=
HCOI_3	MR0096-03	*D. hepta*	14	-	MN442528
HCOI_4	MR0172-07	*D. hepta*	12	-	MN442530
	MR0173-05	*D. hepta*	12	-	=
HCOI_5	MR0173-04	*D. hepta*	12	-	MN442531
HCOI_6	MR0077-01	*D. hepta*	9	-	MN442522
	MR0077-02	*D. hepta*	9	-	=
	MR0078-02	*D. hepta*	9	-	=
	MR0078-05	*D. hepta*	9	-	=
	MR0078-06	*D. hepta*	9	-	=
	MR0090-04	*D. hepta*	6	-	=
	MR0091-01	*D. hepta*	6	-	=
	MR0091-02	*D. hepta*	6	-	=
	MR0091-05	*D. hepta*	6	-	=
	MR0342-01	*D. hepta*	9	n=7	=
	MR0345-01	*D. hepta*	9	n=7	=
	MR0355-01	*D. hepta*	10	n=7	=
	MR0367-01	*D. hepta*	5	n=7	=
HCOI_7	MR0078-03	*D. hepta*	9	-	MN442523
HCOI_8	MR0343-01	*D. hepta*	9	n=7	MN442533
HCOI_9	MR0023-01	*D. benazzii*	1	-	MN442518
HCOI_10	MR0023-02	*D. benazzii*	1	-	MN442519
HCOI_11	MR0088-02	*D. benazzii*	5	-	MN442524
	MR0089-02	*D. benazzii*	5	-	=
	MR0089-08	*D. benazzii*	5	-	=
	MR0090-03	*D. benazzii*	6	-	=
	MR0092-02	*D. sp*	15	-	=
	MR0092-07	*D. benazzii*	15	-	=
	MR0092-08	*D. sp*	15		=
	MR0092-09	*D. sp*	15	-	=
	MR0092-10	*D. sp*	15	-	=
	MR0092-11	*D. sp*	15	-	=

	MR0092-15	*D. hepta*	15	-	=
	MR0092-16	*D. sp*	15	-	=
	MR0092-21	*D. sp*	15	-	=
HCOI_12	MR0022-06	*D. benazzii*	13	-	MN442517
HCOI_13	MR0088-03	*D. benazzii*	5	-	MN442525
HCOI_14	MR0089-03	*D. benazzii*	5	-	MN442526
	MR0089-07	*D. benazzii*	5	-	=
	MR0090-01	*D. benazzii*	6	-	=
	MR0090-02	*D. benazzii*	6	-	=
HCOI_15	MR0089-06	*D. benazzii*	5	-	MN442527
HCOI_16	MR0172-05	*D. sp*	12	-	MN442529
HCOI_17	MR0174-01	*D. benazzii*	12	-	MN442532
HCOI_18	MR0385-02	*D. benazzii*	16	-	MN442534
HCOI_19	MR0385-12	*D. sp*	16	-	MN442535

Table S3. List of haplotypes from cloned sequences from dataset IV (*Dunuc12*).

Haplotype	***Sample***	***Species***	***Locality***	***Karyotype***	***Accession num.***
HDunuc12_1	MR0172-01.C1	*D. benazzii*	12	-	MN565083
HDunuc12_2	MR0172-01.C11	*D. benazzii*	12	-	MN525084
	MR0172-01.C14	*D. benazzii*	12	-	=
	MR0172-01.C20	*D. benazzii*	12	-	=
	MR0172-01.C22	*D. benazzii*	12	-	=
HDunuc12_3	MR0172-01.C12	*D. benazzii*	12	-	MN565085
HDunuc12_4	MR0172-01.C13	*D. benazzii*	12	-	MN565086
HDunuc12_5	MR0172-01.C15	*D. benazzii*	12	-	MN565087
HDunuc12_6	MR0172-01.C16	*D. benazzii*	12	-	MN565088
HDunuc12_7	MR0172-01.C17	*D. benazzii*	12	-	MN565089
HDunuc12_8	MR0172-01.C18	*D. benazzii*	12	-	MN565090
HDunuc12_9	MR0172-01.C19	*D. benazzii*	12	-	MN565091
HDunuc12_10	MR0172-01.C2	*D. benazzii*	12	-	MN565092
HDunuc12_11	MR0172-01.C21	*D. benazzii*	12	-	MN565093
HDunuc12_12	MR0172-01.C3	*D. benazzii*	12	-	MN565094
HDunuc12_13	MR0172-01.C4	*D. benazzii*	12	-	MN565095
HDunuc12_14	MR0172-01.C5	*D. benazzii*	12	-	MN565096
	MR0353-01.C9	*D. benazzii*	10	n=8	=
HDunuc12_15	MR0172-01.C7	*D. benazzii*	12	-	MN565097
HDunuc12_16	MR0172-01.C8	*D. benazzii*	12	-	MN565098
HDunuc12_17	MR0172-01.C9	*D. benazzii*	12	-	MN565099
HDunuc12_18	MR0353-01.C1	*D. benazzii*	10	n=8	MN565100
	MR0353-01.C11	*D. benazzii*	10	n=8	=
	MR0353-01.C12	*D. benazzii*	10	n=8	=

	MR0353-01.C15	*D. benazzii*	10	n=8	=
	MR0353-01.C2	*D. benazzii*	10	n=8	=
HDunuc12_19	MR0353-01.C10	*D. benazzii*	10	n=8	MN565101
HDunuc12_20	MR0353-01.C14	*D. benazzii*	10	n=8	MN565102
HDunuc12_21	MR0353-01.C16	*D. benazzii*	10	n=8	MN565103
HDunuc12_22	MR0353-01.C17	*D. benazzii*	10	n=8	MN565104
HDunuc12_23	MR0353-01.C18	*D. benazzii*	10	n=8	MN565105
HDunuc12_24	MR0353-01.C19	*D. benazzii*	10	n=8	MN565106
HDunuc12_25	MR0353-01.C20	*D. benazzii*	10	n=8	MN565107
HDunuc12_26	MR0353-01.C3	*D. benazzii*	10	n=8	MN565108
HDunuc12_27	MR0353-01.C4	*D. benazzii*	10	n=8	MN565109
HDunuc12_28	MR0353-01.C6	*D. benazzii*	10	n=8	MN565110
HDunuc12_29	MR0353-01.C7	*D. benazzii*	10	n=8	MN565111
HDunuc12_30	MR0353-01-C8	*D. benazzii*	10	n=8	MN565112
HDunuc12_31	MR0352-01.C1	*D. hepta*	10	n=7	MN565113
	MR0352-01.C14	*D. hepta*	10	n=7	=
	MR0352-01.C15	*D. hepta*	10	n=7	=
	MR0352-01.C2	*D. hepta*	10	n=7	=
	MR0352-01.C20	*D. hepta*	10	n=7	=
	MR0352-01.C4	*D. hepta*	10	n=7	=
	MR0352-01.C5	*D. hepta*	10	n=7	=
	MR0352-01.C6	*D. hepta*	10	n=7	=
	MR0352-01.C8	*D. hepta*	10	n=7	=
	MR0368-01.C10	*D. benazzii*	5	n=8	=
	MR0368-01.C2	*D. benazzii*	5	n=8	=
	MR0368-01.C6	*D. benazzii*	5	n=8	=
	MR0368-01.C8	*D. benazzii*	5	n=8	=
	MR0368-01.C9	*D. benazzii*	5	n=8	=
	MR0370-01.C4	*D. benazzii*	5	n=8	=
	MR0370-01.C7	*D. benazzii*	5	n=8	=
	MR0088-02.C1	*D. benazzii*	5	-	=
	MR0088-02.C10	*D. benazzii*	5	-	=
	MR0088-02.C13	*D. benazzii*	5	-	=
	MR0088-02.C14	*D. benazzii*	5	-	=
	MR0088-02.C17	*D. benazzii*	5	-	=
	MR0088-02.C6	*D. benazzii*	5	-	=
	MR0088-02.C8	*D. benazzii*	5	-	=
	MR0092-03.C11	*D. benazzii*	15	-	=
	MR0092-03.C4	*D. benazzii*	15	-	=
	MR0092-04.C1	*D. hepta*	15	-	=
	MR0092-04.C3	*D. hepta*	15	-	=
	MR0092-04.C7	*D. hepta*	15	-	=
	MR0092-11.C12b	*D. sp*	15	-	=

	MR0092-11.C20b	*D. sp*	15	-	=
	MR0092-11.C24b	*D. sp*	15	-	=
	MR0092-11.C26b	*D. sp*	15	-	=
	MR0092-11.C30b	*D. sp*	15	-	=
	MR0092-11.C8	*D. sp*	15	-	=
	MR0092-11.C9	*D. sp*	15	-	=
	MR0092-11.C5	*D. sp*	15	-	=
	MR0025-02.C10	*D. benazzii*	3	-	=
HDunuc12_32	MR0352-01.C10	*D. benazzii*	10	n=8	MN565114
HDunuc12_33	MR0092-11.C27b	*D. sp*	15	-	MN565115
HDunuc12_34	MR0352-01.C11	*D. benazzii*	10	n=8	MN565116
HDunuc12_35	MR0352-01.C13	*D. benazzii*	10	n=8	MN565117
HDunuc12_36	MR0352-01.C16	*D. benazzii*	10	n=8	MN565118
HDunuc12_37	MR0352-01.C18	*D. benazzii*	10	n=8	MN565119
HDunuc12_38	MR0352-01.C19	*D. benazzii*	10	n=8	MN565120
HDunuc12_39	MR0352-01.C3	*D. benazzii*	10	n=8	MN565121
HDunuc12_40	MR0352-01.C7	*D. benazzii*	10	n=8	MN565122
HDunuc12_41	MR0352-01.C9	*D. benazzii*	10	n=8	MN565123
HDunuc12_42	MR0354-01.C1	*D. hepta*	10	n=7	MN565124
HDunuc12_43	MR0354-01.C10	*D. hepta*	10	n=7	MN565125
HDunuc12_44	MR0354-01.C11	*D. hepta*	10	n=7	MN565126
HDunuc12_45	MR0354-01.C13	*D. hepta*	10	n=7	MN565127
HDunuc12_46	MR0354-01.C14	*D. hepta*	10	n=7	MN565128
HDunuc12_47	MR0354-01.C15	*D. hepta*	10	n=7	MN565129
HDunuc12_48	MR0354-01.C2	*D. hepta*	10	n=7	MN565130
HDunuc12_49	MR0354-01.C3	*D. hepta*	10	n=7	MN565131
HDunuc12_50	MR0354-01.C4	*D. hepta*	10	n=7	MN565132
HDunuc12_51	MR0354-01.C5	*D. hepta*	10	n=7	MN565133
HDunuc12_52	MR0354-01.C7	*D. hepta*	10	n=7	MN565134
HDunuc12_53	MR0354-01.C8	*D. hepta*	10	n=7	MN565135
HDunuc12_54	MR0354-01.C9	*D. hepta*	10	n=7	MN565136
HDunuc12_55	MR0355-01.C11	*D. hepta*	10	n=7	MN565137
	MR0355-01.C12	*D. hepta*	10	n=7	=
	MR0355-01.C13	*D. hepta*	10	n=7	=
	MR0355-01.C2	*D. hepta*	10	n=7	=
	MR0355-01.C6	*D. hepta*	10	n=7	=
	MR0355-01.C9	*D. hepta*	10	n=7	=
	MR0091-01.C10	*D. hepta*	6	-	=
	MR0091-01.C14	*D. hepta*	6	-	=
	MR0091-01.C4	*D. hepta*	6	-	=
	MR0091-01.C8	*D. hepta*	6	-	=
HDunuc12_56	MR0355-01.C3	*D. hepta*	10	n=7	MN565138
HDunuc12_57	MR0355-01.C4	*D. hepta*	10	n=7	MN565139

HDunuc12_58	MR0355-01.C5	*D. hepta*	10	n=7	MN565140
HDunuc12_59	MR0355-01.C7	*D. hepta*	10	n=7	MN565141
HDunuc12_60	MR0355-01.C8	*D. hepta*	10	n=7	MN565142
HDunuc12_61	MR0368-01.C1	*D. benazzii*	5	n=8	MN565143
HDunuc12_62	MR0368-01.C12	*D. benazzii*	5	n=8	MN565144
HDunuc12_63	MR0368-01.C14	*D. benazzii*	5	n=8	MN565145
	MR0368-01.C7	*D. benazzii*	5	n=8	=
	MR0370-01.C15	*D. benazzii*	5	n=8	=
	MR0022-05.C7	*D. benazzii*	13	-	=
HDunuc12_64	MR0368-01.C15	*D. benazzii*	5	n=8	MN565146
HDunuc12_65	MR0368-01.C3	*D. benazzii*	5	n=8	MN565147
HDunuc12_66	MR0368-01.C4	*D. benazzii*	5	n=8	MN565148
HDunuc12_67	MR0368-01.C5	*D. benazzii*	5	n=8	MN565149
HDunuc12_68	MR0370-01.C1	*D. benazzii*	5	n=8	MN565150
HDunuc12_69	MR0370-01.C10	*D. benazzii*	5	n=8	MN565151
HDunuc12_70	MR0370-01.C11	*D. benazzii*	5	n=8	MN565152
HDunuc12_71	MR0370-01.C12	*D. benazzii*	5	n=8	MN565153
HDunuc12_72	MR0370-01.C6	*D. benazzii*	5	n=8	MN565154
HDunuc12_73	MR0370-01.C9	*D. benazzii*	5	n=8	MN565155
HDunuc12_74	MR0088-02.C11	*D. benazzii*	5	-	MN565156
HDunuc12_75	MR0088-02.C12	*D. benazzii*	5	-	MN565157
HDunuc12_76	MR0088-02.C15	*D. benazzii*	5	-	MN565158
HDunuc12_77	MR0088-02.C18	*D. benazzii*	5	-	MN565159
HDunuc12_78	MR0088-02.C19	*D. benazzii*	5	-	MN565160
HDunuc12_79	MR0088-02.C20	*D. benazzii*	5	-	MN565161
HDunuc12_80	MR0088-02.C3	*D. benazzii*	5	-	MN565162
HDunuc12_81	MR0088-02.C4	*D. benazzii*	5	-	MN565163
HDunuc12_82	MR0088-02.C5	*D. benazzii*	5	-	MN565164
HDunuc12_83	MR0088-02.C7	*D. benazzii*	5	-	MN565165
HDunuc12_84	MR0088-02.C9	*D. benazzii*	5	-	MN565166
HDunuc12_85	MR0091-01.C1	*D. hepta*	6	-	MN565167
	MR0091-01.C15	*D. hepta*	6	-	=
	MR0091-01.C2	*D. hepta*	6	-	=
	MR0091-01.C6	*D. hepta*	6	-	=
HDunuc12_86	MR0091-01.C11	*D. hepta*	6	-	MN565168
HDunuc12_87	MR0091-01.C12	*D. hepta*	6	-	MN565169
HDunuc12_88	MR0091-01.C13	*D. hepta*	6	-	MN565170
HDunuc12_89	MR0091-01.C3	*D. hepta*	6	-	MN565171
HDunuc12_90	MR0091-01.C5	*D. hepta*	6	-	MN565172
HDunuc12_91	MR0091-01.C7	*D. hepta*	6	-	MN565173
HDunuc12_92	MR0091-01.C9	*D. hepta*	6	-	MN565174
HDunuc12_93	MR0092-03.C1	*D. benazzii*	15	-	MN565175
	MR0092-11.C13b	*D. sp*	15	-	=

	MR0092-11.C14b	*D. sp*	15	-	=
	MR0092-11.C18b	*D. sp*	15	-	=
	MR0092-11.C21b	*D. sp*	15	-	=
HDunuc12_94	MR0092-03.C12	*D. benazzii*	15	-	MN565176
HDunuc12_95	MR0092-03.C15	*D. benazzii*	15	-	MN565177
HDunuc12_96	MR0092-03.C2	*D. benazzii*	15	-	MN565178
HDunuc12_97	MR0092-03.C5	*D. benazzii*	15	-	MN565179
HDunuc12_98	MR0092-03.C6	*D. benazzii*	15	-	MN565180
HDunuc12_99	MR0092-04.C10b	*D. hepta*	15	-	MN565181
HDunuc12_100	MR0092-04.C11	*D. hepta*	15	-	MN565182
HDunuc12_101	MR0092-04.C11b	*D. hepta*	15	-	MN565183
HDunuc12_102	MR0092-04.C12b	*D. hepta*	15	-	MN565184
HDunuc12_103	MR0092-04.C13	*D. hepta*	15	-	MN565185
HDunuc12_104	MR0092-04.C13b	*D. hepta*	15	-	MN565186
	MR0092-04.C17b	*D. hepta*	15	-	=
HDunuc12_105	MR0092-04.C14	*D. hepta*	15	-	MN565187
HDunuc12_106	MR0092-04.C14b	*D. hepta*	15	-	MN565188
HDunuc12_107	MR0092-04.C15	*D. hepta*	15	-	MN565189
HDunuc12_108	MR0092-04.C15b	*D. hepta*	15	-	MN565190
HDunuc12_109	MR0092-04.C16b	*D. hepta*	15	-	MN565191
HDunuc12_110	MR0092-04.C18b	*D. hepta*	15	-	MN565192
	MR0092-04.C7b	*D. hepta*	15	-	=
HDunuc12_111	MR0092-04.C19b	*D. hepta*	15	-	MN565193
HDunuc12_112	MR0092-04.C1b	*D. hepta*	15	-	MN565194
HDunuc12_113	MR0092-04.C2	*D. hepta*	15	-	MN565195
	MR0025-02.C11	*D. benazzii*	3	-	=
HDunuc12_114	MR0092-04.C20b	*D. hepta*	15	-	MN565196
HDunuc12_115	MR0092-04.C2b	*D. hepta*	15	-	MN565197
HDunuc12_116	MR0092-04.C3b	*D. hepta*	15	-	MN565198
HDunuc12_117	MR0092-04.C4b	*D. hepta*	15	-	MN565199
HDunuc12_118	MR0092-04.C5b	*D. hepta*	15	-	MN565200
HDunuc12_119	MR0092-04.C6	*D. hepta*	15	-	MN565201
HDunuc12_120	MR0092-04.C6b	*D. hepta*	15	-	MN565202
HDunuc12_121	MR0092-04.C9b	*D. hepta*	15	-	MN565203
HDunuc12_122	MR0092-05.C10	*D. sp*	15	-	MN565204
HDunuc12_123	MR0092-05.C3	*D. sp*	15	-	MN565205
HDunuc12_124	MR0092-05.C4	*D. sp*	15	-	MN565206
HDunuc12_125	MR0092-05.C6	*D. sp*	15	-	MN565207
HDunuc12_126	MR0022-04.C10	*D. benazzii*	13	-	MN565208
HDunuc12_127	MR0022-04.C11	*D. benazzii*	13	-	MN565209
	MR0022-04.C14	*D. benazzii*	13	-	=
	MR0022-04.C3	*D. benazzii*	13	-	=
	MR0022-04.C4	*D. benazzii*	13	-	=

	MR0022-05.C19	*D. benazzii*	13	-	=
	MR0022-05.C5	*D. benazzii*	13	-	=
	MR0022-05.C20	*D. benazzii*	13	-	=
HDunuc12_128	MR0022-04.C12	*D. benazzii*	13	-	MN565210
HDunuc12_129	MR0022-04.C13	*D. benazzii*	13	-	MN565211
HDunuc12_130	MR0022-04.C15	*D. benazzii*	13	-	MN565212
HDunuc12_131	MR0022-04.C2	*D. benazzii*	13	-	MN565213
HDunuc12_132	MR0022-04.C5	*D. benazzii*	13	-	MN565214
HDunuc12_133	MR0022-04.C6	*D. benazzii*	13	-	MN565215
HDunuc12_134	MR0022-04.C7	*D. benazzii*	13	-	MN565216
HDunuc12_135	MR0022-04.C8	*D. benazzii*	13	-	MN565217
HDunuc12_136	MR0022-04.C9	*D. benazzii*	13	-	MN565218
HDunuc12_137	MR0092-11.C11	*D. sp*	15	-	MN565219
HDunuc12_138	MR0092-11.C12	*D. sp*	15	-	MN565220
HDunuc12_139	MR0092-11.C13b	*D. sp*	15	-	MN565221
HDunuc12_140	MR0092-11.C15b	*D. sp*	15	-	MN565222
	MR0092-11.C17b	*D. sp*	15	-	=
	MR0092-11.C23b	*D. sp*	15	-	=
HDunuc12_141	MR0092-11.C16b	*D. sp*	15	-	MN565223
HDunuc12_142	MR0092-11.C22b	*D. sp*	15	-	MN565224
HDunuc12_143	MR0092-11.C28b	*D. sp*	15	-	MN565225
HDunuc12_144	MR0092-11.C29b	*D. sp*	15	-	MN565226
HDunuc12_145	MR0092-11.C25b	*D. sp*	15	-	MN565227
HDunuc12_146	MR0092-11.C7	*D. sp*	15	-	MN565228
HDunuc12_147	MR0022-05.C4b	*D. benazzii*	13	-	MN565229
HDunuc12_148	MR0022-05.C6b	*D. benazzii*	13	-	MN565230
HDunuc12_149	MR0022-05.C11	*D. benazzii*	13	-	MN565231
HDunuc12_150	MR0022-05.C13	*D. benazzii*	13	-	MN565232
HDunuc12_151	MR0022-05.C1	*D. benazzii*	13	-	MN565233
HDunuc12_152	MR0022-05.C14b	*D. benazzii*	13	-	MN565234
	MR0022-05.C8	*D. benazzii*	13	-	=
HDunuc12_153	MR0022-05.C15	*D. benazzii*	13	-	MN565235
HDunuc12_154	MR0022-05.C16	*D. benazzii*	13	-	MN565236
HDunuc12_155	MR0022-05.C17	*D. benazzii*	13	-	MN565237
	MR0022-05.C18	*D. benazzii*	13	-	=
	MR0022-05.C2	*D. benazzii*	13	-	=
HDunuc12_156	MR0022-05.C3	*D. benazzii*	13	-	MN565238
HDunuc12_157	MR0022-05.C4	*D. benazzii*	13	-	MN565239
HDunuc12_158	MR0022-05.C9	*D. benazzii*	13	-	MN565240
HDunuc12_159	MR0025-02.C1	*D. benazzii*	3	-	MN565241
HDunuc12_160	MR0025-02.C14	*D. benazzii*	3	-	MN565242
HDunuc12_161	MR0025-02.C15	*D. benazzii*	3	-	MN565243
	MR0025-02.C4	*D. benazzii*	3	-	=

	MR0025-02.C5	*D. benazzii*	3	-	=
HDunuc12_162	MR0025-02.C2	*D. benazzii*	3	-	MN565244
HDunuc12_163	MR0025-02.C6	*D. benazzii*	3	-	MN565245
HDunuc12_164	MR0025-02.C7	*D. benazzii*	3	-	MN565246
HDunuc12_165	MR0025-02.C8	*D. benazzii*	3	-	MN565247
HDunuc12_166	MR0025-02.C9	*D. benazzii*	3	-	MN565248
HDunuc12_167	MR0030-06.C1	*D. hepta*	14	-	MN565249
	MR0030-06.C11	*D. hepta*	14	-	=
	MR0030-06.C14	*D. hepta*	14	-	=
HDunuc12_168	MR0030-06.C12	*D. hepta*	14	-	MN565250
HDunuc12_169	MR0030-06.C13	*D. hepta*	14	-	MN565251
HDunuc12_170	MR0030-06.C15	*D. hepta*	14	-	MN565252
HDunuc12_171	MR0030-06.C2	*D. hepta*	14	-	MN565253
	MR0030-06.C4	*D. hepta*	14	-	=
	MR0030-06.C8	*D. hepta*	14	-	=
HDunuc12_172	MR0030-06.C3	*D. hepta*	14	-	MN565254
HDunuc12_173	MR0030-06.C5	*D. hepta*	14	-	MN565255
HDunuc12_174	MR0030-06.C6	*D. hepta*	14	-	MN565256
HDunuc12_175	MR0030-06.C7	*D. hepta*	14	-	MN565257
HDunuc12_176	MR0030-06.C9	*D. hepta*	14	-	MN565258

Chapter II

Chapter II.1

Fantastic beasts and how to delimit them: an integrative approach using multispecies coalescent methods reveals twonew, endemic *Dugesia* species (Platyhelminthes: Tricladida) from Corsica and Sardinia

Dols-Serrate, D., Stocchino, G.A., Sluys, R., Riutort, M.

2023. *Zoological Journal of the Linnean Society*, in press.
doi: 10.1093/zoolinnean/zlad135

Chapter II.2

New insights into the evolution and biogeography of freshwater planarians on islands in the Tyrrhenian Sea, Western Mediterranean Basin, with the integrative description of a new endemic species from Corsica (Platyhelminthes, Tricladida, *Dugesia*)

Dols-Serrate, D., Stocchino, G.A., Nuin-Villabona, P., Sluys, R., Riutort, M.

2024. *Zoological Journal of the Linnean Society*, accepted.
doi: not available

Zoological Journal of the Linnean Society, 2023, XX, 1–29
https://doi.org/10.1093/zoolinnean/zlad135
Advance access publication 18 November 2023
Original Article

OXFORD

Original Article

Fantastic beasts and how to delimit them: an integrative approach using multispecies coalescent methods reveals two new, endemic *Dugesia* species (Platyhelminthes: Tricladida) from Corsica and Sardinia

Daniel Dols-Serrate[1,2,*], Giacinta Angela Stocchino[3], Ronald Sluys[4], Marta Riutort[1,2]

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain
[2]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain
[3]Dipartimento di Medicina Veterinaria, Via Vienna 2, University of Sassari, Sassari I-07100, Italy
[4]Naturalis Biodiversity Center, PO Box 9517, 2300 RA Leiden, The Netherlands

[*]Corresponding author. Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain. E-mail: danieldols@ub.edu/ bcn27101993@gmail.com

ABSTRACT

Dugesia hepta and *D. benazzii* are two species found in Sardinia and Corsica. Previous studies have been unable to resolve their evolutionary relationships and there was doubt about the monophyly of *D. benazzii*. This study used molecular and morphological data to develop a rigorous species-delimitation procedure within an integrative framework. Three different species discovery methods (ABGD, GMYC, and mPTP) were applied on single-locus data to formulate primary species hypotheses. Secondary species hypotheses were proposed based on two Bayesian species-validation methods (BPP and BFD), morphological, and karyological traits. This study presents evidence supporting the recognition of two new species: *Dugesia mariae* from Corsica and *Dugesia hoidi* from Sardinia. A redescription of *D. benazzii* is also reported. The phylogenetic relationships between these four species were resolved with high support, except for *D. hoidi*. The performance of the different molecular-delimitation methodologies is discussed, as well as the presence of ciliate parasites in *D. benazzii* and *D. mariae*.

Keywords: Corsica; *Dugesia*; endemism; integrative taxonomy; morphology; planarian phylogeny; Sardinia; species delimitation

INTRODUCTION

Species play a crucial role in our understanding of the natural world. They serve as the foundation of systematic studies and hold immense value in a variety of other disciplines such as ecology, evolution, conservation biology, and biogeography (Agapow, 2005). Not only do species have a direct impact on our socio-economic activities, but they also offer us a unique window into the intricacies of evolution and the processes that have shaped both species and their environments.

Species delimitation in freshwater triclads has always been rather difficult, notably for species of *Dugesia* Girard, 1850, due to great similarities in external appearance. Generally, detailed study of their anatomy, particularly the reproductive apparatus, supplies sufficient characters for the recognition and delimitation of species. However, even in their reproductive system, species may be very similar, rendering proper identification a painstaking and time-consuming activity that demands expertise and patience. When such anatomical information falls short, karyological data have proven to be a useful additional source of evidence (Leria *et al.*, 2020).

Furthermore, the identification of species of *Dugesia* can become particularly complicated in two situations. First, there are species with fully asexually reproducing populations (Stocchino and Manconi, 2013). Such animals lack a reproductive apparatus and, thus, cannot be identified morphologically. Second, it is well known that species or populations of freshwater planarians may switch between reproductive modes, usually between different seasons. Consequently, during a specific period of the

Received 17 March 2023; revised 4 September 2023; accepted 11 September 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of The Linnean Society of London. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

year, one may exclusively find asexual specimens, which will regrow their reproductive apparatus at another time of the year. In addition, starvation or abiotic stressors may induce the reabsorption of the reproductive apparatus (Ball and Reynoldson, 1981; Newmark and Alvarado, 2002). This reproductive plasticity hampers the identification of the specimens (Álvarez-Presas and Riutort, 2014).

The use of molecular markers and an integrative taxonomic approach may contribute to the resolution of such complicated matters of species delimitation in species of *Dugesia*. Unfortunately, there are only few studies in which delimitation of freshwater planarian species has been performed within an integrative framework, using molecular discovery and/or validation methods (e.g. Sluys *et al.*, 2013; Leria *et al.*, 2020; Stocchino *et al.*, 2021; Wang *et al.*, 2022). In the present paper, we document a new study on integrative species delimitation in the genus *Dugesia*, in which nominal *Dugesia benazzii* Lepori, 1951 and *D. hepta* Pala, Casu & Vacca, 1981 form our model species.

Dugesia hepta is an endemic species of Sardinia found in four northern hydrographic basins, where it often co-occurs with *D. benazzii*, which has a wider Tyrrhenian distribution and is also present on Corsica. *Dugesia hepta* was described on the basis of anatomical and karyological data (Pala *et al.*, 1981; Stocchino *et al.*, 2005). In particular, its peculiar chromosome number (n = 7; 2n = 14) is unique amongst *Dugesia*'s in the Western Palearctic region and, thereby, makes it easily distinguishable from *D. benazzii*. A recent study based on novel molecular data (Dols-Serrate *et al.*, 2020) confirmed that *D. hepta* is monophyletic. In contrast, *D. benazzii* exhibits a large number of karyological biotypes, ranging from diploid, triploid, tetraploid, and hexaploid to aneuploid populations (Lepori, 1951; Pala *et al.*, 1982, 1999; Vacca *et al.*, 1993; Stocchino, 2018), while novel mitochondrial data did not support its presumed monophyletic status (Dols-Serrate *et al.*, 2020). In the last-mentioned study, nominal *D. benazzii* turned out to be a paraphyletic taxon, based on mitochondrial information, with a clear split between Corsican and Sardinian populations. In contrast, nuclear information, albeit short on Corsican representatives, identified *D. benazzii* as a monophyletic group. However, there was a significant level of differentiation between populations of both islands. Nevertheless, a proper species-delimitation analysis was not performed in that study and, therefore, this has become one of the primary goals of the present analysis. Due to the problematic monophyletic status of nominal *D. benazzii*, in the following we will use the term '*D. benazzii* sensu lato (s.l.)' when referring to all its included populations.

Some degree of consensus has arisen on the notion that species should be considered as independently evolving metapopulation lineages, as implied by the general lineage species concept (de Queiroz, 1998). This concept accommodates previous species concepts as sources of evidence that act as recognition criteria when delineating independently evolving lineages (de Queiroz, 2007; Frankham *et al.*, 2012). However, speciation is a continuous process (de Queiroz, 2007), implying that some degree of uncertainty and incongruence may be expected when only one source of evidence is used, e.g. genetic data (Maddison, 1997). Therefore, it is important to include multiple lines of evidence (molecular, morphology, karyology, etc.) in an integrated taxonomic approach to species delimitation (Dayrat, 2005). In this way, the robustness of a species hypothesis will depend on the degree of congruence between the various recognition criteria.

The present paper documents our rigorous species-delimitation efforts on *Dugesia* species from Corsica and Sardinia, based on the procedure detailed in Leria *et al.* (2020) and on the molecular datasets in Dols-Serrate *et al.* (2020), as well as new sequences obtained in the present study for two nuclear markers and four mitochondrial genes. Morphological and karyological data were used to evaluate the validity of our hypotheses on primary species candidates (PSCs). Through our analysis, we were able to compare the performance of three molecular species discovery methods in formulating primary species hypotheses (PSHs) for our datasets and model organisms. We have also used two Bayesian coalescent-based validation methods in order to formulate rigorous secondary species hypotheses (SSHs). Furthermore, a redescription of *D. benazzii* is also reported. Since species discovery and description are fundamental for a wide array of biological disciplines (Agapow, 2005), it is highly important to deepen our knowledge on species diversity and on how to accurately delimit species taxa.

MATERIALS AND METHODS

Taxon sampling

For the present study, *D. benazzii* s.l. and *D. hepta* specimens were sampled between 1993 and 2013 from 17 localities throughout the islands of Corsica and Sardinia (Fig. 1; Appendix, Table A1). For each locality, some specimens were fixed and preserved in absolute ethanol for molecular analyses, whilst others were kept alive and were taken to the laboratory for morphological and karyological analyses. Of the 77 specimens analysed in this study, 50 had already been sampled for a previous study (Dols-Serrate *et al.*, 2020). However, we broadened our dataset by obtaining and adding new sequences for additional nuclear and mitochondrial genes for those previously sampled specimens, as well as for 27 new ones (Appendix, Table A1). Methods for DNA extraction and karyotyping have already been described in Dols-Serrate *et al.* (2020).

In the present study, we avoided using those specimens that had been identified in Dols-Serrate *et al.* (2020) as hybrids between nominal *D. hepta* and *D. benazzii* s.l., as the inclusion of hybrids might blur the phylogenetic signal for distinguishing between species and thus cause misleading results and incorrect conclusions. Furthermore, these hybrids did not produce their own unique nuclear phylogenetic signal, but rather presented haplotypes from both parental species that were only apparent when phased via cloning. With respect to the mitochondrial information, hybrids would result only in individuals being assigned to the ovule donor species or, in some cases, to a clade of uncertain relationships (Dols-Serrate *et al.*, 2020).

Assignment of sampled individuals to species

Individuals of *D. benazzii* s.l. and *D. hepta* from Sardinia (where both species coexist) were identified on the basis of a molecular criterion by using *ITS1* sequences, alongside karyological data (cf. Dols-Serrate *et al.*, 2020). Planarian samples from Corsica

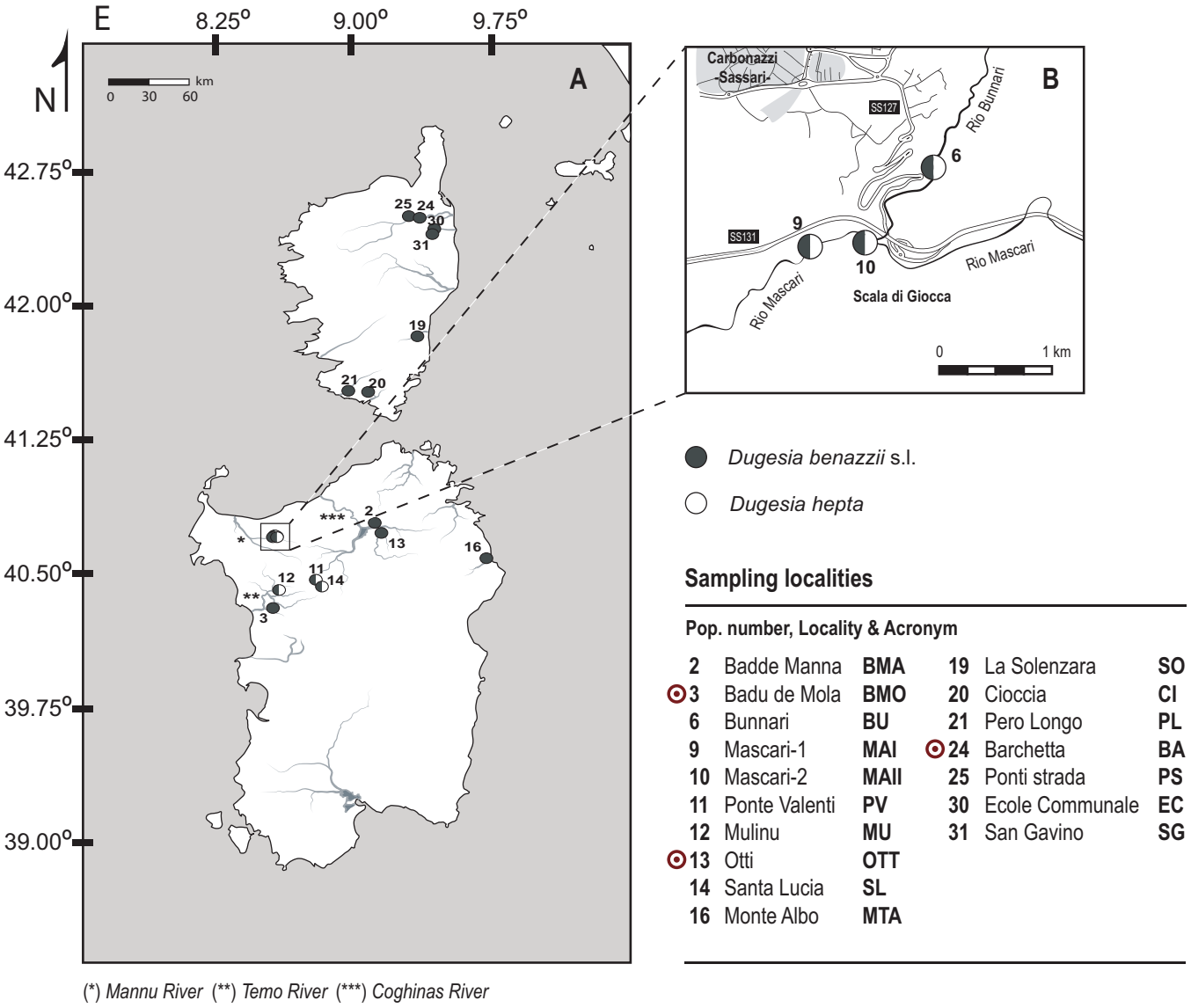


Figure 1. A, sampling localities of the present study. Numbers correspond to population codes listed in Appendix, Table A1 and coincide with those in Dols-Serrate *et al.* (2020). Red circles indicate populations used for morphological analyses. B, rectangular inset: enlargement of the Bunnari–Mascari confluence area. The map was created using Q-GIS v.3.2.2 (https://qgis.org/es/site/ last accessed September 2023) and edited in ILLUSTRATOR CC v.22.0.1 (https://www.adobe.com/products/illustrator.html last accessed September 2023).

were identified according to external morphology, since *D. benazzii* s.l. is distinguishable from the other dugesiid that inhabits the island, viz., *Schmidtea mediterranea* (Lázaro *et al.*, 2011).

Molecular analysis

DNA extraction, quantification, and sequence amplification

For the new samples, genomic DNA was extracted from whole specimens. Samples were digested using proteinase K (5–10 µL/sample) and, subsequently, purified using a Wizard® Genomics DNA Purification Kit (Promega Corporation), following the manufacturer's instructions. DNA quantification was performed for each sample using a spectrophotometer NanoDrop™ 1000 (Thermo Fisher Scientific Inc.), making use of 2 µL/sample.

Specific primers were used to amplify four mitochondrial and two nuclear genes: cytochrome *c* oxidase subunit I (*COI*) in one fragment of approximately 650 bp; NADH dehydrogenase subunit 1 (*NADH1*), tryptophan tRNA and cytochrome *c* oxidase subunit II (*COII*) in a single fragment of approximately 630 bp; the *Dugesia* nuclear gene 10 (*Dunuc10*, anonymous marker) of 450 bp; and *Dugesia* nuclear gene 12 (*Dunuc12*, a transmembrane emp24 domain-containing protein) in a fragment of 670 bp. Both nuclear markers were developed from low-coverage genomic data in a previous study on another group of *Dugesia* species (Leria *et al.*, 2020). Sequences and annealing temperatures of each pair of primers are given in Table 1. Final polymerase chain reaction (PCR) volume for all markers was 25 µL, consisting of: (i) 5 µL of Promega 5x Green GoTaq Flexi Buffer, (ii) 2.5 µL of $MgCl_2$ (25 mM), (iii) 1.5 µL of dNTPs (0.5 mM), (iv) 1 µL of each primer (10 µM), (v) 0.15 µL of *Taq* polymerase (5 u/µL) (GoTaq® Flexi DNA Polymerase, Promega), and (vi) 1 µL of DNA extraction sample. Autoclaved milli Q water was

added to obtain the final PCR volume. PCR conditions consisted of (i) a denaturalization step at 94°C for 2 min; (ii) 39 cycles comprising a 45-s denaturalization step at 94°C, a 50-s annealing step at the specific temperature per set of primers (Table 1), and a 1-min elongation step at 72°C; and (iii) a final elongation step at 72°C for 3 min. The resulting PCR products were checked on a 1% agarose gel to assess the correct amplification of the different molecular markers.

Viable PCR products were purified using a vacuum system (MultiScreen™ HTS Vacuum Manifold of Millipore). Sequencing reactions were performed using Big Dye (3.1, Applied Biosystems) and were run in an automated sequencer ABI Prism 3730 (Unitat de Genòmica dels Serveis Científico-Tècnics de la Universitat de Barcelona) or at Macrogen Inc. (Spain). Both chains were sequenced using the same primers used for amplification. Obtained chromatograms were visually checked with GENEIOUS R8 (Biomatters, http://www.geneious.com/ last accessed September 2022) and then contig and consensus sequences were obtained.

Phylogenetic analysis: gene tree inference

Sequences were aligned using the online software MAFFT v.7 (Katoh and Standley, 2013) using the default parameters and were posteriorly revised in BioEdit v.7.2.5 (Hall, 1999). Prior to analyses, the mitochondrial protein-coding sequences were translated into amino acids to check that there were no stop codons. Furthermore, nucleotide saturation was tested for the first, second, and third positions of the protein-coding mitochondrial genes, using the Xia *et al.* (2003) test implemented in DAMBE v.7.0.28 (Xia and Xie, 2001).

Phylogenetic trees were inferred using both Bayesian inference (BI) and maximum likelihood (ML) methods in MrBayes v.3.2.6 (Ronquist *et al.*, 2012) and RAxML 7.0.3 (Stamatakis, 2006), respectively. Five datasets were prepared for these analyses (Table 2): (i) concatenated mitochondrial genes (dataset I), (ii) nuclear marker *Dunuc10* (dataset II), (iii) *Dunuc12* marker (dataset III), (iv) concatenations of the nuclear genes (dataset IV), and (v) concatenation of nuclear and mitochondrial genes (dataset V). Dataset I was divided into four partitions (by genes), datasets II and III into two partitions (by intron and exon), dataset IV into two partitions (by genes), and dataset V into six partitions (by genes). We did not include a by codon position partition for protein-coding genes in order to avoid an excess of partitions and, consequently, an excess of parameters that may have led to higher error rates (Posada, 2009). Nonetheless, we took into account the putative differences in evolutionary rate among sites within each gene by using a Gamma distribution, as indicated in the AIC criterion test (see below). All datasets included sequences of *D. subtentaculata* (Draparnaud, 1801) as outgroup (Appendix, Table A1). The best substitution model for each partition (Table 2) was selected based on the Akaike information criterion (AIC) in PartitionFinder2 (Lanfear *et al.*, 2017), using a greedy search model (Lanfear *et al.*, 2012). For BI analyses, two runs were performed simultaneously with four Markov chains, starting from random trees and running for 20 000 000 generations and sampling every 1000 generations. Convergence was surveyed

Table 1. List of specific primers used for amplification and sequencing in the present study.

Name		Marker(s)	Sequence 5′–3′	Annealing Temp. (°C)	Source
SamCF	F[1]	COI	GCTAATAATTTGAGTTTTTG	51	Dols-Serrate *et al.* 2020
TarlyCR	R[2]	COI	CATTTTAAAACAACATTACC	51	Dols-Serrate *et al.* 2020
Elo12F	F	Dunuc12	AAGCTATTTATAATTCAGCG	54	Dols-Serrate *et al.* 2020
Hell12R	R	Dunuc12	AAAGAAATTGCTGCTAAAG	54	Dols-Serrate *et al.* 2020
Finsl_D10F	F	Dunuc10	TCCGGTAAGATTAATG	48	Present study
Sous_D10R	R	Dunuc10	TCAACAGATCCGACC	48	Present study
Cupi2	F	Nad1-tRNAW-COII (Mit)	ATTTGRCTYCGTTCTTC	43	Leria *et al.* 2020
coIIP	R	Nad1-tRNAW-COII (Mit)	CCACAAACCTCAGAAC	43	Leria *et al.* 2020

[1]Forward [2]Reverse.

Table 2. Datasets used in the present study, and best fitting substitution model for partition scheme and marker.

Dataset	Marker	Partition scheme	Substitution model
Dataset I	COI-MIT	COI (1-640 bp)	GTR+I+G
		NAD1 (641-699 bp)	HKY+G
		tRNA-W (700-760 bp)	HKY
		COII (761-1268 bp)	GTR+G
Dataset II	Dunuc10	Intron (1-147 bp)	HKY+G
		Exon (148-429 bp)	SYM+I+G
Dataset III	Dunuc12	Exon (1-84 bp)	HKY+G
		Intron (85-661 bp)	HKY+G
Dataset IV	Dataset III + II	–	–
Dataset V	Dataset I + IV	–	–

by checking that the standard deviation of the split frequencies reached a value below 0.01. The first 25% of the sampled trees was discarded as burn-in. ML analyses were performed using a GTR+G substitution model (since GTR is the only available model in RAxML) to all partitions with 1000 replicates in order to compute bootstrap supports for the nodes by conducting a rapid bootstrap analysis. ML search was performed by starting from a random tree. The recovered trees from BI and ML analyses were visualized with FigTree v.1.4.2 (https://github.com/rambaut/figtree/releases last accessed September 2023).

Phylogenetic analysis: alternative topology test

Once the mtDNA gene tree was inferred based on dataset I, a plausible alternative topology hypothesis was tested in which the specimen from Badu de Mola was grouped with the individuals from Sardinia and not with the ones from Corsica. A ML search using a GTR+G substitution model without bootstrapping was performed for 50 runs to find the best tree based on dataset I in RaxML-GUI 1.5b1 (Silvestro and Michalak, 2012). A constraint tree with the alternative topology hypothesis was built and the best constraint tree was inferred according to a ML search for 50 runs. Afterwards, per site log-likelihoods were computed for both best-recovered trees. Ultimately, the corresponding tests for topological analysis were performed in CONSEL (Shimodaira and Hasegawa, 2001).

Integrative taxonomic methodology

Our integrative taxonomic approach was based on the procedure proposed in Leria *et al.* (2020). We formulated PSHs and established PSCs to be tested, based on the results of three discovery methods that fulfilled the following criteria: (i) the split *hepta–benazzii* s.l. must prevail in the partition scheme, and (ii) the partition scheme should be phylogenetically congruent, i.e. we allowed oversplitting to occur only when the resulting groups were monophyletic in our phylogenetic trees. We considered this option to be preferable over an extra-conservative attitude during the formulation of PSHs, since validation methods more quickly collapse two putative species into one than they resolve a complex clade into phylogenetically relevant subclades. For this first step we used dataset I and the following methods: (i) automated barcode gap discovery (ABGD) (Puillandre *et al.*, 2012), (ii) generalized mixed Yule coalescent (GMYC) (Pons *et al.*, 2006), and (iii) multi-rate Poisson tree processes (mPTP) (Kapli *et al.*, 2017). Although ABGD and GMYC are mainly developed for single-locus environments, we still used dataset I, since mitochondrial markers are linked and can be treated as a single locus.

Afterwards, the PSCs were tested by using different sources of evidence: (i) multi-locus data (datasets I, IV, and V), using both the coalescent-based methods implemented in Bayesian phylogenetics and phylogeography software (BPP; Yang and Rannala, 2010), and the Bayes' factors delimitation test (BFD; Grummer *et al.* 2014); (ii) morphological data; (iii) karyological data. Based on the results of these multiple sources of information we formulated the SSHs, from which relevant taxonomical decisions could be derived.

Molecular species delimitation: discovery

Automated barcode gap discovery

ABGD (Puillandre *et al.*, 2012) was used to detect potential barcode gaps and to partition the data into different groups of taxa as putative species for the mitochondrial sequences (dataset I). This analysis was performed on a website interface (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html; last accessed September 2022), using a gap width of 1.4. The prior for the maximum value of intraspecific divergence (P) was set to range from 0.001 to 0.1, with the number of recursive steps set to default (10). The distance matrix was built in MEGA X (Kumar *et al.*, 2018), using the *p*-distance with pairwise deletion for gaps/missing data.

Generalized mixed Yule coalescent

GMYC (Pons *et al.*, 2006) is a tree-based method for species discovery that requires an ultrametric tree as input. This tree was built using BEAST v.1.8.4 (Drummond *et al.*, 2012) based on dataset I. First, the dataset's phylogenetic signal was assessed by running two analyses with default prior parametrization—one with the actual data and one 'without' data and sampling from the priors—for 200 million generations and sampling every 20 000 generations. The distribution of the posterior and prior density probabilities of both runs was checked in TRACER v.1.6 (Rambaut *et al.*, 2014) to assess whether the priors ruled over the data, implying the lack of relevant phylogenetic information within the data itself, or not. Afterwards, eight different models were tested (Supporting Information, Table S1) with different combinations for the tree prior, the molecular clock, and the substitution rate—the latter only for the *COI* partition—by means of marginal likelihood estimates (MLE). MLEs were computed using the path sampling (PS) (Baele *et al.*, 2012) and stepping-stone (SS) (Xie *et al.*, 2011) methods. To enable comparison between different models, Bayes' factors (Kass and Raftery, 1995)—hereafter indicated as *2lnBf*—were calculated according to the Grummer *et al.* (2014) guidelines in order to compare and choose the best-fitting model ($2lnBf = 2 \times$ [best-fitting model's MLE-model *j*'s MLE]). The length of the MCMC chain for all analyses was 200 million, with samplings every 20 000 generations, while the MLE estimation was performed via an additional run of 1 million generations of 100 path-steps (100 million generations) per analysis. All BEAST analyses were run in CIPRES Science Gateway (Miller *et al.*, 2010).

The strength of support from the *2lnBf* comparisons was evaluated on the basis of Kass and Raftery's (1995) framework. Thus, values of *2lnBf* from 0 to 2 were considered as non-significant, from 2 to 10 as positively supported, and >10 as decisively supported. Support, in this case, indicates that the alternative model *j* is indeed different from the best-fitting model and, thus, worse.

Once the best-fitting model was asserted, the maximum clade credibility tree for that analysis was constructed via TreeAnnotator v.1.6.1 (Rambaut and Drummond, 2010) with a default burn-in of the first 25% of the trees sampled from BEAST. GMYC analysis was executed with the SPLITS package (Ezard *et al.*, 2009) in R, using the single threshold time approach.

Multi-rate Poisson tree processes

The mPTP method takes as input a binary, rooted phylogenetic tree inferred with a ML or a BI approach. In this case, the ML analysis of dataset I was used for the execution of the mPTP analysis. Four independent runs starting from random trees were performed with a MCMC chain length of 10 million, sampled every 5000 generations. Minimal branch length was set as

default (0.0001). Convergence of the analysis was surveyed by checking that the average standard deviation of the per-node delimitation support values across the four independent MCMC runs was below 0.01 (a.k.a. average standard deviation of delimitation support values, or ASDDSV).

Integrative species validation

Bayesian methods and molecular validation

Two Bayesian species-delimitation approaches were explored using multi-locus data, viz., BPP and BFD. For both approaches, an a priori hypothesis regarding the population structure, or a plausible putative species hypothesis, is required. This hypothesis was established on the results yielded by ABGD, GMYC, and mPTP.

BPP species delimitation was performed using the software BPP v.4.3 (Yang and Rannala, 2010; Flouri *et al.*, 2020) and conducting the analyses with a fixed guide tree, viz., model A10. In this model a reversible-jump MCMC algorithm (rjMCMC) is used to move amongst the multiple species delimitation models compatible with the fixed guide tree (Yang, 2015). For this analysis, datasets I, IV, and V were used. The outgroup sequences were excluded and each partition for every dataset was treated as a single locus (dataset I = four mitochondrial loci; dataset IV = two nuclear loci; dataset V = six loci overall). BPP includes two types of parameters that follow an inverse gamma prior distribution (IG(α, β)), viz., the ancestral population size (θ), and the species divergence time (τ). Four prior combinations (p.c.) were tested per dataset following the guidelines of Yang (2015). In all cases, we worked with a diffuse prior of $\alpha = 3$ and adjusting β. The first combination of priors assumed a relatively small ancestral population size $\theta_1 \sim$ IG(3, 0.002) and a shallow divergence time $\tau_1 \sim$ IG(3, 0.002). The second combination assumed a relatively large ancestral population size $\theta_2 \sim$ IG(3, 0.1) and a shallow divergence time $\tau_2 \sim$ IG(3, 0.002). The third combination assumed a relatively large ancestral population size $\theta_3 \sim$ IG(3, 0.1) and a deep divergence time $\tau_3 \sim$ IG(3, 0.1). The final combination assumed a relatively small ancestral population size $\theta_4 \sim$ IG(3, 0.002) and a deep divergence time $\tau_4 \sim$ IG(3, 0.1). Analyses under each p.c. were run under both rjMCMC algorithm 0 (with a fine-tunning prior $\varepsilon = 2$) and algorithm 1 (with fine-tuning priors $\alpha = 2$ and $m = 1$). Each analysis was run four times: twice using different random seeds to confirm that the results were stable across runs and twice using two different fixed guide tree topologies. The rjMCMC was run for 500 000 generations, while sampling every five generations. Burn-in was set to 10 000 generations.

BFD was performed following the Grummer *et al.* (2014) framework. This approach is based on a Bayesian coalescent reconstruction of the species trees for a range of delimitation models that involve the lumping and splitting of hypothesized species. MLEs are computed for each hypothetic species-delimitation scenario tested, and models are then evaluated following a *2lnBf* pairwise comparison to assess whether the best-fitting model is significantly better than the rest. For this analysis, datasets I, IV, and V were used. Species tree inference was performed in BEAST 1.8.4 enabling the *BEAST option and setting MCMC chains to run for 200 million generations. Six species-delimitation models were tested. For each model, individuals were assigned to different species with successively increased lumping of the hypothesized groups, thus ranging from model 1 (the species-delimitation hypothesis contemplating *D. benazzii* s.l. split into all PSCs, the same a priori hypothesis for BPP analyses) to model 6 (one single *D. benazzii* species). For dataset I, parametrization was performed applying the best-fitting model specified in the Supporting Information, Table S1. For dataset IV, the HKY+G substitution model and an uncorrelated log-normal relaxed molecular clock with the clock rate fixed at 1.0 was employed for all partitions. Birth-and-death process tree prior was used to infer the species tree. For dataset V, the parametrizations for dataset I and dataset IV were applied to the mitochondrial loci and the nuclear loci, respectively. In all analyses, a piecewise linear and constant root population size model was used, as well as a gamma distribution for the *popMean* prior with mean and standard deviation values set as default. After the standard MCMC chain was finished, MLE calculation was performed using both PS and SS via an additional run of 4 million generations of 100 path-steps, resulting in 400 million generations. To ensure the obtention of values of effective sample size (ESS) for the parameters of MLE higher than 200, and to increase the accuracy of the PS and SS estimates, the analyses were repeated five times for each species-delimitation model in BEAST 1.8.4 on the CIPRES Science Gateway with different random seeds. The final marginal likelihoods were estimated based on the combination of the five runs, and the best-fitting species-delimitation model was chosen based on the *2lnBf* comparisons.

Morphological data

To perform the morphological study, sexual specimens were fixed for 24 h in Bouin's fluid, dehydrated in an ascending ethanol series, cleared in toluene or clove oil, and embedded in synthetic wax. Serial sections were made at intervals of 7–8 μm and were stained with haematoxylin eosin (Harris) or Mallory–Cason (cf. Winsor and Sluys, 2018). Reconstructions of the copulatory apparatus were obtained by using a compound microscope provided with a camera lucida.

Type material and other specimens examined are deposited in the Giacinta A. Stocchino collection (CGAS), at the University of Sassari and at the Naturalis Biodiversity Center, Leiden, The Netherlands (RMNH code).

Karyological data

Chromosome metaphasic plates were obtained by the squashing method. Regenerative blastemas of single caudal fragments were first treated with a solution of colchicine (0.3%) for 4 h, then transferred onto glass slides and treated with a solution of acetic acid (5%) for 5 min. Subsequently, they were stained with acetic orcein for 2 h and squashed under a small coverslip.

Abbreviations used in morphological figures

abc, angled bursal canal; af, atrial fold; bc, bursal canal; cb, copulatory bursa; cg, cement glands; cm, circular muscle; cp, ciliate parasite; d, diaphragm; ed, ejaculatory duct; er, ectal reinforcement; g, gonopore; ga, genital atrium; gd, gonoduct; lm, longitudinal muscle; lod, left oviduct; lvd, left vas deferens; m, macronucleus; od, oviduct; pb, penis bulb; pf, penial fold; ph,

pharynx; pp, penis papilla; rod, right oviduct; rvd, right vas deferens; s, sperm; sg, shell glands; sp, spermatophore, sv, seminal vesicle.

RESULTS

Sequence characteristics and datasets

Five different datasets were built for this study (Table 2). Dataset I was a concatenation of different mitochondrial fragments belonging to four genes (*COI, NADH1*, *tRNA-W*, and *COII*) with an overall sequence length of 1268 bp. Dataset I was comprised of 76 ingroup sequences (*D. benazzii* s.l. and *D. hepta* specimens) and eight outgroup sequences belonging to *D. subtentaculata* individuals. Dataset II comprised *Dunuc10* sequences from 32 ingroup sequences and four outgroup sequences. Dataset III included *Dunuc12* sequences from 29 ingroup sequences and four outgroup sequences. The individuals that had their nuclear markers amplified were selected based on the resulting clades from the mtDNA gene tree inference, so as to make sure that all clades were represented in the nuclear-based analyses. Dataset IV consisted of a concatenation of datasets II and III and had an overall sequence length of 1089 bp. Finally, dataset V (concatenated mitochondrial and nuclear loci) was 2358 bp long and comprised 32 ingroup sequences and four outgroup sequences. Saturation was not observed for dataset I (Supporting Information, Table S2).

Phylogenetic inference

Regarding the mtDNA phylogeny (dataset I), all BI and ML analyses concordantly identified the following four major clades within nominal *D. benazzii* s.l. (Figs 2A, 3; Supporting Information, Fig. S1A): (i) an overall Sardinian clade (BEN1-Sar); (ii) a Sardinian singleton from the locality of Badu de Mola (BEN2-SarB: BMO0025-03; locality 3); (iii) a northern Corsican clade (BEN3-NCor; localities 24, 25, 30, and 31); (iv) a southern Corsican clade (BEN4-SCor; localities 19, 20, and 21), except for individual BA0101-03 (locality 24), which is sister to clade BEN4-SCor with low support (Fig. 3; Supporting Information, Fig. S1A). The singleton from locality 3, albeit originating from Sardinia, is sister to the Corsican clades.

Dugesia hepta is recovered as a monophyletic sister group of *D. benazzii* s.l.. Furthermore, it appears that *D. hepta* exhibits, to a certain extent, signs of genetic structure. At first glance, two subclades can be delineated, one associated with the Mannu River basin (localities 6, 9, and 10, Fig. 1) and the other associated with the Temo River and Coghinas River basins (localities 11, 12, and 14, Fig. 1). Geographical distribution of the *D. benazzii* s.l. clades did not overlap. Overall node support values were higher for BI than for ML analysis, yet the topologies were identical.

Phylogenies recovered for the nuclear datasets II and III (Fig. 2B, C; Supporting Information, Fig. S1B, C) showed topologies that differed from the mitochondrial data. For both datasets, *D. hepta* was recovered as a monophyletic species, albeit that no significant differences amongst basins could be identified. For dataset II, *D. benazzii* s.l. appeared to be a single monophyletic, albeit an unresolved group, in which no distinction could be detected between Corsican and Sardinian specimens. The specimen from Badu de Mola was integrated within that clade. For dataset III, *D. benazzii* s.l. was separated into two clades, viz., one composed of the north-Corsican populations and the other of the southern-Corsican and the Sardinian populations, with no differentiation between the specimens from the two islands. Basal nodes for the phylogenies resulting from datasets II and III were highly supported by BI and ML analyses; no topological differences were present between BI and ML inferred phylogenetic trees.

The phylogenetic tree resulting from the analysis of dataset IV (Fig. 2D; Supporting Information, Fig. S1D) was similar to that of dataset III. The northern Corsican populations were differentiated from the other *D. benazzii* s.l. populations, forming a clade with high support (1/94; pp/bs). With respect to the south-Corsican and Sardinian clades, the concatenation of nuclear data did not yield any new results compared to dataset III results. Regarding *D. hepta*, the distinction between the Mannu basin and the other sampled localities was recovered once again. Nodal support values were higher for BI analyses than for ML analyses, but, overall, there were no large differences (Supporting Information, Fig. S1D). With respect to the topologies, no relevant differences were found between BI and ML analyses.

Finally, the dataset V phylogeny yielded topological results that differed between BI (Fig. 2E1; Supporting Information, Fig. S1E1) and ML analyses (Fig. 2E2; Supporting Information, Fig. S1E2). In both cases, *D. hepta* was monophyletic and presented two internal and fully supported clades that could be matched to the ones identified in the trees resulting from the analysis of datasets I and IV. ML analysis showed that *D. benazzii* s.l. presented almost the same topology recovered with dataset I, although in this case individual BA0101-03 was included in the BEN3-NCor group (Supporting Information, Fig. S1E1, E2) instead of the BEN4-SCor group (Fig. 3; Supporting Information, Fig. S1A). The BI analysis yielded a different topology, in which the specimen from Badu de Mola was sister to the Sardinian clade and not to the Corsican ones (Supporting Information, Fig. S1E1, E2), despite agreeing with the ML analysis on the position of BA0101-03 within the tree. The BI tree was fully supported, whereas the ML tree had overall lower support, albeit still significant.

Alternative topology test

In view of the discrepancies between the results of the nuclear and mitochondrial datasets, as well as between the BI and ML analyses of the concatenated dataset on the phylogenetic position of specimen BMO0025-03, we performed an alternative topology test. Two topologies were considered: (i) topology-1, with specimen BMO0025-03 being related to the Corsican clades, and (ii) topology-2, in which this individual grouped with the Sardinian clades. Dataset I was used to assess the viability of the alternative topology, viz., topology-2. The reasons for this choice are as follows: (i) dataset I resulted in better phylogenetic resolution than datasets II, III, and IV; and (ii) dataset I did not show topological conflict between BI and ML inferences. Results shown in the Supporting Information, Table S3 indicate that the topology-2 hypothesis cannot be discarded, at least based on the present data.

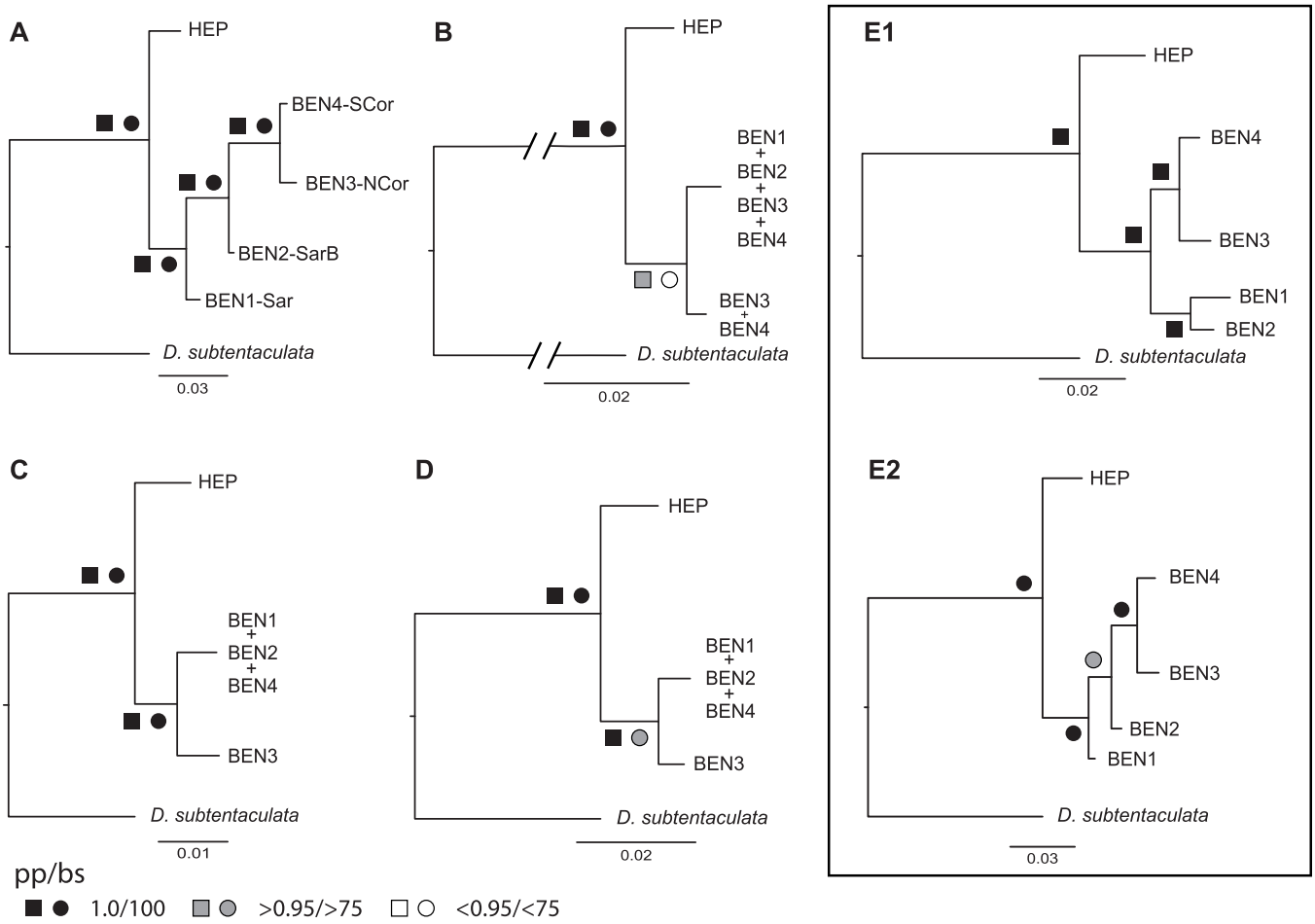


Figure 2. Simplified Bayesian (A–E1) and ML dendrograms (E2). A, dataset I: *COI–NADH1-tRNAW–COII*; B, dataset II: *Dunuc10*; C, dataset III: *Dunuc12*; D, dataset IV: *Dunuc10 + 12*; E, dataset V: mtDNA + nDNA. Posterior probabilities (pp) are indicated by filled squares and bootstrap support values (bs) with filled circles. Full trees are represented in Figure 3 (dataset I) and Supporting Information, Figure S1 (datasets I–V).

Molecular delimitation of candidate species

ABGD obtained two primary groups for all prior *P*-values (Supporting Information, Fig. S2A). One group included the outgroup sequences of *D. subtentaculata*, whilst the other included all ingroup sequences belonging to *D. benazzii* s.l. and *D. hepta*. The recursive analysis recovered a varying number of groups, some of which were highly unlikely to be true (i.e. 27 groups comprised mostly of singletons). Therefore, the recursive analysis with 11 groups (ranging from 0.77% to 1.29% prior *P*-values) was taken as the best possible delimitation (Fig. 3). *Dugesia hepta* was recovered as a single recursive group. Three recursive groups within *D. benazzii* s.l. were singletons: BMO0025-03, BMA0028-01, and BA0101-03. The northern Corsican population (BEN3-NCor) was considered as an entity alongside one of the southern populations (La Solenzara, loc. 19), while the rest of the recursive *D. benazzii* s.l. groups matched to unique sampling localities, i.e. without mixture of individuals from different sampling localities.

The GMYC model was preferred over the null model since the likelihood ratio test was significant (Fig. 3; Supporting Information, Fig. S2B). GMYC's single threshold time approach identified a similar number of clusters and entities, viz., seven clusters and eight entities. *Dugesia hepta* was delimited as a single entity. On the other hand, *D. benazzii* s.l. was separated into five entities. The Sardinian localities from the BEN1-Sar clade were divided into two entities, one including locality 12 (Mulinu; Fig. 1) and the other including all the rest; BEN2-SarB was considered as a singleton. Corsican localities BEN3-NCor and BEN4-SCor were each considered separate entities, with BEN4-SCor including the northern individual BA0101-03.

The mPTP species-delimitation approach identified six entities for the ingroup (Fig. 3; Supporting Information, Fig. S2C). *Dugesia hepta* was separated into three entities: a clade consisting of individuals from the Mannu basin, a singleton from locality 9 (Mascari-1; Fig. 1), and a clade with the individuals from the Temo and Coghinas basins. *Dugesia benazzii* s.l. was separated into three entities as well: a clade with the BEN1-Sar specimens, the singleton from Badu de Mola, and a clade with all Corsican specimens.

Primary species hypotheses

The primary species hypotheses (PSH) are shown in Fig. 3. Each molecular approach yielded different delimitations. ABGD yielded the highest number of delimited groups, but there were incongruities with regard to the inferred phylogeny (Fig. 3; Supporting Information, Fig. S1A). The largest delimited group

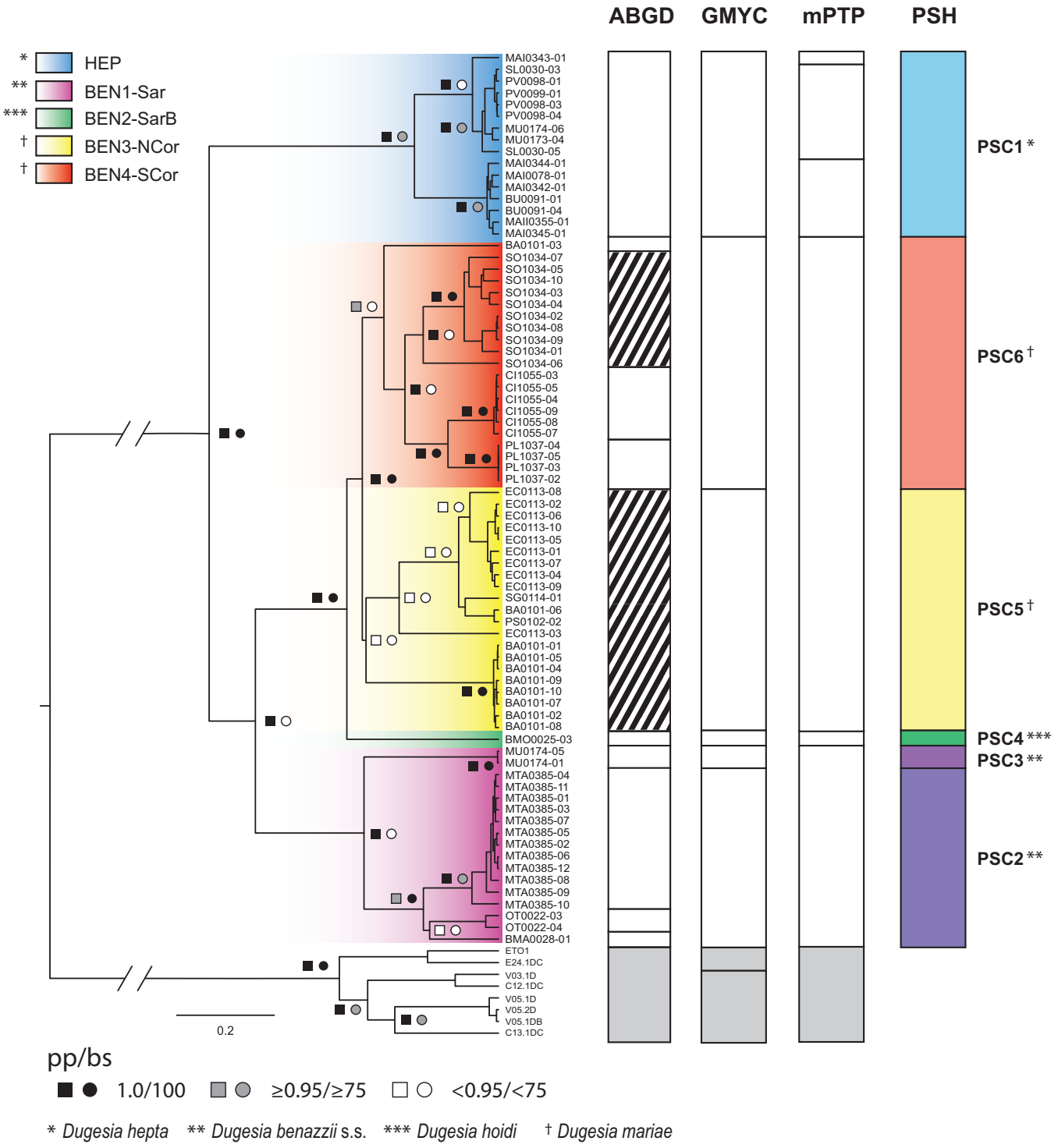


Figure 3. mtDNA (dataset I) phylogenetic tree and species discovery delimitation schemes for ABGD, GMYC, and mPTP, as well as PSHs and PSCs. Ultrametric tree from BEAST is shown only for visual purposes; posterior probabilities (pp) and bootstrap support values (bs) relate to MrBayes and RAxML analyses, respectively; pp and bs node support values represented by squares and circles, filled with white (unsupported), grey (supported), and black (maximum support), respectively.

included individuals from northern Corsica (all sampling localities, except individual BA0101-03) and the southern locality 19 (La Solenzara; Fig. 1), thus delimiting a polyphyletic group according to our phylogenetic analyses. GMYC and mPTP both delimited six ingroup entities but these differed in their composition (Fig. 3). GMYC's delimitation concurred with the phylogenetic trees based on datasets I and V: (i) two major groups within Corsica that could be related to the northern and the southern regions, (ii) the singleton from locality 3, (iii) two groups of Sardinian *D. benazzii* (individuals from locality 12 vs. the rest; Fig. 3), and (iv) a single *D. hepta* group. On the other hand, mPTP's species delimitation showed less diversity in

D. benazzii s.l. and more diversity within *D. hepta*. Given the lack of homogeneity in the results between the different PSH discovery methods, we considered accessory criteria.

Lepori (1951) originally studied *D. benazzii* s.l. samples from several Sardinian and Corsican localities. Unfortunately, our sampling localities do not coincide exactly with those of Lepori (1951), for several reasons. First, the precise sampling localities and coordinates were not reported in Lepori's account. In view of the fact that in some Sardinian rivers there can be up to three different species of *Dugesia* (cf. Stocchino, 2018), the lack of such relevant locality data may easily lead to false interpretations. Second, posterior sampling campaigns in the Crisciuleddu and Viglietu rivers in Sardinia, whose populations were studied by Lepori, have resulted in the collection of only fissiparous specimens, which cannot be identified through morphological study. In point of fact, it had been reported already that these populations were characterized by the coexistence of sexual and fissiparous individuals (Lepori and Pala, 1982). Third, the Temo River does not belong to the Comune di Bolotana locality as originally reported by Lepori and, therefore, these data are not reliable due to what, most likely, is a geographical mistake.

However, we were able to sample some Corsican populations that could be considered as replicates of those sampled by Lepori (1951) and recorded in the original description. The Bastia, Calvi, and Corte localities of Lepori are in the northern region of Corsica and may well be replicated in our northern localities. In the same way, Lepori's Aiaccio locality on the west coast is approximately 90 km away from our southern localities, with both localities being isolated from the northern localities by the midland mountain range.

Lepori (1951) found minor morphological differences between and within the islands. Individuals from Corsica could present an atrial fold that Lepori never found in the Sardinian samples. However, these differences were deemed insufficient to consider the populations of both islands as different species (much less within islands) or even distinct geographical subspecies, due to the variability of this character, which could be present or absent in individuals within the same population (Lepori, 1951).

Stocchino *et al.* (2005) did not specify any morphological differences for *D. hepta* between populations from the different hydrological basins. Considering the minor morphological differences for nominal *D. benazzii* s.l. from Sardinia and the fact that all discovery analyses highlighted differences between Corsica and Sardinia, we felt compelled to test the split of Corsican and Sardinian populations. In the same way, all analyses highlighted the separate status of the specimen from Badu de Mola.

In view of the several lines of evidence as presented above (phylogenetic results, geographical distribution of sampling localities), and the original species descriptions of *D. hepta* and nominal *D. benazzii* s.l., we surmised that the GMYC's species delimitation made the most sense to be tested in the validation step. In turn, this led to six PSC hypotheses in which *D. hepta* represents a single candidate species (PSC1), and nominal *D. benazzii* s.l. falls apart into the following five candidate species: PSC2 (specimens from localities 2, 13, and 16); PSC3 (specimens from locality 12); PSC4 (singleton from locality 3); PSC5 (specimens from localities 24, 25, 30, and 31); PSC6 (specimens from localities 19, 20, and 21 + BA0101-03) (Fig. 3).

Validation of primary species hypotheses

Molecular validation

Since BPP's model A10 required a guide tree and the alternative topology hypothesis could not be rejected, we tested two topologies (Fig. 4; Supporting Information, Fig. S3) and four prior combinations (p.c.) for datasets I, IV, and V. The topologies situated terminal BMO0025-03 as either the sister-clade to the Corsican specimens (topology-1), or as sister to the Sardinian specimens (topology-2). All analyses, regardless of p.c., dataset, or topology, fully supported the separation of *D. hepta* from *D. benazzii* s.l. (nodes A and A' in: Fig. 4; Supporting Information, Fig. S3). Further, all analyses based on dataset I (mtDNA) fully supported the presence of five candidate species within *D. benazzii* s.l., regardless of topology or p.c. tested.

However, the results were less clear when based solely on nuclear data (dataset IV) or on the concatenated dataset V. Based on dataset IV analyses, no p.c. supported the split of the Corsican PSCs (PSC5 and PSC6) into two different species (nodes E and E' in: Fig. 4; Supporting Information, Fig. S3). In contrast, this split was fully supported when dataset V was analysed for all prior combinations and topologies. With respect to the independence of BMO0025-03 (PSC4) as a separate species (nodes D and D' in: Fig. 4; Supporting Information, Fig. S3), slightly different results were recovered. In topology-1, in which PSC4 is sister to the Corsican populations (PSC5 and PSC6), the split was supported under all prior combinations by dataset V and by p.c. 2 for dataset IV (Fig. 4). Prior combination 3 for dataset IV yielded also non-significant results, although these were close to being supported (Supporting Information, Fig. S3). In topology-2, in which PSC4 is sister to the Sardinian populations (PSC2 and PSC3), p.c. 1 and p.c. 4 fully supported the separate status of PSC4 in datasets IV and V. Further, in topology-1 the split of Sardinian PSC2 and PSC3 (nodes C and C' in: Fig. 4; Supporting Information, Fig. S3) was supported by the nuclear data (dataset IV) only under p.c. 1, whereas it was supported under p.c. 1 and p.c. 4 for dataset V. In topology-2, the aforementioned split was supported as well under p.c. 1 and p.c. 4. The split between Corsican and Sardinian populations as different species (nodes B and B' in: Fig. 4; Supporting Information, Fig. S3) was supported by all datasets and prior combinations, except p.c. 4 for topology-1 resulting from dataset IV.

BFD analyses showed that partition model 1 (five species within *D. benazzii* s.l.) was the best-fitting model, according to *2lnBf* comparisons for datasets I and V based on both PS and SS analyses. The best-fitting model for dataset IV was partition model 5, which supported Corsican and Sardinian populations, with the Badu de Mola specimen within the latter group, viz., [PSC2, PSC3, and PSC4] and [PSC5 and PSC6] as two putative species (Fig. 5; Supporting Information, Table S3).

These molecular results were taken as a lead for determining which groups within *D. benazzii* s.l. should be tested and validated by morphological and karyological data. The division between Corsica and Sardinia was fully validated by all molecular methods. The Badu de Mola singleton, instead, received some support as a separate species from the nuclear data in topology-2 (Fig. 4; Supporting Information, Fig. S3.). Further, it was recovered also as a different species when mitochondrial data was added to nuclear data in topology-1,

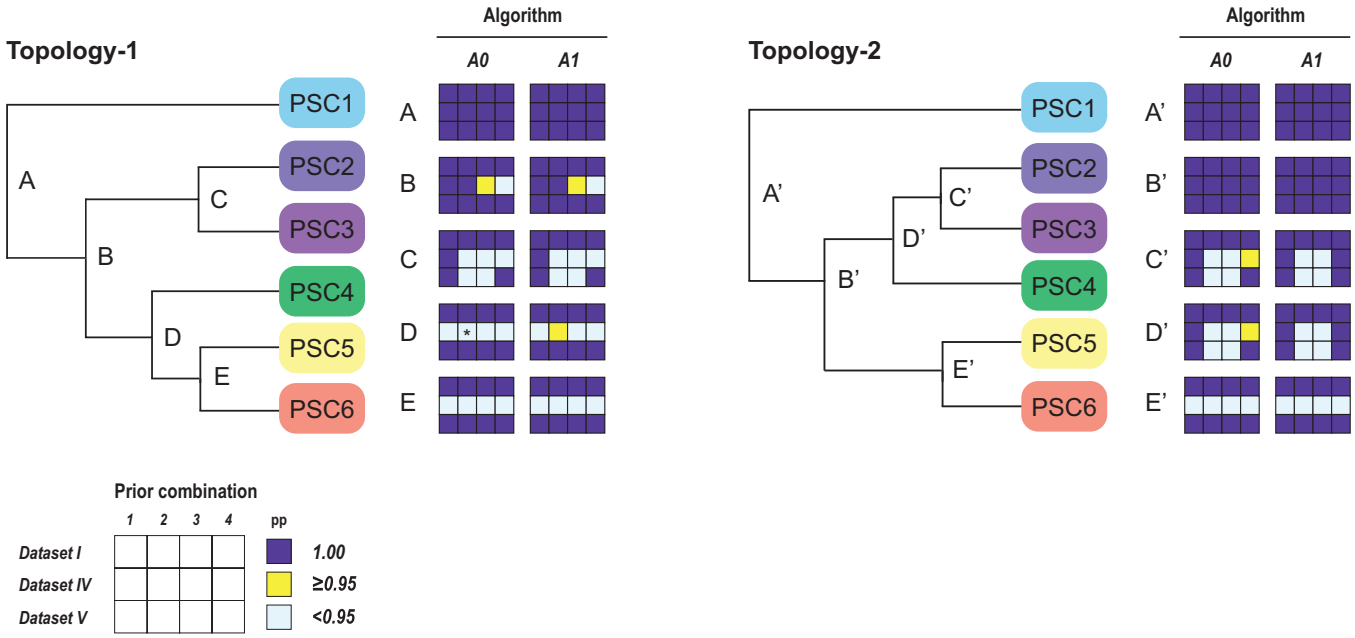


Figure 4. Schematic representation of BPP results on multi-locus data for two topologies. Colour scheme and squares represent posterior probability (pp) for each node under four different prior combinations, using two algorithms (A0 and A1) and three datasets (I, IV, and V). *Unsupported node with a pp of 0.94.

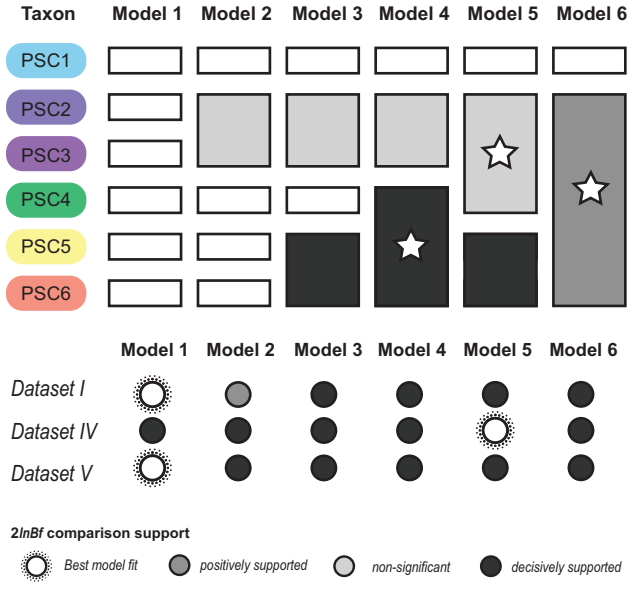


Figure 5. BFD results for the array of models tested (1–6), with different a priori species-delimitation hypotheses tested on three datasets (I, IV, and V). Each column represents a model with a unique combination of lineages (rows). Star symbol indicates reassignment of PSC4. Marginal-likelihood estimates (MLE) and Bayes' factors comparison results (*2lnBf*) from the combined analyses of five independent runs with PS (path-sampling) and SS (stepping stone) are represented with circles and a grey-scale scheme. *2lnBf* comparison support indicates differences from the best model: non-significant indicates no difference in support for the two models; positively supported and decisively-supported indicate clear support in favour of the best-fitting model over its alternative.

but it received no support in the BFD analysis of nuclear data. On the contrary, the division within Sardinia, apart from Badu de Mola, had weak support from nuclear data (dataset IV; topology-2) and was not recovered after addition of mitochondrial information (dataset V). The division within Corsica did not receive any support from the nuclear data. Although in the last-mentioned case addition of mitochondrial data resulted in support for the presence of two separate species, we deemed as more significant the complete lack of support from the more conservative nuclear data. BFD analyses of the nuclear data alone did neither support the presence of separate species within Corsica [PSC5 and PSC6] or within Sardinia [PSC2, PSC3, and PSC4].

In conclusion, existence of the following candidate species was tested by morphological and karyological data: (i) a Sardinian species constituted by PSC2 and PSC3; (ii) the Badu de Mola singleton (PSC4), and (iii) a Corsican species constituted by PSC5 and PSC6.

Morphological validation

A morphological study was performed on two sexual populations from Sardinia (Su Rizzolu River, Otti, loc. 13 and Badu de Mola River, loc. 3; Fig. 1), and one sexual population from Corsica (tributary of Golo River, Barchetta, loc. 24; Fig. 1). Sardinian populations basically showed the main anatomical diagnostic characteristics of *D. benazzii*, as described by Lepori (1951), viz., finger-shaped penis papilla; central course of the ejaculatory duct, which opens terminally at the tip of the penis papilla; pointed diaphragm; symmetrical openings of the oviducts; a single penial fold. In addition to these features, we considered also some diagnostic characters that were not reported in the original description, such as histological details of the epithelia and the underlying musculature, an angled course of the bursal canal, and the presence of cement glands. Moreover, we found discrepancies in some characters described by Lepori, viz., about the course of the vasa deferentia and the characteristics of the penial fold (see below).

Between the two Sardinian candidate species there were no significant morphological differences and, therefore, morphological data did not validate our PSH separating Badu de Mola (PSC4) from the rest of *D. benazzii* s.l. (PSC2 and PSC3) from Sardinia. In contrast, morphological features of the Corsican population exhibited clear differences when compared with Sardinian *D. benazzii* s.l., thus validating our PSH.

Karyological validation

Chromosomal data on Sardinian candidate species were extracted from Vacca *et al.* (1993) and Pala *et al.* (1999). In their karyological analysis on 12 diploid populations, tentatively attributed to *D. benazzii* s.l., with a chromosome complement of 2n = 16, Pala *et al.* (1999) recognized three karyotypes, indicated as A, B, and C. All chromosomes of karyotype A are metacentric (10 populations examined, including that from the Badu de Mola River, which is the only population with karyotype A analysed in the present paper), while in karyotype B chromosome pair 7 is submetacentric (one population). Karyotype C differs from the others in that chromosome pairs 3 and 7 are submetacentric (one population).

Diploid karyotypes (2n = 16) of populations from eight other localities were reported by Vacca *et al.* (1993 and references therein), including Su Rizzolu River (loc. 13), which were assigned to karyotype B. According to Pala *et al.* (1999), the differences between the three karyotypes could be interpreted either as a possible specific differentiation or as intraspecific variation. Later, the single population with karyotype C was recognized as a separate species, *D. leporii* Pala, Stocchino, Corso & Casu, 2000. Therefore, karyological data on the two Sardinian populations from the Badu de Mola River (karyotype A) and Su Rizzolu River (karyotype B) validated our PSH (Fig. 6).

With regard to the Corsican populations, a preliminary karyological analysis revealed a chromosome number of 16, suggesting a diploid complement, for the sexual population of loc. 24, and the sexual individuals of loc. 25 from the Golo River basin. A chromosome number of 32, suggesting a tetraploid condition, was found for the sexual individuals from loc. 30 in the Fiume Alto River. Unfortunately, rapid deterioration of the karyological preparations and the death of all specimens in the laboratory culture prevented a more in-depth karyological study.

Secondary species hypotheses

Molecular, morphological, and karyological evidence, as detailed above and below, strongly suggest that *D. benazzii* s.l. does not form a single monophyletic species. Our empirical data strongly support the differentiation of Corsican and Sardinian populations into two separate phylogenetic entities and also suggest the existence of a third phylogenetic entity at the locality of Badu de Mola (loc. 3). Therefore, we have elevated each of these three groups of populations of the *D. benazzii* s.l. lineage to the status of confirmed candidate species (CCS; cf. Vieites *et al.*, 2009), as follows: PCS2 and PCS3 from Sardinia constitute CCS1; PCS5 and PCS6 from Corsica constitute CCS2; and PCS4 from Sardinia constitutes CCS3. As consequence, this implies the necessary taxonomic description of two new species (see below). The epithet «*benazzii*» is kept for CCS1, since Benazzi (1950) introduced for the first time the name *D. benazzii*, proposed by his assistant teacher and colleague Lepori, when referring to the Sardinian specimens. Later, Lepori (1951) described the species *D. benazzii*, based on a series of comparisons between samples from Corsican and Sardinian localities.

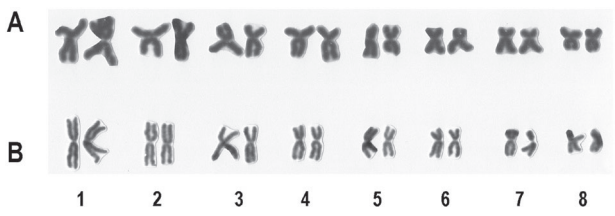


Figure 6. Karyogram of (A) *Dugesia benazzii* s.s. from Su Rizzolu River (Otti, loc. 13) and (B) *Dugesia hoidi*.

Taxonomy section

Order Tricladida Lang, 1884

Family Dugesiidae Ball, 1974

Genus *Dugesia* Girard, 1850

***Dugesia benazzii* Lepori, 1951**

Features of Dugesia benazzii s.s.: Finger-shaped penis papilla; central course of the ejaculatory duct, opening terminally at the tip of the penis papilla; conical diaphragm; symmetrical openings of the oviducts into the bursal canal; common genital atrium; penial fold devoid of a lumen, localized on the left side to the penis papilla.

Material examined: Individuals from populations 2, 6, 9, 10, 11, 12, 13, 14, and 16 (Fig. 1) were analysed molecularly; GenBank accession numbers are detailed in the Appendix Table A1. Animals examined histologically were: CGAS Pla 25. 1, Rio su Rizzolu, locality 13, Coghinas River basin (40°43′55″N, 9°08′36″E), Sardinia, 28 April 1993, coll. M. Pala and G.A. Stocchino, sagittal sections on nine slides; CGAS Pla 25. 2, ibid., sagittal sections on 12 slides; CGAS Pla 25. 3, ibid., transverse sections on 15 slides; CGAS Pla 25. 4, ibid., transverse sections on seven slides; CGAS Pla 25. 5, ibid., horizontal sections on six slides; CGAS Pla 25. 6, ibid., sagittal sections on 20 slides; RMNH.VER.21055.1, ibid., sagittal sections on 16 slides; RMNH.VER.21055.2, ibid., transverse sections on 16 slides. Karyological information was extracted from Vacca *et al.* (1993). A karyogram of the Su Rizzolu population is here reported for the first time (Fig. 6A).

Diagnosis: Molecularly, *Dugesia benazzii* comprises individuals that are identified as a single evolutionary unit together with individuals from populations 2, 6, 9, 10, 11, 12, 13, 14, and 16, when using the coalescence-based method BPP with the loci and settings detailed in the present study's Materials and methods section or equivalent molecular delimitation analyses. Morphologically, *D. benazzii* is characterized by the presence of the following features: finger-shaped penis papilla; central course of the ejaculatory duct, opening terminally at the tip of the penis papilla; conical diaphragm positioned at the base of the penis papilla; asymmetrical course of the vasa deferentia into the penis bulb, with the right sperm duct running dorso-laterally to

the left one, which runs ventro-laterally; asymmetrical openings of the vasa deferentia into the posterior section of the seminal vesicle, close to the diaphragm; symmetrical openings of the oviducts into the bursal canal; ectal reinforcement confined to the posterior tract of the vaginal area of the bursal canal; angled bursal canal; penial fold of the parenchymatic-muscular type. Karyology: 2n = 16; chromosome pairs all metacentric, except chromosome 7, which is submetacentric.

Morphological redescription: Inner and outer pharyngeal musculature bilayered, i.e. without an extra, third, outer longitudinal muscle layer in the inner sheath of muscles. Mouth opening located at the hind end of the pharyngeal pocket. Numerous ciliated protozoan parasites were present in the intestinal tract of most of the specimens examined (seven out of eight).

The anterior portion of the oviducts is expanded into a tuba, which communicates with the dorsal part of the ovaries. From thereon, the infranucleated oviducts run ventrally in a caudal direction, immediately dorsal to the ventral nerve cords. Posterior to the genital pore, the oviducts recurve anteriad to open separately and symmetrically into the vaginal portion of the bursal canal (Fig. 7).

The numerous mature testes are situated dorsally on either side of the body and extend from the level of the ovaries to the posterior end of the body. Vitellaria are located between the testes and the intestinal branches.

The very large, lobated and sac-shaped copulatory bursa occupies the entire dorso-ventral space; it is lined with a vacuolated, columnar, and glandular epithelium with basal nuclei and is surrounded by a layer of muscles. A spermatophore, filled with sperm, as well as free sperm, were present in the bursa of specimen CGAS Pla 25. 1, together with ciliated parasites (Fig. 8). These protozoans infested also the lumen of the copulatory bursa in four of the other seven specimens examined. In specimen CGAS Pla 25. 5, parasites were present also in the bursal canal. On the basis of their general morphology and presence of a macronucleus with three micronuclei, these protozoans, infesting both the gut and the copulatory bursa, can be attributed to the species *Haptophrya planariarum* (von Siebold, 1839) Stein, 1867.

From the postero-dorsal wall of the bursa, the bursal canal runs in a caudal direction to the left side of the copulatory apparatus, extending beyond the penis and the genital pore. Posterior to the gonopore, the bursal canal shows a knee-shaped, antero-ventrally directed bend, thus forming a so-called 'angled bursal canal'; hereafter, it opens through the roof of the common genital atrium. The bursal canal is lined with cylindrical, infranucleated, ciliated cells and is surrounded by a subepithelial layer of longitudinal muscles, followed by a layer of circular muscle. Ectal reinforcement with longitudinal muscles is confined to the posterior tract of the vaginal portion of the bursal canal. Free sperm were present in the bursal canal of specimen CGAS Pla 25. 1. Abundant shell glands discharge their erythrophil secretion into the vaginal section of the bursal canal, just ventrally to the oviducal openings (Figs 7, 9A).

The well-developed penis bulb consists of intermingled longitudinal and circular muscle fibres. The penis bulb houses a pleated seminal vesicle, lined with a nucleated epithelium (Fig. 7). Upon reaching the level of the penis bulb, the vasa deferentia curve dorso-medially and then asymmetrically penetrate the wall of the proximal, anterior section of the penis bulb, with the right sperm duct entering antero-dorsally and the left one antero-ventrally. Once inside the bulb, the ducts run asymmetrically to reach the posterior portion of the seminal vesicle, close to the diaphragm. The sperm ducts open asymmetrically into the seminal vesicle, with the right vas deferens opening more dorsally to the left one (Fig. 7). In all specimens examined, the sperm ducts form well-developed spermiducal vesicles, packed with sperm.

The seminal vesicle opens into the ejaculatory duct via a large, conical diaphragm. This diaphragm, which is located approximately at the base of the penis papilla, receives the openings of penis glands. The finger-shaped, symmetrical penis papilla is covered by an infranucleated epithelium, which is underlain with a subepithelial layer of circular muscles, followed by a layer of longitudinal muscle fibres. The ejaculatory duct follows a central course through the penis papilla and has a terminal opening, while in its anterior tract it receives the openings of erythophil penial glands. The ejaculatory duct, which in all the specimens examined contained an empty spermatophore, is lined by a cuboidal, infranucleated epithelium (Figs 7, 9B).

A very large penial fold of the parenchymatic-muscular type is situated dorso-laterally or laterally on the left side of the base of the penis papilla. The fold, with a length that is approximately three-quarters or one-half of that of the penis papilla, is either (i) almost independent of the penis papilla, or (ii) throughout its length attached to the papilla. The penial fold is covered by an infranucleated epithelium, which is underlain with a subepithelial layer of circular muscles, followed by a layer of longitudinal muscle fibres (Figs 7, 9). In specimen CGAS Pla 25. 3, the penial fold shows lateral grooves in its surface. In specimens CGAS Pla 25. 6 (Fig. 9A) and RMNH.VER.21055.1 there are also two small muscular atrial folds, which are located ventrally and dorsally, respectively, while a single ventral atrial fold is present in specimen CGAS Pla 25. 2.

The genital atrium is lined by an infranucleated epithelium, which is underlain by a thick subepithelial layer of circular muscle, followed by a layer of longitudinal muscle fibres. Abundant cement glands open into the atrium dorsally to the gonopore. The atrium opens ventrally through the genital pore (Fig. 7).

Geographical distribution: Sardinia and Capraia Island (Tuscany Archipelago). Originally, it was also considered to be present on Corsica, where it would be the only species of the genus *Dugesia* inhabiting this Mediterranean island. However, according to our present analysis, its presumed presence on Corsica now has to be evaluated as doubtful.

Reproduction: Sexual and fissiparous. Populations may be exclusively sexual or mixed, the latter characterized by the coexistence of sexual and fissiparous individuals.

Comparative discussion: In all essential details, the specimens examined from the Su Rizzolu River agree with the features described for *D. benazzii* by Lepori (1951). However, with respect to diagnostic characters, there are some differences between our observations and those of Lepori (1951). *Dugesia benazzii* was described with a penial annex or adenodactyl in the form

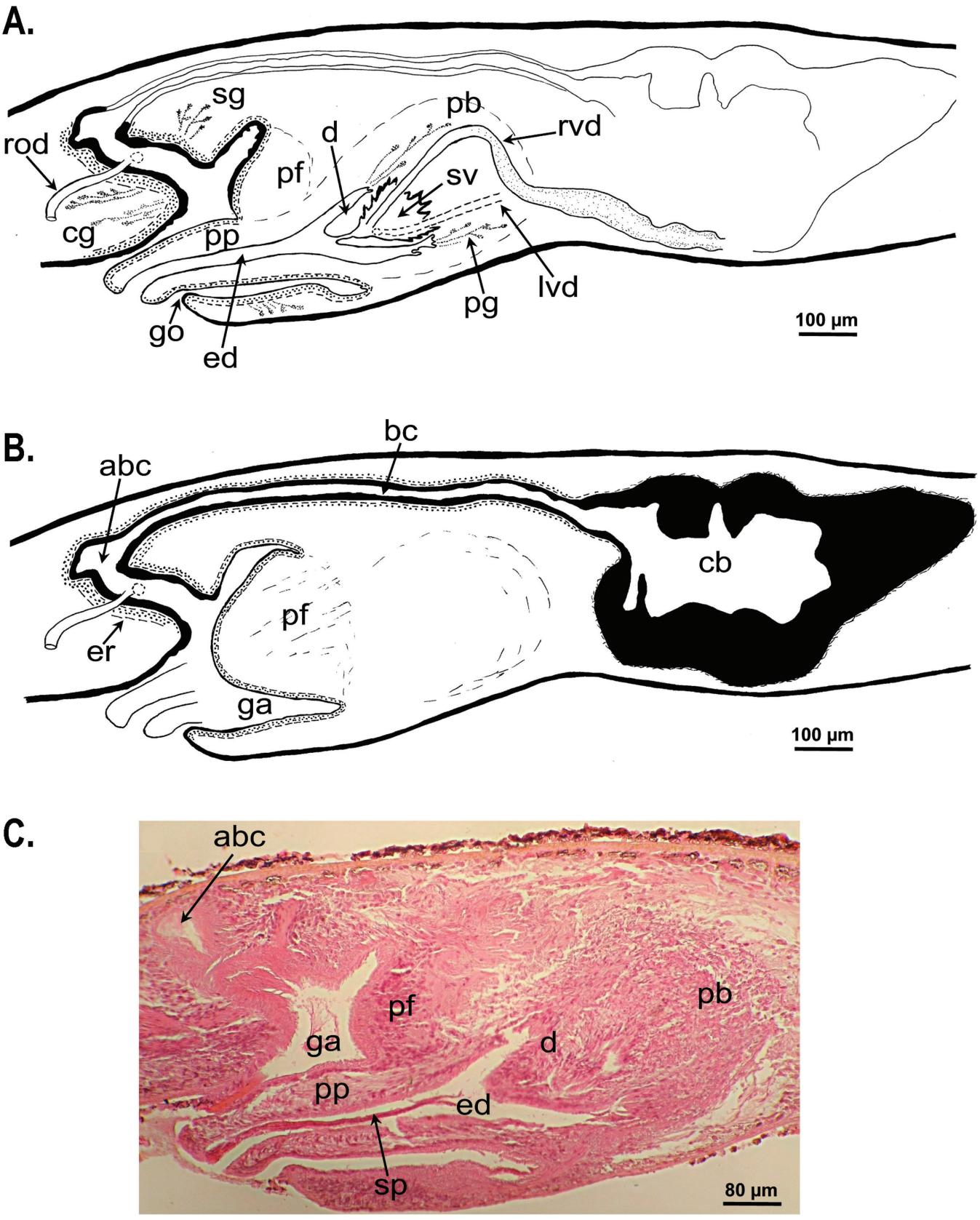


Figure 7. *Dugesia benazzii* s.s., CGAS Pla 25.1: A, sagittal reconstruction of the male copulatory apparatus (anterior to the right); B, sagittal reconstruction of the fold and female copulatory apparatus; C, photomicrograph of sagittal section, showing the penis bulb (pb), penis papilla (pp) with conical, pointed diaphragm (d), ejaculatory duct (ed), penial fold (pf), and 'angled' bursal canal (abc).

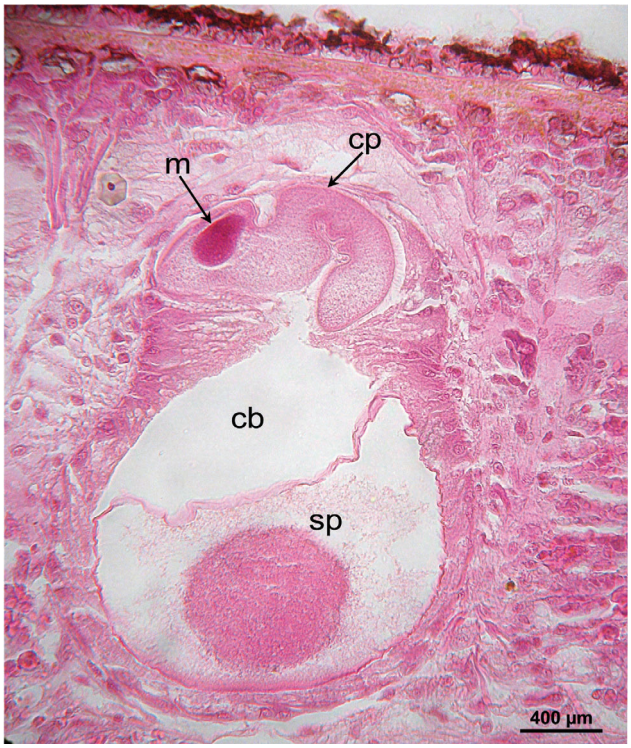


Figure 8. *Dugesia benazzii* s.s., CGAS Pla 25.1. Photomicrograph of the copulatory bursa with a ciliate parasite (cp) and a spermatophore (sp) in a sagittal section.

of a muscular-glandular papilla devoid of a lumen that is situated on the left side of the penis papilla and that even may be larger than the penis itself. Moreover, for some specimens from Bunnari and Mascari rivers, Lepori (1951) reported the presence of a muscular-glandular papilla localized ventro-laterally to the penis papilla. Considering that later in the same rivers *D. hepta* was found, which is characterized by a ventro-lateral fold (see: Stocchino *et al.*, 2005), the description of some specimens could refer precisely to this last-mentioned species instead of nominal *D. benazzii*. However, more recent studies restricted the term adenodactyl to the strongly muscularized, glandular-parenchymatic folds of the copulatory apparatus that are provided with a lumen (Stocchino *et al.*, 2017 and references therein). Our analyses revealed that the penial annexes in the specimens examined actually concern a fold, which is shorter than the penis papilla and conforms to the parenchymatic-muscular type *sensu* Stocchino *et al.* (2017).

The original description of *D. benazzii* does not specify the course of the vasa deferentia or their openings into the seminal vesicle. However, in his horizontal reconstruction drawings of the copulatory apparatus of the Sardinian populations, Lepori (1951: figs 1, 2) depicted a symmetrical course of the vasa deferentia, as well as symmetrical openings into the seminal vesicle. This contrasts with the situation in our animals, in which the vasa deferentia follow clearly asymmetrical trajectories and also open asymmetrically into the seminal vesicle.

Asymmetrical openings of the vasa deferentia have been reported for two other species of the genus *Dugesia* viz., *D. bifida* Stocchino & Sluys, 2014 and *D. circumcisa* Chen & Dong, 2021. *Dugesia benazzii* shares with *D. bifida* the condition with the right duct opening into the seminal vesicle dorsally to the left

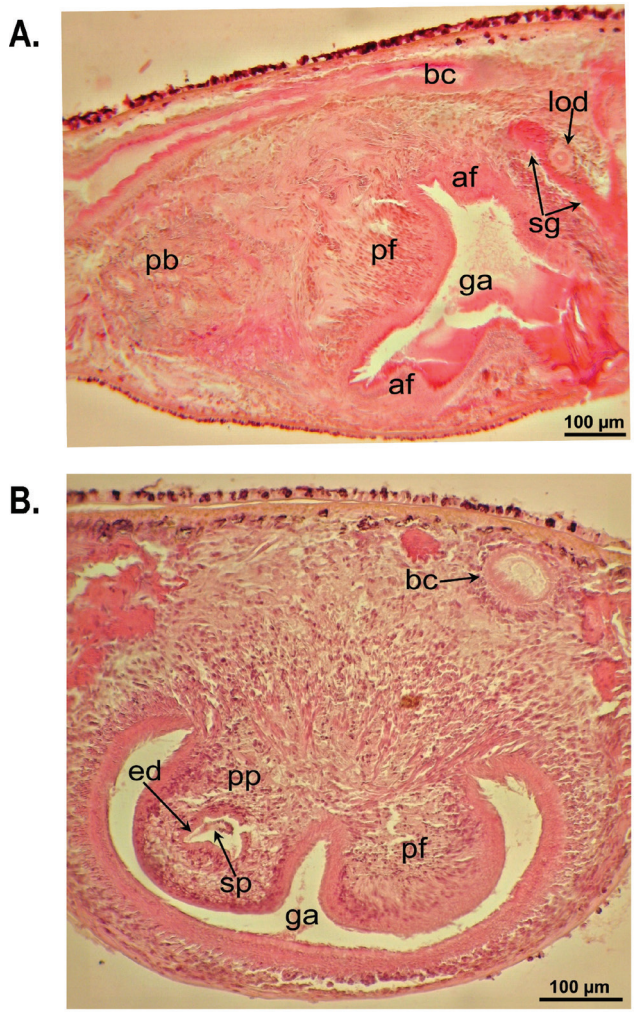


Figure 9. *Dugesia benazzii* s.s.: A, CGAS Pla 25. 6, photomicrograph showing the penial fold (pf) and the two atrial folds (af) in sagittal section; B, CGAS Pla 25. 4, photomicrograph showing the penis papilla (pp) and the penial fold (pf) in transverse section.

one. However, *D. benazzii* differs from the two above-mentioned species in that its vasa deferentia open close to the diaphragm. In *D. bifida*, in contrast, the vasa deferentia open halfway into the vesicle, while in *D. circumcisa* the right sperm duct opens into the anterior portion of the seminal vesicle and the left duct through the mid-lateral wall of the vesicle, with the left sperm duct opening dorsally to the right one (Stocchino *et al.*, 2014; Wang *et al.*, 2021a).

Cement glands are not mentioned in the original description, while they are well evident in all specimens from the Su Rizzolu River. Interestingly, these glands do not discharge their secretions around the gonopore, as in most of the species of the genus *Dugesia*, but they open into the atrium ventrally to the opening of the shell glands and dorsally to the genital pore. A similar condition has been reported only for *D. subtentaculata* (Leria *et al.*; 2020).

In the original description of *D. benazzii*, Lepori (1951) reported the presence of a muscular atrial fold only for the Corsican specimens. We also found this feature in our Sardinian specimens. This atrial annexe has been reported in a few other species of the genus *Dugesia* viz., *D. aenigma* de Vries, 1984, *D. ariadnae* de Vries, 1984, *D. damoae* de Vries, 1984 (de Vries,

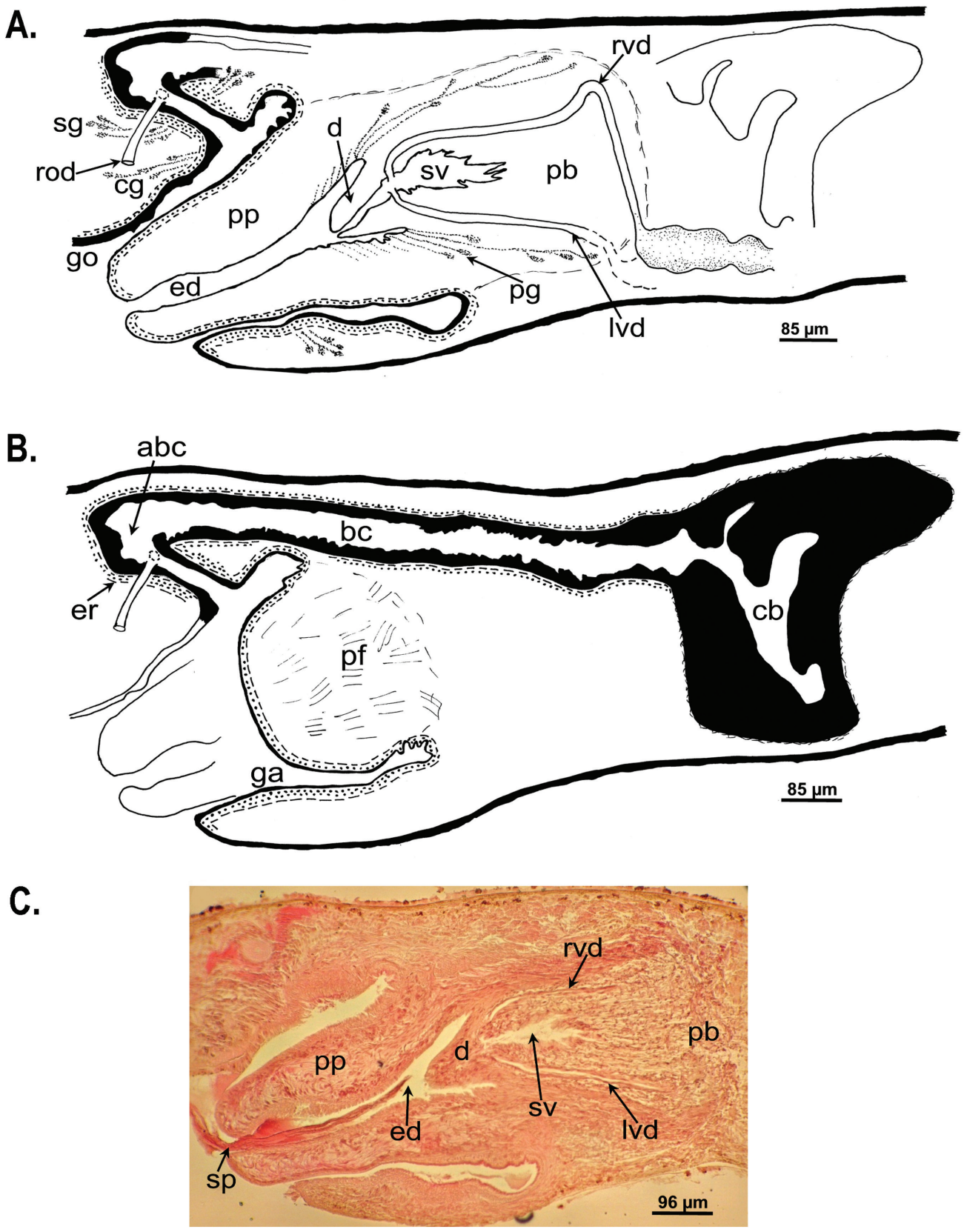


Figure 10. *Dugesia hoidi*. Holotype RMNH.VER.21056.1: A, sagittal reconstruction of the male copulatory apparatus (anterior to the right); B, sagittal reconstruction of the penial fold and female copulatory apparatus; C, photomicrograph of sagittal section, showing penis bulb (pb) with the seminal vesicle (sv), right (rvd) and the left (lvd) vas deferens, penis papilla (pp) with the pointed diaphragm (d), and the ejaculatory duct (ed).

1984; Stocchino *et al.*, 2017), and in *D. mariae* Stocchino, Dols-Serrate & Riutort, sp. nov. (see below).

Unfortunately, Lepori (1951) did not present any details about the bursal canal. Interestingly, all specimens of the population examined are characterized by an 'angled' course of the bursal canal, without exception. It is noteworthy that this character is not common amongst species of *Dugesia*, whereas it forms a diagnostic character of the genus *Girardia* (de Vries and Sluys, 1991; Stocchino *et al.*, 2019).

A short ectal reinforcement confined to the posterior tract of the vaginal portion of the bursal canal is shared with only two species of the genus *Dugesia*, viz., *D. improvisa* Sluys & Solà, 2013 and *D. naiadis* Sluys, 2013 (Sluys *et al.*, 2013).

In several of our specimens of *D. benazzii*, the ciliate parasite *Haptophrya planariarum* (Phylum Ciliophora) was present in the intestinal tract, the bursal canal, and the copulatory bursa. Specimens of *H. planariarum* were found also in the gut and copulatory bursa of *D. mariae* Stocchino, Dols-Serrate & Riutort, sp. nov. (see below), while preliminary observations established the occurrence of these parasites in both the intestinal tract and the copulatory bursa of specimens from other Sardinian populations of *D. benazzii*, as well as in individuals from some populations of *D. hepta* (G.A. Stocchino personal observation). It should be noted that the presence of *H. planariarum* does not seem to affect the development and functionality of the reproductive apparatus of the parasitized individuals.

In addition to these three species of *Dugesia*, presence of *H. planariarum* was previously documented for nine other species of freshwater planarian, from which it was generally reported from the gut, the pharynx lumen, or the pharyngeal pouch: *Planaria torva* (von Graff, 1903; Bishop, 1926; Lom, 1959; de Puytorac, 1963; Keller, 1983), *Schmidtea lugubris* (Lom, 1959; Keller, 1983), *Dendrocoelum lacteum* (von Graff, 1903; Bishop, 1926; de Puytorac, 1963), *Phagocata iwamai* (Kawakatsu, 1970), *Dugesia gonocephala* (von Graff, 1903; Steinböck, 1924; de Beauchamp, 1932; Gamo Garcia and Rodríguez Caabeiro, 1981;Rataj and Vďačný, 2018), *Dugesia monomyoda* (Marcus, 1953) , *Polycelis nigra* (von Graff, 1903; de Pyutorac, 1963), *Paraplanaria dactyligera* (Kenk, 1969), *Paraplanaria occulta* (Kenk, 1969). For only *D. gonocephala* (Steinböck, 1924; de Beauchamp, 1932) it was reported that *H. planariarum* parasites occurred also in the copulatory bursa of some specimens. However, recent molecular studies (Rataj *et al.*, 2022) have highlighted that the morphospecies *H. planariarum* is a complex of three genetically distinct and morphologically cryptic species. The other two ciliate species, *H. dugesiarum* Rataj, Zhang & Vd'ačny, 2022 and *H. schmidtearum* Rataj, Zhang & Vd'ačny, 2022, morphologically correspond to *H. planariarum* and, therefore, these species can be identified only after molecular analysis. While *H. planariarum* and *H. dugesiarum* were found infesting *D. gonocephala*, *H. schmidtearum* was found only in *Schmidtea polychroa*. In all three of the last-mentioned planarian species, the parasites occurred only in the gut and the pharyngeal pouch.

Dugesia hoidi Dols-Serrate, Stocchino & Riutort, sp. nov.

Material examined

Holotype: RMNH.VER.21056.1, Badu de Mola River (Temo River basin) (40°18′1″N, 8°34′57″E), Sardinia, 17 July 1997, coll. M. Pala, sagittal sections on eight slides.

Paratypes: CGAS Pla 26. 1, ibid., sagittal sections on 16 slides; RMNH.VER.21056.2, ibid., transverse sections on 10 slides; CGAS Pla 26. 2, ibid., transverse sections on 10 slides; CGAS Pla 26. 3, ibid., sagittal sections on six slides. Molecular paratype: individual BMO0025-03 from population 3 (Rio Badu de Mola; Fig. 1) was analysed molecularly, corresponding GenBank accession numbers are detailed in Appendix, Table A1. Karyological information was extracted from Pala *et al.* (1999). A karyogram for the Rio Badu de Mola population is here reported for the first time (Fig. 6B).

Diagnosis: Molecularly, *Dugesia hoidi* comprises individuals that are identified as a single evolutionary unit together with individual BMO0025-03 from locality 3 when using the coalescence-based method BPP, or an equivalent molecular-delimitation method, with the loci and settings detailed in the present study's Materials and methods section. Karyology: 2n = 16, with all chromosomes being metacentric.

Etymology: The specific epithet is given to honour the fictional male character Hoid from Brandon Sanderson's epic fantasy series *The Stormlight Archives*. It is a tribute to his mimicry and disguise as one amongst humans, just as *D. hoidi* remained indistinguishable, thus far, amongst specimens of *D. benazzii*.

Geographical distribution: Known only from the type locality: Badu de Mola River (Temo River basin), Sardinia.

Reproduction: Sexual.

Comparative discussion: Dugesia hoidi shows the same main diagnostic morphological features as *D. benazzii* specimens of the Su Rizzolu River population (Figs 7, 10, 11), but there are small differences between the two populations. In particular, in *D. hoidi* holotype RMNH.VER.21056.1 the asymmetry in the course of the vasa deferentia is very distinct, with the course of the right sperm duct through the penis bulb being perfectly dorsal to that of the left one, including its opening into the posterior section of the seminal vesicle (Fig. 10A, C). In the same specimen, the penial fold is situated just on the left side of the penis papilla. Atrial folds were not found in *D. hoidi*, whereas these structures are present in three specimens of *D. benazzii* from Su Rizzolu.

Nevertheless, the conclusion must be that there are no consistent morphological features that can be used as diagnostic characters in the recognition and identification of *D. hoidi* and that, therefore, it can be distinguished from *D. benazzii* only by means of molecular species delimitation analyses and the presence of a differentiated karyotype as described above.

Our molecular analysis unravelled the separate phylogenetic status of *D. hoidi* and, thus, contributes to a better understanding of the evolutionary diversification of the genus *Dugesia* in the Mediterranean region. Nevertheless, taxonomically *D. hoidi* largely remains a cryptic species, since at present there is no consilience of different datasets possible, mostly due to the absence of morphological diagnostic characters through which the species can be characterized and diagnosed. We do not exclude the possibility that this problem may be overcome when new specimens from the Badu de Mola locality become available for analysis. Unfortunately, recent attempts to re-sample

specimens (the most recent one from June 2021) have been unsuccessful.

***Dugesia mariae* Stocchino, Dols-Serrate & Riutort, sp. nov.**

Material examined: Individuals from populations 19, 20, 21, 24, 25, 30, and 31 were analysed molecularly; corresponding GenBank accession numbers are detailed in the Appendix, Table A1.

Holotype: RMNH.VER.21057.1, tributary of Golo River, locality Barchetta, (42°30′26.73″N, 9°04′22.4″E), Corsica, 17 May 2010, coll. M. Pala, G.A. Stocchino and E.M. Lazaro, one set of sagittal sections on seven slides.

Paratypes: CGAS Pla 27. 1, ibid., horizontal sections on three slides; CGAS Pla 27. 2, ibid., horizontal sections on three slides; RMNH.VER.21057.2, ibid., sagittal sections on eight slides; CGAS Pla 27. 3p, ibid., sagittal sections on eight slides; CGAS Pla 27. 4, ibid., transverse sections on 14 slides; RMNH.VER.21057.3, ibid., horizontal sections on six slides.

Diagnosis: Molecularly, *Dugesia mariae* comprises individuals that are identified as a single evolutionary unit together with individuals from populations 19, 20, 21, 24, 25, 30, and 31, when using the coalescence-based method BPP, or equivalent methods, with the loci and settings detailed in the present study's Materials and methods section. Morphologically, *D. mariae* is characterized by the presence of the following features: short and blunt penis papilla; ventrally displaced ejaculatory duct, opening terminally at tip of the penis papilla; ejaculatory duct broad, with pleated wall and irregularly-shaped lumen; large, conical diaphragm; symmetrical openings of the vasa deferentia into the posterior section of the seminal vesicle, i.e. close to the diaphragm; oviducts opening symmetrically into the vaginal portion of the bursal canal; ectal reinforcement musculature on bursal canal extending from vaginal region almost up to the copulatory bursa; penial fold of the parenchymatic-muscular type, located dorsally or dorso-laterally to the penis papilla; cement glands opening into the genital atrium around the dorsal region of the gonoduct.

Etymology: The specific epithet refers to Professor emeritus Maria Pala, in recognition of her contributions to the biology and systematics of freshwater triclads.

Geographical distribution: The species is known from several localities on the island of Corsica: Golo River (Barchetta, loc. 24; Ponti strada, loc. 25), Fium Alto River (Ecole Communale, loc. 30; San Gavino, loc. 31), Solenzara River (loc. 19), Cioccia River (loc. 20), and Castellu River (Pero Longo, loc. 21).

Reproduction: Sexual and fissiparous. At three localities, only sexual individuals were found (loc. 24, 20, and 21), whereas in three other localities both sexual and fissiparous individuals (loc. 25, 30, and 31) were present, while only fissiparous individuals were found at one locality 19 (Fig. 1).

Description: The body of preserved specimens measured up to 9.3 mm in length and 2.8 mm in width (Fig. 12). Two eyes are present in the centre of the low triangular head. Dorsal surface brown; ventral surface pale.

The pharynx is positioned in the mid-region of the body and measures about one-fifth of the body length. Inner and outer pharyngeal musculature bilayered, i.e. without an extra, third, outer longitudinal muscle layer in the inner sheath of muscles. The mouth opening is situated at the hind end of the pharyngeal pocket. In all specimens examined, the intestinal tract was infested by the ciliate parasite *H. planariarum*, which was present in both the pharyngeal pouch and the gut.

The two ovoid ovaries are located ventrally, at a short distance behind the brain, i.e. at about one-third of the distance between the brain and the root of the pharynx. The anterior portion of the oviducts is expanded into a tuba, which communicates with the dorsal part of the ovaries. From the ovaries, the infranucleated oviducts run ventrally in a caudal direction to the level of the genital pore, whereafter they open separately and symmetrically into the vaginal portion of the bursal canal.

The numerous testes are situated dorsally on either side of the body and extend from the posterior level of the ovaries to almost the posterior end of the body. Vitellaria are located between the testes and the intestinal branches.

The very large, more or less cuboidal copulatory bursa occupies the entire dorso-ventral space; it is lined by a columnar, vacuolated, and glandular epithelium with basal nuclei and is surrounded by a layer of muscles. Remnants of a spermatophore and sperm are present in the bursa of specimens RMNH.VER.21057.1, CGAS Pla 27. 1, and CGAS Pla 27. 3. From the postero-dorsal wall of the bursa, the bursal canal runs in a caudal direction to the left side of the copulatory apparatus, then curves ventrally just at the level of the gonoduct and opens through the roof of the genital atrium. The bursal canal is lined with cylindrical, infranucleated, ciliated cells and is surrounded by a thin subepithelial layer of longitudinal muscles, followed by a layer of circular muscle. Ectal reinforcement musculature is present in the form of a well-developed layer of longitudinal muscles running ectally to the circular layer around the bursal canal, extending from the opening of the bursal canal into the atrium to almost the copulatory bursa. Shell glands discharge their erythrophil secretion into the vaginal region of the bursal canal, just ventrally to the oviducal openings (Fig. 13).

The moderately-developed penis bulb has an oblique, ventro-caudal orientation and consists of intermingled longitudinal and circular muscle fibres. The penis bulb houses a spacious seminal vesicle, which is lined with a pleated epithelium (Figs 13, 14A).

At the level of the penis bulb, the vasa deferentia curve dorso-medially and then penetrate the anterior wall of the bulb to open separately and symmetrically into the posterior portion of the seminal vesicle, close to the diaphragm. In specimen CGAS Pla 27. 1 the vasa deferentia are swollen to voluminous spermiducal vesicles.

The seminal vesicle opens into the ejaculatory duct via a well-developed conical diaphragm. The short and blunt penis papilla is covered by an infranucleated epithelium, which is underlain by a subepithelial layer of circular muscles. The ejaculatory duct follows a ventral course and has a terminal opening. The ejaculatory

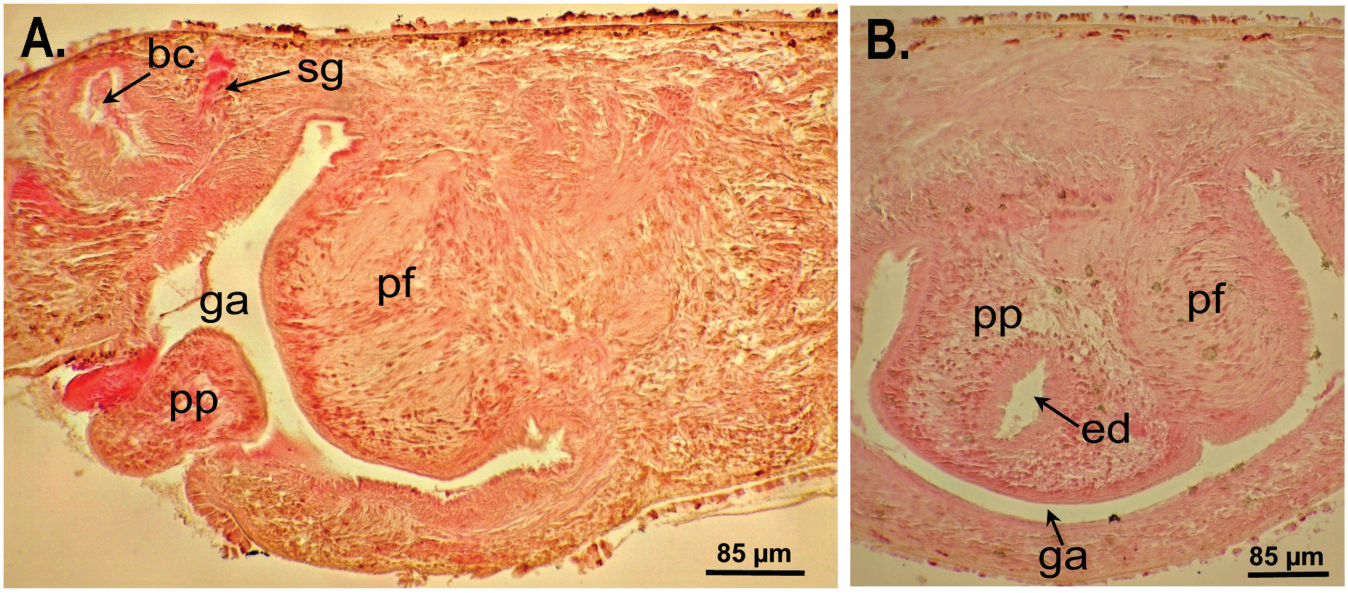


Figure 11. *Dugesia hoidi*: A, holotype RMNH.VER.21056.1, photomicrograph showing the penial fold (pf) in sagittal section; B, paratype RMNH.VER.21056.2, photomicrograph showing the penis papilla (pp) and the penial fold (pf) in transverse section.

Figure 12. *Dugesia mariae*. Photomicrograph of a preserved sexual specimen from the Golo River (Barchetta, loc. 24).

duct, which contained an empty spermatophore in all specimens examined, has an irregularly shaped lumen, due to its considerably pleated wall; it is lined by a cuboidal, infranucleated epithelium and surrounded by a subepithelial layer of longitudinal muscles (Figs 13, 14).

A penial fold of the parenchymatic-muscular type is situated dorsally or dorso-laterally on the left side of the base of the penis papilla. The fold measures approximately one-third or one-half of the length of the penis papilla (Figs 13, 14B).

The genital atrium is lined by a partially infranucleated epithelium, which is underlain by a subepithelial layer of circular muscle, followed by a layer of longitudinal muscle fibres. The genital atrium communicates with a gonoduct that is lined by a nucleated columnar epithelium and which opens to the exterior via the ventral gonopore. Cement glands open into the genital atrium around the dorsal region of the gonoduct (Fig. 13). In specimens CGAS Pla 27. 1, RMNH.VER.21057.2, and RMNH.VER.21057.3 there is a small muscular atrial fold that protrudes ventro-laterally as a semicircular valve (Fig. 14A).

Comparative discussion: Dugesia mariae differs from its congeners in a unique combination of morphological characters concerning the copulatory apparatus: short and blunt penis papilla with a large, conical diaphragm; ventrally displaced ejaculatory duct, with its terminal opening and its broad lumen and pleated wall; symmetrical openings of the vasa deferentia close to the diaphragm; presence of an ectal reinforcement layer, extending almost up to the copulatory bursa; oviducts opening symmetrically into the vaginal portion of the bursal canal; cement glands opening into the atrium around the dorsal region of a gonoduct; a penial fold of the parenchymatic-muscular type located dorsally or dorso-laterally to the penis papilla.

Both sexual and fissiparous Corsican planarians were first considered by Arndt (1922) as belonging to *D. gonocephala*. Later, de Beauchamp (1926) assigned the fissiparous populations to *D. subtentaculata*. Further studies by Benazzi and Lepori led to the description of these populations as a new species, viz., *D. benazzii*, distributed on both Corsica and Sardinia (Lepori, 1951 and references therein). However, *D. mariae* shares with *D. benazzii* s.s. and *D. hoidi* only the large, conical diaphragm and the presence of a penial fold on the left side of the penis papilla, whereas it differs from these two species in the asymmetrical penis papilla, the shape of the ejaculatory duct, the extension of the ectal reinforcement, and the absence of an 'angled' bursal canal.

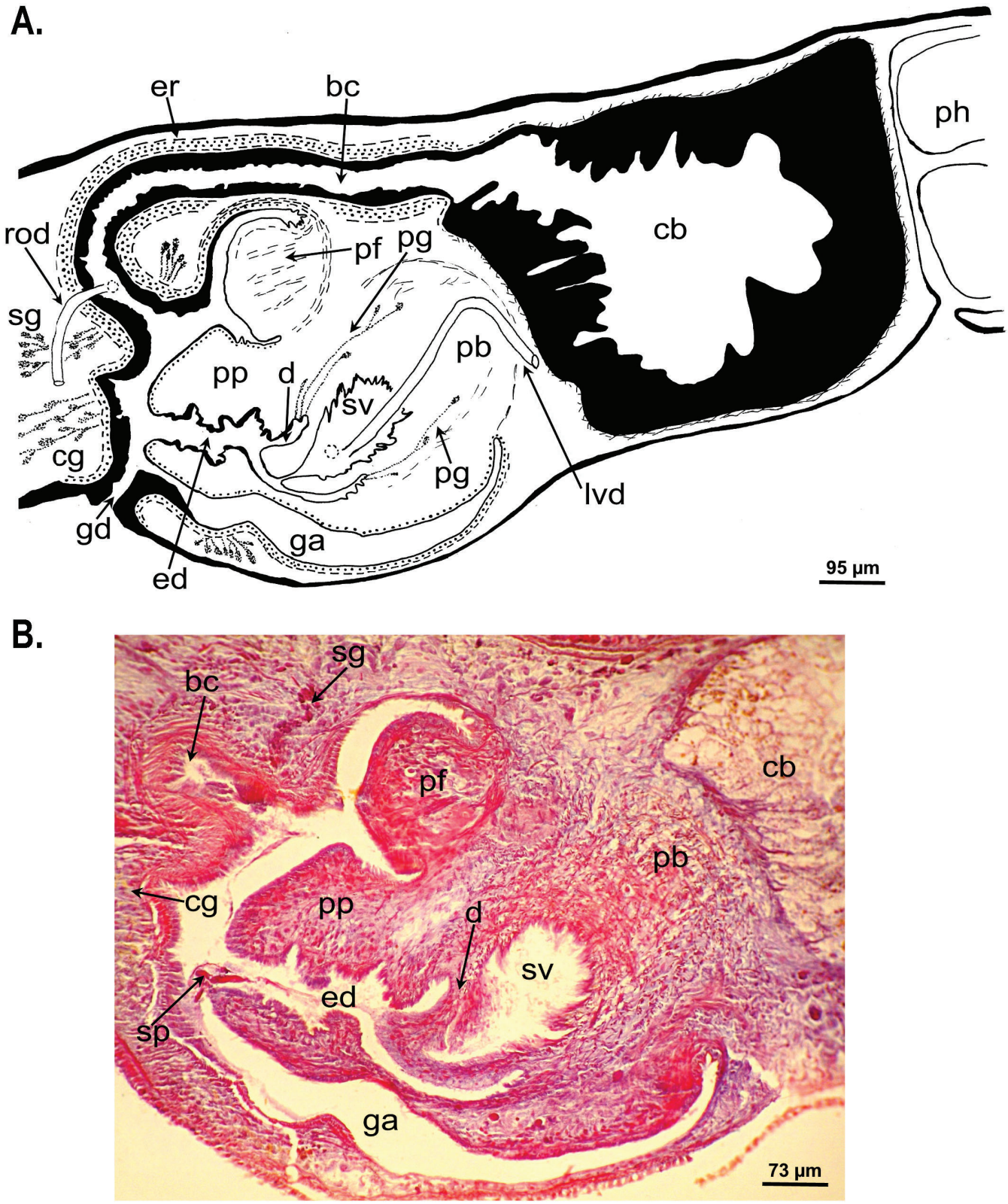


Figure 13. *Dugesia mariae.* Holotype RMNH.VER.21056.1: A, sagittal reconstruction of the copulatory apparatus (anterior to the right); B, photomicrograph showing, in a sagittal section, the penis bulb (pb) with seminal vesicle (sv), penis papilla (pp) with the pointed diaphragm (d), pleated ejaculatory duct (ed), and the penial fold (pf).

With respect to its ventrally displaced ejaculatory duct with terminal opening, *D. mariae* resembles no less than 28 other species of *Dugesia*: *D. afromontana*, *D. andamanensis* (Kaburaki, 1925), *D. annandalei* (Kaburaki, 1918), *D. ariadnae* de Vries, 1984, *D. bifida*, *D. borneana* Kawakatsu, 1972, *D. circumcisa* Chen & Dong, 2021, *D. damoae* de Vries, 1984, *D. deharvengi* Kawakatsu & Mitchell, 1989, *D. didiaphragma* de Vries, 1988, *D. ectophysa* Marcus, 1953,

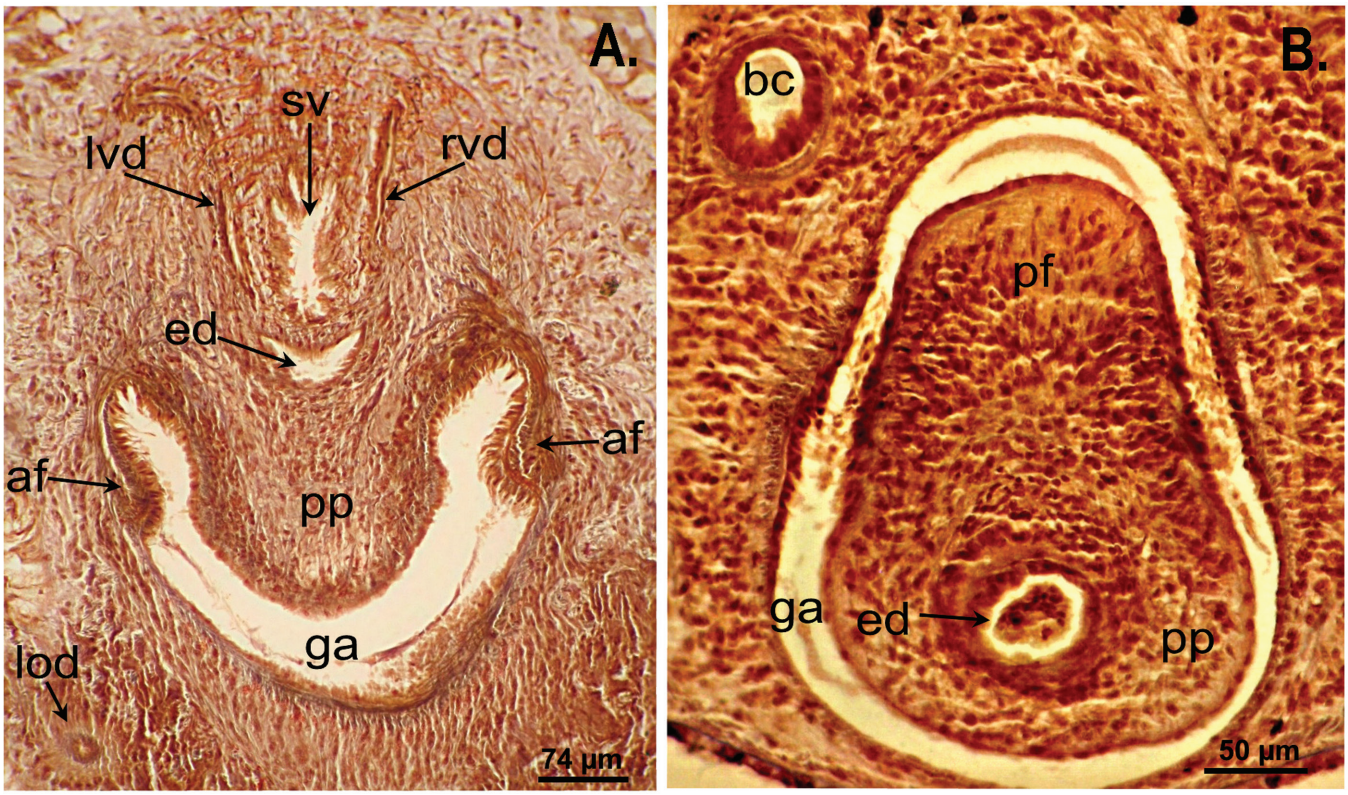


Figure 14. *Dugesia mariae*: A, CGAS Pla 27. 1, photomicrograph showing the penis bulb with the seminal vesicle (sv), left (lvd) and right (rvd) vas deferens, the penial papilla (pp) and the two atrial folds (af) in horizontal section; B, CGAS Pla 27. 4, photomicrograph showing the penis papilla (pp) and the penial fold (pf) in transverse section.

D. iranica (Livanov, 1951), *D. lanzai* Banchetti & Del Papa, 1971, *D. leporii* Pala, Stocchino, Corso & Casu, 2000, *D. libanica* Bromley & Benazzi, 1991, *D. maghrebiana* Stocchino, Manconi, Corso, Sluys, Casu & Pala, 2009, *D. malicky*, *D. mertoni*, *D. monomyoda* Marcus, 1953, *D. naiadis* Sluys, 2013, *D. notogaea* Sluys & Kawakatsu, 1998, *D. semiglobosa* Chen & Dong, 2021, *D. sinensis* Chen & Wang, 2015, *D. pustulata* Harrath & Sluys, 2019, *D. parasagitta* Sluys & Solà, 2013, *D. ryukyuensis* Kawakatsu, 1976, *D. siamana* Kawakatsu, 1980, and *D. tubqalis* Harrath & Sluys, 2012.

In seven of these species (*D. andamanensis*, *D. bifida*, *D. borneana*, *D. mertoni*, *D. notogaea*, *D. ryukyuensis*, and *D. siamana*) a duct is present between the seminal vesicle and the diaphragm that is absent in *D. mariae*. Further, *D. mariae* shares the presence of a penial fold only with *D. iranica*, *D. libanica*, *D. malicky*, *D. parasagitta*, *D. ryukyuensis*, *D. tubqalis*, and *Dugesia* sp. from Lake Ohrid. However, these species have a small or large—but not conical—diaphragm, in contrast to the large, conical or pointed diaphragm of *D. mariae* (cf. de Vries, 1984, 1988; Sluys *et al.*, 1998, 2013; Pala *et al.*, 2000; Stocchino *et al.*, 2009, 2012, 2013, 2014, 2017; Harrath *et al.*, 2012, 2019; Chen *et al.*, 2015; Wang *et al.*, 2021a, b).

The fact that in *D. mariae* the ectal reinforcement extends to almost the copulatory bursa, contrasts with the situation in *D. afromontana*, *D. annandalei*, *D. circumcisa*, *D. didiaphragma*, *D. ectophysa*, *D. iranica*, *D. maghrebiana*, *D. monomyoda*, *D. naiadis*, and *D. semiglobosa*, in which this layer of longitudinal muscles around the bursal canal is completely absent or is confined to the vaginal region of the bursal canal or extends to only to halfway along the canal.

Concurrently, molecular evidence has consistently pointed to the separate specific status of *D. mariae* throughout all phases of our molecular species delimitation, irrespective of the dataset analysed or the particular methodology employed. Further, it is noteworthy that in a recently published study on the phylotranscriptomics of the genus *Dugesia* in the Western Mediterranean, evidence was presented for the reciprocal monophyly of the Corsican and Sardinian nominal *D. benazzii* s.l. populations (Benítez-Álvarez *et al.*, 2023). For *D. hoidi*, molecular species delimitation also showed it as a species different from the rest of *D. benazzii* s.l. populations analysed, a result that is also supported by its differentiated karyotype.

In conclusion, all available evidence, as presented and discussed above, points unequivocally to the separate species status of both *D. mariae* and *D. hoidi*.

DISCUSSION

Phylogenetic relationships between *Dugesia benazzii* s.l. and *Dugesia hepta*

Our phylogenetic analyses strongly support the existence of at least four mitochondrial clades and one singleton within *D. benazzii* s.l., although most of these lineages were not distinguished by the nuclear data (Fig. 2B–D; Supporting Information, Fig. S1B–D). Moreover, phylogenetic trees based on the concatenated mitochondrial and nuclear datasets

retrieved the same four clades, as well as the singleton, which were fully supported (Fig. 2E1, E2; Supporting Information, Fig. S1E). The properties of mtDNA are such that it (i) usually does not recombine, (ii) has smaller effective population sizes than nuclear DNA, and also (iii) has shorter coalescent times (Barr *et al.* 2005, Solà *et al.* 2015). Together with the low dispersal ability of planarians (Álvarez-Presas and Riutort, 2014), these properties explain why mitochondrial data show more structure than nuclear data. The lack of resolution in the latter is a consequence, most probably, of the conservation of the nuclear genes, which may reflect older splitting events. Therefore, it is not surprising that concatenated trees resemble the mitochondrial ones. Nonetheless, improvement of node support values in the concatenated trees suggests that nuclear data synergizes and is congruent with the mitochondrial information.

The topology of our present mitochondrial phylogenetic tree (Fig. 3), differs from that presented in Dols-Serrate *et al.* (2020), particularly in the positions of *D. benazzii* s.l. and *D. hepta*. In Dols-Serrate *et al.* (2020) it was concluded that *D. benazzii* s.l. was not a monophyletic group but a paraphyletic one, and that the populations from Sardinia formed the sister-group of *D. hepta*. That hypothesis seemed to gain weight because of the hybridization that was demonstrated between *D. hepta* and *D. benazzii* s.l. in Sardinia.

Nevertheless, several issues cast some doubt on the conclusions drawn in the study of Dols-Serrate *et al.* (2020): (i) almost complete lack of nuclear sequences from Corsican specimens; (ii) then available taxonomic accounts did not highlight the differences between Corsican and Sardinian *D. benazzii* s.l. populations; and (iii) identical ITS1 sequences for all *D. benazzii* s.l. specimens. Therefore, we increased the number of mitochondrial and nuclear markers, as well as the number of Corsican samples, in order to better resolve the relationship between these two taxa and to be able to perform a thorough species-delimitation analysis.

Our current phylogenetic trees show that *D. benazzii* s.l. is a monophyletic group, which is sister to *D. hepta* and is supported by both mitochondrial and nuclear markers (Figs 2, 3; Supporting Information, Fig. S1A–E). Furthermore, there is a clear distinction, based on mitochondrial data, between Corsican and Sardinian lineages, although nuclear data are less conclusive on this subject. Additionally, when mitochondrial and nuclear data were concatenated, the BI method yielded a new topology (Fig. 2E1; Supporting Information, Fig. S1E1; topology-2 in: Fig. 4; Supporting Information, Fig. S3), in which the *D. benazzii* s.l. populations from each island were recovered as independent monophyletic lineages. From a geographical point of view, topology-2 was the most plausible one and, thus, was tested by using a non-conflictive dataset (dataset I; see above under Materials and methods). The results of this topology test indicated that topology-2 could not be rejected. Consequently, with the present data it was not possible to resolve the relationships of the Badu de Mola population. Taking all of this into account, as well as the fact that Sardinian and Corsican populations can interbreed under laboratory conditions (Benazzi, 1955), the hypothesis gains weight that *D. benazzii* s.l. is monophyletic. Other molecular analyses recovered similar phylogenetic relationships (Lázaro *et al.*, 2009; Solà *et al.*, 2013, 2022; Benítez-Álvarez *et al.*, 2023), but none of these studies aimed to thoroughly study the phylogenetic relationship between nominal *D. benazzii* s.l. and *D. hepta*.

Performance of molecular species delimitation methods

Species delimitation on the basis of mitochondrial data resulted in a high number of predicted species, due to the structured nature of dataset I, in some cases delimiting also species for the deepest splits within the four main mitochondrial clades (Fig. 3). It is noteworthy that mPTP clearly outperformed ABGD and GMYC, with the results of the former being validated. However, we opted to test the GMYC delimitation scheme, instead of mPTP's, and proceeded with the subsequent downstream validation analyses in order to test the population splits of *D. benazzii* s.l. within each island as well. This strategy was chosen because in the validation step the coalescent-based species-delimitation methods may lump candidate species even if the divergence is recent and level of gene flow high (Knowles and Carstens, 2007; Grummer *et al.*, 2014).

In the BPP molecular validation, we considered four different evolutionary scenarios, based on combinations of different values for two priors, namely, the ancestral population size (θ) and the divergence time (τ). Depending on the values of those priors and their various combinations, we built either more permissive or restrictive models to estimate the posterior probabilities of the splitting events in our species hypotheses. Empirical data show that large θ should favour fewer species, as compared with a smaller θ (Leaché and Fujita, 2010; McKay *et al.*, 2013), whilst τ works the other way around (Leaché and Fujita, 2010). Consequently, a large θ and/or a small τ implies that populations may not have evolved into separate species. However, this may not hold true in some cases, especially for τ, for which the effect on the delimitation probability is less predictable (Rannala, 2015). Thus, in the present study, p.c. 2 is the most restrictive value, whereas p.c. 4 is the most permissive (Fig. 4; Supporting Information, Fig. S3).

Further, it is noteworthy that different datasets, especially dataset IV, generated different delimitation results, depending upon p.c. and the topology of the phylogenetic trees. For example, nuclear data for topology-1 show only slight support for the split of PSC4 (Supporting Information, Fig. S3) under the most restrictive scenario (p.c. 2), and fail to support that same split, or other well-established splits, under the most permissive scenario (p.c. 4) in what appears to be an artefactual result. In contrast, results for the same dataset under topology-2 appear more congruent, with PSC4 not being supported under restrictive scenarios (p.c. 2 and p.c. 3), while p.c. 4 supports basal splits (B vs. B'). This suggests that BPP is sensitive to topology and that, therefore, multiple topologies should be tested during the validation step in order to minimize the impact of possible artefacts. We noticed that with our data, θ behaves as expected, as the p.c. scenarios with smaller θ are the ones supporting an increasing number of splits, apparently regardless of τ, of which the effect is more difficult to assess.

BFD analyses yielded results similar to BPP. Analyses involving mitochondrial data (datasets I and V) favoured scenarios with the maximum number of species (Fig. 5; Supporting Information, Table S4). Curiously, under conservative p.c. scenarios for topology-2, the results of BFD for dataset IV were the same as those of BPP for this dataset, thus corroborating the notion that topology-2 is more representative of the actual phylogenetic relationships. The models implemented in BPP are nested statistical hypotheses, in which the one-species model is the null hypothesis and the two-species model is the alternative hypothesis at each node (McKay *et al.*, 2013). Since the null model assumes random mating, any statistically significant deviation from random mating will cause statistical methods, such as BPP, to reject the null model, whereas BFD is non-nested and more conservative than BPP (Yu *et al.*, 2017). Nevertheless, the results of BFD under dataset V results are congruent with the BPP results under less restrictive scenarios (p.c. 1 and p.c. 4).

Dugesia species delimitation

Dugesia hepta and *D. benazzii* s.l. live in shallow, non-eutrophic waters with temperatures ranging between 15 and 20°C (Stocchino *et al.*, 2005). Most of their populations have been steadily displaced to the upper reaches of the rivers, due to the increasing warming of the water in the lower reaches, as well as the presence of species with a greater tolerance to temperature fluctuations, such as *D. sicula* Lepori, 1948 (Pala *et al.*, 1995; Leria *et al.*, 2022). Presently, during mid-summer and early autumn, the Mediterranean climate causes long droughts on the islands, with the result that sections of the river dry up, or even the entire river when waters are already shallow. This causes fragmentation of the habitat of the planarians and, thereby, may isolate their populations. It is not unlikely that in this way populations of *D. hepta* and *D. benazzii* s.l. regularly undergo seasonal bottlenecks, which may explain the observed structure in their mtDNA signal and also the fact that the nuclear data yield higher support values under small θ scenarios. However, it has been argued that BPP may not really be a conservative method for delimiting species, since it is sensitive to the divergent effects of non-random mating caused by intraspecific processes, such as isolation by distance (McKay *et al.*, 2013). Moreover, Sukumaran and Knowles (2017) showed that the multispecies coalescent may delimit genetic structure, instead of species, and that BPP tends to overestimate the real number of species. In view of this, our BPP results may reflect the isolation of *D. benazzii* s.l. populations. Given the structured nature of our mitochondrial data, we followed a conservative approach to species delimitation and took into account also other kinds of data, such as nuclear data, karyology, and morphology, thereby effectively implementing an integrative analysis of the available information.

CONCLUSIONS

Although molecular-based species-delimitation methods have been much improved over the years, our present study strongly suggests that researchers should implement an integrative approach, whenever possible. The reason for this is that exclusive use of molecular genetic data may lead to an overestimation of the true number of species. This effect may be exacerbated by the fact that current genomic resources provide unprecedented power in the detection of increasingly fine-scaled genetic structures (Sukumaran and Knowles, 2017). In addition, our results show that different combinations of priors may have an impact on the number of validated splits in BPP delimitation, at least when analysing presumably recently diverged lineages.

With respect to the taxonomic status of the various *Dugesia* populations considered in the present study, the separate specific status of the former nominal species *D. benazzii* s.l. was rejected, while the specific epithet *benazzii* was restricted to populations with morphological characters shared with those of *D. benazzii* s.s. as reported in the above redescription. Two other, new species were split off from *D. benazzii* s.l., viz., the Sardinian *D. hoidi* and the Corsican *D. mariae*.

SUPPLEMENTARY DATA

Supplementary data are available at *Zoological Journal of the Linnean Society* online.

ACKNOWLEDGEMENTS

We are grateful to Gema Blasco, to whom we owe the excellent maintenance of our samples in the laboratory and the general laboratory organization. We express our gratitude to Laia Leria for assisting with the fieldwork, and Professor Joaquim Dols Rusiñol for making available some of the references cited in this paper. We gratefully thank Professor Maria Pala for her kind support and for having devoted her whole professional life to science and flatworms.

CONFLICT OF INTEREST

None declared.

FUNDING

This research was supported by the Ministerio de Economía y Competitividad (projects CGL2015–63527 and PGC2018-093924-B-100) from Spain. This research was supported as well by the Ajuts de Personal Investigador predoctoral en Formació (APIF) grant from the Universitat de Barcelona (to Daniel Dols-Serrate). Giacinta A. Stocchino was supported by a grant from SYNTHESYS (European Union-funded Integrated Activities, grant: NL-TAF 8228), and from Fondazione di Sardegna (grant FdS/RAS-2016/CUP J86C1800082005) and Regione Autonoma Sardegna (grant RAS2012-LR7/2007-CRP-60215).

DATA AVAILABILITY

The data underlying this article are available in the GenBank Nucleotide Database, and can be accessed with the accession numbers provided in Appendix Table A1.

REFERENCES

Agapow, P-M.. Species: demarcation and diversity. In: Purvis A, Gittleman JL (eds), *Phylogeny Conservation*. Cambridge: Cambridge University Press, 2005;57–75.

Álvarez-Presas M, Riutort M. Planarian (Platyhelminthes, Tricladida) diversity and molecular markers: a new view of an old group. *Diversity* 2014;**6**:323–338. https://doi.org/10.3390/d6020323

Arndt W. Untersuchungen an Bachtrikladen. Ein Beitrag zur Kenntnis der Paludicolen Korsikas, Rumäniens und Sibiriens. *Zeitschrift für Wissenschaftliche Zoologie* 1922;**120**:98–146.

Baele G, Lemey P, Bedford T *et al*. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* 2012;**29**:2157–2167. https://doi.org/10.1093/molbev/mss084

Ball, I.R. A contribution to the phylogeny and biogeography of freshwater triclads (Platyhelminthes: Turbellaria). In: Riser NW, Morse MP (eds), *Biology of the Turbellaria*. New York: McGraw–Hill, 1974;339–401.

Ball, I.R., Reynoldson, T.B., 1981. *British Planarians*. Cambridge University Press, Cambridge.

Barr CM, Neiman M, Taylor DR. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* 2005;**168**:39–50. https://doi.org/10.1111/j.1469-8137.2005.01492.x

de Beauchamp P. Les Turbellariés. Histoire du peuplement de la Corse. *Bulletin de la Société des Sciences historiques et naturelles de la Corse* 1926;**45**:149–151.

de Beauchamp P. Biospeologica LVIII. Turbellariés, Hirudinées, Branchiobdellidés. *Archives de Zoologie Expérimentale et Générale* 1932;**73**:113–380.

Benazzi M. Ginogenesi in tricladi di acqua dolce. *Chromosoma* 1950;**3**:474–482.

Benazzi M. Appunti sulla distribuzione dei triclacli in italia. *Bollettino di Zoologia* 1955;**22**:149–164. https://doi.org/10.1080/11250005509439195

Benítez-Álvarez L, Leria L, Fernández R *et al*. Phylotranscriptomics interrogation uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean. *Molecular Phylogenetics and Evolution* 2023;**178**:107649. https://doi.org/10.1016/j.ympev.2022.107649

Bishop A. Notes upon *Sieboldiellina planariarum* (Siebold), a ciliate parasite of *Planaria torva*. *Parasitology* 1926;**18**:187–194.

Chen Y, Chen X, Wu C *et al*. A new species of the genus *Dugesia* (Tricladida: Dugesiidae) from China. *Zoological Systematics* 2015;**40**:237–249. https://doi.org/10.11865/zs.20150301

Dayrat B. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 2005;**85**:407–415.

Dols-Serrate D, Leria L, Aguilar JP *et al*. *Dugesia hepta* and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex? *Organisms Diversity and Evolution* 2020;**20**:369–386. https://doi.org/10.1007/s13127-020-00438-z

Draparnaud JPR. Tableau des Mollusques terrestres et fluviatiles de la France. Montpellier & Paris: Renaud. 1801.

Drummond AJ, Suchard MA, Xie D *et al*. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 2012;**29**:1969–1973. https://doi.org/10.1093/molbev/mss075

Ezard T, Fujisawa T, Barraclough TG. 2009. SPLITS: species' limits by threshold statistics. *R package version 1.0–11/r29*. Available from: http://R-Forge.R-project.org/projects/splits/ (last accessed September 2023).

Flouri, T., Rannala, B., Yang, Z. A tutorial on the use of BPP for species tree estimation and species delimitation. In: Delsuc F, Galtier N, Scornavacca C (eds), *Phylogenetics in the Genomic Era* 2020;1–16. No commercial publisher. Authors open access book, hal02536475.

Frankham R, Ballou JD, Dudash MR *et al*. Implications of different species concepts for conserving biodiversity. *Biological Conservation* 2012;**153**:25–31. https://doi.org/10.1016/j.biocon.2012.04.034

Gamo García J, Rodríguez Caabeiro F. *Dugesia gonocephala* (Turbellaria, Tricladida) nueovo hospedador para *Sieboldiellina planariarum* (Ciliata, Holotricha, Astomata). *Revista Ibérica de Parasitología* 1981;**41**:313–315.

Girard C. A brief account of the fresh water planariae of the United States. *Proceedings of the Boston Society of Natural History* 1850;**3**:264–265.

von Graff L. Die Turbellarien als Parasiten und Wirte. *Leuschner and Lubensky's Universitäts-Buchhandlung, Graz* 1903:1–66+ 3 pls.

Grummer JA, Bryson RW, Reeder TW. Species delimitation using bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* 2014;**63**:119–133. https://doi.org/10.1093/sysbio/syt069

Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 1999;**41**:95–98. https://thalljiscience.github.io/. Last accessed: September 2023.

Harrath AH, Sluys R, Merzoug D *et al*. Freshwater planarians (Platyhelminthes, Tricladida) from the Palearctic section of the African continent: new records, with the description of a new species. *Zootaxa* 2012;**3182**:1–15. https://doi.org/10.11646/zootaxa.3182.1.1

Harrath AH, Sluys R, Mansour L *et al*. Molecular and morphological identification of two new African species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from Cameroon. *Journal of Natural History* 2019;**53**:253–271. https://doi.org/10.1080/00222933.2019.1577508

Kapli P, Lutteropp S, Zhang J *et al*. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* 2017;**33**:1630–1638. https://doi.org/10.1093/bioinformatics/btx025

Kass RE, Raftery AE. Bayes Factors. *Journal of the American Statistical Association* 1995;**90**:773–795.

Katoh K, Standley DM. MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* 2013;**30**:772–780. https://doi.org/10.1093/molbev/mst010

Kawakatsu M. Parasites of Turbellarians: records of several species of endoparasites of freshwater planarians. *The Research Bulletin of the Meguro Parasitological Museum* 1970;**3**:37–47 [in Japanese, with English abstract].

Keller JM. Les planaires épigées d'eau douce. *Revue française d'Aquariologie* 1983;**10**:25–32.

Kenk R. Freshwater triclads (Turbellaria) of North America. I. The genus *Planaria*. *Proceedings of the Biological Society of Washington* 1969;**82**:539–558.

Knowles LL, Carstens BC. Delimiting species without monophyletic gene trees. *Systematic Biology* 2007;**56**:887–895. https://doi.org/10.1080/10635150701701091

Kumar S, Stecher G, Li M *et al*. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 2018;**35**:1547–1549. https://doi.org/10.1093/molbev/msy096

Lanfear R, Calcott B, Ho SYW *et al*. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 2012;**29**:1695–1701. https://doi.org/10.1093/molbev/mss020

Lanfear R, Frandsen PB, Wright AM *et al*. Partitionfinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 2017;**34**:772–773. https://doi.org/10.1093/molbev/msw260

Lázaro E, Sluys R, Pala M *et al*. Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution* 2009;**52**:835–845. https://doi.org/10.1016/j.ympev.2009.04.022

Lázaro EM, Harrath A, Stocchino GA *et al*. *Schmidtea mediterranea* phylogeography: an old species surviving on a few Mediterranean islands? *BMC Evolutionary Biology* 2011;**11**:1–15. https://doi.org/10.1186/1471-2148-11-274

Leaché AD, Fujita MK. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B: Biological Sciences* 2010;**277**:3071–3077. https://doi.org/10.1098/rspb.2010.0662

Lepori NG. Sulle caratteristiche morfologiche e sulla posizione sistematica della planaria di Sardegna e Corsica già ascritta a *Dugesia* (=*Euplanaria*) *gonocephala* (Dugès). *Atti Società Toscana di Scienze Naturali* 1951;**58**:28–47.

Lepori NG, Pala M. Fissioning in planarians. 1. Karyological analysis of fissiparous strains of *Dugesia gonocephala* s. l. (Turbellaria, Tricladida) collected in Sardinia, in order to determine the factors responsible for fissioning. *Monitore Zoologica Italiano* 1982;**16**:105–131.

Leria L, Vila-Farré M, Álvarez-Presas M *et al.* Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhelminthes, Tricladida): extreme intraindividual genetic diversity, morphological stasis, and karyological variability. *Molecular Phylogenetics and Evolution* 2020;**143**:106496. https://doi.org/10.1016/j.ympev.2019.05.010

Leria L, Riutort M, Romero R *et al.* Microplate tectonics and environmental factors as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography* 2022;**49**:1124–1136. https://doi.org/10.1111/jbi.14373

Lom J. Beiträge zur Kenntnis der parasitischen Ciliaten aus Evertebraten IV. Neue Ciliaten aus der Familie Haptophryidae Cépède 1923, nebst einigen Bemerkungen zum heutigen Stand dieser Gruppe. *Archiv für Protistenkunde* 1959;**104**:133–154.

Maddison WP. Gene trees in species trees. *Systematic Biology* 1997;**46**:523–536. https://doi.org/10.1093/sysbio/46.3.523

Marcus E. Turbellaria, Tricladida. In: Exploration du Parc National de l'Upemba, *Mission G. F. de Witte*. Bruxelles: M. Hayez, Fasc. 1953;**21**:1–62.

McKay BD, Mays HL, Wu Y *et al.* An empirical comparison of character-based and coalescent-based approaches to species delimitation in a young avian complex. *Molecular Ecology* 2013;**22**:4943–4957. https://doi.org/10.1111/mec.12446

Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)* 2010;1–8. https://doi.org/10.1109/GCE.2010.5676129

Newmark PA, Alvarado AS. Not your father's planarian: a classic model enters the era of functioning genomics. *Nature Reviews Genetics* 2002;**3**:210–219. https://doi.org/10.1038/nrg759

Pala M, Casu S, Vacca RA. *Dugesia hepta*, nuova specie di planaria di acqua dolce di Sardegna appartenente alla superspecie *Dugesia gonocephala* (Dugès). *Bollettino della Società Sarda di Scienze Naturali* 1981;**20**:97–107. https://core.ac.uk/reader/11689490 (last accessed 7 June 2023).

Pala M, Casu S, Lepori NG. Stabilized natural interspecific hybrid population of the fresh water planarians *Dugesia gonocephala* s.l. (Turbellaria, Tricladida). *Caryologia* 1982;**35**:247–256. https://doi.org/10.1080/00087114.1982.10796929

Pala M, Vacca RA, Casu S *et al.* The freshwater planarian *Dugesia sicula* Lepori from Sardinia (Platyhelminthes, Tricladida). *Hydrobiologia* 1995;**310**:151–156. https://doi.org/10.1007/BF00015533

Pala M, Casu S, Stocchino GA. Karyology and karyotype analysis of diploid freshwater planarian populations of the *Dugesia gonocephala* group (Platyhelminthes, Tricladida) found in Sardinia. *Hydrobiologia* 1999;**392**:113–119. https://doi.org/10.1023/A:1003534507632

Pala M, Stocchino GA, Corso G *et al.* A new species of freshwater planarian belonging to the genus *Dugesia* (Platyhelminthes, Tricladida) from Sardinia. *Italian Journal of Zoology* 2000;**67**:379–383. https://doi.org/10.1080/11250000009356343

Pons J, Barraclough TG, Gomez-Zurita J *et al.* Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology* 2006;**55**:595–609. https://doi.org/10.1080/10635150600852011

Posada D. 2009. Selecting models of evolution. In: Lemey P, Salemi M, Vandamme AM (eds), *The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd edn. Cambridge: Cambridge University Press, 345–361.

Puillandre N, Lambert A, Brouillet S *et al.* ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology* 2012;**21**:1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

de Puytorac P. Contribution à l'étude des ciliés astomes Haptophryidae Cépède, 1903 (cytologie, ultrastructure, taxinomie). *Annales des Sciences naturelles, Zoologie et Biologie Animale* 1963;**5**:173–210.

de Queiroz, K. The general lineage concept of species, species criteria, and the process of speciation. A conceptual unification and terminological recommendations. In: Howard DJ, Berlocher S (eds), *Endless Forms: Species and Speciation*. New York: Oxford University Press, 1998;57–75.

de Queiroz K. Species concepts and species delimitation. *Systematic Biology* 2007;**56**:879–886. https://doi.org/10.1080/10635150701701083

Rambaut A, Drummond AJ. *TreeAnnotator v.1.6.1*, 2010. Available from http://beast.bio.ed.ac.uk (last accessed September 2022).

Rambaut A, Suchard MA, Xie D, Drummond AJ. *Tracer v.1.6*, 2014. Available from http://beast.bio.ed.ac.uk/Tracer (last accessed September 2022).

Rannala B. The art and science of species delimitation. *Current Zoology* 2015;**61**:846–853. https://doi.org/10.1093/czoolo/61.5.846

Rataj M, Vd'ačny P. Dawn of astome ciliates in light of morphology and time-calibrated phylogeny of *Haptophrya planariarum*, an obligate endosymbiont of freshwater turbellarians. *European Journal of Protistology* 2018;**64**:54–71. https://doi.org/10.1016/j.ejop.2018.03.004

Rataj M, Zhang R, Vd'ačny P. Nuclear and mitochondrial SSU rRNA genes reveal hidden diversity of *Haptophrya* endosymbionts in freshwater planarians and challenge their traditional classification in Astomatia. *Frontiers in Microbiology* 2022;**13**:830951.

Ronquist F, Teslenko M, Van Der Mark P *et al.* MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 2012;**61**:539–542. https://doi.org/10.1093/sysbio/sys029

Shimodaira H, Hasegawa M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 2001;**17**:1246–1247. https://doi.org/10.1093/bioinformatics/17.12.1246

Silvestro D, Michalak I. RaxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution* 2012;**12**:335–337. https://doi.org/10.1007/s13127-011-0056-0

Sluys R, Kawakatsu M, Winsor L. The genus *Dugesia* in Australia, with its phylogenetic analysis and historical biogeography (Platyhelminthes, Tricladida, Dugesiidae). *Zoologica Scripta* 1998;**27**:273–290.

Sluys R, Solà E, Gritzalis K *et al.* Integrative delineation of species of Mediterranean freshwater planarians (Platyhelminthes: Tricladida: Dugesiidae). *Zoological Journal of the Linnean Society* 2013;**169**:523–547. https://doi.org/10.1111/zoj.12077

Solà E, Sluys R, Gritzalis K *et al.* Fluvial basin history in the northeastern Mediterranean region underlies dispersal and speciation patterns in the genus *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution* 2013;**66**:877–888. https://doi.org/10.1016/j.ympev.2012.11.010

Solà E, Álvarez-Presas M, Frías-López C *et al.* Evolutionary analysis of mitogenomes from parasitic and free-living flatworms. *PLoS One* 2015;**10**:e0120081–20. https://doi.org/10.1371/journal.pone.0120081

Solà E, Leria L, Stocchino GA *et al.* Three dispersal routes out of Africa: a puzzling biogeographical history in freshwater planarians. *Journal of Biogeography* 2022;**49**:1219–1233. https://doi.org/10.1111/jbi.14371

Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 2006;**22**:2688–2690. https://doi.org/10.1093/bioinformatics/btl446

Steinböck O. Untersuchungen über die Geschlechts-Darmverbindung bei Turbellarien nebst einem Beitrag zur Morphologie des

Trikladendarmes. *Zeitschrift für Morphologie und Ökologie der Tiere* 1924;**2**:461–504.
Stocchino GA. 80 years of research on planarians (Platyhelminthes, Tricladida) from Sardinia, Italy: an annotated checklist. *Zootaxa* 2018;**4532**:539–552. https://doi.org/10.11646/zootaxa.4532.4.5
Stocchino GA, Manconi R. Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology* 2013;**803**:319–328. https://doi.org/10.1080/11250003.2013.822025
Stocchino GA, Corso G, Manconi R *et al.* Endemic freshwater planarians of Sardinia: redescription of *Dugesia hepta* (Platyhelminthes, Tricladida) with a comparison of the Mediterranean species of the genus. *Journal of Natural History* 2005;**39**:1947–1960. https://doi.org/10.1080/00222930500060025
Stocchino GA, Manconi R, Corso G *et al.* African planarians: Morphology and karyology of *Dugesia maghrebiana* sp. n. (Platyhelminthes, Tricladida) from Tunisia. *Italian Journal Zoology* 2009;**76**:83–91. https://doi.org/10.1080/11250000802141683
Stocchino GA, Sluys R, Manconi R. A new species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from the Afromontane forest in South Africa, with an overview of freshwater planarians from the African continent. *Zootaxa* 2012;**3551**:43–58. https://doi.org/10.11646/zootaxa.3551.1.3
Stocchino GA, Sluys R, Deri P *et al.* Integrative taxonomy of a new species of planarian from the Lake Ohrid basin, including an analysis of biogeographical patterns in freshwater triclads from the Ohrid region (Platyhelminthes, Tricladida, Dugesiidae). *ZooKeys* 2013;**313**:25–43. https://doi.org/10.3897/zookeys.313.5363
Stocchino GA, Sluys R, Manconi R. A new and aberrant species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from Madagascar. *ZooKeys* 2014;**425**:71–88. https://doi.org/10.3897/zookeys.425.7811
Stocchino GA, Sluys R, Riutort M *et al.* Freshwater planarian diversity (Platyhelminthes: Tricladida: Dugesiidae) in Madagascar: new species, cryptic species, with a redefinition of character states. *Zoological Journal of the Linnean Society* 2017;**181**:727–756. https://doi.org/10.1093/zoolinnean/zlx017
Stocchino GA, Sluys R, Harrath AH *et al.* The invasive alien freshwater flatworm *Girardia tigrina* (Girard, 1850) (Platyhelminthes, Tricladida) in Western Europe: new insights into its morphology, karyology and reproductive biology. *Contributions to Zoology* 2019;**88**:1–21. https://doi.org/10.1163/18759866-20191406
Stocchino GA, Dols-Serrate D, Sluys R, *et al.* Amphibioplanidae: a new branch and family on the phylogenetic tree of the triclad flatworms (Platyhelminthes: Tricladida), represented by a species from Sardinian caves with a remarkable lifestyle. *Zoological Journal of the Linnean Society* 2021;**193**:1364–1391.
Sukumaran J, Knowles LL. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences* 2017;**114**:1607–1612. https://doi.org/10.1073/pnas.1607921114
Vacca RA, Casu S, Pala M. Popolamento planariologico dei fiumi del nord Sardegna: 2. I cariotipi dei tricladi d'acqua dolce rinvenuti nel bacino idrografico del fiume Coghinas. *Bollettino della Società Sarda di Scienze Naturali* 1993;**29**:59–73.
Vieites DR, Wollenberg KC, Andreone F *et al.* Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences* 2009;**106**:8267–8272. https://doi.org/10.1073/pnas.0810821106
de Vries EJ. On the species of the *Dugesia gonocephala* group (Platyhelminthes, Turbellaria, Tricladida) from Greece. *Bijdragen tot de Dierkunde* 1984;**54**:101–126.
de Vries EJ. Further contributions to the taxonomy and biogeography of the subgenus *Dugesia* (Platyhelminthes: Tricladida: Paludicola) in the Mediterranean region and the Middle East. *Israel Journal of Zoology* 1988;**35**:109–316.
de Vries EJ, Sluys R. Phylogenetic relationships of the genus *Dugesia* (Platyhelminthes, Tricladida, Paludicola). *Journal of Zoology* 1991;**223**:103–116.
Wang L, Chen JZ, Dong ZM *et al.* Two new species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from the tropical monsoon forest in southern China. *ZooKeys* 2021a;**1059**:89–116. https://doi.org/10.3897/zookeys.1059.65633
Wang L, Dong ZM, Chen GW *et al.* Integrative descriptions of two new species of *Dugesia* from Hainan Island, China (Platyhelminthes, Tricladida, Dugesiidae). *ZooKeys* 2021b;**1028**:1–28. https://doi.org/10.3897/zookeys.1028.60838
Wang L, Wang YX, Dong ZM *et al.* Integrative taxonomy unveils a new species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from the southern portion of the Taihang Mountains in northern China, with the description of its complete mitogenome and an exploratory analysis of mitochondrial gene order as a taxonomic character. *Integrative Zoology* 2022;**17**:1193–1214. https://doi.org/10.1111/1749-4877.12605
Winsor, L., Sluys, R. Basic histological techniques for planarians. In: Rink, J. C. (ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology*, Methods in Molecular Biology, Vol. **1774**. Springer Science+Business Media, New York, 2018;285-351. https://doi.org/10.1007/978-1-4939-7802-1_9
Xia X, Xie Z. DAMBE: Software package for data analysis in molecular biology and evolution. *Journal of Heredity* 2001;**92**:371–373. https://doi.org/10.1093/jhered/92.4.371
Xia X, Xie Z, Salemi M *et al.* An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* 2003;**26**:1–7. https://doi.org/10.1016/S1055-7903(02)00326-3
Xie W, Lewis PO, Fan Y *et al.* Improving marginal likelihood estimation for bayesian phylogenetic model selection. *Systematic Biology* 2011;**60**:150–160. https://doi.org/10.1093/sysbio/syq085
Yang Z. The BPP program for species tree estimation and species delimitation. *Current Zoology* 2015;**61**:854–865. https://doi.org/10.1093/czoolo/61.5.854
Yang Z, Rannala B. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 2010;**164**:1645–1656. https://doi.org/10.1073/pnas.0913022107
Yu G, Rao D, Matsui M *et al.* Coalescent-based delimitation outperforms distance-based methods for delineating less divergent species: the case of *Kurixalus odontotarsus* species group. *Scientific Reports* 2017;**7**:1–13. https://doi.org/10.1038/s41598-017-16309-1

Appendix Table A1. Complete list of specimens used for the present study and pertinent Gen Bank accession numbers. Population numbers from Dols-Serrate et al. (2020) are presented to facilitate comparisons with that study.

Locality		Coordinates				Population number		
Specimen	**Species**	**Karyotype**	**COI**	**nad1-tRNAW-COII**	**Dunuc10**	**Dunuc12**	**Collectors***	**Date of collection**
Badde Manna, Sardinia		**40°46'52"N, 9°06'59"E**				**2**		
BMA0028-01	*D. benazzii*	-	MN525640	OP720241	-	MN442438	P	17/07/1997
Badu de Mola, Sardinia		**40°18'11"N, 8°34'57"E**				**3**		
BMO0025-03	*D. hoidi*	-	MN525637	OP720240	OP795470	MN442435	P	17/07/1997
Bunnari, Sardinia		**40°42'17"N, 8°35'46"E**				**6**		
BU0091-01	*D. hepta*	-	MN525674	OP720245	OP795473	MN442450	P, S, L	06/05/2010
BU0091-04	*D. hepta*	-	MN525677	OP720246	-	MN442451	P, S, L	06/05/2010
Mascari 1, Sardinia		**40°41'57"N, 8°35'03"E**				**9**		
MAI0078-01	*D. hepta*	-	MN525650	OP720244	OP795472	MN442443	P, S, L	06/05/2010
MAI0342-01	*D. hepta*	n = 7	MK385923	OP720277	-	MK713158	S	01/01/2013
MAI0343-01	*D. hepta*	n = 7	MN525730	OP720278	OP795484	MN442483	S	01/01/2013
MAI0344-01	*D. hepta*	n = 7	MK385924	OP720279	-	MK713159	S	01/01/2013
MAI0345-01	*D. hepta*	n = 7	MN525731	OP720280	-	MN442484	S	01/01/2013
Mascari 2, Sardinia		**40°41'60"N, 8°35'23"E**				**10**		
MAII0355-01	*D. hepta*	n = 7	MN525736	OP720281	OP795485	MN442494	S	01/01/2013
Ponte Valenti, Sardinia		**40°28'09"N, 8°48'21"E**				**11**		
PV0098-01	*D. hepta*	-	MN525699	OP720247	-	MN442463	P, S, L	10/05/2010
PV0098-03	*D. hepta*	-	OP561859	OP720248	-	-	P, S, L	10/05/2010
PV0098-04	*D. hepta*	-	MN525700	OP720249	-	MN442465	P, S, L	10/05/2010
PV0099-01	*D. hepta*	-	MN525701	OP720250	OP795474	MN442466	P, S, L	10/05/2010
Mulinu, Sardinia		**40°24'25"N, 8°37'35"E**				**12**		
MU0173-04	*D. hepta*	-	MN525722	OP720273	-	MN442475	P, S, L	13/05/2010
MU0174-01	*D. benazzii*	-	MN525725	OP720274	OP795482	MN442478	P, S, L	13/05/2010
MU0174-05	*D. benazzii*	-	MN252726	OP720275	OP795483	MN442481	P, S, L	13/05/2010
MU0174-06	*D. hepta*	-	MN525727	OP720276	-	MN442482	P, S, L	13/05/2010
Su Rizzolu River, Otti, Sardinia		**40°43'55"N, 9°08'36"E**				**13**		
OT0022-03	*D. benazzii*	-	MN525630	OP720238	-	OP795453	P, S	28/04/1993
OT0022-04	*D. benazzii*	-	MN525631	OP720239	-	-	P, S	28/04/1993
Santa Lucia, Sardinia		**40°25'53"N, 8°50'09"E**				**14**		
SL0030-03	*D. hepta*	-	MN525641	OP720242	-	MN442439	P	01/02/2006
SL0030-04	*D. hepta*	-	MN525642	-	OP795471	MN442440	P	01/02/2006
SL0030-05	*D. hepta*	-	MK712638	OP720243	-	MK713160	P	01/02/2006
Monte Albo, Sardinia		**40°34'30"N, 9°40'31"E**				**16**		
MTA0385-01	*D. benazzii*	-	MK385925	OP720282	OP795486	MK713162	S	01/05/2012
MTA0385-02	*D. benazzii*	-	MK385926	OP720283	-	MK713163	S	01/05/2012
MTA0385-03	*D. benazzii*	-	MN525740	OP720284	OP795487	MN442509	S	01/05/2012
MTA0385-04	*D. benazzii*	-	MN525741	OP720285	OP795488	-	S	01/05/2012
MTA0385-05	*D. benazzii*	-	MN525742	OP720286	-	MN442510	S	01/05/2012
MTA0385-06	*D. benazzii*	-	MN525743	OP720287	-	MN442511	S	01/05/2012
MTA0385-07	*D. benazzii*	-	MN525744	-	-	MN442512	S	01/05/2012
MTA0385-08	*D. benazzii*	-	MN525745	OP720288	-	MN442513	S	01/05/2012
MTA0385-09	*D. benazzii*	-	MN525746	OP720289	-	MN442514	S	01/05/2012
MTA0385-10	*D. benazzii*	-	MN525747	OP720290	-	MN442515	S	01/05/2012
MTA0385-11	*D. benazzii*	-	MN525748	OP720291	-	-	S	01/05/2012
MTA0385-12	*D. benazzii*	-	MN525749	OP720292	-	MN442516	S	01/05/2012

Appendix Table A1. Continued

Locality		Coordinates				Population number		
La Solenzara, Corsica		**41°50'39.6"N, 9°20'34"E**				**19**		
SO1034-01	*D. mariae*	-	MN525750	OP720293	-	-	P, S, L	01/05/2009
SO1034-02	*D. mariae*	-	MN525751	OP720294	OP795489	OP795461	P, S, L	01/05/2009
SO1034-03	*D. mariae*	-	MN525752	OP720295	OP795490	OP795462	P, S, L	01/05/2009
SO1034-04	*D. mariae*	-	MN525753	OP720296	-	-	P, S, L	01/05/2009
SO1034-05	*D. mariae*	-	MN525754	OP720297	-	-	P, S, L	01/05/2009
SO1034-06	*D. mariae*	-	OP561878	OP720298	-	-	P, S, L	01/05/2009
SO1034-07	*D. mariae*	-	OP561879	OP720299	OP795491	OP795463	P, S, L	01/05/2009
SO1034-08	*D. mariae*	-	OP561880	OP720300	-	-	P, S, L	01/05/2009
SO1034-09	*D. mariae*	-	OP561881	OP720301	-	-	P, S, L	01/05/2009
SO1034-10	*D. mariae*	-	OP561882	OP720302	-	-	P, S, L	01/05/2009
Cioccia, Corsica		**41°32'41.7"N, 9°01'24.1"E**				**20**		
CI1055-03	*D. mariae*	-	MN525765	OP720307	OP795496	OP795468	P, S, L	01/05/2009
CI1055-04	*D. mariae*	-	MN525766	OP720308	-	-	P, S, L	01/05/2009
CI1055-05	*D. mariae*	-	MN525767	-	-	-	P, S, L	01/05/2009
CI1055-07	*D. mariae*	-	OP561883	OP720309	OP795497	OP795469	P, S, L	01/05/2009
CI1055-08	*D. mariae*	-	-	OP720310	-	-	P, S, L	01/05/2009
CI1055-09	*D. mariae*	-	OP561884	OP720311	-	-	P, S, L	01/05/2009
Pero Longo, Corsica		**41°32'27.6"N, 8°58'07.2"E**				**21**		
PL1037-02	*D. mariae*	-	MN525756	OP720303	OP795492	OP795464	P, S, L	01/05/2009
PL1037-03	*D. mariae*	-	MN525757	OP720304	OP795493	OP795465	P, S, L	01/05/2009
PL1037-04	*D. mariae*	-	MN525758	OP720305	OP795494	OP795466	P, S, L	01/05/2009
PL1037-05	*D. mariae*	-	MN525759	OP720306	OP795495	OP795467	P, S, L	01/05/2009
Barchetta, Corsica		**42°30'27.0"N, 9°22'25.9"E**				**24**		
BA0101-01	*D. mariae*	-	MN525702	OP720251	OP795475	OP795454	P, S, L	17/05/2010
BA0101-02	*D. mariae*	-	MN525703	OP720252	OP795476	OP795455	P, S, L	17/05/2010
BA0101-03	*D. mariae*	-	OP561860	OP720253	OP795477	OP795456	P, S, L	17/05/2010
BA0101-04	*D. mariae*	-	OP561861	OP720254	-	-	P, S, L	17/05/2010
BA0101-05	*D. mariae*	-	OP561862	OP720255	-	-	P, S, L	17/05/2010
BA0101-06	*D. mariae*	-	OP561863	OP720256	-	-	P, S, L	17/05/2010
BA0101-07	*D. mariae*	-	OP561864	OP720257	-	-	P, S, L	17/05/2010
BA0101-08	*D. mariae*	-	OP561865	OP720258	-	-	P, S, L	17/05/2010
BA0101-09	*D. mariae*	-	OP561866	OP720259	-	-	P, S, L	17/05/2010
BA0101-10	*D. mariae*	-	OP561867	OP720260	OP795478	OP795457	P, S, L	17/05/2010
Ponti strada, Corsica		**42°30'34.41"N, 9°21'16.15"E**				**25**		
PS0102-02	*D. mariae*	-	MN525705	OP720261	-	-	P, S, L	17/05/2010
Ecole Communale, Corsica		**42°26.648'N, 9°26.386'E**				**30**		
EC0113-01	*D. mariae*	-	OP561868	OP720262	-	-	P, S, L	18/05/2010
EC0113-02	*D. mariae*	-	OP561869	OP720263	-	-	P, S, L	18/05/2010
EC0113-03	*D. mariae*	-	OP561870	OP720264	-	-	P, S, L	18/05/2010
EC0113-04	*D. mariae*	-	OP561871	OP720265	-	-	P, S, L	18/05/2010
EC0113-05	*D. mariae*	-	OP561872	OP720266	OP795479	OP795458	P, S, L	18/05/2010
EC0113-06	*D. mariae*	-	OP561873	OP720267	OP795480	OP795459	P, S, L	18/05/2010
EC0113-07	*D. mariae*	-	OP561874	OP720268	-	-	P, S, L	18/05/2010
EC0113-08	*D. mariae*	-	OP561875	OP720269	-	OP795460	P, S, L	18/05/2010
EC0113-09	*D. mariae*	-	OP561876	OP720270	-	-	P, S, L	18/05/2010
EC0113-10	*D. mariae*	-	OP561877	OP720271	-	-	P, S, L	18/05/2010

Appendix Table A1. Continued

Locality		Coordinates				Population number		
San Gavino, Corsica		**42°25.52'N, 9°25.972'E**				**31**		
SG0114-01	*D. mariae*	-	MK712641	OP720272	OP795481	MK713166	P, S, L	19/05/2010
Outgroup								
Trelles, Spain		**43°28'33.6"N, 6°44'00.9"W**						
C12.1DC	*D. subtentaculata*		MK712596	MK712667 + MK713188 + MK713061	MK712870	MK713106	V, So, Le	2013
Luarca, Spain		**39°17'20.2"N, 7°23'01.8"W**						
C13.1DC	*D. subtentaculata*		MK712612	MK712669 + MK713190 + MK713063	MK712871	-	V, So, Le	2013
Monte de Vento, Portugal		**39°17'20.2"N, 7°23'01.8"W**						
E24.1DC	*D. subtentaculata*		MK712611	MK712682 + MK713200 + MK713074	MK712887	MK713134	A, R, Le	04/06/2013
Los Barrios, Spain		**36°09'10.0"N, 5°34'43.0"W**						
ETO1	*D. subtentaculata*		MK712621	MK712684 + MK713202 + MK713076	MK712890	MK713111	M, So	2013
Albarracín, Spain		**36°09'10.0"N, 5°34'43.0"W**						
V03.1D	*D. subtentaculata*		MK712571	MK712703 + MK713221 + MK713095	-	MK713126	V, So, Le	2013
Peralejos de las Truchas, Spain		**40°35'35.5"N, 1°55'32.2"W**						
V05.1D	*D. subtentaculata*		MK712608	MK712704 + MK713222 + MK713096	MK712921	MK713128	V, So, Le	2013
V05.1DB	*D. subtentaculata*		MK712609	MK712705 + MK713223 + MK713097	MK712922	MK712933	V, So, Le	2013
V05.2D	*D. subtentaculata*		MK712610	MK712706 + MK713224 + MK713098	-	-	V, So, Le	2013

* **Collectors:** (A) M. Álvarez-Presas, (M) E. Mateos, (L) E. Lázaro, (Le) L. Leria, (P) M. Pala, (S) G. A. Stocchino, (So) E. Solà, (R) M. Riutort, (V) M. Vila-Farré

Supporting Information: Figures & Tables

Fantastic beasts and how to delimit them: an integrative approach using multi-species coalescent methods reveals two new, endemic *Dugesia* species (Platyhelminthes, Tricladida) from Corsica and Sardinia

Daniel Dols-Serrate[1,2], Giacinta Angela Stocchino[3], Ronald Sluys[4] & Marta Riutort[1,2]

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[3]Dipartimento di Medicina Veterinaria, Via Vienna 2, University of Sassari, Sassari I-07100, Italy

[4]Naturalis Biodiversity Center, P. O. Boxz 9517, 2300 RA Leiden, The Netherlands

ORCIDs:
0000-0002-2134-7674 Marta Riutort
0000-0003-0820-7324 Daniel Dols-Serrate
0000-0002-7005-208X Giacinta Angela Stocchino
0000-0002-9776-3471 Ronald Sluys

Corresponding author: D. Dols-Serrate danieldols@ub.edu / bcn27101993@gmail.com

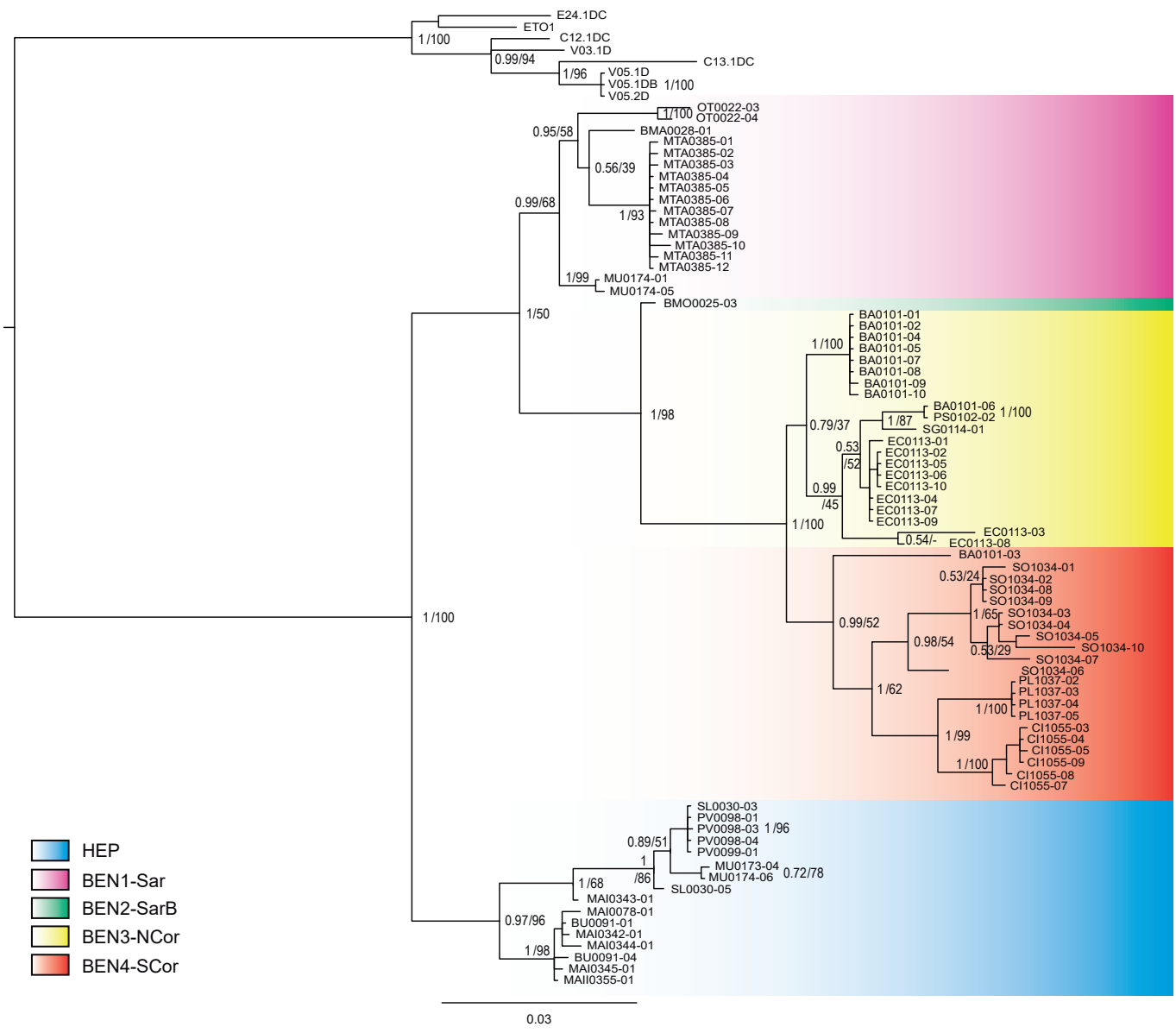
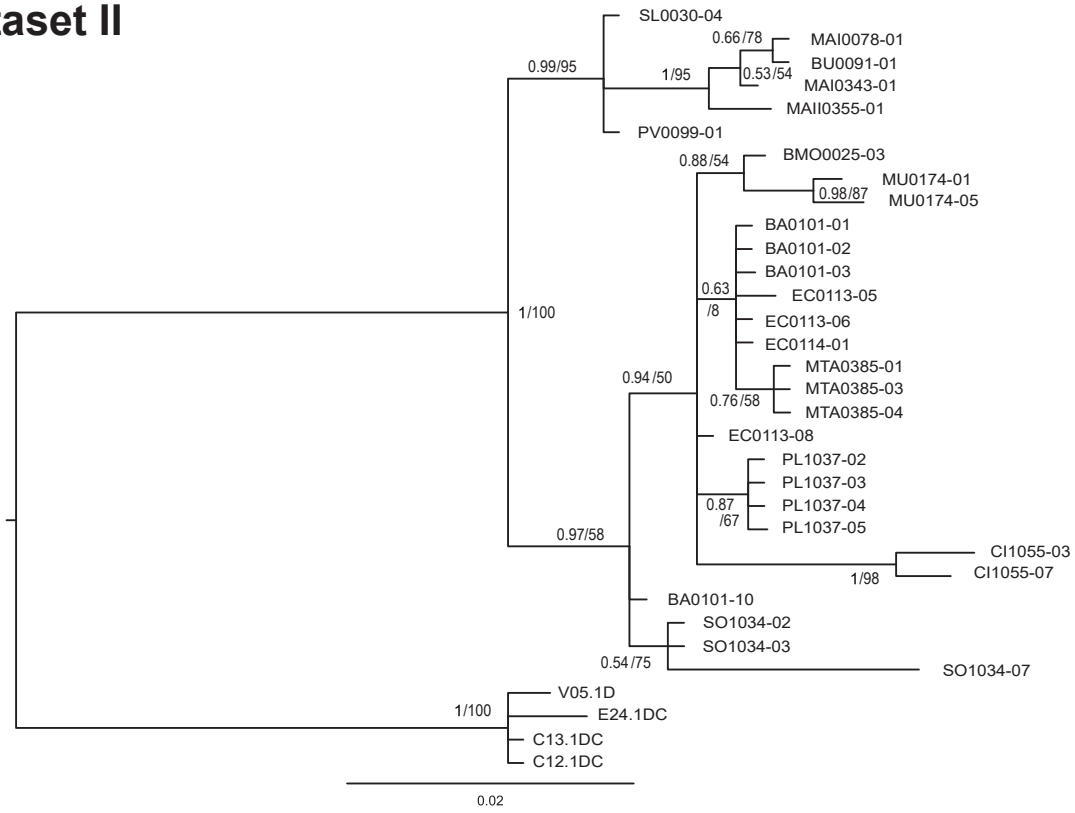
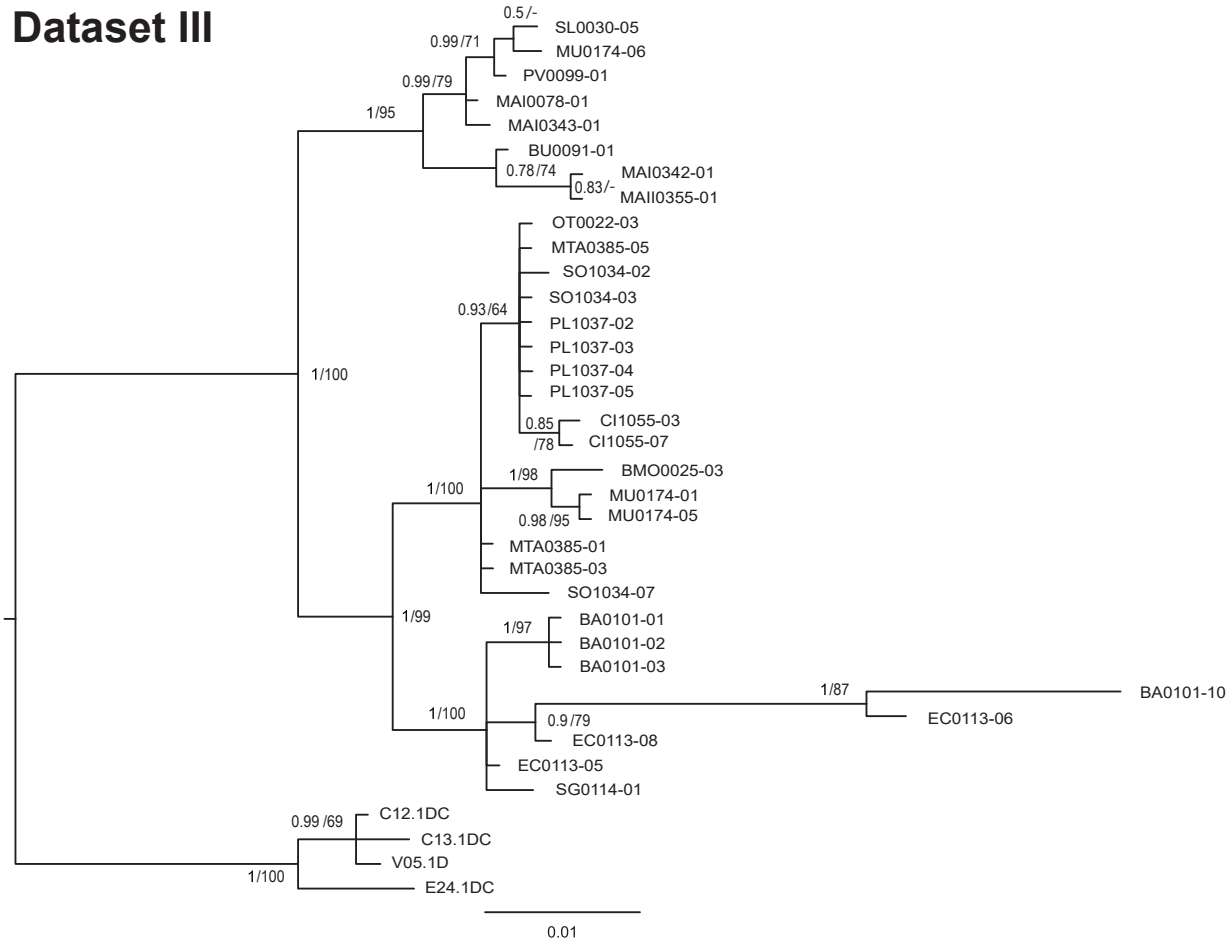


Fig. S1. Bayesian phylogenetic trees for (**A**) dataset I (CoxI-nad1-tRNAW-CoxII); (**B**) dataset II (Dunuc10); (**C**) dataset III (Dunuc12); (**D**) dataset IV (Dunuc12+10); and (**E**) dataset V (mitochondrial and nuclear concatenation). Node support values are indicated as posterior probabilities (0-1) and bootstrap support (0-100). BMO0025-03 phylogenetic position is highlighted with a red arrow in relation to the Corsican clade (highlighted with a red box).

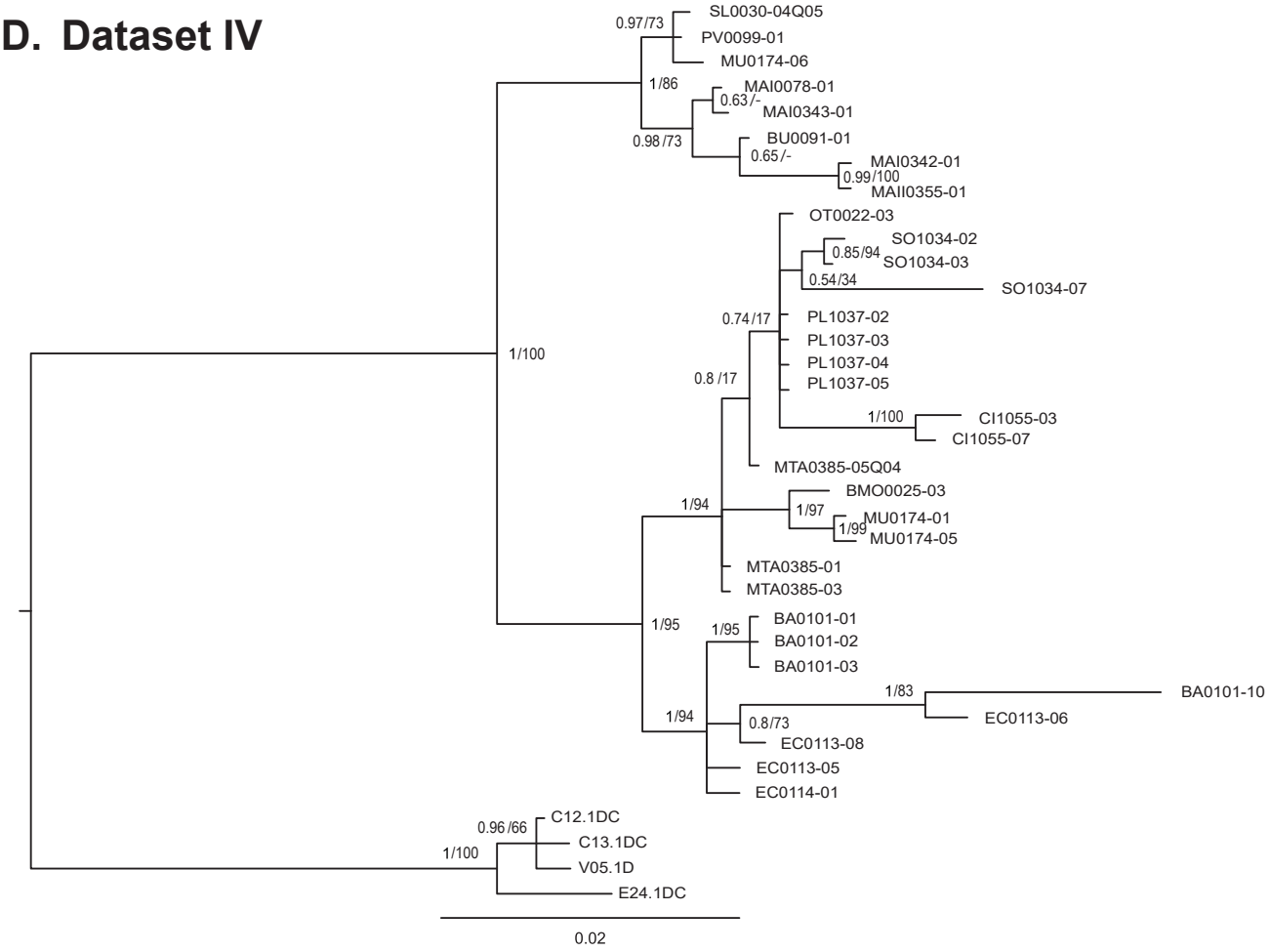
B. Dataset II



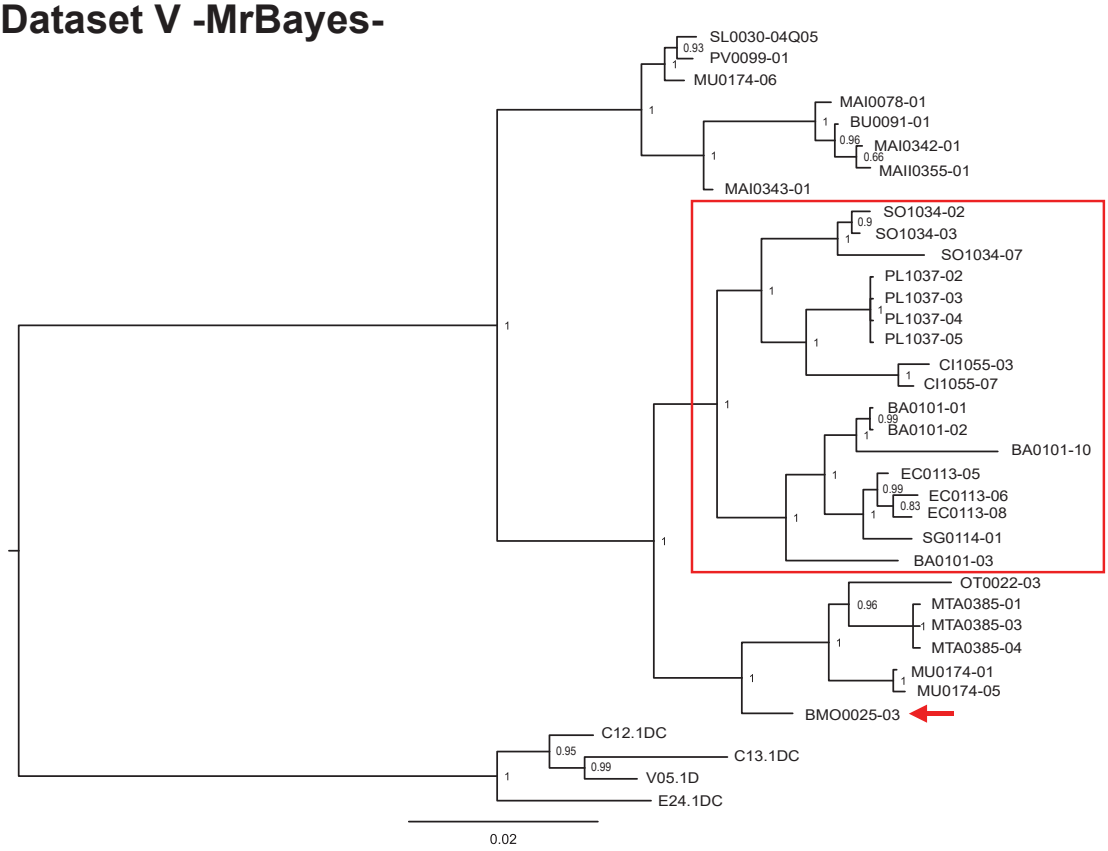
C. Dataset III



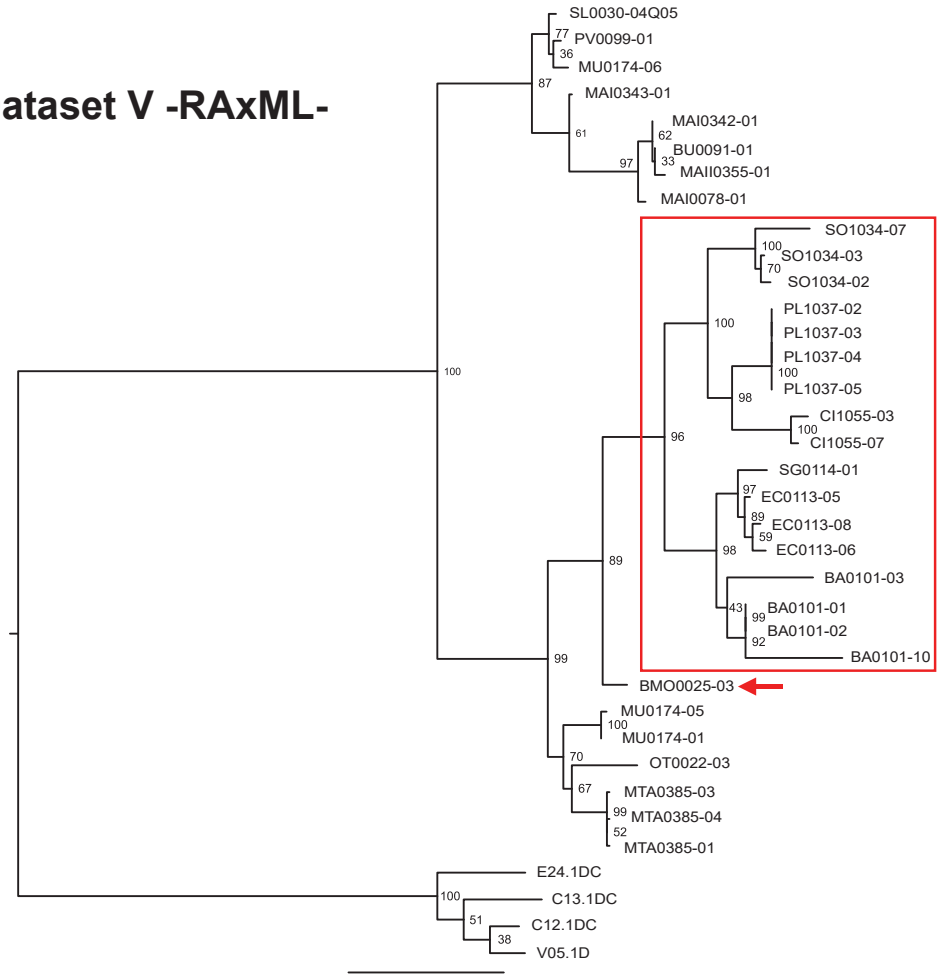
D. Dataset IV


SL0030-04Q05
PV0099-01
MU0174-06
MAI0078-01
MAI0343-01
BU0091-01
MAI0342-01
MAII0355-01
OT0022-03
SO1034-02
SO1034-03
SO1034-07
PL1037-02
PL1037-03
PL1037-04
PL1037-05
CI1055-03
CI1055-07
MTA0385-05Q04
BMO0025-03
MU0174-01
MU0174-05
MTA0385-01
MTA0385-03
BA0101-01
BA0101-02
BA0101-03
BA0101-10
EC0113-06
EC0113-08
EC0113-05
EC0114-01
C12.1DC
C13.1DC
V05.1D
E24.1DC
0.97/73
1/86
0.63/-
0.98/73
0.65/-
0.99/100
0.85/94
0.54/34
0.74/17
1/100
0.8/17
1/94
1/97
1/99
1/95
1/95
1/83
1/94
0.8/73
0.96/66
1/100
0.02

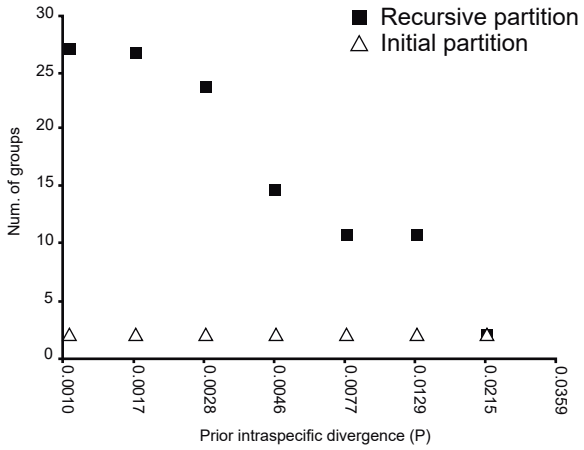
E1. Dataset V -MrBayes-



E2. Dataset V -RAxML-



A.



B.

GMYC test summary	
Method: single	
Clusters	7(6-8)
Entities	8(7-9)
Likelihood null model	443.96
Likelihood GMYC model	452.36
Likelihood ratio	16.79
Likelihood ratio test	2.2×10^{-4}

C.

mPTP test summary	
Null model	263.37
Single coalescent rate	271.41
Multi coalescent rate	275.96
	ASV*
MCMC1	0.8175
MCMC2	0.8176
MCMC3	0.8183
MCMC4	0.8184
ASDDSV	0.0003

* Average support value

D.

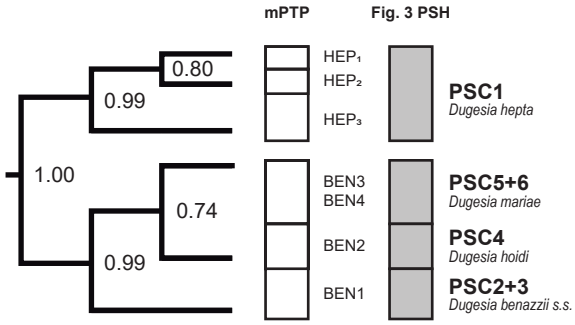
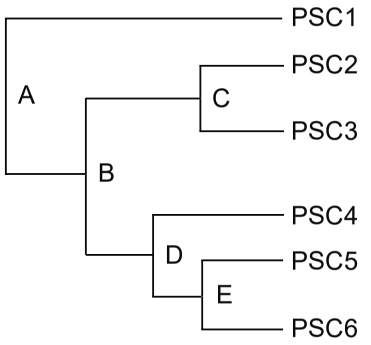
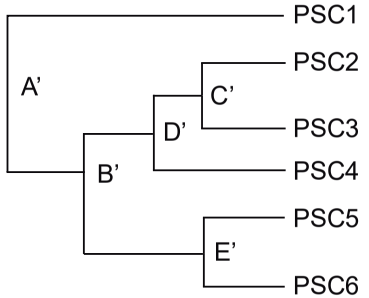


Fig. S2. Summary for ABGD **(A)**, GMYC **(B)** and mPTP **(C, D)** analyses. **A**: ABGD partitions obtained for dataset I. **B**: summary of GMYC test results. **C**: the average support value (ASV; 0-1) is given for each independent run. The closer ASV is to one, the better the support values agree with the ML delimitation -agreement between sampled delimitations and the best scoring delimitation-. ASDDSV or *average standard deviation of delimitation support values* indicate converge of the multiple independent MCMC runs to the same distribution of delimitations. **D**: nodal support for mPTP's partition scheme and comparison with proposed PSH in Fig. 3.

Topology-1



Topology-2



Delimited species	Model 1 θ_1~IG (3, 0.002) τ_1~IG (3, 0.002)				Model 2 θ_2~IG (3, 0.1) τ_2~IG (3, 0.002)				Model 3 θ_3~IG (3, 0.1) τ_3~IG (3, 0.1)				Model 4 θ_4~IG (3, 0.002) τ_4~IG (3, 0.1)			
	A0		A1		A0		A1		A0		A1		A0		A1	
	S = -145	S = -1	S = -145	S = -1	S = -145	S = -1	S = -145	S = -1	S = -145	S = -1	S = -145	S = -1	S = -145	S = -1	S = -145	S = -1
Topology 1: dataset I																
Nodes																
A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
D	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
E	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Topology 1: dataset IV																
Nodes																
A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.87	0.93	0.85	0.84
C	0.99	0.99	0.96	0.94	0.39	0.39	0.40	0.39	0.24	0.24	0.24	0.24	0.58	0.63	0.53	0.58
D	0.70	0.72	0.69	0.65	0.94	0.94	0.95	0.95	0.87	0.85	0.85	0.85	0.03	0.03	0.03	0.03
E	0.51	0.54	0.53	0.50	0.80	0.80	0.81	0.81	0.63	0.63	0.62	0.63	0.01	0.00	0.00	0.01
Topology 1: dataset V																
Nodes																
A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	1.00	1.00	1.00	0.76	0.77	0.78	0.77	0.77	0.76	0.77	0.76	1.00	1.00	1.00	1.00
D	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
E	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Topology 2: dataset I																
Nodes																
A'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
D'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
E'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Topology 2: dataset IV																
Nodes																
A'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C'	1.00	1.00	1.00	1.00	0.18	0.17	0.17	0.18	0.07	0.07	0.07	0.07	0.98	0.98	1.00	1.00
D'	1.00	1.00	1.00	1.00	0.46	0.45	0.45	0.45	0.28	0.28	0.28	0.28	0.99	0.98	1.00	1.00
E'	0.38	0.37	0.38	0.38	0.85	0.85	0.85	0.86	0.77	0.77	0.77	0.77	0.13	0.06	0.11	0.12
Topology 2: dataset V																
Nodes																
A'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C'	1.00	1.00	1.00	1.00	0.51	0.52	0.50	0.50	0.50	0.51	0.48	0.47	1.00	1.00	1.00	1.00
D'	1.00	1.00	1.00	1.00	0.70	0.71	0.70	0.70	0.69	0.70	0.68	0.67	1.00	1.00	1.00	1.00
E'	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Fig. S3. Results of guided Bayesian species delimitation (BPP, model A10) based on datasets I, IV and V. Posterior probabilities are presented for each node (A-E; A'-E') and non-significant results are highlighted in red. A0: algorithm 0; A1: algorithm 1; S: seed.

Table S1. Comparison of the different models and their prior parametrization assessed before performing the ultrametric tree reconstruction in BEAST as well as the GMYC species delimitation. Marginal-likelihood estimates were calculated using the path-sampling (PS) and stepping stone (SS) methods.

Tree prior	Molecular clock	Substitution rate[a]	Substitution model	PS		SS	
				MLE	*2lnBf*	MLE	*2lnBf*
Birth-and-Death	Strict	0.017[b]		-5461.413	202.950	-5461.350	202.530
Birth-and-Death	Uncorrelated relaxed*	0.017[b]		-5359.939	N/A	-5360.086	N/A
Yule	Strict	0.017[b]		-5519.387	318.600	-5520.344	320.520
Yule	Uncorrelated relaxed*	0.017[b]	COI, nad1, COII: HKY+G	-5398.387	76.900	-5400.494	80.820
Birth-and-Death	Strict	Default	tRNA-W: HKY	-5470.096	220.314	-5470.882	221.592
Birth-and-Death	Uncorrelated relaxed*	Default		-5369.754	19.631	-5370.287	20.401
Yule	Strict	Default		-5512.748	305.618	-5514.68	309.224
Yule	Uncorrelated relaxed*	Default		-5390.836	61.795	-5392.231	64.290

[a] Number of substitutions per site divided by length of tree

[b] Only applied to CoxI partition; normal distribution with mean 0.017 and stdev 0.05

* All partitions except for tRNAW which had a strict clock

Table S2. Results for the saturation tests on the first, second and third codon positions of dataset I. Significant results (P < 0.05 highlighted with *) when Iss < Iss.c indicate no significant saturation.

Num. OTU	Iss			Iss.cSym			Iss.cAsym			P		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
4	0.174	0.158	0.345	0.790	0.790	0.790	0.758	0.758	0.758	0.000*	0.000*	0.000*
8	0.155	0.154	0.325	0.744	0.744	0.744	0.633	0.633	0.633	0.000*	0.000*	0.000*
16	0.161	0.151	0.310	0.707	0.707	0.707	0.497	0.497	0.497	0.000*	0.000*	0.000*
32	0.174	0.164	0.315	0.693	0.693	0.693	0.365	0.365	0.365	0.000*	0.000*	0.000*

Table S3. Results for the alternative topology test. *Rank* classification of the best trees according to he likelihood score. Constrained tree (*Rank* 2) is compared to the best scoring unconstrained tree (*Rank* 1). P-values > 0.05 indicate that our data can de supported by the alternative topology hypothesis.

Rank	Item	Obs*	au	kh	sh
1	Topology 1	-17.7	0.949	0.936	0.936
2	Topology 2	17.7	0.051	0.64	0.64

au: approximately unbiased test

kh: Kishino-Hasegawa test

sh: Shimodaira-Hasegawa test

*log-likelihood difference

Table S4. Marginal-likelihood estimates (MLE) and Bayes factors comparisons (*2lnBf*) for BFD. N/A indicates the best model per data set and estimate method (PP/SS).

Dataset	Mitochondrial (dataset I)				Nuclear (dataset IV)				Nuc + Mit (dataset V)			
Model	PS		SS		PS		SS		PS		SS	
	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*
1	**-5224.0109**	**N/A**	**-5224.0109**	**N/A**	-2876.2606	31.8246	-2876.3478	32.0224	**-7245.318**	**N/A**	**-7245.3361**	**N/A**
2	-5225.3638	2.7148	-5225.3638	2.552	-2884.1742	47.6518	-2884.1659	47.6586	-7251.907	13.1778	-7252.024	13.3758
3	-5278.5298	109.0378	-5278.5298	108.8818	-2884.9327	49.1688	-2884.8962	49.1192	-7274.881	59.1258	-7274.9502	59.2282
4	-5278.0929	108.1640	-5278.0929	-108.5058	-2885.5511	50.4056	-2885.5429	50.4126	-7273.779	56.9212	-7273.7648	56.8574
5	-5319.8252	188.9138	-5319.8252	189.1234	**-2860.3483**	**N/A**	**-2860.3366**	**N/A**	-7273.063	55.4906	-7273.4091	55.2810
6	-5366.2052	284.3886	-5366.2052	285.5164	-2889.3483	57.6518	-2889.0625	57.4518	-7307.063	124.3904	-7307.9766	124.1460

Accepted in Zoological Journal of the Linnean Society

New insights into the evolution and biogeography of freshwater planarians on islands in the Tyrrhenian Sea, Western Mediterranean Basin, with the integrative description of a new endemic species from Corsica (Platyhelminthes, Tricladida, *Dugesia*)

Daniel Dols-Serrate[1,2], Giacinta Angela Stocchino[3], Paula Nuin-Villabona[1], Ronald Sluys[4] & Marta Riutort[1,2]

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[3] Dipartimento di Medicina Veterinaria, Via Vienna 2, University of Sassari, Sassari I-07100, Italy

[4] Naturalis Biodiversity Center, P. O. Box 9517, 2300 RA Leiden, The Netherlands

ORCIDs:
0000-0002-2134-7674 Marta Riutort
0000-0003-0820-7324 Daniel Dols-Serrate
0000-0002-7005-208X Giacinta Angela Stocchino
0000-0002-9776-3471 Ronald Sluys

Corresponding author: D. Dols-Serrate danieldols@ub.edu/bcn27101993@gmail.com

Abstract

A recent study on the freshwater planarian fauna of Corsica and Sardinia established that the formerly presumed single species *Dugesia benazzii* actually subsumed a complex of species. In that study, a thorough integrative taxonomic approach, combining molecular, morphological, and karyological data, uncovered the presence of two new endemic species. For the present study, additional samplings were conducted on Corsica, thus covering several new localities. The obtained data was added to our previous datasets, which included information on specimens collected from both islands. By using a similar integrative approach as in the prior study, we here describe another new endemic species from the island of Corsica, *Dugesia xeropotamica*. Although our results support the separate status of this new species, it turned out also that the evolutionary relationships among species of the Corso-Sardinian clade remain unclear, as the relationships are not well resolved. Additionally, we corroborate the presence of *Dugesia mariae* on the Tyrrhenian islets of Montecristo and Molara, representing new records of this group of *Dugesia* species in the region of the Tyrrhenian Sea. In all, our findings indicate that the evolutionary and biogeographic history of *Dugesia* species in this region of the Western Mediterranean is more intricate than previously thought.

ADDITIONAL KEYWORDS: Integrative taxonomy – morphology – phylogeny – species delimitation – BPP – BFD – multispecies coalescent – triclad flatworms

1. Introduction

Through a series of recent studies focusing on various genera of the Dugesiidae Ball, 1974 (Sluys et al., 2013; Solà et al., 2013, 2022; Dols-Serrate et al., 2020, 2024; Leria et al., 2020; Benítez-Álvarez et al., 2023b), it has become increasingly apparent that a significant challenge lies in accurately delimiting the species within this family. Most likely, our current understanding of species diversity underestimates the actual number of species. This is primarily beacause the majority of the approximately 110 known species within the freshwater planarian genus *Dugesia* Girard, 1850 have been described on the basis of anatomical characters, particularly relying on the analysis of the copulatory apparatus (Sluys and Riutort, 2018), due to the lack of reliable external morphological apomorphies (Stocchino et al., 2005; Dols-Serrate et al., 2020). Unfortunately, the existence of permanently asexual populations, the capability of sexual individuals to reabsorb the reproductive apparatus in response to abiotic stressors (Berninger, 1911; Ball and Reynoldson, 1981; Newmark and Alvarado, 2002; Stocchino and Manconi, 2013; Álvarez-Presas and Riutort, 2014), and the presence of cryptic species (Stocchino et al., 2017; Leria et al., 2020; Dols-Serrate et al., 2024) may interfere with the proper identification and recognition of species. This state of affairs may be tackled only with the help of complementary sources of information. For example, inclusion of molecular markers has proven effective in the assignation of asexual individuals to known species (Lázaro et al., 2009), while the use of karyological data has helped tracing species boundaries when morphological and/or genetic results were inconclusive (Pala et al., 1981, 1999; Ribas, 1990; Leria et al., 2018, 2020; Dols-Serrate et al., 2024). Thus, recent studies using molecular-based analyses resulted in the discovery of new species, some of which cryptic, within various genera of the Dugesiidae, including the type genus *Dugesia* (Chen et al., 2015; Leria et al., 2020; Wang et al., 2021a, 2021b; Dols-Serrate et al., 2024).

Having a proper understanding of the extant species and the independently evolving lineages within a region, is one crucial step towards achieving some insight into the evolutionary biology of a group of species. In a previous study (Dols-Serrate et al., 2024), we recounted and further unravelled the complex and confusing situation that the presumed single species *Dugesia benazzii* Lepori, 1951 (hereafter referred to as *D. benazzii* "sensu lato" (s.l.)) was thought to be distributed across the Mediterranean islands of Corsica and Sardinia. This notion of a single species distributed across both islands was first challenged when it turned out that a new biotype of *D. benazzii* s.l. with a haploid complement of 7 chromosomes (2n = 14) actually represented a new species, viz., *Dugesia hepta* Pala, Casu & Vacca, 1981, which occurs in four Sardinian hydrographic basins (Fig. 1) (Stocchino et al. 2005). Thereafter, another new species from Sardinia was described, based on morphological and karyological evidence, viz., *Dugesia leporii* Pala, Stocchino, Corso & Casu, 2000 (Pala et al., 2000). Furthermore, the presence of *Dugesia sicula* Lepori, 1948 was reported also from several Sardinian localities (Pala et al., 1995).

More recently, an integrative taxonomic study of several populations of *D. benazzii* s.l.

from both Corsica and Sardinia led to the description of two new species (Dols-Serrate et al., 2024), viz., *Dugesia mariae* Stocchino, Dols-Serrate & Riutort, 2024 from Corsica, and *Dugesia hoidi* Dols-Serrate, Stocchino & Riutort, 2024, from Sardinia and only known from their type locality. Additionally, a re-description was given for *D. benazzii* "sensu stricto" (s.s., hereafter simply referred to as *D. benazzii*), which from then on would only comprise the Sardinian populations of former *D. benazzii* s.l., excluding animals from *D. hoidi*'s type locality (Fig. 1).

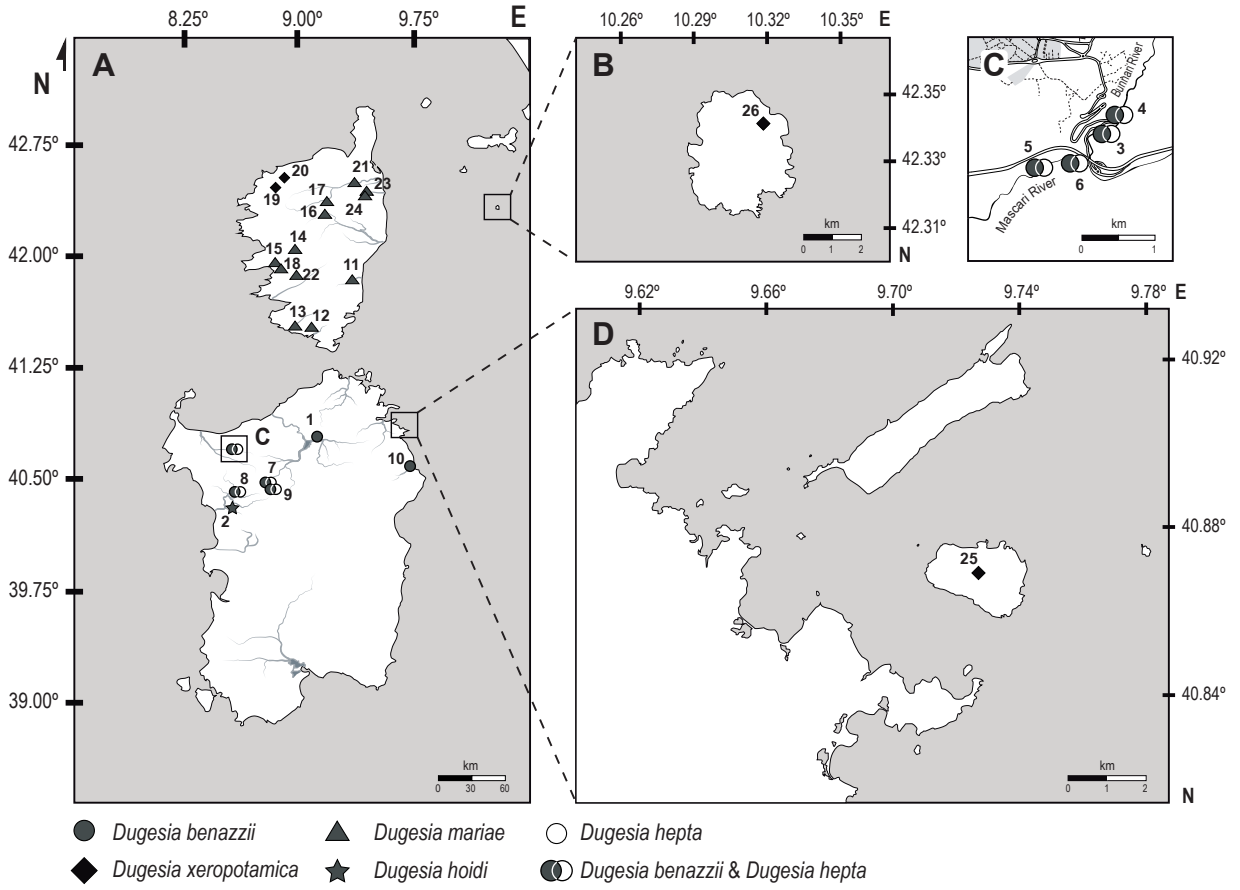


Figure 1. (A) Sampling sites and geographic distribution of several species of *Dugesia* on Corsica and Sardinia, including records detailed in the present paper; rectangular inset marked with C on Sardinia corresponds with panel C. Rectangular insets show: (B) Montecristo Island, in the Tuscany Archipelago; (C) confluence of the Bunnari and Mascari rivers, near Sassari, Sardinia; (D) islet of Molara, Sardinia.

However, in our previous study (Dols-Serrate et al. 2024) we were unable to resolve some issues concerning the taxonomic status of the species in both islands. For example, Lepori's (1951) original paper did not only refrain from specifying the type locality for nominal *D. benazzii* s.l. but even failed to give precise coordinates for the various sampling localities and only documented their geographic names, which in some cases proved to be erroneous (Dols-Serrate et al., 2024). Furthermore, at some point, the original slides with histological sections were lost, which thus complicated comparisons between the results obtained by Lepori (1951) and Dols-Serrate et al. (2024). Moreover, some of the localities mentioned by Lepori (1951) for *D. benazzii* s.l. were later found to house mixed

populations of *D. hepta* and *D. benazzii* (Pala et al., 1981, 1982). In addition, the recent description of *D. mariae*, although not based on the same localities that Lepori examined, raised questions regarding the taxonomic status of the original material from Corsica. Had Lepori (1951), unknowingly, perhaps used mixed samples of *D. benazzii*, *D. hepta*, and *D. mariae* in his description of nominal *D. benazzii* s.l. or even other undescribed species?

In an attempt to tackle such questions, we conducted another sampling campaign in Corsica, with the particular aim of resampling the original localities detailed in Lepori's (1951) paper. The newly collected specimens were combined in a new comprehensive species delimitation analysis encompassing all known species previously falling under the *D. benazzii* s.l. group. At the same time, this new study offered us the opportunity to address the status of unidentified *Dugesia* specimens from the islets of Molara (Protected Marine Area since 1997, WDPA code 13174) and Montecristo (Tuscan Archipelago National Park since 1988, WDPA code 32674). A recently published phylogeographic study of the entire genus *Dugesia* (Solà et al., 2022) included an unidentified individual from the island of Montecristo, which was found to be the sister species of *D. benazzii*, while together these two species shared a sister-group relationship with *D. hepta*. This suggests that again a species of *Dugesia* belonging to the Corso-Sardinian clade occurs on a different island in the Tuscan Archipelago, besides Capraia (Benazzi and Benazzii-Lentati, 1976; de Vries, 1985). In the present study we used the dataset of Dols-Serrate et al. (2024), supplemented with newly obtained molecular sequences resulting from specimens collected at the new sampling sites. As a result, our phylogenetic analysis revealed a new and independent clade among the Corsican populations. Subsequent integrative species delimitation corroborated the discovery of a new species, which is herein described. Last, but not least, our detailed phylogenetic analysis provided a first glimpse of the evolutionary and historical biogeographic history of *Dugesia* species inhabiting this group of islands in the Tyrrhenian Sea.

2. Materials and methods

2.1. Collection of specimens

For the present study, specimens of *D. hepta*, *D. benazzii*, *D. hoidi*, *D. mariae*, and *Dugesia xeropotamica* Stocchino, Dols-Serrate & Riutort, sp. nov. from 13 Corsican and 8 Sardinian localities were used (Fig. 1, Tables 1 & 2, Supplementary Information, Table S1). In an attempt to provide a better account of the distribution of species of *Dugesia* in Corsica, and to check the presence or absence of *D. mariae* in areas of the island that were not covered by previous studies (Dols-Serrate et al., 2020, 2024), we retrieved data for five Corsican localities from Dols-Serrate et al. (2024), and, additionally, conducted fieldwork in eight Corsican localities that were never visited before (Table 1). In the comparison of our results with Lepori's (1951) Corsican material, we used 112 individuals, 48 of which were newly collected for this study. Additionally, two specimens from Montecristo and two from Molara were included in the present work. Specimens from each locality were

fixed and preserved in absolute ethanol for subsequent molecular analyses. For morphological analyses, the posterior portions of the sexual specimens were fixed for 24 h in Bouin's fluid for subsequent histological study while their head portion and the tip of the tails were removed and preserved for molecular analyses immediately after collection.

Table 1. Final species identification, geographical locations and their acronyms, population number, geographic coordinates, and number of individuals (*n*) examined concerning each newly studied population from localities on Corsica, Molara, and Montecristo.

Species	Locality	Code	Population number	Latitude	Longitude	*n*
D. mariae	U Crucolli	UC	14	42°03'31.9"N	8°57'54.5"E	10
D. mariae	Aiaccio	AI	15	41°58'25.0"N	8°49'53.4"E	2
D. mariae	Corte	CT	16	42°18'18.3"N	9°09'55.9"E	5
D. mariae	C. Campita	CC	17	42°23'34.2"N	9°10'34.8"E	4
D. mariae	Setti Polli	SP	18	41°55'43.9"N	8°52'15.6"E	5
D. xeropotamica	Calvi	CA	19	42°29'10.8"N	8°48'09.2"E	4
D. xeropotamica	Montegrosso	MG	20	42°32'17.4"N	8°50'38.5"E	3
D. mariae	Barchetta	BA	21	42°30'24.2"N	9°22'27.3"E	5
D. mariae	Turghja	TG	22	41°51'09.1"N	8°58'00.1"E	6
D. mariae	Molara	IM	25	40°52'04.2"N	9°43'36.6"E	2
D. mariae	Montecristo	IMC	26	42°20'26.0"N	10°19'05.8"E	2

2.2. High quality DNA extraction, quantification, and sequence amplification

Total genomic DNA extraction was carried out by performing an overnight digestion step at 37ºC with proteinase K (10 µl/sample; 20 mM) to process the individuals' tissues, followed by isolation and purification of DNA using a Wizard® Genomics DNA Purification Kit (Promega Corporation), following the manufacturer's instructions. Quality and quantity of total DNA was examined using a spectrophotometer NanoDrop™ 1000 (Thermo Fisher Scientific Inc.).

Specific primers (Supplementary Information, Table S2) were used to amplify the following gene fragments: (a) *Dugesia* nuclear gene 10 (*Dunuc10*, anonymous marker) in a fragment of approximately 450 base pairs (bp); (b) *Dugesia* nuclear gene 12 (*Dunuc12*, a transmembrane emp24 domain-containing protein) in a fragment of approx. 670 bp; (c) a fragment of approx. 650 bp of the cytochrome c oxidase I (*COI*); (d) an approx. 630 bp sequence containing a fragment of the NADH subunit I (*NADH1*), the tryptophan tRNA, and a fragment of the cytochrome c oxidase II (*COII*). Primer sequences and annealing temperature specifics are given in Supplementary Information, Table S2. Final PCR reaction volume per sample was 25 µl, consisting of: (1) 5 µl of Promega 5x Green GoTaq Flexi Buffer, (2) 2.5 µl of MgCl2 (25 mM), (3) 1.5 µl of dNTPs (0.5 mM), (4) 1 µl of each primer (10 µM), (5) 0.15 µl of Taq polymerase (5 u/µL) (GoTaq® Flexi DNA Polymerase, Promega) and (6) 1 µl of DNA extraction sample. Autoclaved milli Q water was added to obtain the final PCR volume. The resulting PCR products were run in a 1% agarose gel to check the correct amplification for the different molecular markers.

Table 2. Population numerical code, sampling locality, and pertinent acronym used in the present paper.

1	Badde Manna	**BMA**	**14**	U Crucolli	**UC**
2	Badu de Mola	**BMO**	**15**	Aiaccio	**AI**
3	Bunnari-2	**BUII**	**16**	Corte	**CT**
4	Bunnari-3	**BUIII**	**17**	Camping Campita	**CC**
5	Mascari-1	**MAI**	**18**	Setti Polli	**SP**
6	Mascari-2	**MAII**	**19**	Calvi	**CA**
7	Ponte Valenti	**PV**	**20**	Montegrosso	**MG***
8	Mulinu	**MU**	**21**	Barchetta	**BA**
9	Santa Lucia	**SL**	**22**	Turghja	**TG**
10	Monte Albo	**MTA**	**23**	Ecole Communale	**EC**
11	La Solenzara	**SO**	**24**	San Gavino	**SG**
12	Cioccia	**CI**	**25**	Molara	**IM**
13	Pero Longo	**PL**	**26**	Montecristo	**IMC**

* Populations used for morphological analysis

Viable amplification products were purified before sequencing using a vacuum system (MultiScreen™HTS Vacuum Manifold of Millipore). Sequencing reactions were performed using Big Dye (3.1, Applied Biosystems) and ran in an automated sequencer ABI Prism 3730 (Unitat de Genòmica dels Serveis Científico-Tècnics de la Universitat de Barcelona) or at Macrogen Inc. (Madrid). Both chains were sequenced using the same primers used for amplification. Complementary strands of DNA were assembled into contigs, and established consensus sequences using Geneious R10 (Biomatters, http://www.geneious.com/ last accessed Janaury 2023).

2.3. Phylogenetic analyses: Sequence alignments and datasets

Sequence alignments for all gene fragments were performed using the online version of MAFFT v.7 (Katoh and Standley, 2013) using default parameters, and were subsequently checked and revised in BioEdit v.7.2.5 (Hall, 1999). Protein coding genes were translated to proteins in order to check the reading frame (codon table 9 was used for mitochondrial genes and codon table 1 for nuclear genes; https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi, last accessed January 2023). Nucleotide saturation was tested for the first, second, and third positions of the protein-coding mitochondrial genes using the Xia et al. (2003) test implemented in DAMBE v. 7.0.28 (Xia and Xie, 2001), as well as for the intronic regions of both nuclear markers as a whole. Three datasets were assembled for downstream analyses: (a) a concatenation of mitochondrial markers (dataset I); (b) a concatenation of nuclear markers (dataset II), and (c) a concatenation of the aforementioned datasets (dataset III).

2.4. Phylogenetic analyses: Gene tree inference

Gene tree inference was carried out using methods of Bayesian inference (BI), as implemented in MrBayes v 3.2.7 (Ronquist et al., 2012) and Maximum Likelihood (ML), as implemented in IQ-Tree 2.0 (Minh et al., 2020; Nguyen et al., 2015) and RAXML-NG v 0.80 (Kozlov et al., 2019). For the BI and the RAXML-NG trees, best substitution models were inferred using MEGA v.11 (Tamura et al., 2021) and selected based on the lowest BIC (Bayesian Information Criterion) score (Table 3).

Table 3. Datasets used in the present study, and best-fitting substitution model for partition scheme and marker.

Dataset	Marker	Partition scheme	MrBayes	RAXML-NG	IQ-Tree
Dataset I	COI-MIT	COI (1-640 bp)	HKY+G	HKY+FO+G4m+B	HKY+FI+G4
		NADH1 (641-697 bp)	HKY+G	HKY+FO+G4m+B	HKY+F+I
		tRNAW (698-759 bp)	HKY	HKY+FO+B	TPM2+F+I
		COII (760-1265 bp)	HKY+G+I	HKY+FO+I+G4m+B	TPM3+F+I+G4
Dataset II	Dunuc12+10	Dunuc12 exon (1-102 bp)	GTR	GTR+FC+G4m+B	HKY+F
		Dunuc12 intron (103-664 bp)	GTR	GTR+FC+G4m+B	TPM3+F+R2
		Dunuc10 intron (665-811 bp)	GTR	GTR+FC+G4m+B	F81+F+R2
		Dunuc10 exon (812-1094 bp)	GTR	GTR+FC+G4m+B	K2P+R2
Dataset III	Dataset I+II	-	-		-

In the eventuality that the chosen best model had no implementation in neither MrBayes nor RAXML-NG, the next best implementable model was used. BI was carried out by simultaneously performing two parallel runs with four Markov chains –one cold, three heated- starting from random trees and running for 50,000,000 generations, with a sampling frequency of one sample every 1000 generations and a burn-in of 25%. Parallel runs were considered to converge when the standard deviation value of the split frequencies was equal or less than 0.01. ML inference in RAXML-NG was conducted starting from a random tree and using 1000 replicates to compute bootstrap supports (bs). Concurrently, best substitution models were inferred too, by using ModelFinder (Kalyaanamoorthy et al., 2017) as implemented in IQ-Tree 2.0 software, while nodal supports were computed using an ultrafast bootstrap (UFboot) approximation (Hoang et al., 2018), as well as the SH-aLRT test (Guindon et al., 2010) with 10,000 replicates per approach. Sequences of *Dugesia subtentaculata* (Draparnaud, 1801) were used as outgroup for the phylogenetic analyses. All recovered trees were visualized with FigTree v 1.4.4 (https://github.com/rambaut/figtree/releases; last accessed October 2023).

2.5. Integrative taxonomic methodology: Discovery and validation

Our methodology followed the approach described in Leria et al. (2020) and Dols-Serrate et al. (2024), albeit with some accommodations. The establishment of a primary species hypothesis (PSH) to propose primary species candidates (PSCs) was not conducted based on a molecular discovery method using a single locus dataset. Instead, a phylogenetic tree was constructed using dataset I, which included representatives of all previously handled *Dugesia* species in Dols-Serrate et al. (2024), as well as all of our newly obtained specimens. Thus, new PSCs were proposed based on the emergence of new clades in the tree that could not be assigned to any known *Dugesia* species in the region of the Tyrrhenian Sea.

Next, PSCs were tested by means of complementary sources of evidence, viz., (a) multilocus datasets analysed using BPP and Bayes factor delimitation (BFD), and (b) morphological data. Subsequent results led to the proposal of secondary species hypotheses (SSHs) and further pertinent taxonomical decisions.

2.5.1. Molecular validation

BPP v.4.3, model A10 (Yang and Rannala, 2010; Flouri et al., 2020), was used to perform species delimitation analyses for datasets I, II and III using a fixed guide tree based on dataset III phylogenetic results (Fig. 2A, Supplementary Information Fig. S1C). Outgroup sequences were excluded and each partition for every dataset was treated as a single locus. Hence, dataset I contained four mitochondrial loci (in four partitions, one per gene), dataset II two nuclear loci (in two partitions, one for introns and one for exons), and dataset III six loci (using the same partition scheme). Following Yang's (2015) guidelines, four different evolutionary model scenarios were considered using distinct prior combinations (pc) for the ancestral population size (θ) and the species divergence time (τ) parameters to test the splits' support. Both parameters are modelled after an inverse gamma prior distribution IG(α, β). We used a fixed diffuse α = 3 prior and worked by adjusting β. The following scenarios were tested: (a) relatively small ancestral population size θ_1 ~ IG(3, 0.002) and shallow divergence time τ_1 ~ IG(3, 0.002); (b) relatively large ancestral population size θ_2 ~ IG(3, 0.1) and shallow divergence time τ_2 ~ IG(3, 0.002); (c) relatively large ancestral population size θ_3 ~ IG(3, 0.1) and deep divergence time τ_3 ~ IG(3, 0.1); and (d) relatively small ancestral population size θ_4 ~ IG(3, 0.002) and deep divergence time τ_4 ~ IG(3, 0.1). All analyses were performed twice to assess the robustness of results using the two reversible-jump MCMC algorithms (rjMCMC) available and starting from random seeds, viz., algorithm 0 (with fine-tunning prior ε = 2) and algorithm 1 (with fine-tunning priors α = 2 and m = 1). rjMCMC chains ran for 500.000 generations with a sample frequency of five generations and a burn-in of 10.000 generations.

BFD was performed following the guidelines of Grummer et al. (2014). Marginal likelihood estimates (MLE) were computed for several species delimitation hypotheses and

assessed by means of pairwise Bayes factor comparisons (hereafter referred to as *2lnBf*; Kass and Raftery, 1995). The *2lnBf* were calculated as in Grummer et al. (2014) (*2lnBf* = 2 x [best-fitting model's MLE − model *j*'s MLE]), considering the hypothesis with the highest MLE as the best-fitting model. After calculation of *2lnBf* values for competing models, the strength of support for a given comparison was evaluated on the basis of Kass and Raftery's (1995) recommendations. Thus, *2lnBf* = 0 − 2 indicates non-significant support, *2lnBf* = 2 − 10 indicates positive support, and *2lnBf* > 10 indicates decisive support in distinguishing competing species delimitation hypotheses. Datasets I, II and III were used for this analysis. MLE were computed using the path sampling (PS) (Baele et al., 2012) and stepping-stone (SS) (Xie et al., 2011) methods as implemented in BEAST v.1.8.4 (Drummond et al., 2012) and enabling the *BEAST module. Five species delimitation models were tested (Fig. 3B). Parametrization used for each dataset is specified in Supporting Information, Table S3. In all cases, a *piecewise linear & constant root* population size model and a gamma distribution for the *popMean* prior with *mean* and *standard deviation* values set as default were used. MCMC chains ran for 200 million generations, after which MLE were calculated using both PS and SS via an additional run of four million generations of 100 path-steps.

2.5.2. Morphological validation

For morphological study, sexual specimens that had been fixed for 24 h in Bouin's fluid were dehydrated in an ascending ethanol series, cleared in clove oil, and embedded in synthetic wax. Serial sections were made at intervals of 8 µm and were stained with Mallory-Cason (cf. Winsor and Sluys, 2018). Reconstructions of the copulatory apparatus were obtained by using a compound microscope provided with a camera lucida.

Type material is deposited in the Giacinta A. Stocchino collection (CGAS), at the University of Sassari and at the Naturalis Biodiversity Center, Leiden, The Netherlands (RMNH code).

2.5.3 Abbreviations used in morphological figures

bc: bursal canal; ca; copulatory apparatus; cb: copulatory bursa; cm: circular muscle; d: diaphragm; e: epithelium; ed: ejaculatory duct; epg: extra bulbar penial glands; er: ectal reinforcement; g: gonopore; ga: genital atrium; lm: longitudinal muscle; lod: left oviduct; lvd: left vas deferens; pb: penis bulb; ph: pharynx; pp: penis papilla; rod: right oviduct; sg: shell glands; sv: seminal vesicle.

3. Results

3.1. Molecular datasets

A total of 166 new sequences were obtained for the present work and complemented with a dataset encompassing 215 sequences extracted from Dols-Serrate et al. (2020, 2024) (see the Supporting Information, Table S1). Overall, 116 individuals were included in the present study, representing the following species: *D. benazzii* (15 individuals), *D. hepta* (29 individuals), *D. mariae* (59 individuals, four of them from Molara and Montecristo), *Dugesia xeropotamica* (7 individuals), *D. hoidi* (2 individuals), and *D. subtentaculata* (outgroup, 4 individuals). In total, three datasets were built (Table 3). Dataset I was composed of 112 ingroup individuals, and datasets II and III of 66 ingroup individuals. In almost all cases, no gene, gene fragment or position showed relevant signs of saturation, but *Dunuc10* intron did show low phylogenetic signal (Supporting Information, Table S4).

3.2. Phylogenetic tree inference

Our initial phylogenetic approach using the mitochondrial dataset I displayed all major clades within the *D. benazzii* s.l. species complex as established in Dols-Serrate et al. (2024), and recovered, overall, the same topological relationships found in the latter work (Supplementary Information Fig. S1A). In the tree, *Dugesia hepta* formed the sister taxon of a group including *D. benazzii*, *D. hoidi*, and *D. mariae*. Furthermore, the phylogenetic inference based on the mitochondrial dataset indicated that the specimens of *Dugesia* collected from the islets of Molara and Montecristo (loc. 25, 26), as well as the individuals of *Dugesia* from the new localities 14 to 18 and 22 (Fig. 1, Table 1), all belonged to the clade formed by terminals representing *D. mariae*. Additionally, a new Corsican clade with good overall support (SH-aLRT/UFboot/pp/bs; 83.2/90/1/86) was apparent, consisting of specimens of *Dugesia* from localities 19 and 20 (see Fig. 2, Supplementary Information, Fig. S1A). Although the new Corsican clade was sister to the *hepta-hoidi-mariae-benazzii* clade, it should be noted that the monophyly of the latter clade is unsupported, thus leaving the precise phylogenetic relationship of the new Corsican clade unresolved (Supplementary Information Fig. S1A). Overall, the same topology was recovered for all dataset I phylogenetic trees, regardless of the inference methodology or software used.

Phylogenetic trees recovered for the nuclear dataset II (Supplementary Information Fig. S1B), were consistent with previous reports for *Dunuc10* and *Dunuc12*, exemplifying their lack of phylogenetic signal for closely related species, such as *D. benazzii*, *D. hoidi*, *D. mariae*, and *D. hepta* (Dols-Serrate et al., 2024). The reconstructed phylogeny presents a topology that is distinct from that based on dataset I. Nevertheless, *D. hepta*, *D. hoidi*, and also the new Corsican clade (loc. 19, 20) (see above) were recovered as monophyletic taxa. Although *D. benazzii* and *D. mariae* seem to be polyphyletic, due to the lack of phylogenetic signal (Supplementary Information, Table S4) and the phylogenetic relationships among *D. benazzii*, *D. hoidi*, and *D. mariae* were unsupported and inconclusive (Supplementary

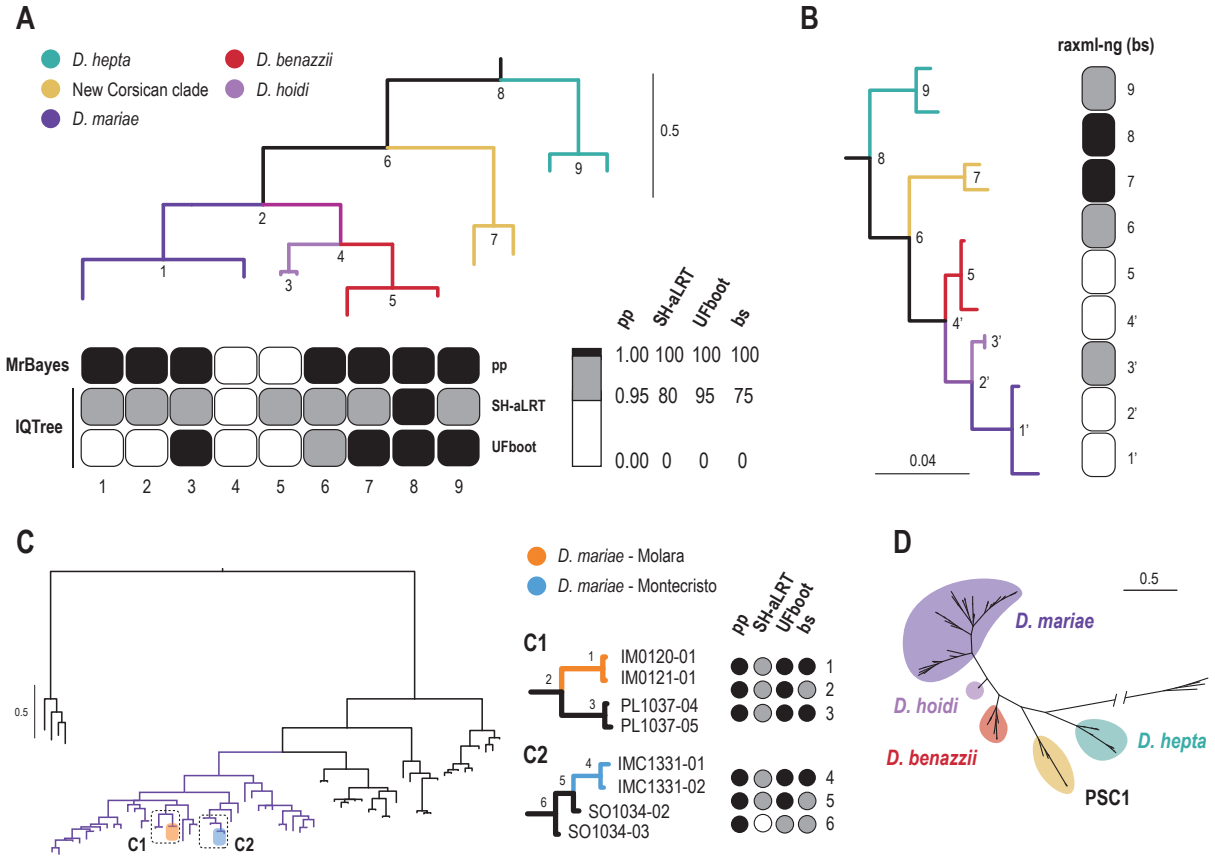


Figure 2. (A-B) Simplified phylogenetic trees based on dataset III (concatenation of mitochondrial and nuclear data), depicting the conflicting phylogenetic position of *D. hoidi*. (C) Detail of Bayesian phylogenetic tree based on dataset III, showing the phylogenetic positions of the unidentified specimens of Molara (C1) and Montecristo (C2). The complete phylogenetic trees inferred from all datasets are available in Supplementay Information, Fig. S1. (D) Unrooted phylogenetic tree based on dataset III, with valid nominal species and candidate species (PSC1) highlighted in various colours. Bootstrap support (bs), posterior probability (pp), ultrafast bootstrap (UFboot), and SH-aLRT test support values are colour-coded to indicate whether a node is unsupported (white), supported (grey), or fully supported (black).

Information Fig. S1B), it should be noted that *D. hepta* was recovered, with high support, as the sister-taxon of the remaining Corso-Sardinian clade, comprising the species *D. benazzii*, *D. hoidi*, *D. mariae*, and the new Corsican clade, similar to the results obtained by Dols-Serrate et al. (2024). Specimens from Montecristo and Molara did not show any signs of independent taxonomic status, as they are all well embedded within the clade of *D. mariae*. In line with the phylogenetic trees based on dataset I, phylogenies generated on the basis of dataset II consistently yielded identical topological results, independent of the inference methodology or software employed.

Phylogenetic trees based on dataset III yielded two distinct topologies dependent on the approach used (Fig. 2A-B, Supplementary Information Fig. S1C-D), thus affecting the relationships of *D. hoidi*. Nevertheless, the following features were shared between both topologies: (a) *D. hepta* was recovered as the sister taxon of the clade comprising *D. xeropotamica*, *D. benazzii*, *D. hoidi*, and *D. mariae* with high support; (b) the new Corsi-

can clade (loc. 19, 20), featuring in the trees based on datasets I and II, was recovered with almost full support (SH-aLRT/UFboot/pp/bs; 99.6/100/1/100) and found to be sister to the clade formed by *D. benazzii*, *D. hoidi*, and *D. mariae* with almost full support (SH-aLRT/UFboot/pp/bs; 96/99/1/97); (c) all newly collected individuals from previously unsampled sites were placed within the clade formed by terminals representing *D. mariae*, excepting specimens from the Corsican localities 19 and 20; (d) specimens of *Dugesia* from the islets of Montecristo and Molara (loc. 25, 26) were unequivocally positioned within the *D. mariae* clade (Fig. 2C); (e) the clade comprising the terminals of Corsican *D. mariae* exhibited higher genetic structure as compared with the Sardinian clades.

Phylogenetic tree reconstructions using MrBayes and IQ-Tree 2.0 recovered *D. hoidi* as sister taxon to *D. benazzii*, whereas results for RAXML-NG placed *D. hoidi* as sister to *D. mariae* (Supplementary Information, Fig. S1C-D). However, in both cases support for the node relating *D. hoidi* either to *D. mariae* or to *D. benazzii* was insufficient, so that its proper phylogenetic position at present remains unresolved. Once again, overall support was higher for all nodes in the reconstructed trees based on MrBayes and IQ-Tree than in the tree inferred by the RAXML-NG approach.

3.3. Primary species hypothesis

Based on the results from the phylogenetic analyses of dataset I (Supplementary Information, Fig. S1A), we identified one new primary species candidate, denoted as PSC1, represented by specimens from localities 19 and 20 on Corsica. This finding was further supported in the phylogenetic trees inferred with dataset III (Fig. 2D, Supplementary Information, Fig. S1C-D). Additionally, the analyses of the nuclear genes, despite their low phylogenetic signal, highlighted the plausible separate species status of the same specimens, as suggested by their inferred monophyly (Supplementary Information, Fig. S1B).

3.4. Primary species hypothesis validations

3.4.1. Molecular validation

Given that in the above only one PSC was proposed for validation, and the position of *D. hoidi* in the tree was of little consequence to the validity of the split for PSC1, we used a single fixed guide tree based on the topology of the tree resulting from dataset III under MrBayes/IQ-Tree 2.0 for the BPP analysis (Fig. 2A, Supplementary Information Fig. S1C). However, it should be noted that a topology test performed on mitochondrial data in Dols-Serrate et al. (2024) showed that the phylogenetic position of *D. hoidi* cannot be resolved with the available datasets, as used in the present study, and that, thus, the alternative topology (Fig. 2B, Supplementary Information Fig. S1D) could have been used as well. The BPP analyses clearly suggested that the differentiation of PSC1 might be the result of speciation processes rather than coalescent ones, as indicated by the full support

of PSC1's node split (Fig. 3A, node B, Supplementary Information, Table S5). The use of different prior combinations (pc 1-4) and rjMCMC algorithms had no effect on the support of node B, whose pp remained as high as 1.00 for all datasets analysed.

With regard to the BFD analyses, results based on *2lnBf* comparisons for datasets I and III clearly suggested model 1 (i.e., independent status for all species as well as PSC1) as the best-fitting model, based on both PS and SS methods (Fig. 3B, Supplementary Information, Table S6). Although the results based on dataset II supported both model 1 and model 2 (i.e., *D. hoidi* and *D. benazzii* forming a single clade) as plausible, both models indicated the separate status of PSC1.

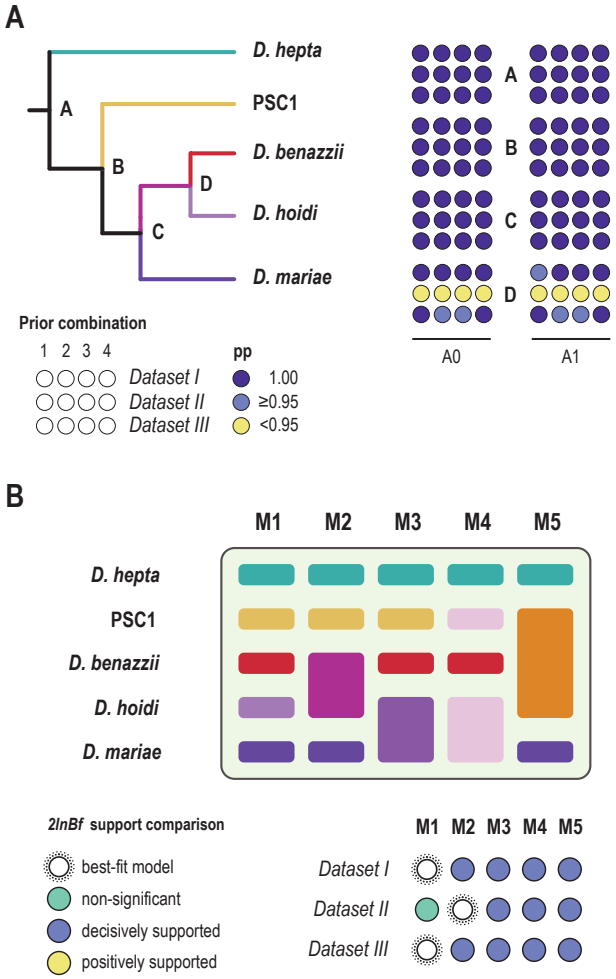


Figure 3. (A) Schematic representation summarizing the guided tree used for BPP analysis. Results based on multi-locus data using two algorithms (A0, A1), three datasets (I, II, III), and four evolutionary models (prior combinations) are shown. (B) BFD results for the array of models tested (M1–5), with different a priori species-delimitation hypotheses.

3.4.2. Morphological validation

A morphological study was performed on two sexual individuals (the only ones found at collection) from Seccu River, Corsica (Montegrosso, loc. 20 in Fig. 1, Table 1,

Supplementary Information, Table S1), while their heads and the tip of the tails were removed for DNA extraction. With respect to the anatomy of the copulatory apparatus, these individuals exhibited a unique combination of morphological features (see below), constituting clear differences with all the other known species of the genus *Dugesia*, thus validating our PSH.

3.5. Secondary species hypothesis

All molecular analyses undoubtedly suggested the separate status of PSC1. Additionally, morphological data corroborated the molecular results. Therefore, based on all presently available evidence, we here consider PSC1 to constitute a Confirmed Candidate Species (CCS), according to the nomenclature of Vieites et al. (2009).

3.6. Taxonomic account

Order Tricladida Lang, 1884
Suborder Continenticola Carranza, Littlewood, Clough, Ruiz-Trillo, Baguñà & Riutort, 1998
Family Dugesiidae Ball, 1974
Genus *Dugesia* Girard, 1850
***Dugesia xeropotamica* Stocchino, Dols-Serrate & Riutort, sp. nov.**

Material examined:
Individuals CA1291-08, CA1291-09, CA1291-11, CA1291-13 from loc. 19, and MG1292-08, MG1292-09, MG1292-13 from loc. 20 were analysed molecularly; corresponding GenBank accession numbers are detailed in Supplementary Information, Table S1. Holotype: RMNH.VER.21534.1, Seccu River, locality Montegrosso, (42°32'17.4"N, 8°50'38.5"E), Corsica, 29 November 2019, coll. D. Dols-Serrate, L. Leria, one set of sagittal sections on 9 slides. Holotype: RMNH.VER.21534.1, Seccu River, locality Montegrosso, (42°32'17.4"N, 8°50'38.5"E), Corsica, 29 November 2019, coll. D. Dols-Serrate, L. Leria, one set of sagittal sections on 9 slides.

Paratype: CGAS Pla 29.1, ibid., sagittal sections on 7 slides.

Diagnosis:
Molecularly, *Dugesia xeropotamica* comprises individuals that are identified as a single evolutionary unit alongside individuals from localities 19 and 20 of the present work when using a multispecies coalescent-based method such as BPP with the loci and settings detailed in the Materials and Methods section of the present paper, or equivalent molecular species delimitation analyses.

Morphologically, *D. xeropotamica* is characterised by the presence of the following fea-

tures: short, more or less barrel-shaped and asymmetrical penis papilla; dorsal lip of the penis papilla bigger than the ventral one; ventrally displaced ejaculatory duct, opening terminally at tip of the penis papilla; short ejaculatory duct; large, conical diaphragm; symmetrical openings of the vasa deferentia into the posterior section of the seminal vesicle, i.e., close to the diaphragm; oviducts opening symmetrically into the vaginal portion of the bursal canal; ectal reinforcement musculature on bursal canal extending from vaginal region almost up to the copulatory bursa; numerous extra-bulbar penial glands.

Etymology:

The specific epithet is derived from the Greek words '*ξερός*' (xeros), meaning 'arid', and '*ποταμός*' (potamos), meaning 'river'. The epithet alludes to the name of the type locality Seccu River, whose Corsican name translates to 'Arid River'. This river runs dry during the summer season and its course disappears as it approaches the sea.

Geographical distribution:

This species is only known from the type locality, Seccu River (Montegrosso, loc. 20 in Fig. 1), Haute-Corse, Corsica, and from a small stream running parallel to the Figarella River, known as Lioli Rivulet (Calvi, loc. 19 in Fig. 1), Haute-Corse, Corsica.

Reproduction:

Sexual.

Description:

Two eyes are present in the centre of the low triangular head. Dorsal surface brown with two longitudinal dark brown stripes on either side of the body (Fig. 4); ventral surface pale. Unfortunately, it was not possible to take photos of live animals in the field, while the head was removed for DNA extraction.

Inner and outer pharyngeal musculature bilayered, i.e., without an extra, third, outer longitudinal muscle layer in the inner sheath of muscles. The mouth opening is situated at the hind end of the pharyngeal pocket.

The infranucleated oviducts run ventrally in a caudal direction to the level of the genital pore, whereafter they open separately and symmetrically into the vaginal portion of the bursal canal.

The numerous testes are situated dorsally on either side of the body and extend to almost the posterior end of the body. Vitellaria are located between the testes and the intestinal branches.

The large, more or less cuboidal copulatory bursa occupies the entire dorso-ventral space; it is lined by a columnar, vacuolated, and glandular epithelium with basal nuclei and is surrounded by a layer of muscles. From the postero-dorsal wall of the bursa, the bursal

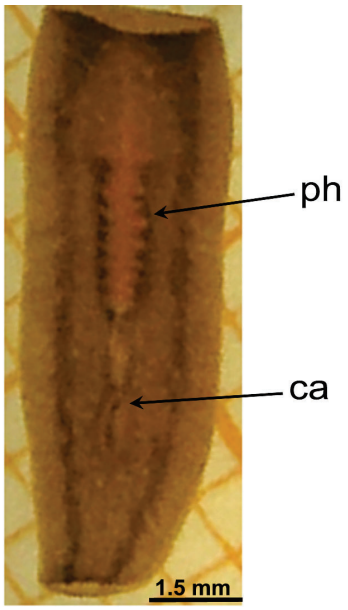


Figure 4. *Dugesia xeropotamica*. Photomicrograph of a preserved sexual specimen (anterior to the top) from Seccu River (Montegrosso, loc. 20, Corsica).

canal runs in a caudal direction to the left side of the copulatory apparatus and then, just at the level of the gonopore, exhibits a knee-shaped bend, whereafter it opens through the roof of the genital atrium. The bursal canal is lined with cylindrical, infranucleated, ciliated cells and is surrounded by a thin subepithelial layer of longitudinal muscles, followed by a layer of circular muscle. Ectal reinforcement musculature is present in the form of a well-developed layer of longitudinal muscles running ectally to the circular layer around the bursal canal, extending from the opening of the bursal canal into the atrium to the copulatory bursa (Figs 5A, 6A). Shell glands discharge their erythrophil secretion into the vaginal region of the bursal canal, just at the level of the oviducal openings (Fig. 5).

The genital atrium is lined by an infranucleated, columnar epithelium, which is underlain by a subepithelial layer of circular muscle, followed by a layer of longitudinal muscle fibres. The genital atrium opens to the exterior via the ventral gonopore. No cement glands were discerned in either of the specimens examined (Fig. 5).

Comparative discussion:

In addition to its genetic differentiation from the rest of species analysed in the present work, *Dugesia xeropotamica* differs from its congeners in a unique combination of morphological characters concerning the copulatory apparatus. Notably, this concerns the following characters: short and asymmetrical penis papilla with its dorsal lip being bigger than the ventral one; a ventrally displaced ejaculatory duct with its terminal opening; a large, conical diaphragm; symmetrical openings of the vasa deferentia into the seminal vesicle close to the diaphragm; presence of ectal reinforcement along the entire length of the bursal canal; symmetrical openings of the oviducts into the vaginal portion of the bursal canal; numerous extra-bulbar penial glands.

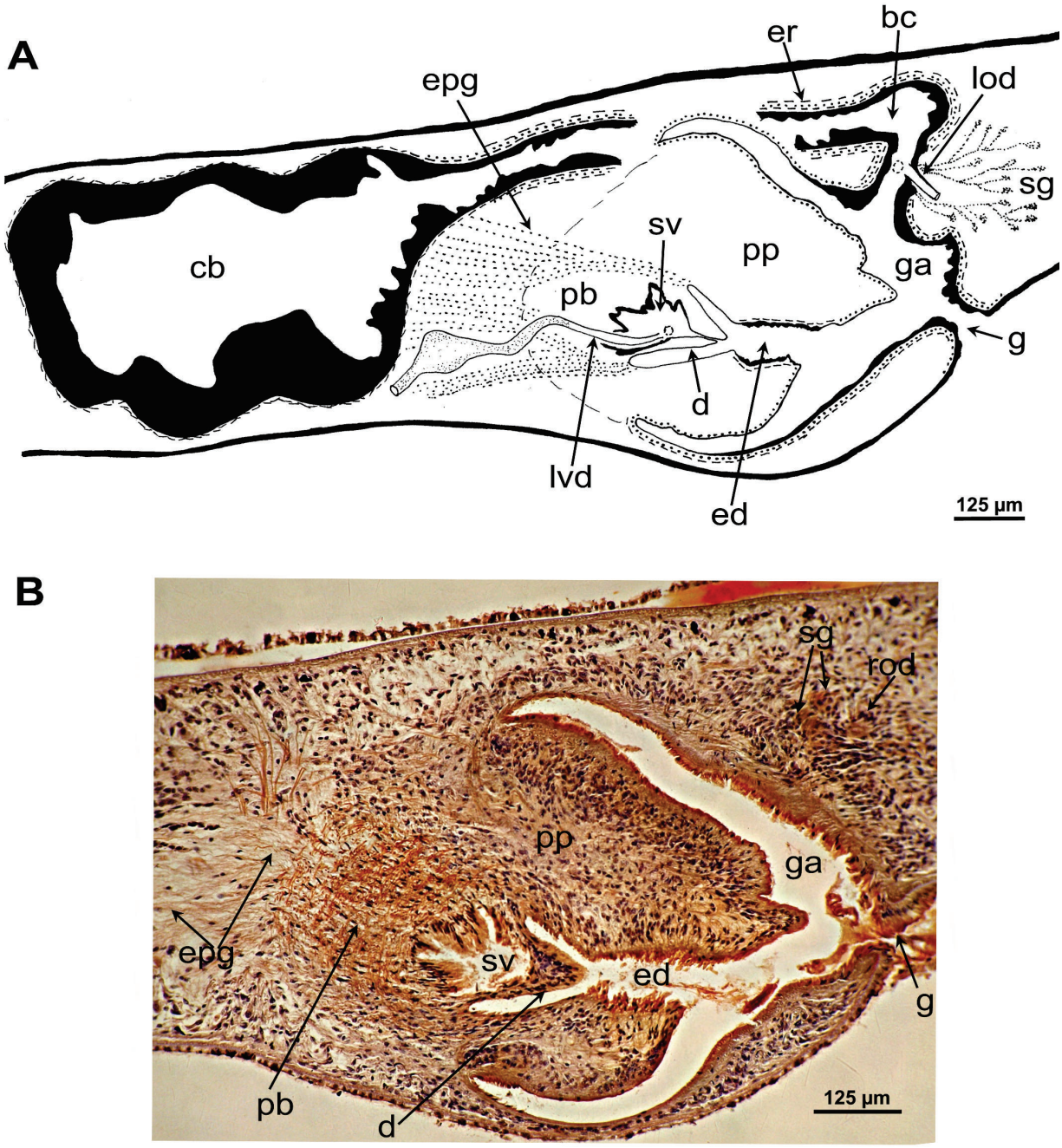


Figure 5. *Dugesia xeropotamica*. Holotype RMNH.VER.21534.1 (A) Sagittal reconstruction of the copulatory apparatus (anterior to the left); (B) photomicrograph showing, in a sagittal histological section, the penis bulb (pb) with seminal vesicle (sv); penis papilla (pp) with the pointed diaphragm (d); ejaculatory duct (ed); and extrabulbar glands (epg).

From the island of Corsica thus far only two species of the genus *Dugesia* have been reported, viz., *D. benazzii* and *D. mariae*, although the presence of *D. benazzii* must be evaluated as doubtful, according to a recent study (cf. Dols-Serrate et al., 2024). *Dugesia xeropotamica* shares with *D. benazzii* and *D. mariae* only the presence of a large, conical diaphragm, whereas it differs from these two species in the absence of a penial fold and the presence of numerous extra-bulbar glands. *Dugesia xeropotamica* shares with *D. mari-*

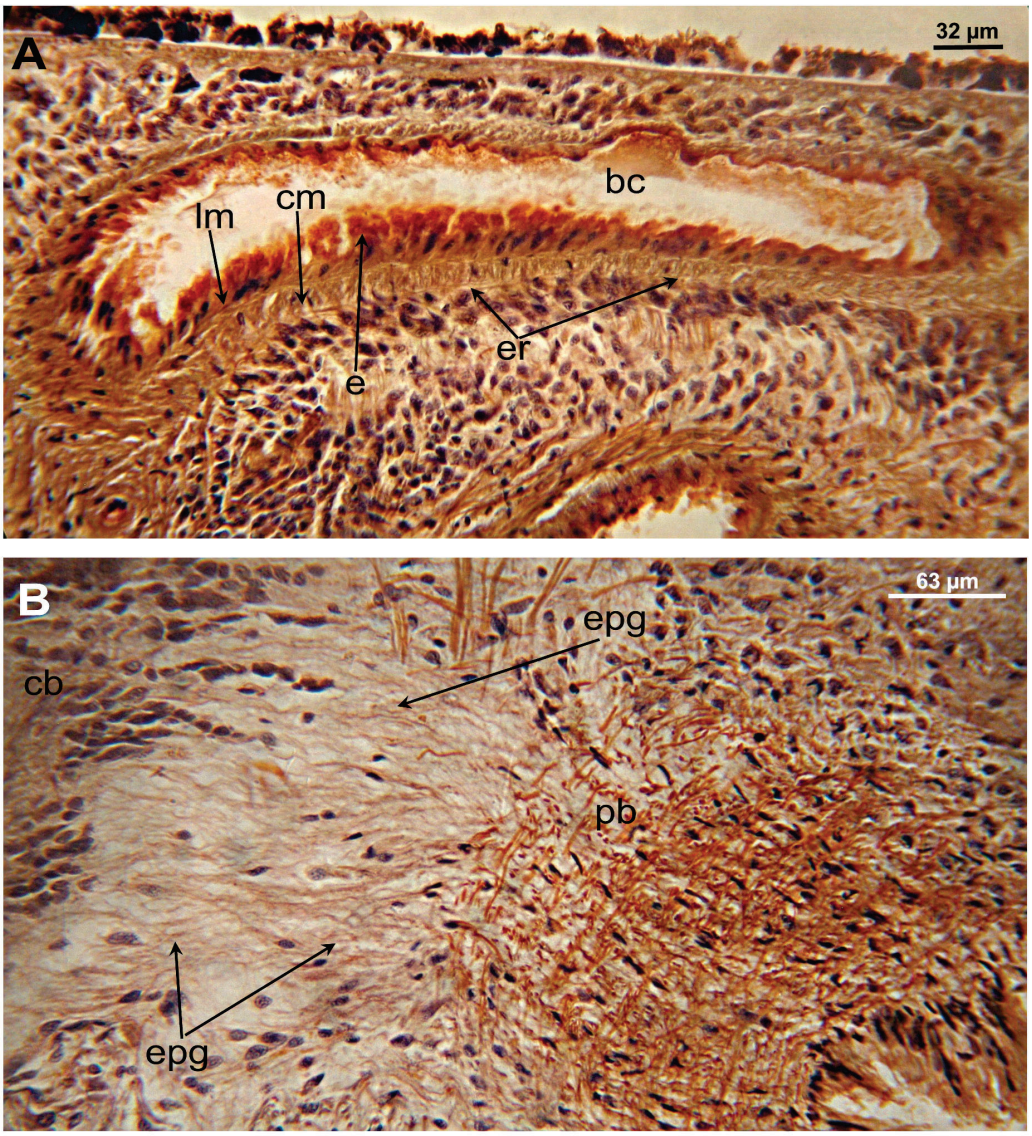


Figure 6. *Dugesia xeropotamica*. Photomicrographs of sagittal histological sections. (A) Paratype GAS Pla 29.1, showing the bursal canal with the ectal reinforcement (er); (B) Holotype RMNH.VER.21534.1, showing the extra-bulbar penial glands (epg), and the penis bulb (pb).

ae the ventral course of the ejaculatory duct and the extension of the ectal reinforcement up to the copulatory bursa (Dols-Serrate et al., 2024).

The stubby, more or less barrel-shaped penis papilla of *D. xeropotamica* resembles that of *D. leporii* from Sardinia (Pala et al., 2000). However, it differs from this last-mentioned species in the absence of a penial valve, the presence of a short ejaculatory duct, and presence of extrabulbar penial glands.

In its peculiar shape of the penis papilla specimen RMNH.VER.21534.1 of *D. xeropotamica*, with its dorsal lip being longer than the ventral one, resembles *D. japonica* Ichikawa & Kawakatsu, 1964 from Matsumoto City (Honshû, Japan) and *D. ryukyuensis* Kawakatsu,

1976 (Ichikawa & Kawakatsu, 1964; Kawakatsu et al., 1976). However, *D. xeropotamica* has a large, pointed diaphragm, in contrast to the small diaphragm present in the two above-mentioned Oriental species. Moreover, *D. japonica* and *D. ryukyuensis* have their openings of vasa deferentia located at the anterior section of the seminal vesicle, whereas in *D. xeropotamica* the sperm ducts open close to the diaphragm.

Extra-bulbar penial glands are not rare in *Dugesia*, as these have been reported for a number of species, such as: *D. annandalei* (Kaburaki, 1918), *D. andamanensis* (Kaburaki, 1925), *D. mirabilis* de Vries, 1988, *D. debeauchampi* de Vries, 1988, *D. myopa* de Vries, 1988, *D. neumanni* (Neppi, 1904), *D. milloti* de Beauchamp, 1952, *D. monomyoda* Marcus, 1953 *D. bifida* Stocchino & Sluys, 2014, *D. burmaensis* (Kaburaki, 1925), and *D. sigmoides* Stocchino & Sluys, 2017 (Kaburaki, 1918, 1925; Marcus, 1953; de Vries, 1988; Stocchino et al., 2014, 2017). Although extra-bulbar glands are not explicitly reported in the literature on several other species, some of the drawings in the related descriptions, however, do suggest the presence of such glands. This is the case, for example, in accounts on *D. deharvengi* Kawakatsu & Mitchell, 1989 (Kawakatsu & Mitchell, 1989: fig. 3A), *D. indica* Kawakatsu, 1969 (Kawakatsu, 1969: fig. 2), *D. lanzai* Banchetti & Del Papa, 1971 (Sluys, 2007: fig 9), *D. leclerci* Kawakatsu & Mitchell, 1995 (Kawakatsu & Mitchell, 1995: fig. 10A), and *D. uenorum* Kawakatsu & Mitchell, 1995 (Kawakatsu & Mitchell, 1995: fig. 6). Absence of cement glands in *D. xeropotamica* appears to be peculiar among species of *Dugesia*, and even in triclads in general, since it is the secretion of these glands that allows cocoons to be fastened to the substratum (Hyman, 1951; Corso et al., 2006; Stocchino et al., 2014).

4. General discussion

4.1. *Dugesia* in the islets of Montecristo and Molara

Just as our phylogenetic analysis clearly suggested separate species status for *D. xeropotamica*, it also decisively indicated that the specimens from our new Corsican localities (14, 15, 16, 17, 18, and 22; Fig. 1, Table 1, Supplementary Information, Table S1), as well as the *Dugesia* individuals from Montecristo and Molara, must be considered as members of the species *D. mariae*.

Benazzi (1969) considered the animals from Molara to belong to *D. benazzii*, characterized by a chromosomal complement of 2n = 16; n = 8. However, a morphological study by Pala et al. (1980) established affiliation with *D. etrusca monoadenodactyla* Lepori, 1947, a subspecies previously reported only from Tuscany. Furthermore, a detailed study of the karyotype by these authors revealed a chromosome portrait that was completely identical to that of *D. etrusca labronica* from Tuscany (which is interfertile with *D. etrusca monoadenodactyla*), with respect to chromosomal length and centromere position. The authors concluded that presence of *D. etrusca monoadenodactyla* on Molara was a testament to ancient connections between Tuscany and Sardinia, probably through the Tuscan Archi-

pelago and Corsica. Our present results complement those reported by Pala et al. (1980), and introduce a new record of a *Dugesia* species on Molara, indicating that species from two phylogenetically different groups, viz., the Corso-Sardinian clade (i.e., *D. mariae*) and the Iberia-Apennines-Alps clade (i.e., *D. etrusca monoadenodactyla*), can be found on Molara (Leria et al., 2022; see Benítez-Álvarez et al., 2023b for the definition of the clades). Regarding Montecristo, a molecular study based on the presence of a highly repetitive DNA sequence and random amplified polymorphic DNA (RAPD) markers revealed that individuals of an asexual population from that island were not related to specimens of *D. etrusca* Benazzi, 1946 collected from several localities in Tuscany. In addition, comparison of the animals from Montecristo with specimens of *D. benazzii* s.l. from a southern Corsican population (Roccapina) enfeebled any similarity to the latter species (Batistoni et al., 2001). Given Montecristo's geographic proximity to Tuscany, one would expect to find species that are more closely related to the Iberia-Apennines-Alps clade, than a relative of the Corso-Sardinian clade. Nevertheless, our present results on the individuals from Montecristo point to affiliation with the species *D. mariae*, a member of the Corso-Sardinian clade.

In this context, it is important to note that the island of Montecristo is an isolated landmass, covering approximately 10 km^2, located near the southern limit of the shallow Tuscan platform (with water depths less than 200 m). It is positioned about 40 km south of Elba Island and more than 60 km from the Italian and Corsican coasts (Innocenti et al., 1997; Giusti et al., 2023). Although geological literature on the island is scarce, it is well established that Montecristo is primarily composed of a Late Miocene monzogranitic pluton that formed about 7.3 million years ago (Mya), during the late Tortonian, just before the Messinian period (Ferrara and Tonarini, 1985; Innocenti et al., 1997). Therefore, during the Messinian salinity crisis (5.9-5.3 Mya) Corsica and Montecristo, alongside mainland Tuscany, were most likely interconnected, while possibly the same conditions were present during the Pleistocene glaciations. Similarly, given the close proximity of the islet of Molara to Sardinia, only 2.3 km from its coast, it is not unreasonable to assume that during these two geological periods Molara was connected to the Sardo-Corsican complex.

Present eco-geographical factors are expected to shape the distribution patterns of high vagility species (Dapporto et al., 2007; Dapporto and Cini, 2007; Dapporto and Strumia, 2008), whereas distribution patterns of low vagility species, such as freshwater planarians, are more influenced by historical and, possibly, anthropogenic factors (Fattorini, 2009 ; Barbato et al., 2018). With respect to the historical biogeographic evolution of species of *Dugesia* in the Western Mediterranean region, a recent study suggested that the age of the ancestral node of the Corso-Sardinian clade, and thus the onset of the group's diversification, could be dated between 5.7 and 2.2 Mya, while the ancestral node of the evolutionary lineages involving *D. etrusca* was estimated to date back to about 1.9 Mya (Solà et al., 2022). Consequently, this temporal framework implies that any contact between *D. etrusca* and *D. mariae* would have occurred during the Pleistocene. Various studies have discussed the glacio-eustatic regressions of the sea level during the glacial periods in that

epoch, in some cases reaching -200 m (Guilcher, 1969), although many estimates suggest that the sea level would not have dropped below 110-150 m (Lambeck et al., 2002, 2004; Jouet et al., 2006; Zecchin et al., 2015; Dumitru et al., 2021). This implies that throughout most of the Pleistocene epoch Montecristo was isolated from the Sardo-Corsican complex, mainland Tuscany, as well as the remaining Tuscan Archipelago, while Molara was in contact solely with Corsica and Sardinia (cf. Fattorini, 2010). Some studies based on fossils of Quaternary fauna suggest that during the Early Würm Corsica, and, consequently Sardinia and Molara, could have been connected to mainland Tuscany via an isthmus that subsided (Conchon, 1976).

However, other possibilities should be taken into consideration also in the explanation of present co-occurrence of *D. mariae* and *D. etrusca* on Molara, and *D. mariae*'s presence on Montecristo. Notably, anthropochorous dispersal may have caused present-day distributional patterns. The oldest known human remains in Sardinia date back to ~20,000 years ago (Melis, 2002), and it has been argued that most of the Corsican and Sardinian mammalian fauna resulted from human introductions from the mainland during the Neolithic age (Vigne, 1992). Furthermore, Molara has been inhabited since ancient times and has been subjected to anthropochorous faunal and floral introductions (Bocchieri and Iiriti, 2005). In this way, specimens of *D. mariae* or *D. etrusca monoadenodactyla* could have been introduced, inadvertently, via water transported for human consumption or for agricultural use. Regarding Montecristo, the island has been intermittently colonized by humans and may have hosted a Neolithic settlement (Radmilli, 1976), and, thus, has experienced numerous faunal introductions (Randi et al., 1990; Masseti, 2016; Torrini et al., 2016; Barbato et al., 2018; Giusti et al., 2021).

Considering the timeframe proposed by Solà et al. (2022), *D. mariae* could have reached the islet of Molara through the Sardo-Corsican complex during a period of sea level regression, and go extinct in the rest of Sardinia (perhaps being displaced by *D. benazzii* or *D. sicula*). However, in light of the information discussed above, the presence of *D. etrusca monoadenodactyla* on Molara is best explained when contemplating anthropochorous dispersal, instead of interpreting it to be a relictual distribution of a once much broader ancestral geographic range. On the other hand, the unlikely connection of Montecristo Island with nearby islands of the Tuscan Archipelago, the Sardo-Corsican complex, or any continental masses during the Pleistocene, suggests a likely human-driven arrival of *D. mariae* on Montecristo. This assumption relies on the idea that the diversification of the Corso-Sardinian clade occurred after the end of the Messinian salinity crisis. However, unlike *D. etrusca*'s presence on Molara, *D. mariae*'s presence in Montecristo could be seen as evidence of a broader ancestral distribution in the Western Mediterranean basin if, conversely, we acknowledge an earlier onset of the Corso-Sardinian clade's diversification (~5.7 Mya or earlier).

Unfortunately, earlier studies of *Dugesia* on Molara and Montecristo (Benazzi, 1969; Pala et al., 1980; Batistoni et al., 2001) were performed at different moments in time, and

often utilized specimens that were collected during independent sampling campaigns that were conducted years apart, and probably involved different locations. Much to our regret, these studies provided no information on the exact localities that were visited on the islands, which makes it practically impossible to ascertain whether the same populations have been studied, either by us or by previous workers. Therefore, it is only through dedicated new sampling campaigns, conducted within an integrative taxonomic framework, that we will be able to accurately study the freshwater planarian diversity and their biogeographical evolution in the Tuscan Archipelago as well as on the numerous Sardinian islets.

4.2. High genetic structuring within Corsica

The identification of a new species in the present study offers a new perspective on the evolutionary history of the genus *Dugesia* in the Western Mediterranean region. This new species, *Dugesia xeropotamica*, turned out to be closely related to four other species that evolved on Sardina and/or Corsica, viz., *D. hepta*, *D. benazzii*, *D. hoidi*, and *D. mariae*. The elusive *D. leporii*, a Sardinian endemic whose conservation status is uncertain (Stocchino, pers. info.), remains enigmatic, as all efforts to find new specimens have been unsuccessful thus far. Nevertheless, all things considered, it is to be expected that if molecular data becomes available in the future it will support taxonomic assignment of *D. leporii* to the above-mentioned group of species.

Unfortunately, precise relationships between *D. benazzii*, *D. hoidi*, *D. mariae*, and *D. xeropotamica* remain uncertain despite the fact that all the species clades were decisively supported. For example, concatenated dataset III (Fig. 2A-B; Supplementary material Fig. S2C-D) did not yield a single monophyletic group per island. Instead, *D. xeropotamica* emerged as the sister taxon of *D. benazzii*, *D. hoidi*, and *D. mariae*. Furthermore, the position of *D. hoidi* still remains debatable, as it did not unequivocally group with either *D. benazzii* or *D. mariae*, as reported in a previous study (Dols-Serrate et al., 2024). This lack of resolution of interspecific relationships may be due to the limited resolving power of some of the markers used in the present study. This is reflected in the results of the saturation test of the intronic regions of *Dunuc10* and *Dunuc12* (see Supplementary Information, Table S4), showing their lack of sufficient phylogenetic signal to properly resolve the evolutionary relationships among closely related species within the Corso-Sardinian clade. However, this is consistent with the findings of our previous study (Dols-Serrate et al., 2024) and, thus, only stresses the need to obtain genomic data if we wish to resolve the evolutionary relationships within this clade.

Our phylogenetic results based on dataset III suggested genetic structure within the *D. mariae* clade, so that, perhaps, at least three distinct sub-clades may be recognized (Supplementary Information, Fig. S1C-D). Although these sub-clades were not strongly supported, they are largely geographically restricted, albeit there are some exceptions to this regional pattern, as some of the specimens from Corte and San Gavino (loc. 16 and

24 each) fall into two different clades. However, this may be due to the limited resolving power of the nuclear markers used, as mentioned earlier. On the other hand, it may be erroneous to assume that any given sampling site on Corsica would harbour only a single species, as at least in Sardinia this assumption does not always hold true (see Stocchino, 2018). Furthermore, although the same markers that we used were applied successfully in a study delimiting species of *Dugesia* in the Iberian Peninsula and Balearic Islands (Leria et al., 2020) and these markers have resolved the taxonomic diversity subsumed under former *D. benazzzii* s.l., here they seemed unable to clearly delimit groups within *D. mariae*. Nevertheless, the genetic structure that appeared to be present within *D. mariae* may reflect the existence of additional, still unidentified, species, or signal initial stages of speciation.

5. Conclusions

Although our current dataset was not able to fully resolve all taxonomic issues, nevertheless, it enabled us to develop a new perspective on the evolution of species of *Dugesia* on the islands of Corsica and Sardinia and nearby islets. These new insights will enable us to delve deeper into the mechanisms driving the evolutionary and biogeographical diversity in this group of flatworms. Although some doubt remains on the phylogenetic placement of some of the species, some preliminary conclusions are already possible. Our present results and insights signal a necessary re-evaluation of the putative hybridisation event between *D. hepta* and *D. benazzii* (Pala et al., 1982; Dols-Serrate et al., 2020), which should be performed by using a more comprehensive genetic representation from all species, alongside an upgrade of the dataset to a genomic level. This would allow us to investigate whether hybridisation or other factors, such as incomplete lineage sorting, have influenced the evolution of this group of *Dugesia* species in the Western Mediterranean Basin.

Furthermore, we document here for the first time the presence of a member of the Corso-Sardinian clade, viz., *D. mariae*, on isolated Montecristo Island and on the Sardinian islet of Molara. Although our results did not yet point to a well-supported biogeographical hypothesis explaining the presence of *D. mariae* in both islets, or for the occurrence of members of the Iberia-Apennines-Alps clade, viz., *D. etrusca*, on Molara, we here developed for the current spatial distribution of *D. mariae* a preliminary explanation that takes into account both historical biogeographic factors and anthropochorous dispersal.

Conflicts of interes

None declared.

Funding

This research was supported by the Ministerio de Economía y Competitividad (project PGC2018-093924-B-100) from Spain. This research was supported also by the Ajuts de Personal Investigador predoctoral en Formació (APIF) grant from the Universitat de Barcelona (to Daniel Dols-Serrate). Giacinta A. Stocchino was supported by a grant from SYNTHESYS (European Union-funded Integrated Activities, grant: NL-TAF 8228), and from Fondazione di Sardegna (grant FdS/RAS-2016/CUP J86C1800082005) and Regione Autonoma Sardegna (grant RAS2012-LR7/2007-CRP-60215).

Acknowledgements

We are very grateful to Gema Blasco for her exceptional dedication to maintaining with impeccable care our planarian samples in the laboratory. We thank Dr. Laia Leria for her invaluable support and guidance during the fieldwork, and Dr. Roger Vila for kindly providing planarian specimens from Montecristo Island. We express our sincere gratitude to Professor Maria Pala for providing the planarian specimens from the islet of Molara and for devoting her whole professional life to the study of flatworms and science.

Data availability

The data underlying this article are available in the GenBank Nucleotide Database and can be accessed with the accession numbers provided in Supplementary Information, Table S1.

References

Aguilar JP. 2011, *Genética Evolutiva de Dugesia benazzii y Dugesia hepta de las islas de Córcega y Cerdeña*. Master thesis, Universitat de Barcelona. Research Gate: https://doi.org/10.13140/RG.2.2. 23010.27849

Álvarez-Presas M, Riutort M. Planarian (Platyhelminthes, Tricladida) diversity and molecular markers: a new view of an old group. *Diversity* 2014;6:323–338. https://doi.org/10.3390/d6020323

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* 2012;29:2157–2167. https://doi.org/10.1093/molbev/mss084

Ball IR. A contribution to the phylogeny and biogeography of freshwater triclads (Platyhelminthes: Turbellaria). In: Riser NW, Morse MP (eds), *Biology of the Turbellaria*. New

York: McGraw–Hill, 1974;339–401

Ball IR, Reynoldson TB, 1981. *British Planarians*. Cambridge University Press, Cambridge.

Barbato D, Benocci A, Manganelli G. The biogeography of non-marine molluscs in the Tuscan Archipelago reveals combined effects of current eco-geographical drivers and paleogeography. *Organisms Diversity and Evolution* 2018;18:443–457. https://doi.org/10.1007/s13127-018-0378-2

Batistoni R, Rossi L, Cori F, Deri P. A molecular characterization of species and populations of *Dugesia gonocephala* s.l. *Belgian Journal of Zoology* 2001;131:97–100.

Benazzi M. Nuovi reperti planariologici in isole tirreniche. *Archivio Botanico e Biogeografico Italiano* 1969;45:153–159.

Benazzi M, Benazzi-Lentati G. Platyhelminthes. In: John B (Ed.), *Animal Cytogenetics*. Berlin-Suttgart: Gebründer Borntraeger, 1976;1–182.

Benítez-Álvarez L, Leria L, Fernández R, Mateos E, El Ouanighi Y, Bennas N, El Alami M, Yacoubi-Khebiza M, Ougougdal HA, Riutort M. Phylotranscriptomics interrogation uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean. *Molecular Phylogenetics and Evolution* 2023b;178:107649. https://doi.org/10.1016/j.ympev.2022.107649

Benítez-Álvarez L, Sluys R, Leal-Zanchet AM, Leria L, Riutort R. First molecular phylogeny of the freshwater planarian genus *Girardia* (Platyhelminthes: Tricladida) unveils hidden taxonomic diversity and initiates resolution of its historical biogeography. *Zoological Journal of the Linnean Society* 2023a;197:1–19. https://doi.org/10.1093/zoolinnean/zlac065

Berninger J. Über die Einwirkung des Hungers auf Planarien. *Zoologische Jahrbücher, Abteilung für allgemeine Zoologie und Physiologie der Tiere* 1911;30:181–216.

Bocchieri E, Iiriti G. Flora and vegetation landscape of Molara Island (North-Eastern Sardinia). *Lagascalia* 2005;25:15–89.

Chen Y, Chen X, Wu C, Wang A. A new species of the genus *Dugesia* (Tricladida: Dugesiidae) from China. *Zoological Systematics* 2015;40:237–249. https://doi.org/10.11865/zs.20150301

Conchon O. The human settlements of Corsica: Paleogeographic and tectonic considerations. *Journal of Human Evolution* 1976;5:241–248. https://doi.org/10.1016/0047-2484(76)90028-2

Corso G, Manconi R, Stocchino GA. A histochemical study of the reproductive structures in the flatworm *Dugesia leporii* (Platyhelminthes, Tricladida). *Invertebrate Biology* 2006;125:91–105. https://doi.org/10.1111/j.1744-7410.2006.00044.x

Dapporto L, Cini A. Faunal patterns in Tuscan archipelago butterflies: The dominant influence is recent geography not paleogeography. *European Journal of Entomology* 2007;104:497–503. https://doi.org/10.14411/eje.2007.070

Dapporto L, Wolf H, Strumia F. Recent geography determines the distribution of some flying Hymenoptera in the Tuscan Archipelago. *Journal of Zoology* 2007;272:37–44. https://doi.org/10.1111/j.1469-7998.2006.00233.x

Dapporto L, Strumia F. A rapid assessment of Lepidoptera biogeography in the Tuscan Archipelago using economical Malaise traps. *Italian Journal of Zoology* 2008;75:219–224. https://doi.org/10.1080/11250000701867735

Dols-Serrate D, Leria L, Aguilar JP, Stocchino GA, Riutort M. *Dugesia hepta* and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex? *Organisms Diversity and Evolution* 2020;20:369–386. https://doi.org/10.1007/s13127-020-00438-z

Dols-Serrate D, Stocchino GA, Sluys R, Riutort R. Fantastic beasts and how to delimit them: an integrative approach using multispecies coalescent methods reveals two new, endemic *Dugesia* species (Platyhelminthes: Tricladida) from Corsica and Sardinia. *Zoological Journal of the Linnean Society* 2024; in press. https://doi.org/10.1093/zoolinnean/zlad135

Draparnaud JPR. Tableau des Mollusques terrestres et fluviatiles de la France. Montpellier & Paris: Renaud. 1801.

Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUTi and the BEAST 1.7. *Molecular Biology and Evolution* 2012;29:1969–1973. https://doi.org/10.1093/molbev/mss075

Dumitru OA, Austermann J, Polyak VJ, Fornós JJ, Asmerom Y, Ginés J, Onac BP. Sea-level stands from the Western Mediterranean over the past 6.5 million years. *Scientific Reports* 2021;11:1–10. https://doi.org/10.1038/s41598-020-80025-6

Ferrara G, Tonarini S. Radiometric geochronology in Tuscany: results and problems. *Rendiconti della Società Italiana di Mineralogia e Petrologia* 1985;40:111–124.

Fattorini S. Faunal patterns in tenebrionids (Coleoptera: Tenebrionidae) on the Tuscan Islands: The dominance of paleogeography over recent geography. *European Journal of*

Entomology 2009;106:415–423. https://doi.org/10.14411/eje.2009.053

Fattorini S. Influence of recent geography and paleogeography on the structure of reptile communities in a land-bridge archipelago. *Journal of Herpetology* 2010;44:242–252. https://doi.org/10.1670/09-046.1

Flouri T, Rannala B, Yang Z. A tutorial on the use of BPP for species tree estimation and species delimitation. In: Delsuc F, Galtier N, Scornavacca C (eds), *Phylogenetics in the Genomic Era* 2020;1–16. No commercial publisher. Authors' open access book, hal02536475.

Girard C. A brief account of the fresh water planariae of the United States. *Proceedings of the Boston Society of Natural History* 1850;3:264–265.

Giusti F, Lesicki A, Benocci A, Pienkowska JR, Manganelli G. *Weltersia obscura*, a new slug from the island of Montecristo (Tuscan Archipelago, Italy): a hitherto undiscovered endemic or a recent alien? (Mollusca, Pulmonata, Limacidae). *Systematics and Biodiversity* 2021;19:648–664. https://doi.org/10.1080/14772000.2021.1908442

Giusti R, Colombo G, Pandeli E, Elter FM, Casalini M, Orti L. The plutonic-dyke system and host ligurian successions of the island of Montecristo (Tuscan Archipelago, Italy): New stratigraphic, petrographic and structural data. *Ofioliti* 2023;48:47–73. https://doi.org/10.4454/ofioliti.v48i1.561

Grummer JA, Bryson RW, Reeder TW. Species delimitation using bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* 2014;63:119–133. https://doi.org/10.1093/sysbio/syt069

Guilcher A. Pleistocene and Holocene sea level changes. *Earth Science Reviews* 1969;5:69–97.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 2010;59:307–321. https://doi.org/10.1093/sysbio/syq010

Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 1999;41:95–98. https://thalljiscience.github.io/. Last accessed: September 2023.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the Ultrafast Bootstrap approximation. *Molecular Biology and Evolution* 2018;35:518–522. https://doi.org/10.5281/zenodo.854445

Hyman LH. 1951. The invertebrates: Platyhelminthes and Rhynchocoela - *The acoelomate bilateria*, Vol. II. New York: McGraw-Hill Book Company, Inc.

Ichikawa A, Kawakatsu M. A new freshwater planarian, *Dugesia japonica*, commonly but erroneously known as *Dugesia gonocephala* (Dugès). *Annotationes Zoologicae Japonenses* 1964;37:185–194.

Innocenti F, Westerman DS, Rocchi S, Tonarini S. The Montecristo monzogranite (Northern Tyrrhenian Sea, Italy): a collisional pluton in an extensional setting. *Geological Journal* 1997;32:131–151. https://doi.org/10.1002/(SICI)1099-1034(199706)32:2<131::AID-GJ735>3.0.CO;2-F

Jouet G, Berné S, Rabineau M, Bassetti MA, Bernier P, Dennielou B, Sierro FJ, Flores JA, Taviani M. Shoreface migrations at the shelf edge and sea-level changes around the Last Glacial Maximum (Gulf of Lions, NW Mediterranean). *Marine Geology* 2006;234:21–42. https://doi.org/10.1016/j.margeo.2006.09.012

Kaburaki T. Freshwater triclads from the Basin of the Inle Lake. *Records of the Indian Museum* 1918;14:187–194.

Kaburaki T. Planarians from the Andamans. *Records of the Indian Museum* 1925;27:29–32.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017;14:587–589. https://doi.org/10.1038/nmeth.4285

Kass RE, Raftery AE. Bayes Factors. *Journal of the American Statistical Association* 1995;90:773–795.

Katoh K, Standley DM. MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* 2013;30:772–780. https://doi.org/10.1093/molbev/mst010

Kawakatsu M. Report on freswhater planaria from India. *Annotationes Zoologicae Japonenses* 1969;42:210–215.

Kawakatsu M, Oki I, Tamura S, Sugino H. Studies on the morphology, karyology and taxonomy of the Japanese freshwater planarian *Dugesia japonica* Ichikawa and Kawakatsu, with a description of a new subspecies, *Dugesia japonica ryukyensis* subspec. nov. *Bulletin of Fuji Women's College* 1976; no. 14, ser. II:81–126.

Kawakatsu M, Mitchell RW. *Dugesia deharvengi* sp. nov., new troglobitic freshwater planarians from Tham Kubio Cave, Thailand (Turbellaria, Tricladida, Paludicola). *Bulletin of*

the Biogeographical Society of Japan 1989;44:175–182.

Kawakatsu M, Mitchell RW. Two new freshwater cavernicolous planarians (Turbellaria, Tricladida, Paludicola) from Sulawesi (Celebes), Indonesia. *Special Bulletin of the Japanese Society of Coleopterology* 1995;4:81–103.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;1–3. https://doi.org/10.1093/bioinformatics/btz305

Lambeck K, Yokoyama Y, Purcell T. Into and out of the Last Glacial Maximum: sea-level change during Oxygen Isotope Stages 3 and 2. *Quaternary Science Reviews* 2002;21:343–360. https://doi.org/10.1016/S0277-3791(01)00071-3

Lambeck K, Antonioli F, Purcell A, Silenzi S. Sea-level change along the Italian coast for the past 10,000 yr. *Quaternary Science Reviews* 2004;23:1567–1598. https://doi.org/10.1016/j.quascirev.2004.02.009

Lang A. *Die Polycladen (Seeplanarien) des Golfes von Neapel und der angrenzenden Meeresabschnitte*. Leipzig: W. Engelmann. 1884.

Lázaro E, Sluys R, Pala M, Stocchino GA, Baguñà J, Riutort M. Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution* 2009;52:835–845. https://doi.org/10.1016/j.ympev.2009.04.022

Lepori NG. Descrizione di *Dugesia etrusca monoadenodactyla*, nuova razza di planaria di acqua dolce. *Monitore Zoologico Italiano* 1947;33:183–193.

Lepori NG. Descrizione di *Dugesia sícula*, nuova specie di Triclade d'acqua dolce dei dintorni di Catania. *Archivio Zoologico Italiano* 1948;33:461–472.

Lepori NG. Sulle caratteristiche morfologiche e sulla posizione sistematica della planaria di Sardegna e Corsica già ascritta a *Dugesia* (=*Euplanaria*) *gonocephala* (Dugès). *Atti Società Toscana di Scienze Naturali* 1951;58:28–47.

Leria L, Sluys R, Riutort M. Diversification and biogeographic history of the Western Palearctic freshwater flatworm genus *Schmidtea* (Tricladida: Dugesiidae), with a redescription of Schmidtea nova. *Journal of Zoological Systematics and Evolutionary Research* 2018;56:335–351. https://doi.org/10.1111/jzs.12214

Leria L, Vila-Farré M, Álvarez-Presas M, Sánchez-Gracia A, Rozas J, Sluys R, Riutort M. Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhel-

minthes, Tricladida): Extreme intraindividual genetic diversity, morphological stasis, and karyological variability. *Molecular Phylogenetics and Evolution* 2020;143:106496. https://doi.org/10.1016/j.ympev.2019.05.010

Marcus E. Turbellaria, Tricladida. In: Exploration du Parc National de l'Upemba, *Mission G. F. de Witte*. Bruxelles: M. Hayez, Fasc. 1953;21:1–62.

Massetti M. On the elusive origin of the wild goat, *Capra aegagrus* Erxleben, 17771 on the island of Montecristo (Italy). *Mammalia* 2016;80:124–141. https://doi.org/10.1515/mammalia-2014-0168

Melis P. Un Approdo della costa di Castelsardo, fra età nuragica e romana. *Carocci*, 2002.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 2020;37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Newmark PA, Alvarado AS. Not your father's planarian: a classic model enters the era of functional genomics. *Nature Reviews Genetics* 2002;3:210–219. https://doi.org/10.1038/nrg759

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 2015;32:268–274. https://doi.org/10.1093/molbev/msu300

Pala M, Casu S, Vacca RA. Rinvenimento di una planaria ascrivibile a *Dugesia etrusca monoadenodactyla* Lepori (Turbellaria, Tricladida) nell'Isola di Molara (Sardegna). *Bollettino della Società Sarda di Scienze Naturali* 1980;19:177–181.

Pala M, Casu S, Vacca RA. *Dugesia hepta*, nuova specie di planaria di acqua dolce di Sardegna appartenente alla superspecie *Dugesia gonocephala* (Duges). *Bollettino della Società Sarda di Scienze Naturali* 1981;20:97–107.

Pala M, Casu S, Lepori NG. Stabilized natural interspecific hybrid population of the fresh water planarians *Dugesia gonocephala* s.l. (Turbellaria, Tricladida). *Caryologia* 1982;35:247–256. https://doi.org/10.1080/00087114.1982.10796929

Pala M, Vacca R, Casu S, Stocchino GA. The freshwater planarian *Dugesia sicula* Lepori from Sardinia (Platyhelminthes, Tricladida). *Hydrobiologia* 1995;310:151–156. https://doi.org/10.1007/bf00015533

Pala M, Casu S, Stocchino GA. Karyology and karyotype analysis of diploid freshwater pla-

narian populations of the *Dugesia gonocephala* group (Platyhelminthes, Tricladida) found in Sardinia. *Hydrobiologia* 1999;392:113–119. https://doi.org/10.1023/A:1003534507632

Pala M, Stocchino GA, Corso G, Casu S. A new species of freshwater planarian belonging to the genus *Dugesia* (Platyhelminthes, Tricladida) from Sardinia. *Italian Journal of Zoology* 2000;67:379–383. https://doi.org/10.1080/11250000009356343

Radmilli AM. Il popolamento umano dell'Arcipelago Toscano prima dell'età romana. *The Journal of Integrative Biogeography* 1976;5:899–916. https://doi.org/10.21426/b65110080

Randi E, Tosi G, Toso S, Lorenzini R, Fusco G. Genetic variability and conservation problems in Alpine ibex, domestic and feral goat populations (genus *Capra*). *Zeitschrift für Säugetierkunde* 1990;55:413–420.

Ribas M. *Cariologia, sistemàtica i biogeografia de les planàries d'aigües dolces als Països Catalans.* Ph.D. Thesis, Universitat de Barcelona, 1990.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 2012;61:539–542. https://doi.org/10.1093/sysbio/sys029

Sluys R. Annotations on freshwater planarians (Platyhelminthes, Tricladida, Dugesiidae) from the Afrotropical Region. *Tropical Zoology* 2007;20:229–257.

Sluys R, Riutort M. Planarian diversity and phylogeny. In: J. C. Rink (Ed.), *Planarian regeneration. Methods and protocols. Methods in Molecular Biology*, Vol. 1774, Humana Press, Springer Science+Business Media, New York, 2018, 1– 56. https://doi.org/10.1007/978-1-4939-7802-1_1

Sluys R, Solà E, Gritzalis K, Vila-Farré M, Mateos E, Riutort M. Integrative delineation of species of Mediterranean freshwater planarians (Platyhelminthes: Tricladida: Dugesiidae). *Zoological Journal of the Linnean Society* 2013;169:523–547. https://doi.org/10.1111/zoj.12077

Solà E, Leria L, Stocchino GA, Bagherzadeh R, Balke M, Daniels SR, Harrath AH, Khang TF, Krailas D, Kumar B, Li M-H, Maghsoudlou A, Matsumoto M, Naser N, Oben B, Segev O, Thielicke M, Tong X, Zivanovic G, Manconi R, Baguñà J, Riutort M. Three dispersal routes out of Africa: A puzzling biogeographical history in freshwater planarians. *Journal of Biogeography* 2022;49:1219–1233. https://doi.org/10.1111/jbi.14371

Solà E, Sluys R, Gritzalis K, Riutort, M. Fluvial basin history in the northeastern Mediterranean region underlies dispersal and speciation patterns in the genus *Dugesia* (Platyhel-

minthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution* 2013;66:877–888. https://doi.org/10.1016/j.ympev.2012.11.010

Stocchino GA, Corso G, Manconi R, Casu S, Pala M. Endemic freshwater planarians of Sardinia: redescription of *Dugesia hepta* (Platyhelminthes, Tricladida) with a comparison of the Mediterranean species of the genus. *Journal of Natural History* 2005;39:1947–1960. https://doi.org/10.1080/00222930500060025

Stocchino GA, Manconi R. Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology* 2013;803:319–328. https://doi.org/10.1080/11250003.2013.822025

Stocchino GA, Sluys R, Manconi R. A new and aberrant species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from Madagascar. *ZooKeys* 2014;425:71–88. http://dx.doi.org/10.3897/zookeys.425.7811

Stocchino GA, Sluys R, Riutort M, Solà E, Manconi R. Freshwater planarian diversity (Platyhelminthes, Tricladida, Dugesiidae) in Madagascar: new species, cryptic species, with a redefinition of character states. *Zoological Journal of the Linnean Society* 2017;181:727–756. http://dx.doi.org/10.1093/zoolinnean/zlx017

Stocchino GA. 80 years of research on planarians (Platyhelminthes, Tricladida) from Sardinia, Italy: an annotated checklist. *Zootaxa* 2018;4532:539–552. https://doi.org/10.11646/zootaxa.4532.4.5

Sukumaran J, Knowles LL. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences* 2017;114:1607–1612. https://doi.org/10.1073/pnas.1607921114

Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution* 2021;38:3022–3027. https://doi.org/10.1093/molbev/msab1200

Torrini G, Mazza G, Strangi A, Barabaschi D, Landi S, Mori E, Menchetti M, Sposimo P, Giuliani C, Zoccola A, Lazzaro L, Ferretti G, Foggi B, Roversi PF. Oscheius tipulae in Italy: Evidence of an alien isolate in the integral Natural Reserve of Montecristo Island (Italy). *Journal of Nematology* 2016;48:8–13. https://doi.org/10.21307/jofnem-2017-003

Vieites DR, Wollenberg KC, Andreone F, Köhler J, Glaw F, Vences M. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. *Proceedings of the National Academy of Sciences* 2009;106:8267–8272. https://doi.org/10.1073/pnas.0810821106

Vigne J-D. Zooarchaeology and the biogeographical history of the mammals of Corsica and Sardinia since the last ice age. *Mammal Review* 1992;22:87–96. https://doi.org/10.1111/j.1365-2907.1992.tb00124.x

de Vries EJ. The biogeography of the genus *Dugesia* (Turbellaria, Tricladida, Paludicola) in the Mediterranean Region. *Journal of Biogeography* 1985;12:509–518. https://doi.org/10.2307/2844906

de Vries EJ. A synopsis of the nominal species of the subgenus *Dugesia* (Platyhelminthes, Tricladida, Paludicola) from Africa and Madagascar. *Zoological Journal of the Linnean Society* 1988;92:345–382.

Wang L, Chen JZ, Dong ZM, Chen GW, Sluys R, Liu DZ. Two new species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from the tropical monsoon forest in southern China. *ZooKeys* 2021a;1059:89–116. https://doi.org/10.3897/zookeys.1059.65633

Wang L, Dong ZM, Chen GW, Sluys R, Liu DZ. Integrative descriptions of two new species of *Dugesia* from Hainan Island, China (Platyhelminthes, Tricladida, Dugesiidae). *ZooKeys* 2021b;1028:1–28. https://doi.org/10.3897/zookeys.1028.60838

Winsor, L., Sluys, R. Basic histological techniques for planarians. In: Rink JC (ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology, Methods in Molecular Biology*, Vol. 1774, Humana Press, Springer Science+Business Media, New York, 2018;285–351. https://doi.org/10.1007/978-1-4939-7802-1_9

Xia X, Xie Z. DAMBE: Software package for data analysis in molecular biology and evolution. *Journal of Heredity* 2001;92:371–373. https://doi.org/10.1093/jhered/92.4.371

Xia X, Xie Z, Salemi M, Chen L, Wang Y. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* 2003;26:1–7. https://doi.org/10.1016/S10557903(02)00326-3

Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* 2011;60:150–160. https://doi.org/10.1093/sysbio/syq085

Yang Z. The BPP program for species tree estimation and species delimitation. *Current Zoology* 2015;61:854–865. https://doi.org/10.1093/czoolo/61.5.854

Yang Z, Rannala B. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 2010;164:1645–1656. https://doi.org/10.1073/pnas.0913022107

Zecchin M, Ceramicola S, Lodolo E, Casalbore D, Chiocci FL. Episodic, rapid sea-level rises on the central Mediterranean shelves after the Last Glacial Maximum: A review. *Marine Geology* 2015;369:212–223. http://dx.doi.org/10.1016/j.margeo.2015.09.002

Supporting Information: Figures & Tables

New insights into the evolution and biogeography of freshwater planarians on islands in the Tyrrhenian Sea, Western Mediterranean Basin, with the integrative description of a new endemic species from Corsica (Platyhelminthes, Tricladida, *Dugesia*)

Daniel Dols-Serrate[1,2], Giacinta Angela Stocchino[3], Paula Nuin-Villabona[1], Ronald Sluys[4] & Marta Riutort[1,2]

[1]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[3]Dipartimento di Medicina Veterinaria, Via Vienna 2, University of Sassari, Sassari I-07100, Italy

[4]Naturalis Biodiversity Center, P. O. Box 9517, 2300 RA Leiden, The Netherlands

ORCIDs:

0000-0002-2134-7674 Marta Riutort

0000-0003-0820-7324 Daniel Dols-Serrate

0000-0002-7005-208X Giacinta Angela Stocchino

0000-0002-9776-3471 Ronald Sluys

Corresponding author: D. Dols-Serrate danieldols@ub.edu/bcn27101993@gmail.com

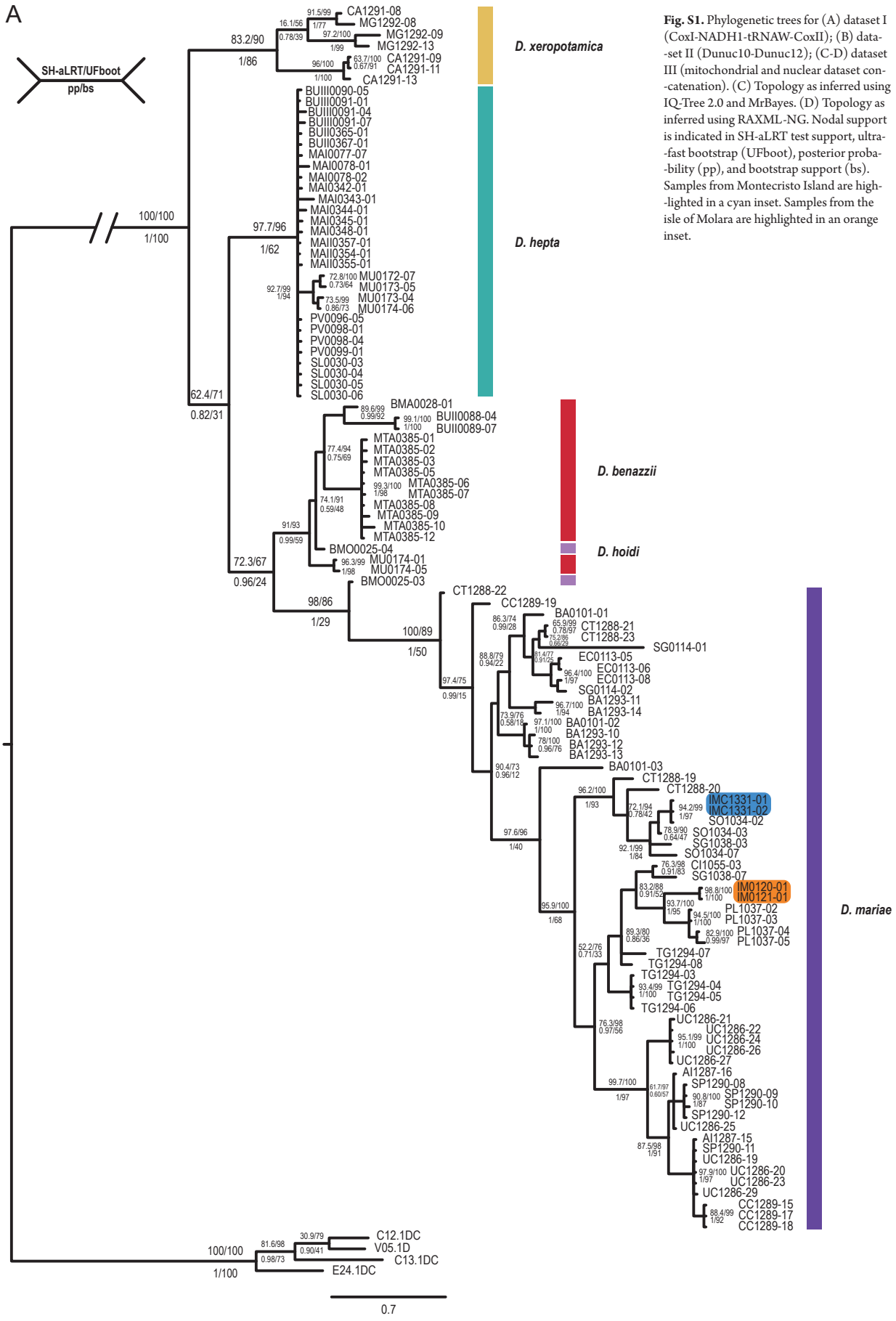
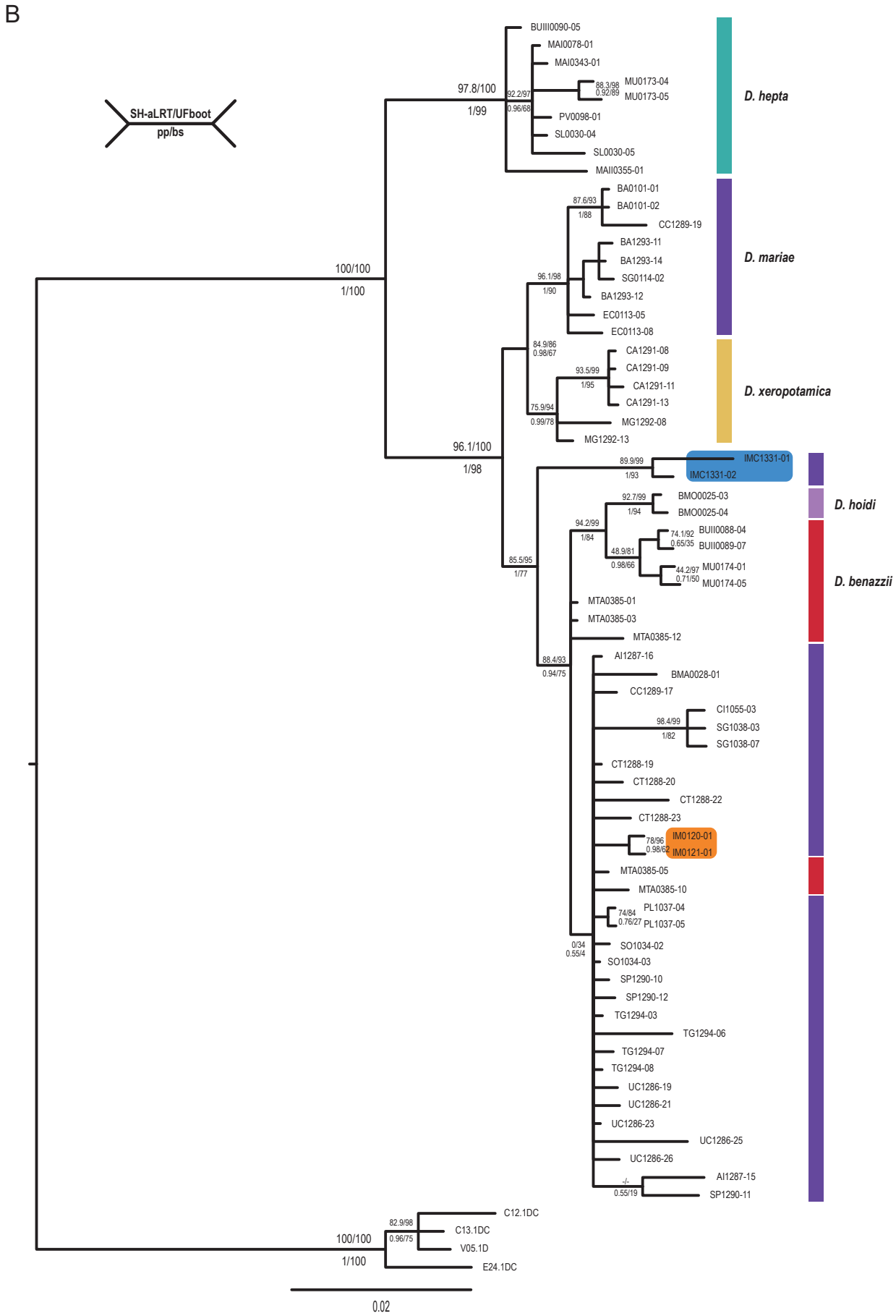


Fig. S1. Phylogenetic trees for (A) dataset I (CoxI-NADH1-tRNAW-CoxII); (B) data--set II (Dunuc10-Dunuc12); (C-D) dataset III (mitochondrial and nuclear dataset con--catenation). (C) Topology as inferred using IQ-Tree 2.0 and MrBayes. (D) Topology as inferred using RAXML-NG. Nodal support is indicated in SH-aLRT test support, ultra--fast bootstrap (UFboot), posterior proba--bility (pp), and bootstrap support (bs). Samples from Montecristo Island are high--lighted in a cyan inset. Samples from the isle of Molara are highlighted in an orange inset.


B
SH-aLRT/UFboot
pp/bs
BUIII0090-05
MAI0078-01
MAI0343-01
MU0173-04
MU0173-05
PV0098-01
SL0030-04
SL0030-05
MAII0355-01
D. hepta
BA0101-01
BA0101-02
CC1289-19
BA1293-11
BA1293-14
SG0114-02
BA1293-12
EC0113-05
EC0113-08
D. mariae
CA1291-08
CA1291-09
CA1291-11
CA1291-13
MG1292-08
MG1292-13
D. xeropotamica
IMC1331-01
IMC1331-02
D. hoidi
BMO0025-03
BMO0025-04
BUII0088-04
BUII0089-07
MU0174-01
MU0174-05
D. benazzii
MTA0385-01
MTA0385-03
MTA0385-12
AI1287-16
BMA0028-01
CC1289-17
CI1055-03
SG1038-03
SG1038-07
CT1288-19
CT1288-20
CT1288-22
CT1288-23
IM0120-01
IM0121-01
MTA0385-05
MTA0385-10
PL1037-04
PL1037-05
SO1034-02
SO1034-03
SP1290-10
SP1290-12
TG1294-03
TG1294-06
TG1294-07
TG1294-08
UC1286-19
UC1286-21
UC1286-23
UC1286-25
UC1286-26
AI1287-15
SP1290-11
C12.1DC
C13.1DC
V05.1D
E24.1DC
0.02

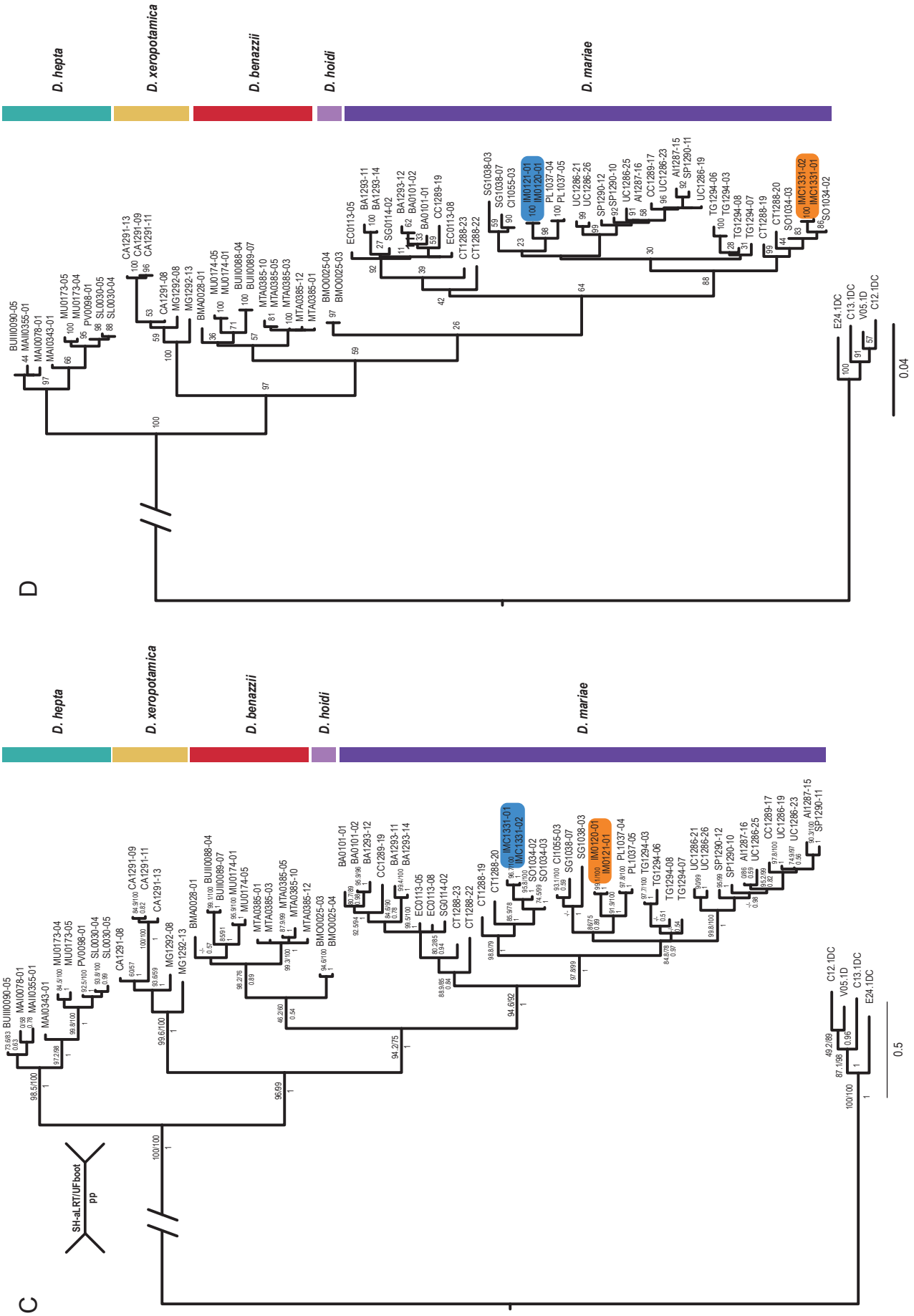

C
D
D. hepta
D. xeropotamica
D. benazzii
D. holoi
D. mariae
SH-aLRT/UFboot
pp
0.5
0.04

Table S1. Complete list of specimens used for the present study and pertinent Gen Bank accession numbers.

Locality		Coordinates				Population number		
Specimen	**Species**	**Karyotype**	**COI**	**NADH1-tRNAW-COII**	**Dunuc10**	**Dunuc12**	**Collectors[a]**	**Date of collection**
Badde Manna, Sardinia			**40°46'52"N, 9°06'59"E**			**1**		
BMA0028-01	*D. benazzii*	-	MN525640	OP720241	-	MN442438	P	17/07/1997
Badu de Mola, Sardinia			**40°18'11"N, 8°34'57"E**			**2**		
BMO0025-03	*D. hoidi*	-	MN525637	OP720240	OP795470	MN442435	P	17/07/1997
BMO0025-04	*D. hoidi*	-	OR650791	-	-	OR640643	P	17/07/1997
Bunnari 2, Sardinia			**40°42'05"N, 8°35'30"E**			**3**		
BUII0088-04	*D. benazzii*	-	MK712640	-	OR640540	MK713164	P, S, L	05/05/2010
BUII0089-07	*D. benazzii*	-	MN525665	-	-	MN442448	P, S, L	05/05/2010
BUII0365-01	*D. hepta*	n = 7	MN525738	-	-	MN442504	S	01/01/2013
BUII0367-01	*D. hepta*	n = 7	MN525739	-	-	MN442506	S	01/01/2013
Bunnari 3, Sardinia			**40°42'17"N, 8°35'46"E**			**4**		
BUIII0090-05	*D. hepta*	-	MN525671	-	-	MN442449	P, S, L	06/05/2010
BUIII0091-01	*D. hepta*	-	MN525674	OP720245	OP795473	MN442450	P, S, L	06/05/2010
BUIII0091-04	*D. hepta*	-	MN525677	OP720246	-	MN442451	P, S, L	06/05/2010
BUIII0091-07	*D. hepta*	-	MN525680	-	-	MN442452	P, S, L	06/05/2010
Mascari 1, Sardinia			**40°41'57"N, 8°35'03"E**			**5**		
MAI0077-07	*D. hepta*		MN525649	-	-	MN442442	P, S, L	06/05/2010
MAI0078-01	*D. hepta*	-	MN525650	OP720244	OP795472	MN442443	P, S, L	06/05/2010

MAI0078-02	*D. hepta*	-	MN525651	-	-	MN442444	P, S, L	06/05/2010
MAI0342-01	*D. hepta*	n = 7	MK385923	OP720277	-	MK713158	S	01/01/2013
MAI0343-01	*D. hepta*	n = 7	MN525730	OP720278	OP795484	MN442483	S	01/01/2013
MAI0344-01	*D. hepta*	n = 7	MK385924	OP720279	-	MK713159	S	01/01/2013
MAI0345-01	*D. hepta*	n = 7	MN525731	OP720280	-	MN442484	S	01/01/2013
MAI0348-01	*D. hepta*	n = 7	MN525732	-	-	MN442487	S	01/01/2013
Mascari 2, Sardinia			**40°41'60"N, 8°35'23"E**			**6**		
MAII0354-01	*D. hepta*	n = 7	MN525735	OR640510	-	MN442493	S	01/01/2013
MAII0355-01	*D. hepta*	n = 7	MN525736	OP720281	OP795485	MN442494	S	01/01/2013
MAII0357-01	*D. hepta*	n = 7	MN525737	-	-	MN442496	S	01/01/2013
Ponte Valenti, Sardinia			**40°28'09"N, 8°48'21"E**			**7**		
PV0096-05	*D. hepta*	-	MN525697	-	-	-	P, S, L	10/05/2010
PV0098-01	*D. hepta*	-	MN525699	OP720247	-	MN442463	P, S, L	10/05/2010
PV0098-04	*D. hepta*	-	MN525700	OP720249	-	MN442465	P, S, L	10/05/2010
PV0099-01	*D. hepta*	-	MN525701	OP720250	OP795474	MN442466	P, S, L	10/05/2010
Mulinu, Sardinia			**40°24'25"N, 8°37'35"E**			**8**		
MU0172-07	*D. hepta*	-	MK712639	-	-	MK713161	P, S, L	13/05/2010
MU0173-04	*D. hepta*	-	MN525722	OP720273	-	MN442475	P, S, L	13/05/2010
MU0173-05	*D. hepta*	-	MN525723	OR640514	-	MN442476	P, S, L	13/05/2010
MU0174-01	*D. benazzii*	-	MN525725	OP720274	OP795482	MN442478	P, S, L	13/05/2010
MU0174-05	*D. benazzii*	-	MN252726	OP720275	OP795483	MN442481	P, S, L	13/05/2010
MU0174-06	*D. hepta*	-	MN525727	OP720276	-	MN442482	P, S, L	13/05/2010
Santa Lucia, Sardinia			**40°25'53"N, 8°50'09"E**			**9**		
SL0030-03	*D. hepta*	-	MN525641	OP720242	-	MN442439	P	01/02/2006

SL0030-04	*D. hepta*	-	MN525642	-	OP795471	MN442440	P	01/02/2006
SL0030-05	*D. hepta*	-	MK712638	OP720243	-	MK713160	P	01/02/2006
SL0030-06	*D. hepta*	-	MN525643	OR640516	-	MN442441	P	01/02/2006
Monte Albo, Sardinia			**40°34'30"N, 9°40'31"E**			**10**		
MTA0385-01	*D. benazzii*	-	MK385925	OP720282	OP795486	MK713162	S	01/05/2012
MTA0385-02	*D. benazzii*	-	MK385926	OP720283	-	MK713163	S	01/05/2012
MTA0385-03	*D. benazzii*	-	MN525740	OP720284	OP795487	MN442509	S	01/05/2012
MTA0385-05	*D. benazzii*	-	MN525742	OP720286	-	MN442510	S	01/05/2012
MTA0385-06	*D. benazzii*	-	MN525743	OP720287	-	MN442511	S	01/05/2012
MTA0385-07	*D. benazzii*	-	MN525744	-	-	MN442512	S	01/05/2012
MTA0385-08	*D. benazzii*	-	MN525745	OP720288	-	MN442513	S	01/05/2012
MTA0385-09	*D. benazzii*	-	MN525746	OP720289	-	MN442514	S	01/05/2012
MTA0385-10	*D. benazzii*	-	MN525747	OP720290	-	MN442515	S	01/05/2012
MTA0385-12	*D. benazzii*	-	MN525749	OP720292	-	MN442516	S	01/05/2012
La Solenzara, Corsica			**41°50'39.6"N, 9°20'34"E**			**11**		
SO1034-02	*D. mariae*	-	MN525751	OP720294	OP795489	OP795461	P, S, L	01/05/2009
SO1034-03	*D. mariae*	-	MN525752	OP720295	OP795490	OP795462	P, S, L	01/05/2009
SO1034-07	*D. mariae*	-	OP561879	OP720299	OP795491	OP795463	P, S, L	01/05/2009
Cioccia, Corsica			**41°32'41.7"N, 9°01'24.1"E**			**12**		
CI1055-03	*D. mariae*	-	MN525765	OP720307	OP795496	OP795468	P, S, L	01/05/2009
Pero Longo, Corsica			**41°32'27.6"N, 8°58'07.2"E**			**13**		
PL1037-02	*D. mariae*	-	MN525756	OP720303	OP795492	OP795464	P, S, L	01/05/2009
PL1037-03	*D. mariae*	-	MN525757	OP720304	OP795493	OP795465	P, S, L	01/05/2009
PL1037-04	*D. mariae*	-	MN525758	OP720305	OP795494	OP795466	P, S, L	01/05/2009

PL1037-05	*D. mariae*	-	MN525759	OP720306	OP795495	OP795467	P, S, L	01/05/2009
U Crucolli, Corsica			**42°03'31.9"N, 8°57'54.5"E**			**14**		
UC1286-19	*D. mariae*	n = 8*	OR650820	OR640528	OR640563	OR640664	L, DS	26/11/2019
UC1286-20	*D. mariae*	-	OR650821	OR640529	-	-	L, DS	26/11/2019
UC1286-21	*D. mariae*	-	OR650822	OR640530	OR640564	OR640665	L, DS	26/11/2019
UC1286-22	*D. mariae*	-	OR650823	OR640531	-	-	L, DS	26/11/2019
UC1286-23	*D. mariae*	-	OR650824	OR640532	OR640565	OR640666	L, DS	26/11/2019
UC1286-24	*D. mariae*	-	OR650825	OR640533	-	-	L, DS	26/11/2019
UC1286-25	*D. mariae*	n = 8*	OR650826	OR640534	OR640566	OR640667	L, DS	26/11/2019
UC1286-26	*D. mariae*	n = 8*	OR560827	-	OR640567	OR640668	L, DS	26/11/2019
UC1286-27	*D. mariae*	-	OR650828	OR640535	-	-	L, DS	26/11/2019
UC1286-29	*D. mariae*	n = 8*	OR650829	-	-	-	L, DS	26/11/2019
Aiaccio, Corsica			**41°58'25.0"N, 8°49'53.4"E**			**15**		
AI1287-15	*D. mariae*	n = 8*	OR650784	OR640490	OR640536	OR640638	L, DS	26/11/2019
AI1287-16	*D. mariae*	n = 8*	OR650785	OR640491	OR640537	OR640639	L, DS	26/11/2019
Corte, Corsica			**42°18'18.3"N, 9°09'55.9"E**			**16**		
CT1288-19	*D. mariae*	-	OR650800	OR640505	OR640548	OR640650	L, DS	26/11/2019
CT1288-20	*D. mariae*	n = 8*	OR650801	OR640506	OR640549	OR640651	L, DS	26/11/2019
CT1288-21	*D. mariae*	n = 8*	OR650802	OR640507	-	-	L, DS	26/11/2019
CT1288-22	*D. mariae*	-	OR650803	OR640508	OR640550	OR640652	L, DS	26/11/2019
CT1288-23	*D. mariae*	-	OR650804	OR640509	OR640551	OR640653	L, DS	26/11/2019
Camping Campita, Corsica			**42°23'34.2"N, 9°10'34.8"E**			**17**		
CC1289-15	*D. mariae*	n = 8*	OR650796	OR640501	-	-	L, DS	26/11/2019
CC1289-17	*D. mariae*	n = 8[t]	OR650797	OR640502	OR640546	OR640648	L, DS	26/11/2019
CC1289-18	*D. mariae*	n = 8[t]	OR650798	OR640503	-	-	L, DS	26/11/2019

CC1289-19	*D. mariae*	-	OR650799	OR640504	OR640547	OR640649	L, DS	26/11/2019
Setti Polli, Corsica			**41°55'43.9"N, 8°52'15.6"E**			**18**		
SP1290-08	*D. mariae*	n = 8	OR650809	OR640517	-	-	L, DS	26/11/2019
SP1290-09	*D. mariae*	n = 8	OR650810	OR640518	-	-	L, DS	26/11/2019
SP1290-10	*D. mariae*	n = 8	OR650811	OR640519	OR640556	OR640657	L, DS	26/11/2019
SP1290-11	*D. mariae*	n = 8	OR650812	OR640520	OR640557	OR640658	L, DS	26/11/2019
SP1290-12	*D. mariae*	n = 8	OR650813	OR640521	OR640558	OR640659	L, DS	26/11/2019
Calvi, Corsica			**42°29'10.8"N, 8°48'09.2"E**			**19**		
CA1291-08	*D. xeropotamica*	n = 8	OR650792	OR640497	OR640542	OR640644	L, DS	27/11/2019
CA1291-09	*D. xeropotamica*	n = 8	OR650793	OR640498	OR640543	OR640645	L, DS	27/11/2019
CA1291-11	*D. xeropotamica*	n = 8	OR650794	OR640499	OR640544	OR640646	L, DS	27/11/2019
CA1291-13	*D. xeropotamica*	n = 8	OR650795	OR640500	OR640545	OR640647	L, DS	27/11/2019
Montegrosso, Corsica			**42°32'17.4"N, 8°50'38.5"E**			**20**		
MG1292-08	*D. xeropotamica*	n = 8	OR650805	OR640511	OR640552	OR640654	L, DS	27/11/2019
MG1292-09	*D. xeropotamica*	n = 8	OR650806	OR640512	-	-	L, DS	27/11/2019
MG1292-13	*D. xeropotamica*	n = 8	OR650807	OR640513	OR640553	OR640656	L, DS	27/11/2019
Barchetta, Corsica			**42°30'27.0"N, 9°22'25.9"E**			**21**		
BA0101-01	*D. mariae*	-	MN525702	OP720251	OP795475	OP795454	P, S, L	17/05/2010
BA0101-02	*D. mariae*	-	MN525703	OP720252	OP795476	OP795455	P, S, L	17/05/2010
BA0101-03	*D. mariae*	-	OP561860	OP720253	OP795477	OP795456	P, S, L	17/05/2010
BA0101-10	*D. mariae*	-	OP561867	OP720260	OP795478	OP795457	P, S, L	17/05/2010
BA1293-11	*D. mariae*	n = 8	OR650787	OR640493	OR640538	OR640640	L, DS	27/11/2019
BA1293-12	*D. mariae*	n = 8	OR650788	OR640494	OR640539	OR640641	L, DS	27/11/2019
BA1293-13	*D. mariae*	n = 8	OR650789	OR640495	-	-	L, DS	27/11/2019
BA1293-14	*D. mariae*	n = 8	OR650790	OR640496	OR640540	OR640642	L, DS	27/11/2019

Turghja, Corsica		**41°51'09.1"N, 8°58'00.1"E**				**22**		
TG1294-03	*D. mariae*	n = 8*	OR650814	OR640522	OR640559	OR640660	L, DS	28/11/2019
TG1294-04	*D. mariae*	n = 8*	OR650815	OR640523	-	-	L, DS	28/11/2019
TG1294-05	*D. mariae*	n = 8*	OR650816	OR640524	-	-	L, DS	28/11/2019
TG1294-06	*D. mariae*	n = 8*	OR650817	OR640425	OR640560	OR640661	L, DS	28/11/2019
TG1294-07	*D. mariae*	-	OR650818	OR640426	OR640561	OR640662	L, DS	28/11/2019
TG1294-08	*D. mariae*	n = 8*	OR650819	OR640527	OR640562	OR640663	L, DS	28/11/2019
Ecole Communale, Corsica		**42°26.648'N, 9°26.386'E**				**23**		
EC0113-05	*D. mariae*	-	OP561872	OP720266	OP795479	OP795458	P, S, L	18/05/2010
EC0113-06	*D. mariae*	-	OP561873	OP720267	OP795480	OP795459	P, S, L	18/05/2010
EC0113-08	*D. mariae*	-	OP561875	OP720269	-	OP795460	P, S, L	18/05/2010
San Gavino, Corsica		**42°25.52'N, 9°25.972'E**				**24**		
SG0114-01	*D. mariae*	-	MK712641	OP720272	OP795481	MK713166	P, S, L	19/05/2010
SG0114-02	*D. mariae*	-	MK712642	OR640515	-	MK713167	P, S, L	19/05/2010
SG1038-03	*D. mariae*	-	MN525761	-	OR640554	-	P, S, L	01/05/2009
SG1038-07	*D. mariae*	-	OR650808	-	OR640555	-	P, S, L	01/05/2009
Molara, Sardinia		**40°52'04.2"N, 9°43'36.6"E**				**25**		
IM0120-01	*D. mariae*	-	PP087950	PP100177	-	PP100181	P	01/07/1997
IM0121-01	*D. mariae*	-	PP087951	PP100178	-	PP100182	P	01/07/1997
Montecristo, Italy		**42°20'26.0"N, 10°19'05.8"E**				**26**		
IMC1331-01	*D. mariae*	-	PP087950	PP100179	-	PP100183	RV	2014
IMC1331-02	*D. mariae*	-	PP087951	PP100180	-	PP100184	RV	2014

Outgroup

Trelles, Spain		**43°28'33.6"N, 6°44'00.9"W**					
C12.1DC	*D. subtentaculata*	MK712596	MK712667 + MK713188 + MK713061	MK712870	MK713106	V, So, Le	2013
Luarca, Spain		**43°32'30.8"N, 6°32'07.4"W**					
C13.1DC	*D. subtentaculata*	MK712612	MK712669 + MK713190 + MK713063	MK712871	-	V, So, Le	2013
Monte de Vento, Portugal		**39°17'20.2"N, 7°23'01.8"W**					
E24.1DC	*D. subtentaculata*	MK712611	MK712682 + MK713200 + MK713074	MK712887	MK713134	A, R, Le	04/06/2013
Peralejos de las Truchas, Spain		**40°35'35.5"N, 1°55'32.2"W**					
V05.1D	*D. subtentaculata*	MK712608	MK712704 + MK713222 + MK713096	MK712921	MK713128	V, So, Le	2013

[a] **Collectors:** (A) M. Álvarez-Presas, (DS) D. Dols-Serrate, (M) E. Mateos, (L) E. Lázaro, (Le) L. Leria, (P) M. Pala, (S) G. A. Stocchino, (So) E. Solà, (R) M. Riutort, (RV) R. Vila, (V) M. Vila-Farré. *** Aneuploidy** [t] **Triploidy**

Table S2. List of specific primers used for amplification and sequencing in the present study.

Name		Marker(s)	Sequence 5'-3'	Annealing Temp. (°C)	Source
SamCF	F[1]	CoxI	GCTAATAATTTGAGTTTTTG	51	Dols-Serrate et al., 2020
TarlyCR	R[2]	CoxI	CATTTTAAAACAACATTACC	51	Dols-Serrate et al., 2020
Elo12F	F	Dunuc12	AAGCTATTTATAATTCAGCG	54	Dols-Serrate et al., 2020
Hell12R	R	Dunuc12	AAAGAAATTGCTGCTAAAG	54	Dols-Serrate et al., 2020
Finsl_D10F	F	Dunuc10	TCCGGTAAGATTAATG	48	Dols-Serrate et al., 2023
Sous_D10R	R	Dunuc10	TCAACAGATCCGACC	48	Dols-Serrate et al., 2023
Cupi2	F	NADH1-tRNAW-CoxII	ATTTGRCTYCGTTCTTC	43	Leria et al., 2020
coIIP	R	NADH1-tRNAW-CoxII	CCACAAACCTCAGAAC	43	Leria et al., 2020

[1] Forward [2] Reverse

Table S3. Comparison of the different models and their prior parametrization used for the Bayes Factors Delimitation (BFD) analyses.

Dataset	Tree Prior	Molecular clock	Substitution rate[a]	Substitution model
Dataset I	Birth-and-Death	Uncorrelated relaxed*	0.017[b]	CoxI, NADH1, CoxII: HKY+G tRNA-W: HKY
Dataset II	Birth-and-Death	Uncorrelated relaxed	Default	Dunuc10, Dunuc12 : HKY+G
Dataset III	Birth-and-Death	Uncorrelated relaxed*	0.017[b]	CoxI, NADH1, CoxII, Dunuc10, Dunuc12: HKY+G tRNA-W: HKY

[a] Number of substitutions per site divided by length of tree

[b] Only applied to CoxI partition; normal distribution with mean 0.017 and stdev 0.05

* All partitions except for tRNAW which had a strict clock

Table S4. Results for the saturation tests on the first, second and third codon positions of dataset I and dataset II introns. Significant results (P < 0.05 highlighted with *) when Iss < Iss.c indicates no significant saturation.

Num. OTU	Iss			Iss.cSym			Iss.cAsym			P Sym			P Asym		
Dataset I	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
4	0.301	0.291	0.444	0.790	0.790	0.790	0.758	0.758	0.758	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
8	0.266	0.255	0.404	0.744	0.744	0.744	0.633	0.633	0.633	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
16	0.257	0.244	0.384	0.706	0.706	0.706	0.497	0.497	0.497	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*
32	0.254	0.240	0.383	0.693	0.693	0.693	0.365	0.365	0.365	0.000*	0.000*	0.000*	0.000*	0.000*	0.000*

Num. OTU				
Dunuc10 intron	Iss	Iss.cSym	Iss.cAsym	P Sym	P Asym
4	0.558	0.786	0.803	0.000*	0.000*
8	0.492	0.759	0.683	0.000*	0.001*
16	0.469	0.609	0.460	0.021*	0.877[a]
32	0.465	0.737	0.465	0.000*	0.999[a]

Num. OTU				
Dunuc12 intron	Iss	Iss.cSym	Iss.cAsym	P Sym	P Asym
4	0.112	0.799	0.767	0.000*	0.000*
8	0.105	0.757	0.646	0.000*	0.000*
16	0.103	0.730	0.520	0.000*	0.000*
32	0.105	0.708	0.380	0.000*	0.000*

[a] Poor for phylogenetics

Table S5. Results of guided Bayesian species delimitation in BPP (model A10) based on datasets I, II, and III, using four different evolutionary and demographic models. Posterior probabilities are presented for each node (A-D, in Fig. 3A). A0: algorithm 0; A1: algorithm 1. Non-supported splits are highlighted in bold.

	Model 1		Model 2		Model 3		Model 4	
	$\theta_1 \sim IG(3, 0.002)\ \tau_1 \sim IG(3, 0.002)$		$\theta_2 \sim IG(3, 0.1)\ \tau_2 \sim IG(3, 0.002)$		$\theta_3 \sim IG(3, 0.1)\ \tau_3 \sim IG(3, 0.1)$		$\theta_4 \sim IG(3, 0.002)\ \tau_4 \sim IG(3, 0.1)$	
	A0	A1	A0	A1	A0	A1	A0	A1
Dataset I								
A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
D	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
Dataset II								
A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
D	**0.93**	**0.92**	**0.63**	**0.64**	**0.52**	**0.52**	**0.91**	**0.91**
Dataset III								
A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
D	1.00	1.00	0.99	0.99	0.99	0.99	1.00	1.00

Table S6. Marginal-likelihood estimates (MLE) and Bayes factors comparisons (*2lnBf*) for BFD. N/A indicates the best model per data set and estimation method (PP/SS).

Dataset	Mitochondrial (dataset I)				Nuclear (dataset II)				Nuc + Mit (dataset III)			
Model	PS		SS		PS		SS		PS		SS	
	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*
1	**-5570.8671**	**N/A**	**-5571.9164**	**N/A**	-2925.4943	0.1300	-2925.5562	0.2886	**-7839.7955**	**N/A**	**-7841.9383**	**N/A**
2	-5602.5721	63.4100	-5602.9216	62.0104	**-2925.4293**	**N/A**	**-2925.4219**	**N/A**	-7854.5340	29.4770	-7855.9801	28.0386
3	-5590.6569	39.5796	-5589.5441	35.2554	-2934.4042	17.9498	-2934.4255	18.0072	-7846.6745	13.7580	-7848.7515	13.6264
4	-5647.3603	152.9864	-5648.8421	153.8514	-2954.4898	58.1392	-2954.7889	58.7340	-7911.1769	142.7628	-7913.9558	144.0350
5	-5656.0444	170.3546	-5657.1517	170.4706	-2944.3524	37.8462	-2944.4659	38.0880	-7919.3761	159.1612	-7922.5462	161.2158

Chapter III

Mixed, not stirred: Genomic analyses of *Dugesia* (Platyhelminthes, Tricladida) shed light on the evolutionary history and diversification of the Corso-Sardinian clade and reveal the first case of interspecific hybridization in freshwater flatworms

Dols-Serrate, D., Longhua, G., Kruglyak, L., Riutort, M.

Under review in Molecular Phylogenetics and Evolution

Mixed, not stirred: Genomic analyses of *Dugesia* (Platyhelminthes, Tricladida) shed light on the evolutionary history and diversification of the Corso-Sardinian clade and reveal the first case of interspecific hybridization in freshwater flatworms

Daniel Dols-Serrate[1,2], Longhua Guo[3,4], Leonid Kruglyak[5,6] & Marta Riutort[1,2]

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[3] Geriatrics Center and Institute of Gerontology, University of Michigan, Ann Arbor, Michigan, USA

[4] Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, Michigan, USA

[5] Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

[6] Department of Human Genetics, Department of Biological Chemistry, University of California, Los Angeles, California, USA

ORCIDs:
0000-0003-0820-7324 Daniel Dols-Serrate
0000-0001-9690-9750 Longhua Guo
0000-0002-8065-3057 Leonid Kruglyak
0000-0002-2134-7674 Marta Riutort

Corresponding author: D. Dols-Serrate danieldols@ub.edu/bcn27101993@gmail.com

Abstract

Speciation is a complex process where many evolutionary forces interplay. The Mediterranean is acknowledged as one of the most relevant biodiverse areas in the Palearctic region and researchers have long studied the species inhabiting it to pursue the goals of evolutionary biology. Here, we study a complex of freshwater flatworm species of the genus *Dugesia* from Corsica and Sardinia using restriction site-associated DNA sequencing (3RAD) data to unravel their evolutionary history and tackle the processes driving it. We assess the phylogenetic relationships and population structure within the group and evaluate new species boundaries using multispecies coalescent approaches. Furthermore, we offer insights into the environmental niche model of the group and use said model to guide our sampling efforts and collect and present molecular evidence for the first time of *Dugesia leporii* specimens, endemic from Sardinia last spotted in 1999. Our results indicate that paleoclimatic conditions rather than microplate tectonic dynamics were likely an important driver of diversification for the Corso-Sardinian group. Furthermore, our results warrant the taxonomic re-evaluation of the group as eight primary species candidates are established based on molecular data. Our study also reveals the first case of interspecific natural hybridization reported in Dugesiidae and, to our knowledge, in Tricladida.

Keywords: *Dugesia* flatworms; 3RAD; phylogeny; species delimitation; hybridization; biogeography.

1. Introduction

The understanding of how populations diverge, and species come to be, is one of the fundamental questions in evolutionary biology, and supposes a core link of all disciplines that benefit from such knowledge. With the ever-increasing availability of genome-wide data and the advances in sequencing technologies (e.g., Baird et al., 2008; Bayona-Vásquez et al., 2019; Etter et al., 2011; Peterson et al., 2012; Schweyen et al., 2014) and analytical methods (e.g., Yang and Rannala, 2010), studies on phylogeography, introgression, hybridization, and species delimitation have been revolutionized over the last two decades (Bourgeois and Warren, 2021). As a result, biologists have gained new insights into the processes leading to speciation, such as gene flow, mutation, recombination, and selection, although their interplay remains challenging (Ravinet et al., 2017).

Over the past two decades, freshwater triclads have served as valuable models for understanding the processes contributing to species diversity. In particular, the genus *Dugesia* Girard, 1850 in the Western Mediterranean region has garnered special attention in recent studies for their taxa's potential as models for evolutionary biology, phylogeography and chromosomal speciation (Benítez-Álvarez et al., 2023; Dols-Serrate et al., 2020, 2024; Leria et al., 2022). This research has been rooted in the assumption that the dispersal of freshwater planarians is constrained by their limited vagility (Ball, 1974), attributed to their dependency on the availability of freshwater resources, their lack of protection against desiccation, and their direct mode of development. Some noteworthy studies (e.g. Lázaro et al. 2009, 2013; Solà et al. 2013, 2022; Leria et al. 2022) have drawn interesting conclusions about how geological history, climatic and hydro-environmental variables, and anthropogenic activities have influenced the diversity and current distribution of these organisms, thus contrasting their intrinsic value as models of study.

However, these studies faced two main limitations. Firstly, classical species descriptions based solely on morphological data had underestimated the number of species present in the analyzed regions, which prompted subsequent studies using molecular data to reveal and describe multiple new species using an integrative approach (Leria et al., 2020; Sluys et al., 2013; Solà et al., 2013). Secondly, the markers used thus far often lacked phylogenetic resolving power (Dols-Serrate et al., 2025).

Corsica and Sardinia are two islands of the Western Mediterranean wherein several endemic *Dugesia* species cluster in a monophyletic group known as the Corso-Sardinian clade (see Benítez-Álvarez et al., 2023). This clade most likely originated during the Oligocene-Miocene period, when a microplate including what would later become the Balearic Islands, Corsica, and Sardinia, among other geographical regions, detached from southwestern Europe (Leria et al., 2022). In recent years, multiple phylogenetic studies have strived to uncover the diversity among endemic *Dugesia* populations of the Corso-Sardinian clade. These populations were previously presumed to be a single species distributed across both islands, viz. *Dugesia benazzii* Lepori, 1951 (hereafter referred to as *D. benazzii*

"sensu lato" (s.l.)). This notion would be first challenged in the early 1980s, when a new biotype with a haploid complement of 7 chromosomes (2n = 14) was discovered and described as a new species, viz. *Dugesia hepta* Pala, Casu & Vacca, 1981, restricted to four Sardinian hydrographic basins (Stocchino et al., 2005). Thereafter, a study by Pala et al. (2000) would elevate a southern Sardinian diploid population (n = 8; 2n = 16) to species status based on morphological and karyological data, viz. *Dugesia leporii* Pala, Stocchino, Corso & Casu, 2000. Noteworthy, *D. leporii* has not been recovered in the wild ever since its description.

The peculiar karyotype of *D. hepta*, which is unique among *Dugesia* in the Western Palearctic region, sparked interest in the Corso-Sardinian clade for studying the chromosomal rearrangement as a potential driver of speciation. An initial molecular study by Dols-Serrate et al. (2020) found evidence that warranted a taxonomic re-evaluation of *D. benazzii* s.l. In subsequent studies, Dols-Serrate et al. (2024, 2025) conclusively established the reciprocal monophyly of *D. hepta* and *D. benazzii* s.l., and, at the same time, formalized the description of three new species, viz. *Dugesia hoidi* Dols-Serrate, Stocchino & Riutort, 2024, from Sardinia and only known from its type locality, *Dugesia mariae* Stocchino, Dols-Serrate & Riutort, 2024, and *Dugesia xeropotamica* Stocchino, Dols-Serrate & Riutort, 2025, two Corsican endemics. In addition, a re-description was given for *D. benazzii* "sensu stricto" (s.s., hereafter referred to as *D. benazzii*), which from then on would only encompass the Sardinian populations of former *D. benazzii* s.l., barring *D. hoidi*'s type locality. Furthermore, the initial study by Dols-Serrate et al. (2020) presented molecular data which pointed to the possibility of ongoing natural hybridization between *D. hepta* and *D. benazzii* s.l.. The question of natural hybridization between both species was first posed by Pala et al. (1982) based on karyological data. However, it was dismissed by Benazzi-Lentati and Benazzi (1985), on the basis of karyological data as well. Unfortunately, even though Dols-Serrate et al. (2020) rekindled the idea of natural hybridization between *D. hepta* and *D. benazzii*, with the available markers at the time, the possibility of incomplete lineage sorting (ILS) being the underlying cause for the reported hybridization patterns could not be discarded.

The lack of or limited resolving phylogenetic power of the used nuclear markers (Dols-Serrate et al., 2020, 2024, 2025), as well as mito-nuclear discrepancies, have also limited our ability to disentangle and understand the evolutionary history of the Corso-Sardinian clade. Problems regarding the resolution of the relative phylogenetic position of particular taxa persist, as the actual phylogenetic positions of *D. hoidi* and *D. xeropotamica*. Furthermore, unanswered questions lingered. First, the placement of *D. leporii* remains unknown due to the absolute absence of molecular data for the species. Second, the status of the putative hybrids reported in Dols-Serrate et al. (2020) needed to be confirmed.

In this study, we integrate genomic data obtained through restriction site-associated DNA sequencing (RADseq) and karyologic information from a broad representation of *Dugesia* populations across both islands to: (a) elucidate the evolutionary history of the

Corso-Sardinian clade, and advance towards a robust and resolved phylogeny of the group; (b) explore the number of independently evolving lineages by applying a MSC species delimitation approach; (c) validate the hybridization hypothesis between *D. hepta* and *D. benazzii*; (d) characterize the ploidy of the putative hybrids; (e) and propose a first biogeographical hypothesis for the group.

2. Materials and methods

In the present study, we followed a comprehensive strategy to ensure a representative sampling of the diversity of endemic *Dugesia* species from Corsica and Sardinia. We aimed to sample approximately 10 individuals per locality. To do so, a thorough sampling campaign was carried out (Supplementary Fig. S1 and Table S1), wherein each potential autochthonous *Dugesia* individual was meticulously characterized by obtaining its karyotype and barcoding sequences.

To this end, animals were maintained in the laboratory, and in batches of approximately 12 specimens at a time, they were individually housed in specialized containers to ensure their isolation while their karyotype was determined, and species assignment was established through a phylogenetic analysis of mitochondrial data. Following this initial characterization, high-quality genomic DNA was extracted from selected individuals of each batch and preserved in a frozen state until the preparation of the RADseq library.

Samples for RADseq analyses were selected based on these criteria: samples had to belong to distinct and different monophyletic groups obtained in the mitochondrial based phylogenetic analysis, ensuring a comprehensive representation for the selected localities across Corsica and Sardinia. To mitigate sampling biases, we refrained from including localities that were close and were part of the same monophyletic group. Additionally, we tried to include differing karyologies whenever they occurred within the same monophyletic group.

2.1. Environmental niche modelling to estimate potential distribution

Historically, studies focused on the Corso-Sardinian clade have favoured northern Sardinian localities. While *D. hepta* is known to have a restricted distribution range comprising four hydrographic basins in the northwestern part of the island (Pala et al., 1981; Stocchino, 2018; Stocchino et al., 2005), the expansion of *D. benazzii* s.l. southwards is less explored in comparison. The elusive *D. leporii* constitutes the southernmost endemic species of Sardinia. However, its range is unknown given how the only specimens ever reported were from the type locality (Pala et al., 2000), and posterior efforts to sample the species have been fruitless. In a new attempt to (i) sample new localities from southern Sardinia and (ii) collect specimens of *D. leporii*, we modeled the potential distribution

of the *D. benazzii* s.l. using the software Maxent 3.4.0 (Phillips et al., 2017; Phillips and Dudík, 2008) performing 100 bootstrap replicates. Distribution data of 15 populations of *D. benazzii* s.l. published in Dols-Serrate et al. (2020) (Supplementary Table S2) and 5 hydro-environmental variables extracted from the HydroATLAS database (Linke et al., 2019) with a spatial resolution of 15 or 30 arc-seconds were used. Selected variables were; *actual evapotranspiration* (aet_mm_cyr), *air temperature* (tmp_dc_cyr), *natural discharge* (dis_m3_cyr), *precipitation* (pre_mm_cyr), and *land cover classes* (glc_cl_cmj) (see RiverATLAS Catalog in HydroATLAS database for detailed information), as they have been shown to correlate with the environmental niche of *Dugesia* species (Leria et al., 2022).

2.2. Taxon sampling

A thorough sampling campaign was carried out between 2019 and 2021 to collect specimens of the autochthonous *Dugesia* species from Corsica and Sardinia (Supplementary Fig. S1). A total of 839 individuals (Supplementary Table S1) were collected from 33 localities, which collectively represent almost the entire known distribution range of Corso-Sardinian clade. From each locality, hereafter referred to as populations, some of the individuals were fixed in absolute ethanol and stored at -20 °C for future genomic analyses, whilst others were kept alive and transported to the laboratory for karyological analysis before proceeding with subsequent DNA extractions. In the present study, we processed for RADseq library preparation a sample of 90 selected specimens from 17 populations across both islands (Fig. 1, Supplementary Table S3).

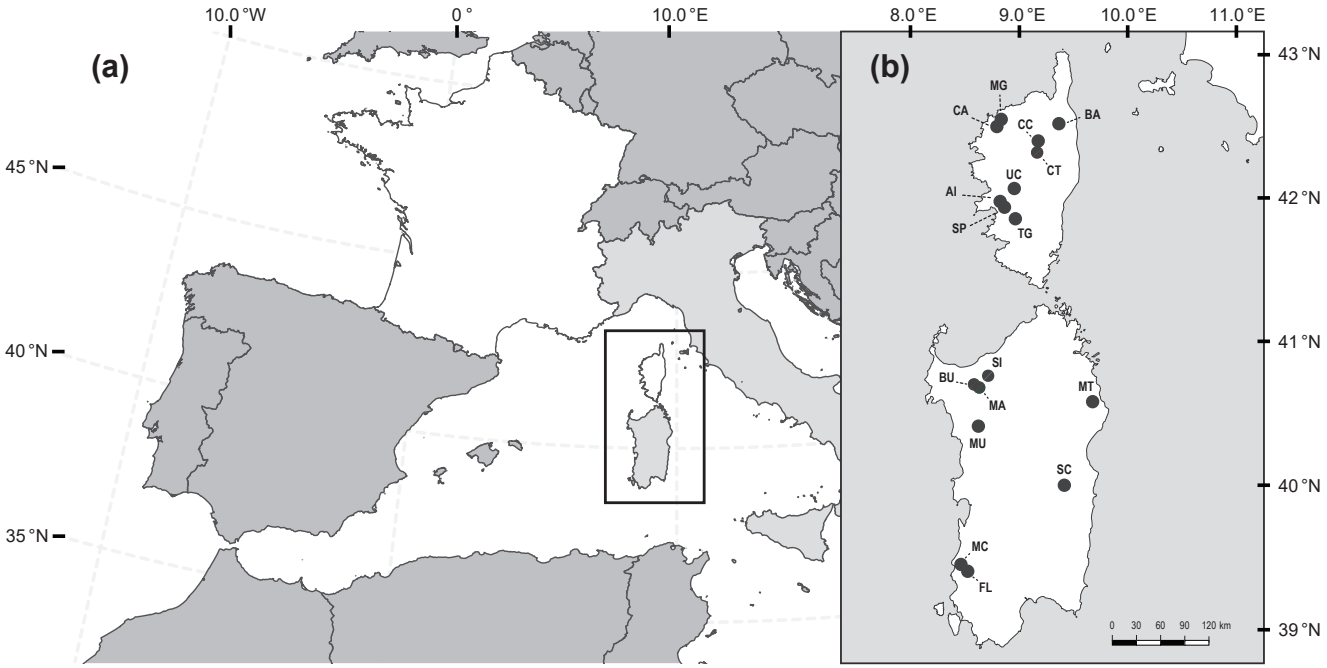


Figure 1. (a) Map of the Western Mediterranean region. Rectangular inset shows Corsica (white) and Sardinia (light gray), Tyrrhenian Islands where samples were collected for the present study. (b) Enlargement of Corsica and Sardinia showing the geographic distribution of sampling localities with their pertinent acronyms.

2.3. Animal rearing in the laboratory

Collected specimens were initially kept in conventional glasses filled with 1× Montjuïc water (1.6 mM NaCl, 1.0 mM $CaCl_2$, 1.0 mM $MgSO_4$, 0.1 mM $MgCl_2$, 0.1 mM KCl, and 1.2 mM $NaHCO_3$ in DI H2O, pH 6.9–8.1) at a constant temperature of 14 ºC. Each glass accommodated 10 to 15 individuals sampled from the same locality. Following karyological characterization, specimens were isolated in separate 5 mL polypropylene tubes (401100, Deltalab). All specimens were fed ecological chicken liver on a weekly basis and were subjected to a one-week starvation period before DNA extraction.

2.4. Karyological data

Chromosomal metaphasic plates were obtained by cutting either the tail or the head of individuals and letting the tissue regenerate for 7 h embedded in a 0.25% colchicine solution. Afterwards, the severed fragments were placed in a Petri dish, punctured with a needle, and incubated in distilled water for 20 min at room temperature (RT). After incubation, each sample was carefully placed into an Eppendorf tube using forceps and fixed with freshly prepared Carnoy's fixative (3:1 dilution of methanol:glacial acetic acid), and incubated for 5 min. Subsequently, each sample was incubated for 20 min in a 50% glacial acetic acid solution in a separate Eppendorf, and then placed onto a glass slide using a Pasteur pipette. Blastema and post-blastema were minced using a razor blade and suspended in 80-100 µl of the glacial acetic acid solution. The macerated cell suspension was dropped onto a 65º C preheated glass slide (3-4 drops per slide) and left to air-dry. Thereafter, karyological preparations were stained with 1:20 Giemsa:tap water mixture for 1 min, air-dried, and sealed under a cover slip using DPX.

2.5. Sanger sequencing, gene tree, and molecular species identification

New samples were assigned to a species based on their placement in a mitochondrial phylogenetic gene tree in which data from previous studies was included (all samples used in Supplementary Table S4), as inferred using the markers proposed by Dols-Serrate et al. (2024). Upon completion of the genomic DNA's extraction, and prior to the RAD library preparation, the following gene fragments were amplified: (a) a fragment of approximately 650 bp of the cytochrome c oxidase I (*COI*); (b) a sequence of approximately 630 bp containing a fragment of the NADH subunit I (*NADH1*), the tryptophan tRNA, and a fragment of the cytochrome c oxidase II (*COII*). Polymerase chain reaction (PCR) and primer specifics are given in Supplementary Tables S5 and S6. The thermal cycle for all primers was: (i) 94.0 ºC for 2 min; (ii) 39 cycles encompassing a 94.0 ºC denaturation step for 45 sec., a 50 sec. annealing step, and a 1 min elongation step at 72.0 ºC; followed by (iii) a final elongation step at 72.0 ºC for 3 min. PCR amplification products were purified using a vaccum system (MultiScreen™HTS Vacuum Manifold of Millipore). Sequencing

reactions were performed using Big Dye (3.1, Applied Biosystems) and ran in an automated sequencer ABI Prism 3730 (Unitat de Genòmica dels Serveis Científico-Tècnics de la Universitat de Barcelona) or at Macrogen Inc. (Madrid). Raw reads were assembled into contigs, and established consensus sequences using Geneious R10 (Biomatters, http://www.geneious.com/ last accessed January 2023). Subsequently, we constructed sequence alignments, incorporating the newly assembled consensus sequences and sequences from the mitochondrial dataset by Dols-Serrate et al. (2025). Default parameters were used for MAFFT v7 (Katoh and Standley, 2013), as implemented in the online version. To assign all individuals to a species, a gene tree was inferred using IQ-Tree 2.0 (Minh et al., 2020; Nguyen et al., 2015). The best substitution model was estimated using ModelFinder (Kalyaanamoorthy et al., 2017), as implemented in IQ-Tree 2.0. Nodal supports were computed using an ultrafast bootstrap (UFboot) approximation (Hoang et al., 2018) and the SH-aLRT test (Guindon et al., 2010) with 10,000 replicates each. Sequences of *D. subtentaculata* (Draparnaud, 1801) were used as outgroup. The inferred tree was visualized using FigTree v.1.4.4 (https://github.com/rambaut/figtree/releases last accessed September 2023).

In this study, 59 of the presently studied specimens were newly identified using the aforementioned procedure (see Supplementary Fig. S1), since the remainder had already been featured and identified in Dols-Serrate et al. (2025). In cases where obtaining mitochondrial sequences was not possible, we instead amplified and sequenced the internal transcribed spacer 1 (*ITS-1*) as described in Dols-Serrate et al. (2020).

2.6. DNA extraction and RADseq library preparation

High-quality genomic DNA was extracted from a tissue fragment of each the 90 selected animals (Supplementary Table S3) performing an overnight digestion step at 37 ºC with proteinase K (5 µl/sample; 20 mM), followed by purification with the Easy-DNA gDNA Purification Kit (K1800-01, ThermoFisher), following the manufacturer's instructions. Quality and quantity of total gDNA was examined using a Qubit 4 Fluorometer (Invitrogen). Sequencing RADseq libraries were prepared following the procedures of Adapterama III (Bayona-Vásquez et al., 2019) with little modifications, and DNA was digested using *EcoRI* and *NheI* as cutters. After digestion, Illumina adapters were ligated onto the DNA fragments and individually barcoded samples were pooled. Indexed samples were subjected to PCR for 12 cycles of 98.0 ºC for 20 sec., 60 ºC for 15 sec., and 72 ºC for 30 sec., preceded by a 98 ºC step for 1 min, and finished with a 72 ºC step for 5 min. Library was size selected (between 400 and 600 bp) by running a 1% agarose gel, recovered using a QIAquick® Gel Extraction Kit (28704, Qiagen), and sequenced in an Illumina NextSeq v2 paired-end (PE) platform (300 cycles, mid-output) using 150 bp PE sequencing. Sequencing was carried out at the UCLA Technology Center for Genomics & Bioinformatics.

2.7. RADseq data processing, de novo assembly, and parametrization

The quality of the sequenced RAD reads was checked using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ last accessed October 2023). RADseq data were processed using STACKS v2.52 (Rochette et al., 2019). Paired-end raw reads were quality-filtered, demultiplexed, and truncated to a length of 135 bp using process_radtags. Subsequently, a de novo pipeline was used to assemble RAD loci using the *denovo_map.pl* wrapper. All samples with depth of coverage below 10x were discarded.

To assess the robustness of the results, three sets of loci were assembled. The first set was assembled following a method for optimizing de novo assembly parameters using STACKS (Paris et al., 2017; Rochette and Catchen, 2017), which consists of varying separately each of the three key parameters using *denovo_map.pl* (-m: minimum stack depth, -M: within-individual number of mismatches between stacks, and -n: number of mismatches allowed to merge putative loci of the catalog) to determine the values at which the tendency of new polymorphic loci becomes linear or decreases when increasing *m*, *M* or *n* (optimal set, OP; m2, M1, n2). The second set was assembled using the default parameters (default set, DF; m3, M2, n1), and the third set was assembled using higher M and n values (over-merged set, OVM; m3, M5, n6). We opted for following this approach for several reasons. First, mosaicism, a genetic consequence of the prevalence of asexuality (viz. mosaic-Meselson effect; see Leria et al., 2019), has been described in *Dugesia* species and reported as well for *D. hepta* and the species of the *benazzii* clade (Dols-Serrate et al., 2020). Second, aneuploidy and polyploidy are recurrent in most populations of asexually reproducing or facultative fissiparous individuals, a condition that may add noise when calling variants and exacerbate the mosaicism. Such type of specimens is present in this study. Third, correctly setting *M* and *n* requires a balance between not collapsing alleles of the same locus or failing to find homologous loci in different samples (low settings) and risking the merger of alleles or loci into falsely polymorphic loci (high settings). The optimal set had overall low settings, notably for *M*, and we had concerns regarding such approach when dealing with the problematics mentioned above.

Once the three loci catalogs were constructed, contaminants were filtered by mapping assembled loci against the nucleotide database of the NCBI using BLAST+ *blastn* (2.12.0) (Altschul et al., 1997; Camacho et al., 2009) and Blobtools v.3.4.0, as implemented in BlobToolKit (Challis et al., 2020; Laetsch and Blaxter, 2017), to create respective blacklists for downstream analyses. Only hits against Platyhelminthes and no-hit were retained (given the recurring lack of information on Platyhelminthes in databases). We removed potential paralogs by identifying loci that possessed more than two haplotypes within a single sample. The module *populations* was used to export data in various formats for downstream analyses, requiring a minimum allele frequency (MAF) of 0.04, to filter out singleton SNPs, and a maximum observed heterozygosity of 0.99, to remove potential remaining paralogous loci (Stobie et al., 2019) and process a SNP. For SNP-based analyses, a single

random SNP per loci was retained (--write-snp-random) and VCF file were filtered with VCFtools v.0.1.16 (Danecek et al., 2011) to include only biallelic SNPs.

2.8. Dataset assembly

We constructed different datasets by processing non-contaminant loci for each loci catalog utilizing populations. We used different criteria to sort and retain loci, such as the minimum percentage of individuals in a population sharing a locus (-r) and the minimum number of populations presenting it (-p) required to be processed. Thereafter, we built whitelists of polymorphic loci that met our intra- and inter-population prevalence requirements and used them to run *populations* anew to output the dataset for each intended analysis. For some analyses we used different subsets of populations and/or individuals and in such cases, we processed loci anew. All datasets build for the present study, and which analyses they were used for, are shown in Supplementary Table S7.

2.9. Phylogenetic analyses

We used non-partitioned concatenated RADseq datasets (--phylip-var-all) to infer the phylogenetic relationships among studied taxa, as this approach has been shown to be more accurate for this type of phylogenomic dataset (Molloy and Warnow, 2018). To explore the phylogenetic resolution of our data, multiple phylogenetic alignments were prepared, each with varying amounts of missing data distributed within and across populations for each loci catalog. Furthermore, we used two different sets of samples, one including all individuals (n = 63) and another focused on the Corsican samples (n = 29) (see Supplementary Tables S7 and S8). Loci were retained based on their presence in at least 40%, 60%, and 70% of the individuals of any given population (-r), and concurrently present in 50%, 75% and 100% of populations (-p) resulting in a total of 27 datasets per sample set (Supplementary Table S7).

Maximum-likelihood (ML) phylogenies were constructed using IQ-Tree 2.0 (Minh et al., 2020; Nguyen et al., 2015), with 10,000 replicates of UFboot and SH-aLRT to calculate node supports, and estimating the best-fit evolutionary model through the ModelFinder Plus (-m MFP) option. Specimens of *D. hepta* were used as outgroup and treated as a single population.

In addition, we inferred species trees with matrices of unlinked SNPs of each dataset (viz. optimal, default, and over-merged) using a multispecies coalescent method: the quartet-based method SVDQuartets (Chifman and Kubatko, 2014) as implemented in PAUP* v.4 (Swofford, 2003). We used a single SNP per loci that was present in at least 70% of the individuals within a population and in at least 50% of the populations (OP-0.7P50, DF-0.7P50, OVM-0.7P50), but excluding population 27 (MC). For each matrix, we evaluated all possible quartets conducting 100 bootstrap replicates and summarized the results in a 50% majority-rule consensus tree.

2.10. Topology tests

While estimating ML trees using the optimal set, a consistent topological conflict involving individuals from two Corsican populations (viz. CC and CT) emerged. We performed topology tests based on two phylogenetic hypotheses regarding the sister group relationship of these populations to BA, CA, and MG, potentially forming a northern clade, and to TG, SP, UC, and AI, populations regarded as meridional. To do so, we conducted constrained tree searches (-g) using a subsample that only included Corsican specimens (n = 25) and a Sardinian outgroup (population MT, n = 4). In total, 27 datasets were assembled and analyzed (Supplementary Table S7, see *Topology test column*). First, using IQ-Tree 1.6.4 (Nguyen et al., 2015), we enforced the monophyly of CC and CT with the northern populations (topology-N) in cases where the unconstrained topology grouped CC and CT with southern Corsican populations (topology-S; datasets OP-0.4CP50, OP-0.6CP50, and OP-0.7CP50). Alternatively, we enforced topology-S for all remaining cases, were topology-N was recovered in the unconstrained state. Then, log-likelihoods were computed for the constrained trees (-z) and evaluated using 10,000 replicates of the RELL approximation (Kishino et al., 1990), as implemented in IQ-Tree 1.6.4, and the approximately unbiased (AU) test (Shimodaira, 2002).

2.11. Population genetic structure

We explored the genomic variation among and within taxa using SNPs from loci that were present in at least 70% of the samples within a population (r = 0.7) and in at least 50% of the populations (Supplementary Table S7, OP-0.7P50, DF-0.7P50, and OVM-0.7P50). Resulting VCF files contained a single random SNP per retained loci to minimize the impact of linkage in our data.

We computed eigenvalues and eigenvectors from the datasets mentioned above using PLINK 1.9, and performed Principal Component Analysis (PCA), as implemented in the R package PCAtools (Blighe and Lun, 2023), to elucidate the spatial patterns of genetic admixture in our samples at different hierarchical levels (Supplementary Table S7).

Additionally, we performed Bayesian model-based clustering analyses to characterize individuals using STRUCTURE v.2.3.4 (Pritchard et al., 2000), as implemented in structure_threader (Pina-Martins et al., 2017), under the admixture and the correlated allele frequency models. We tested values for K = 1-8 and performed 10 independent runs per each K value running 100,000 Monte Carlo Markov Chains (MCMC) iterations following a previous burn-in step of 100,000 MCMC iterations. Optimal K was determined using the ΔK method (Evanno et al., 2005), however, additional K values were assessed. To examine finer levels of genetic structure, hierarchical analyses were performed.

We used fineRADstructure v.0.3.2 (Malinsky et al., 2018) to infer shared ancestry among all individuals. Since our loci were unmapped, as per the developer's recommenda-

tion, we used the script *sampleLD.R* to reorder our loci according to linkage disequilibrium (LD) using the recommended default parameters. Coancestry matrices were inferred using RADpainter, and fineSTRUCTURE's clustering algorithm was used to assign individuals into clusters, running 100,000 MCMC iterations sampled every 1,000 iterations (following a 100,000 iterations burn-in period). We used *fineRADstructurePlot.R* to visualize the results (available at https://github.com/millanek/fineRADstructure, last accessed November 2023).

2.12. Hybrid detection using INTROGRESS and ancestry-painting

We identified potential hybrids in our data set on the basis of criteria proposed by Dols-Serrate et al. (2020) and preliminary results reported in the present study (see Results). We evaluated the relationships between individual heterozygosity (H_{ind}) and genome wide mean ancestry (hybrid index: *HI*) for each specimen of the candidate parental species-hybrid trios. We estimated both H_{ind} and *HI* and represented them in a triangle plot using INTROGRESS version 1.2.3 (Gompert and Buerkle, 2010). In this analysis, we selected SNPs that were fixed differentially between parental species, and present in the pertinent hybrids, although not necessarily in all of them. Additionally, we generated ancestry-painting plots for the same evaluated trios based on fixed SNPs between candidate parental species that were present in all potential hybrids as well (completeness = 1.0). Scripts used to generate the latter plots are available at https://github.com/millanek/tutorials/tree/master/analysis_of_introgression_with_snp_data (last accessed, December 2023).

For these analyses, we used a random SNP per loci that were present in at least 70% of the samples within a population, and in at least 50% of the populations of a subsample consisting of the specimens in the aforementioned trios (Supplementary Table S7; OP-HxBP50, DF-HxBP50, OVM-HxBP50, OP-LxBP50, DF-LxBP50, OVM-LxBP50).

2.13. Testing for introgression

To further explore tree discordances and the temporal framework/extend of hybridization/introgression, we performed ABBA-BABA tests for all taxa and populations quartets involving *D. benazzii* s.s., FL, and MC. We assigned a subsample (n = 5) of Corsican individuals from populations 1 and 3 (AI, UC) to the outgroup. Calculations were performed using Dsuite Dtrios (Malinsky et al., 2021) with a SNP dataset that included loci present in at least 70% of the samples within a population, and in at least 50% of the populations of a subsample (OP-ABBAP50, DF-ABBAP50, OVM-ABBAP50; Supplementary Table S7).

2.14. MSC species delimitation

We determined the number of independently evolving lineages using two coalescent-based species delimitation approaches, viz. BPP v.4.0 (Flouri et al., 2018; Yang and Rannala, 2010) and BFD* (Leaché et al., 2014). For these analyses, we excluded samples from MC, as hybridization would represent a violation of the multispecies coalescent model (see Results), and *D. hepta*, the outgroup. BPP was run using model A10, which performs a comparison of species delimitation models induced on a given fixed guide tree. To ensure computational tractability, we performed BPP analyses using a subset of 500 randomly selected loci from datasets OP-0.7P50, DF-0.7P50, OVM-0.7P50 (Supplementary Table S7), which has been shown to provide sufficient resolving power (Tonzo et al., 2019). To test the impact of different evolutionary and demographic scenarios, we followed Yang (2015) approach and used different inverse-gamma distributed priors, IG(α, β), for the population size (θ) and the species divergence time (τ). We used a fixed diffused α = 3, and worked by adjusting β. The following scenarios were tested: (a) relatively small ancestral population size θ_1 ~ IG(3, 0.002) and shallow divergence time τ_1 ~ IG(3, 0.002); (b) relatively large ancestral population size θ_2 ~ IG(3, 0.1) and shallow divergence time τ_2 ~ IG(3, 0.002); (c) relatively large ancestral population size θ_3 ~ IG(3, 0.1) and deep divergence time τ_3 ~ IG(3, 0.1); and (d) relatively small ancestral population size θ 4 ~ IG(3, 0.002) and deep divergence time τ_4 ~ IG(3, 0.1). All analyses were performed twice to assess the robustness of results using the two reversible-jump MCMC algorithms (rjMCMC) available and starting from random seeds: viz. algorithm 0 (with fine-tunning prior ε = 2) and algorithm 1 (with fine-tunning priors α = 2 and m = 1). rjMCMC chains run for 100,000 generations with a sample frequency of two generations and a burn-in of 50,000 generations.

For the BFD* analyses, we extracted a single random SNP per loci and built a matrix per dataset (OP-bfdP50, DF-bfdP50, OVM-bfdP50; Supplementary Table S7) filtering SNPs by the proportion of missing data using VCFtools v.0.1.16 (--max-missing = 0.95). Consequently, we proposed six competing species delimitation hypotheses (SDH) based on our results, and compared them using pairwise Bayes factors comparisons (hereafter referred to as *2lnBf*; Kass and Raftery, 1995). To do so, we calculated marginal likelihood estimates (MLE) using SNAPP v.1.6.1 (Bryant et al., 2012). We used the *vcf2snapp.py* script (available at https://github.com/CoBiG2/RAD_Tools/blob/master/vcf2snapp.py, last accessed December 2023) to format the input data for SNAPP. For MLE calculation, we performed path-sampling analyses with 40 steps for 100,000 iterations after 12,000 burn-in iterations and setting alpha to 0.3.

2.15. Divergence time estimation

To estimate divergence times, we used a MSC approach as implemented in SNAPP v.1.6.1. We used a single random SNP per locus from datasets OP-0.7P50, DF-0.7P50, and OVM-0.7P50, and excluded sites that presented more than 15% of missing data

(--max-missing 0.85) and an allele count of 1 (--mac 1) with VCFtools v.0.1.16. We downsampled the resulting VCF files (--keep) to not include samples from MC and only 2-3 individuals per PSC. We specified an age constraint on the root using a normal distribution with a mean of 5.7 Mya and a standard deviation (SD) of 1.45. SD values were calculated to fit the 95% high posterior density (HPD) interval of the Corso-Sardinian clade basal node time estimate, which was calculated by Solà et al. (2022). We conducted two replicate runs of 1,500,00 MCMC iterations each after 150,000 burn-in iterations.

3. Results

3.1. Guided sampling using environmental data

The environmental niche models for *D. benazzii* s.l. on Sardinia yielded AUC values higher than 0.7 for both training and testing data (Supplementary Table S9), indicating a relatively high predictability of the species distribution model. The environmental variables' contribution to the model was heavily influenced by the evapotranspiration and natural discharge variables (Supplementary Table S10). Suitable environmental conditions for *D. benazzii* s.l. positively correlate with low natural water discharge (annual average < 200 m3 per second), moderate evapotranspiration and soil water deficit (annual average > 600 mm), temperate temperatures (annual average of 14-15º C), and a relatively high precipitation regime (annual average > 900 mm). Habitat suitability of *D. benazzii* s.l. was at its highest (0.77-0.92) across septentrional Sardinia, notably in tributaries of the Temo River, Tirso River, and Mannu di Ozieri River. Areas that, traditionally, have accumulated sampling efforts. However, two relatively high suitability areas (0.69-0.82) were inferred in the Gennargentu Massif (eastern Sardinia), and a mountainous area located at the west coast of the Province of South Sardinia. In consequence, we visited for the first time six sampling sites distributed across both areas (loc. 22 to 27, Supplementary Fig. S1). Among these, four sites were found to harbor freshwater flatworms, with *Dugesia* specimens present in three of them (Supplementary Table S1).

3.2. Identification of individuals and putative *hepta* × *benazzii* hybrids based on a mitochondrial gene tree

We used Dols-Serrate et al. (2025) mitochondrial dataset, along with sequences of individuals from Dols-Serrate et al. (2020) identified as putative hybrids, to preliminarily assign all presently studied individuals to recognized species of the Corso-Sardinian clade. Following this approach, we aimed to avoid the potential inclusion of individuals from the allochthonous species *Dugesia sicula* Lepori, 1948, while recognizing putative *D. hepta* × *D. benazzii* hybrids. Based on the mitochondrial tree (Supplementary Fig. S3), 33 samples were assigned to *D. mariae*, 12 samples to *D. xeropotamica*, 27 samples to *D. benazzii*, 9 to *D. hepta*, and 9 individuals collected across populations 15-17 (SI, BU, and MA) as potential hybrids between *D. hepta* and *D. benazzii* (Supplementary Fig. S3).

3.3. *Dugesia leporii*, a welcomed surprise

When sampling population 22 (FL, Fig. 1; Supplementary Fig. S1), individuals with an external morphology seemingly matching the pattern of *D. leporii*'s dorsal surface were collected (viz., uniform dark brown dorsal surface with a thin clear peripheral line). To our surprise and delight, careful morphological examination of the anatomy of said individuals confirmed their affiliation with *D. leporii* (G. A. Stocchino, pers. comm.).

3.4. Ploidy of populations

We obtained karyological data for 72 individuals out of the total 90 (Supplementary Table S3). All outgroup *D. hepta* individuals were diploid with a haploid complement of n = 7 and were identified from 4 different Sardinian populations (i.e. BU, MA, MU, and SI). Regarding ingroup samples, results indicated that all examined specimens from Sardinian populations FL, SI, and Corsican populations CA, MG, and SP were diploid with a haploid complement of n = 8 (2n = 16). Only three individuals from three different populations were found to be triploid with haploid complement n = 8 (populations BU, CC, and MU; Supplementary Table S3). Corsican populations BA and CT were found to harbor diploid and aneuploid individuals, with haploid complement n = 8 as well. Aneuploids were present in BU, MA, SC, AI, BA, CC, CT, TG, and UC, and showed high intraindividual variability of the chromosome number. Recounts of the chromosomal metaphasic plates of any given aneuploid individual often yielded a number between 28 and 32 chromosomes, although occasionally the range of variability extended from 16 to more than 32 chromosomes. Finally, four individuals from MU, along with all individuals from MC, were identified as potentially mixoploids, exhibiting a combination of seemingly triploid and diploid chromosomal sets (Supplementary Fig. S2). These samples also displayed intraindividual chromosomal number variability, as chromosomal plates often yielded a number of chromosomes of either about 16 or 24.

3.5. RADseq data

We produced RADseq data from 90 individuals (Supplementary Table S3). Illumina sequencing yielded 376.8 million sequence reads. After demultiplexing and filtering for quality, we retained 83.1% of the reads, and obtained on average ~4.2 million raw reads per sample. Based on Paris et al. (2017) approach, we estimated the optimal parameters for de novo loci assembly (m2 M1 n2) and, afterwards, discarded those samples with a depth of coverage below 10× to minimize the impact of unreliable genotype calling (Rochette and Catchen, 2017). As a result, 63 individuals were chosen for subsequent downstream analyses and loci were assembled for the optimal, default, and over-merged sets based on these specimens. The mean depth per locus across individuals averaged 17.13 ± 6.64 for the optimal set, 23.02 ± 8.61 for the default set, and 26.09 ± 8.28 for the over-merged set.

Following contaminant filtering, the final datasets comprised 691,471 loci (39.42% singletons) for the optimal set, 631,815 loci (42.63% singletons) for the default set, and 257,282 loci (33.30% singletons) for the over-merged set.

3.6. Phylogenetic analyses

Phylogenetic trees reconstructed using IQ-Tree 2.0, when including all samples (n = 63), yielded, overall, similar topological results independent of the data set used (Fig. 2; Supplementary Fig. S4). All trees had the following traits in common: (i) most individuals were clustered in monophyletic groups by their respective populations, except for Corsican individuals from population 6 (CT in Fig. 1; Supplementary Table S1), and Sardinian specimens from BU, MA, and SI; (ii) a novel, not previously reported, clade comprised exclusively of southern Sardinian populations (*D. leporii* and MC) is recovered, generally, as the sister group to *D. benazzii*; (iii) *D. mariae* is not recovered as a monophyletic taxon; (iv) *D. xeropotamica* is comprised of two clades separated by long branches, which was not reported in Dols-Serrate et al. (2025); (v) no evidence for the putative hybrid clade described in Dols-Serrate et al. (2020). Concurrently, individual MU1394-09, which had been identified as *D. hepta* based on mitochondrial data (Supplementary Fig. S3), was unequivocally grouped with *D. benazzii* samples. Its karyological reconstruction (Supplementary Table S3) supported its affiliation to *D. benazzii* as well, indicating that the initial identification was attributable to a cross-contamination during legacy Sanger loci amplification.

Clear topological differences arise when comparing trees inferred using loci present in at least 50% to 75% of the populations (P50, P75) and loci present in all populations (P100) (Fig. 2d; Supplementary Fig. S4). Trees inferred using P50-75 loci, regardless of the loci prevalence within a population (0.4, 0.6 or 0.7) and, generally, the loci set (OP, DF, or OVM), yielded two well-supported monophyletic groups encompassing Sardinian populations and Corsican populations, respectively. However, trees inferred with over-merged P75 loci supported *D. leporii* and MC as sister group to the remaining Sardinian and Corsican populations (3 out of 18 trees). Moreover, there are minor topological differences concerning Corsican populations. Default and over-merged P50-P75 datasets yielded two subclades ascribable to northern (CA, MG, BA, CC, and CT) and southern (AI, SP, TG, and UC) populations, suggesting a potential biogeographic pattern. However, optimal P50-P75 datasets showed divisiveness in supporting this pattern, suggesting that CC and CT could be ascribed to a clade containing southern populations as well. As of note, support for the node determining CC and CT relative position in the tree showed discordances between SH-aLRT test and UFboot supports, suggesting topological conflict.

With a number of loci two to three orders of magnitude lower than the datasets mentioned above, trees inferred with P100 loci datasets often recovered *D. leporii* and MC as the sister group to the remaining ingroup (8 out of 9 trees) but lost the monophyly of

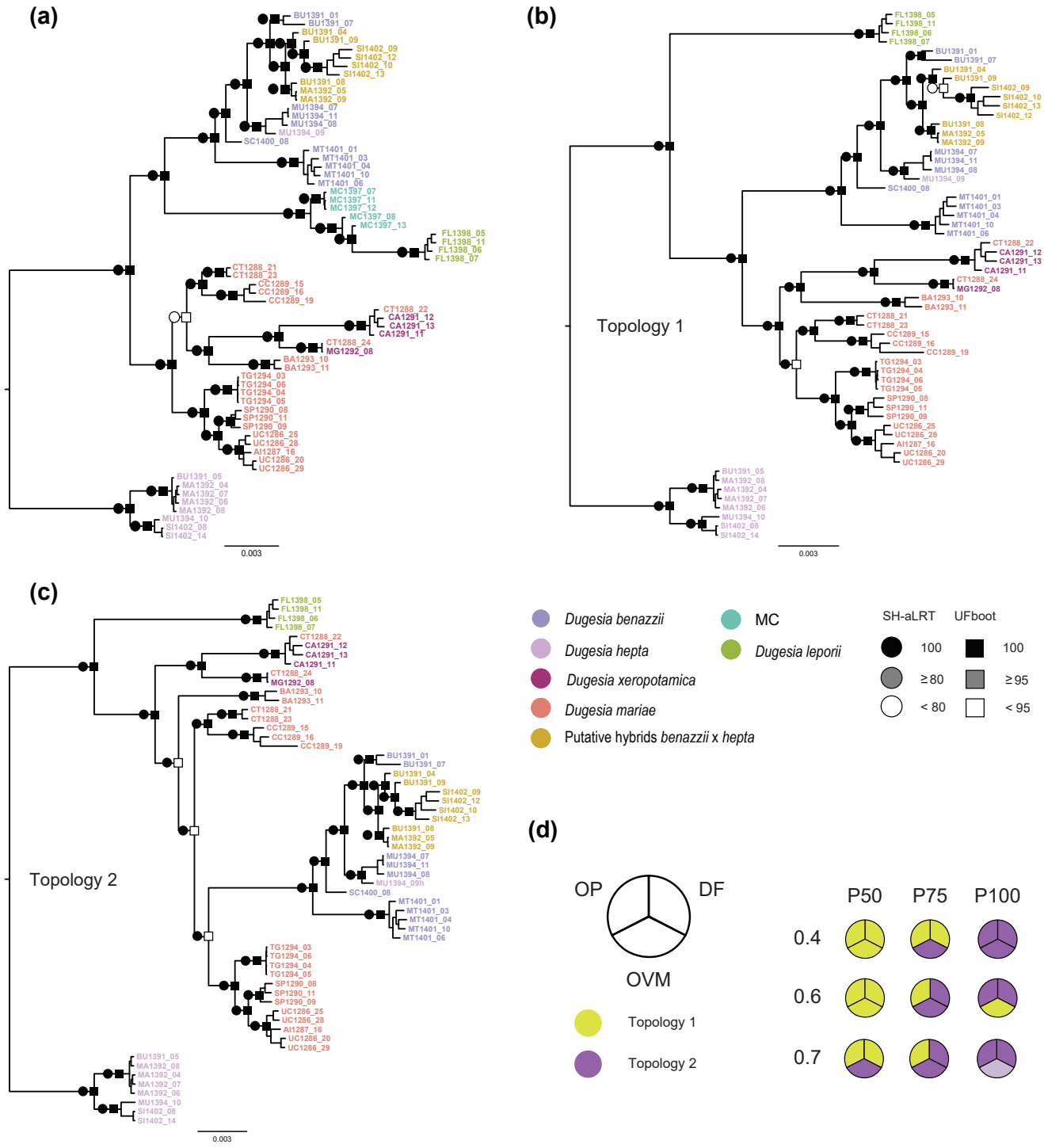


Figure 2. Summary of phylogenetic results and topological alterations when including or excluding samples from population MC, which is highlighted in teal. Samples are color coded based on mitochondrial (see Supplementary Fig. S3) and morphological species assignation. (a) Phylogenetic tree estimated with OVM-0.7P50 loci representing the majority topology recovered when performing phylogenetic inference including all samples (n = 63). (b-c) Alternate topological results when phylogenetic inference is performed excluding samples from MC (n = 58), here represented by trees inferred using OVM-0.4P50 loci (b, topology 1) and OVM-0.7P50 loci (c, topology 2). (d) Schematic representation of supported tree topologies dependent on the taxon coverage (-r and -p) and assembly parameters used (OP, DF, OVM). In light lilac: abnormal topology (see Supplementary Fig. S4).

the Corsican populations' clade. The disarrangement of subclades is pervasive in almost all P100 trees, especially affecting the northern Corsican populations, yielding short internodal branches and low supports. However, monophyly of *D. benazzii* is consistently recovered across trees, with full support (100/100), and always highlighting MT as the sister group to the remaining *D. benazzii* populations. This last aspect is common for all trees (P50, P75, and P100), regardless of the loci sets used to infer them.

To further explore the conflicting phylogenetic relationships within Corsica, phylogenetic trees were reconstructed using loci selected from a subsample (n = 29) encompassing all Corsican specimens (n = 25) and MT as outgroup (Supplementary Tables S7 and S8, and Fig. S5). All inferred trees (27 in total) supported *D. mariae* as non-monophyletic, suggested the presence of two diverging lineages in *D. xeropotamica*, and consistently recovered a monophyletic group comprised of all southern Corsican populations with full support (100/100). Regarding a potential northern clade, most trees (24 out of 27) recovered said clade with high support. However, trees inferred with the optimal set and loci present in at least 50% of the subsample Corsican populations (OP-CP50) yielded a clade that grouped CC and CT with the southern populations (topology-S). Support for this group decreased with the increase of −r (i.e. 0.4, 0.6, and 0.7). However, barring OP-CP50 trees, the impact of filtering loci based on different intra-population prevalence criteria was hardly noticeable for the analyzed subsample (Supplementary Fig. S5).

Considering the genetic clustering results (see below), we explored the possible changes in topology by excluding the samples of MC. We inferred new trees using the same loci employed in the analyses with all populations (Supplementary Table S7). In all, these trees had in common most of the features mentioned above. As of note, almost all obtained trees (26 out of 27) recovered *D. leporii* as sister group to the remaining ingroup populations, similarly to the results for OVM-P75 loci when using all populations (Supplementary Fig. S4c). However, topological discordances were evident between analyses when using different sets of loci (OP, DF, and OVM) and differing amounts of missing data. Notably, the results supported two topologies (Fig. 2b,c; Supplementary Fig. S6). On one hand, *D. leporii* as sister taxon to a clade formed by two distinct subclades comprising all remaining Sardinian populations and all Corsican populations, respectively (topology 1, 13 out of 27 trees), while the alternative topology recovered *D. leporii* as sister group to a clade comprising *D. xeropotamica*, *D. mariae*, and *D. benazzii*, with short internodes, where Corsican populations were not grouped monophyletically (topology 2, 14 out of 27 trees). Topology 1 was often recovered when using P50 loci (8 out of 9 trees), whereas P75-100 loci often yielded topology 2 (13 out of 21 trees) (Fig. 2d; Supplementary Fig. S6).

In addition, we used a MSC method, SVDQuartets, for inferring phylogenetic species trees based on SNPs of 0.7P50 loci, and excluded MC from the analyses, to evaluate the relative phylogenetic position of *D. leporii*. Phylogenetic results based on the quartet-based approach recovered discordant topologies dependent on the set of loci used (Supplementary Fig. S7). Inferred trees using OP and DF-0.7P50 loci suggested *D. leporii* was sister to Corsican populations, a result not obtained in any previous tree, while the tree based on OVM-0.7P50 loci supported *D. leporii* as sister group to the remaining ingroup populations.

3.7. Topology tests results

Phylogenetic results highlighted topological discrepancies involving certain Corsican populations, viz. CC and CT, dependent on the dataset used. Two topologies were considered: i) topology-S, yielding a southern clade comprising AI, SP, TG, UC, CC, and CT, and supported by OP-CP50 loci; ii) topology-N, yielding a northern clade comprising BA, CA, MG, CC, and CT, and supported by the remaining loci datasets. We performed topology tests by enforcing either topology-N or –S, and then used them as the alternative topology of their respective complementary for each tested dataset (i.e. if, for a given dataset, topology-N was the unconstrained topology, we tested topology-S as the alternative hypothesis, and vice versa). Results shown in Supplementary Table S11 indicate that, overall, unconstrained trees yielding topology-S cannot reject the alternative topology-N, while unconstrained trees yielding topology-N reject the alternative topology-S.

3.8. Clustering analyses

STRUCTURE analysis based on all samples (OP, DF, OVM-0.7P50, n = 63) yielded "optimal" values for $K = 2$ (OP, DF) and $K = 3$ (OVM), based on the ΔK criterion, although the log probabilities of the data (LnP (X|K)) steadily increased with greater K values. The inferred genetic groups were inconsistent with the well-established taxonomic assignment of some populations, grouping, for instance, *D. xeropotamica* and *D. mariae* in a single cluster, or presenting *D. leporii* and MC as admixed populations with *D. hepta*, a species not even present in southern Sardinia (Fig. 3; Supplementary Fig. S8a). In consequence, we explored additional K values in all subsequent hierarchical STRUCTURE analyses. Assignment values to additional clusters for $K > 4$ were meaningless ($Q < 0.001$) and, thus, rendered the respective bar plots virtually identical to those obtained for $K = 4$. Results for $K = 4$ recovered the following clusters: (i) *D. hepta*, (ii) a cluster comprising all Corsican populations, (iii) a cluster grouping *D. benazzii* alongside their putative hybrids with *D. hepta*, and (iv) a cluster comprised of *D. leporii* specimens. MC population was inferred as an admixture of the *D. leporii* and the *D. benazzii* genetic clusters. In addition, no signs of admixture were detected in any of the populations were *D. hepta* and *D. benazzii* are known to co-occur alongside their putative hybrids (BU, MA, MU, and SI).

STRUCTURE analysis using the same datasets but excluding *D. hepta* (OP, DF, OVM-0.7P50, n = 55) yielded "optimal" values for $K = 2$ (OP, DF) and $K = 3$ (OVM), while cluster representation in the bar plot capped at $K = 3$ (OP, DF) and $K = 4$ (OVM). Results based on OP and DF datasets had little biological sense, especially the "optimal" results, which grouped *D. leporii* and *D. benazzii* into a single genetic cluster (Supplementary Fig. S8b). However, in both OVM cases, *D. leporii* and *D. benazzii* were inferred as separate genetic groups, and MC showed strong genetic admixture from both clusters. For $K = 3$, all Corsican populations were inferred as a single genetic cluster, while results for $K = 4$ discriminated *D. xeropotamica*, although admixture was detected in specimens MG1292-08 and CT1288-24 (Fig. 3). STRUCTURE analyses with Corsican (OP, DF, OVM-CorP50,

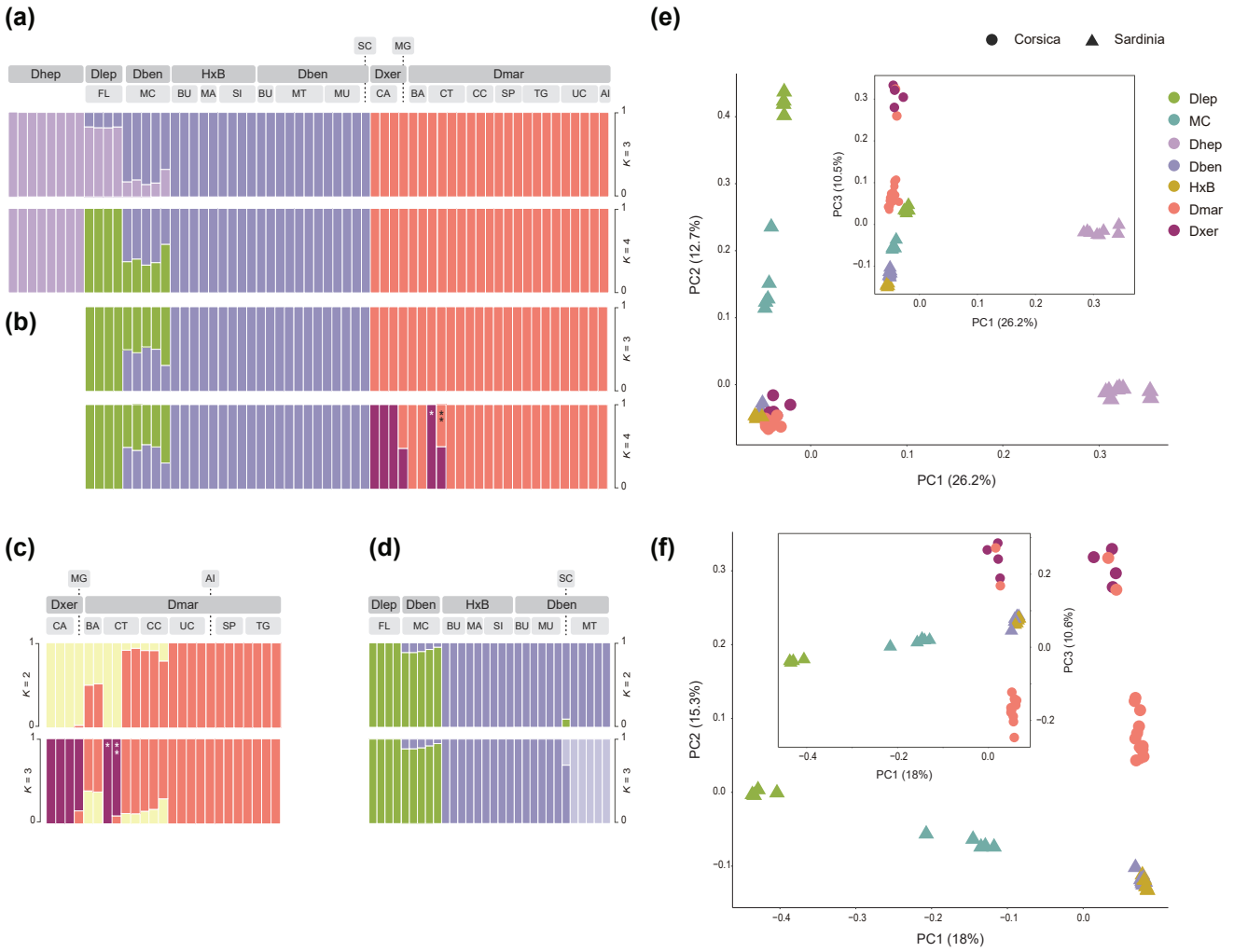


Figure 3. Genetic population structure of the five recognized species of the Corso-Sardinian clade featured in this study. (a) STRUCTURE results for $K = 3$ and $K = 4$ of all samples (n = 63), and (b) excluding *D. hepta* (n = 55). (c) Hierarchical STRUCTURE results for $K = 2$ and $K = 3$ of Corsican samples (n = 25) and (d) Sardinian samples (n = 30). Upper facet labels above the plots represent: *D. benazzi* (Dben), *D. hepta* (Dhep), *D. leporii* (Dlep), *D.mariae* (Dmar), *D. xeropotamica* (Dxer), and the putative *hepta* x *benazzii* hybrids (HxB). Lower facet labels above the plots correspond to populations. (e) Principal Component Analysis (PCA) of all samples showing PC1 (26.2% variance) and PC2 (12.7% variance). Inset shows PC1 versus PC3 (10.5% variance). (f) PCA excluding *D. hepta* showing PC1 (18% variance) versus PC2 (15.3% variance). Inset shows PC1 versus PC3 (10.6% variance). Results shown for datasets OVM-0.7P50 (a,e), OVM-woHP50 (b,f), OVM-CorP50 (c), and OVM-SarP50 (d). Samples (*) CT1288-22 and (**) CT1288-24 are highlighted.

n = 25) and Sardinian (OP, DF, OVM-SarP50, n = 30) subsamples revealed further genetic structure. Results for the Corsican subsample were congruent between datasets, yielding an "optimal" value for $K = 2$, while clustering capped at $K = 3$. In both cases, *D. xeropotamica* was completely discriminated and southern Corsican populations, viz. AI, SP, TG, and UC, were inferred as a distinct genetic group. Remaining northern populations, viz. BA, CC, and CT, showed signals of genetic admixture either between *D. xeropotamica* and the southern Corsican genetic cluster ($K = 2$) or a "ghost" population ($K = 3$). Concurrently, results for the Sardinian subsample yielded "optimal" values for $K = 2$ (OVM) and $K = 3$ (OP, DF), and cluster representation capped at $K = 3$ (OP, OVM) and $K = 4$ (DF). Only OVM analyses recovered *D. leporii*'s genetic cluster and detected signals of genetic

admixture in MC samples, following the same pattern mentioned above. However, overall, results for $K = 3$ showed clear discrimination of samples from MT.

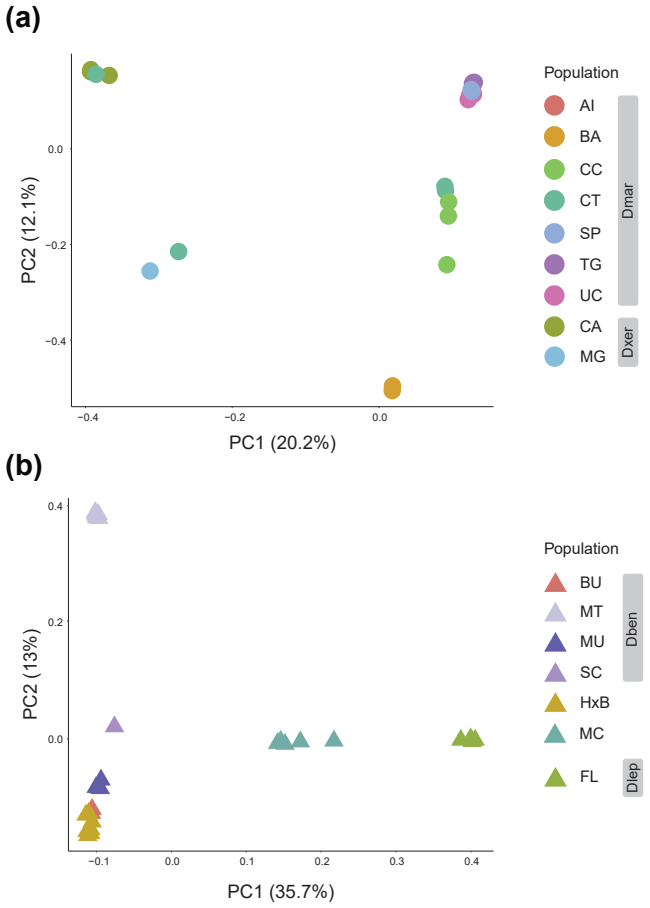


Figure 4. Hierarchical PCA results for (a) Corsican populations (OVM-CorP50, circles) showing PC1 (20.2% variance) and PC2 (12.1% variance). (b) Hierarchical results for Sardinian populations (OVM-SarP50, triangles) showing PC1 (35.7% variance) versus PC2 (13% variance).

PCA results (Fig. 3e,f; Supplementary Fig. S9) for all samples (OP-0.7P50, DF-0.7P50, OVM-0.7P50) showed a clear separation between *D. hepta* and the rest of samples along PC1 (OP, 22.1%; DF, 21.9%; OVM, 26.2% of the variance). Furthermore, PC2 (OP, 11.3%; DF, 12.4%; OVM, 12.7%) clearly separated *D. leporii* from the remaining Corsican and Sardinian populations, while situating MC in an intermediate position between *D. leporii* and the Corso-Sardinian group. PC3 (OP, 10.7%; DF, 11%; OVM, 10.5%) showed a clear separation between *D. xeropotamica* (CA, MG) and the remaining Corsican populations, while at the same time highlighted the differentiation of *D. leporii* and MC from the rest of *D. benazzii*. PCA results after excluding *D. hepta* did not vary significantly, this time, however, PC1 showed a clear separation of *D. leporii* and MC from *D. benazzii* along its axis (OP, 15.7%; DF, 16%; OVM, 18%), and PC2 (OP, 14.6%; DF, 15.1%; OVM, 15.3%) separated Corsican from Sardinian populations, and further separated *D. xeropotamica* from *D. mariae*. Further clustering can be observed when analyzing the Corsican and Sardinian subsamples (Fig. 4; Supplementary Fig. S10). Results using OP-CorP50, DF-CorP50, and OVM-CorP50 datasets clearly discriminated at least five genetic clusters across PC1 (OP,

17.1%; DF, 17.7%; OVM, 20.2%) and PC2 (OP, 11.8%; DF, 12%; OVM, 12.1%), viz. the two diverging *D. xeropotamica* lineages, the specimens from BA, a group of individuals from CC and CT, and, lastly, the meridional populations AI-SP-TG-UC. As per the Sardinian subsample, results based on OP-SarP50, DF-SarP50, and OVM-SarP50 datasets clearly separated *D. leporii* and MC from the remaining *D. benazzii* populations along PC1 (OP, 27.9%; DF, 29%; OVM, 35.7%), while PC2 (OP, 14.4%; DF, 14.1%; OVM, 13%) markedly discriminated MT from the rest of *D. benazzii* populations.

Due to the differences observed in clustering patterns between STRUCTURE and PCA analyses, we performed fineRADstructure analyses, which emphasizes recent coancestry, of the Corsican and Sardinian subsamples, using OP, DF, and OVM-P50 loci, as well as of the total ingroup sample (i.e. excluding *D. hepta*) (Fig. 5; Supplementary Fig. S11).

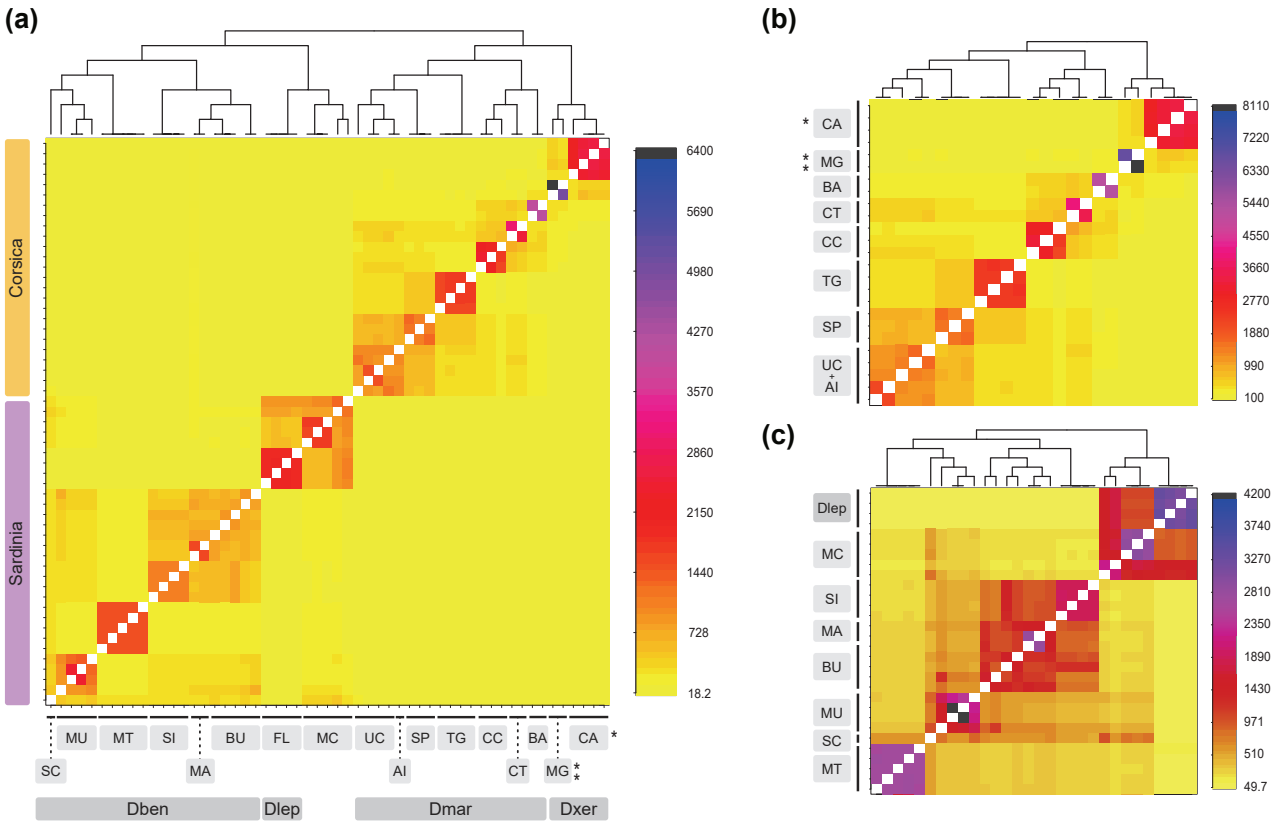


Figure 5. fineRADstructure results showing patterns of shared recent coancestry for (a) all ingroup samples (OVM-0.7P50, *D. hepta* excluded), (b) Corsican subsample set (OVM-CorP50), and (c) Sardinian subsample set (OVM-SarP50). Coancestry coefficients are color coded from low (yellow) to high (blue-black) following the values in the legends. fineRADstructure pertinent clustering dendograms based on the coancestry coefficient matrix are also shown. Branch supports are not shown as all posterior probabilities were 1. (*) CT1288-22 and (**) CT1288-24 are highlighted.

Overall, fineRADstructure recovered the main groups inferred by STRUCTURE analyses with $K = 3$ of both subsample sets but detected finer-scale genetic structure. Noteworthy, all results concurred irrespective of the dataset used. Results on Corsica (Fig. 5b) yielded a clear separation within *D. xeropotamica*, in accordance with PCA and phylogenetic results. Within *D. mariae*, the northern populations (BA, CC, and CT) and southern populations

(AI, SP, TG, and UC) formed two blocks that showed low signs of coancestry. Specimens from BA appeared as a distinct cluster, and two additional clusters were present within the northern block with low levels of coancestry between them, viz. CC and CT. Within the southern block, fineRADstructure showed an incipient separation between TG and the rest of southern populations. Results on Sardinia showed (Fig. 5c), at least, three distinct genetic clusters. Individuals of *D. leporii* shared high recent coancestry with individuals from MC, as seen in phylogenetic results, and, conversely, showed no signs of coancestry with the rest of Sardinian samples. Within *D. benazzii*, specimens from MT were clearly discriminated, sharing low recent coancestry with the remaining populations of the species, and, furthermore, an incipient separation between MU and a cluster comprised by BU, MA and SI was detected. It is worth mentioning that MC shared, overall, low coancestry with *D. benazzii*. Lastly, fineRADstructure results for all ingroup samples were largely similar between datasets and showed no recent coancestry between Corsican and Sardinian populations (Fig. 5a).

3.9. Assessing signals of hybridization

Phylogenetic and clustering results led us to consider the status of MC specimens (n = 5) as potential hybrids between *D. leporii* (FL, n = 4) and *D. benazzii*. Thus, on one hand, we studied *D. benazzii*, *D. hepta*, and their alledged hybrids from SI, BU, and MA (H×B), and on the other, *D. benazzii*, *D. leporii*, and their potential hybrids from MC population (L×B) (Fig. 6).

STRUCTURE analysis for the H×B parental-hybrid trio yielded an optimal value for K = 3, using the OVM-HxBP50 dataset (Fig. 6b), and did not detect signals of genetic admixture in the putative hybrids. Furthermore, MT was discriminated, as in previous results mentioned above, as a separate genetic cluster. To ascertain that MT specimens were not polarizing our results and masking any signs of genetic admixture between *D. hepta* and *D. benazzii*, we excluded them from new analyses. Results showed, once again, no signs of genetic admixture. STRUCTURE results based on optimal and default sets of loci (Supplementary Fig. S8g,h) yielded an "optimal" K value of 2, discriminating MT but grouping *D. hepta* and the remaining specimens in a single genetic group, which was taxonomically inconsistent. Increasing K provided further resolution, discriminating MT (K = 3) and MU (K = 4), but genetic admixture was not detected in the putative H×B hybrids.

STRUCTURE analyses for the L×B parental-hybrid trio yielded an optimal value for K = 2, irrespective of the dataset used (Fig. 6a; Supplementary Fig. S8e,f). Admixture was not detected, as MC was grouped with *D. leporii*. However, after excluding samples from MT and re-running the analyses genetic admixture was detected in MC.

PCA analyses of the H×B parental-hybrid trio showed a clear separation between *D. hepta* and *D. benazzii* along PC1 (OP, 30.8%; DF, 32%; OVM, 42.7%), and between MT

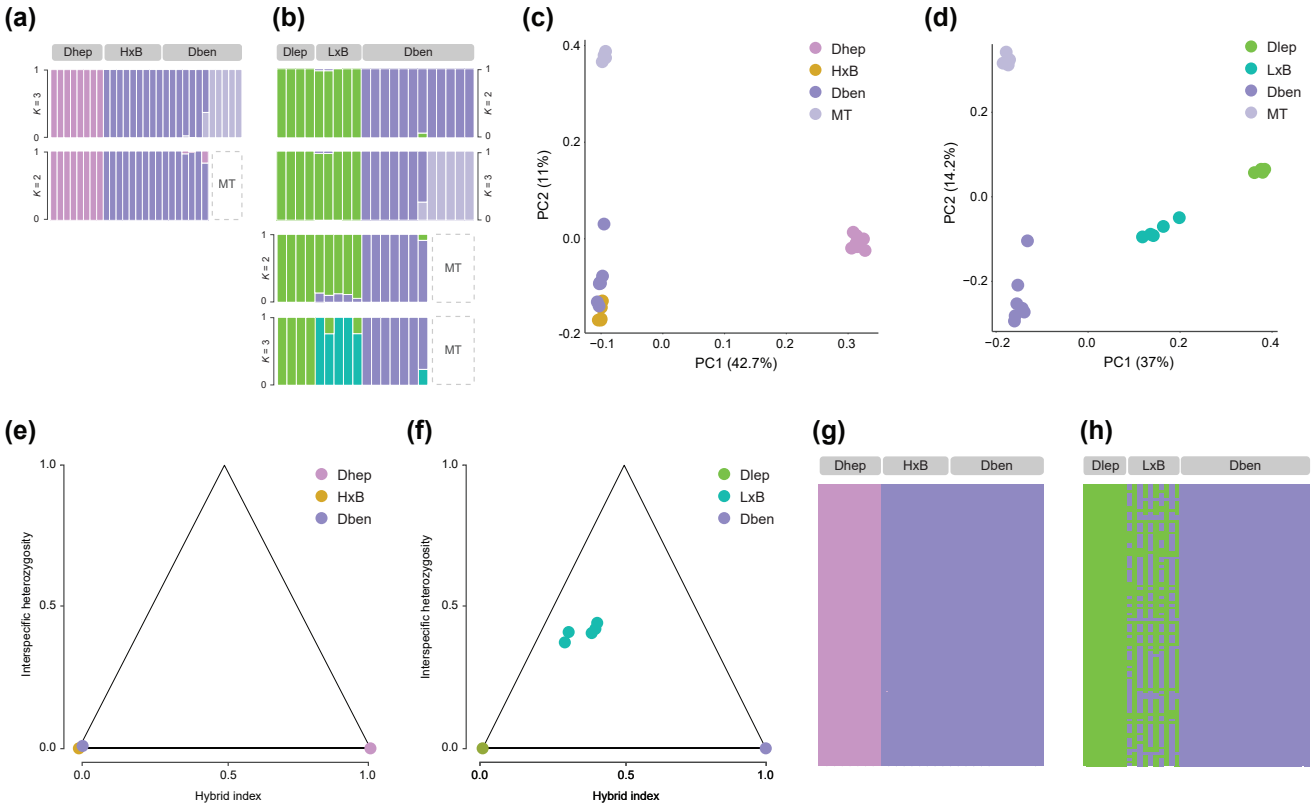


Figure 6. Clustering analysis, INTROGRESS, and ancestry-painting plot results for both candidate parental-hybrid trios analyzed in this study. (a-b) STRUCTURE results for L×B (a; OVM-LxBP50) and H×B (b; OVM-HxBP50) parental-hybrid trios with and without MT. (c-d) Pertinent PCA results for both aforementioned datasets showing PC1 versus PC2. (e-f) Triangle plots showing the relationships between individual heterozygosity (H_{ind}) and hybrid index (*HI*) of both parental-hybrid trios. (g-h) Ancestry-painting plots showing lack of admixture in H×B putative hybrids, but not in L×B hybrids.

and *D. benazzii* along PC2 (OP, 13.1%; DF, 12.7%; OVM, 11%) (Fig. 6; Supplementary Fig. S12). All alleged H×B hybrids were grouped with *D. benazzii* specimens from BU and MU. On the other hand, PCA analyses of the L×B parental-hybrid trio clearly separated *D. leporii* and *D. benazzii* along PC1 (OP, 29%; DF, 30%; OVM, 37%), and discriminated a genetic cluster comprised of MC individuals whose intermediate position along PC1, when considering all analyses mentioned above, might be considered as a sign of admixture. Additionally, a clear separation between MT, *D. leporii*, and *D. benazzii* was recovered along PC2 (OP, 15.9%; DF, 15.8%; OVM, 14.2%).

To further explore the relationships within both parental-hybrid trios, we visualized H_{ind} – *HI* relationships using different sets of loci, while filtering SNPs that were fixed differentially in both candidate parental species and their hybrids (Fig. 6e,f; Supplementary Fig. S13). After filtering, a total of 754 (OP-HxBP50), 290 (DF-HxBP50), and 1478 (OVM-HxBP50) ancestry-informative sites were retained for the H×B parental-hybrid trio. Likewise, the L×B parental-hybrid trio retained 2154 (OP-LxBP50), 1189 (DF-LxBP50), and 2692 (OVM-LxBP50) ancestry-informative sites. In the L×B triangle plots, the configurations of specimens resembled those observed in the PCA plots, in that MC did not cluster with any of the candidate parental species. Contrastingly, H×B triangle plots showed near

full overlap of the putative hybrids with *D. benazzii*'s cluster. Finally, we reconstructed ancestry-painting plots by filtering the ancestry-informative sites to retain only those that were present in all individuals of the pertinent trio (Fig. 6g,h; Supplementary Fig. S13). After filtering anew, the H×B parental-hybrid trio retained 71 (OP-HxBP50), 31 (DF-HxBP50), and 214 (OVM-HxBP50) sites, while 75 (OP-LxBP50), 54 (DF-LxBP50), and 101 (OVM-LxBP50) sites were retained for the L×B parental-hybrid trio. Overall, results of L×B ancestry-painting plots clearly showed signals compatible with hybridization in all MC specimens, whereas H×B results showed that the potential hybrids between *D. hepta* and *D. benazzii* could not be distinguished from *D. benazzii* specimens.

3.10. ABBA-BABA test results

Results of *D*-statistics tests revealed significant introgression involving MC and *D. benazzii* (OP-D_s = 0.47 ± 0.07; DF-D_s = 0.55 ± 0.08; OVM-D_s = 0.59 ± 0.03), as well as between MC and D. leporii (OP-D_s = 0.86 ± 0.02; DF-D_s = 0.88 ± 0.02; OVM-D_s = 0.92 ± 0.01) (Fig. 7; Supplementary Fig. S14). Additionally, the f4-ratio tests indicated that MC specimens have higher *D. leporii* than *D. benazzii* genetic component (Supplementary Table S12).

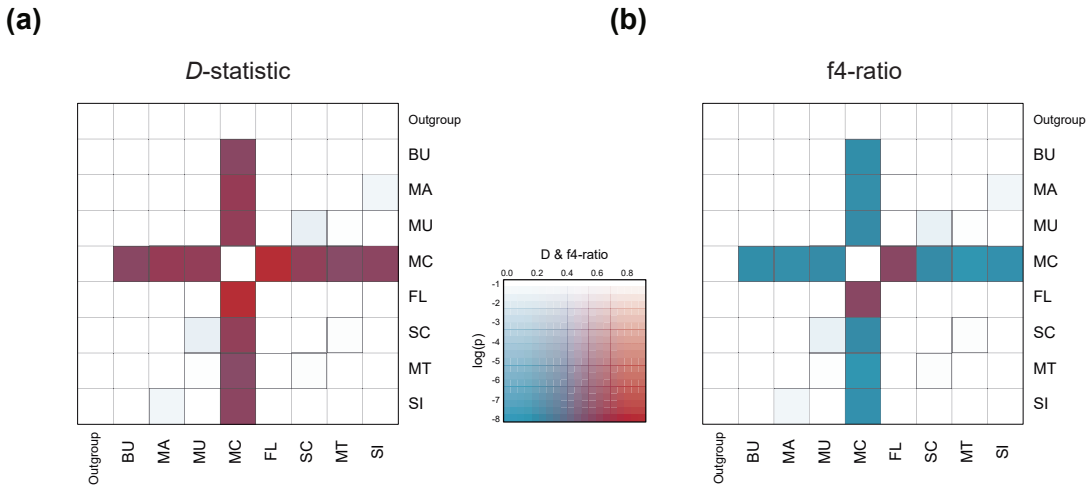


Figure 7. (a) Results for the ABBA-BABA tests and (b) f4-ratio tests based on OVM-ABBAP50. Heatmap cells: blue hues indicate low values for *D*-statistics and f4-ratio tests. Conversely, red hues indicate high values. Color intensity indicates varying degree of p-value significance after applying the Benjamini-Hochberg correction for multiple testing.

3.11. Primary species candidates

Based on our phylogenetic and clustering analyses results, we proposed the following primary species candidates (PSC) to be tested and validated using two coalescent-based approaches: (i) PSC1, comprised of specimens currently identified as *D. benazzii* from Sardinian populations BU, MA, MU, SI, and SC; (ii) PSC2, comprised of specimens from MT; (iii) PSC3, comprised of specimens from CA and individual CT1288-22; (iv) PSC4, comprising individuals CT1288-24 and MG1292-08; (v) PSC5, comprised of spec-

imens from population BA; (vi) PSC6, comprising Corsican individuals CT1288-21 and CT1288-23; (vii) PSC7, comprised of specimens from CC; and (viii) PSC8, comprised of specimens from Corsican populations AI, SP, TG, and UC. Although not strictly necessary, we included *D. leporii* as well in the BPP analyses to formalize its molecular validation but excluded it from subsequent BFD* analyses. This method is computationally intensive, especially when assessing too many species, and *D. leporii* is already a formally described species, which has been shown to be genetically distinct in all our analyses.

Considering the number of PSCs, we opted for optimizing the number of hypotheses to be tested with BFD* following the rationale detailed below. Given that *D. benazzii* (PSC1-2), *D. xeropotamica* (PSC3-4) and *D. mariae* (PSC5-8) are well established species that have been delineated in previous works using integrative approaches, it fell out of reason to assess hypotheses that merged interspecific PSCs. Furthermore, 6 out of the 8 PSCs occurred in Corsica, and their rebuttal or validation had virtually no impact on the delimitation of Sardinian PSCs (PSC1 and PSC2). Thus, we first tested SDH1 vs SDH2 (Table 1) to evaluate whether the Sardinian PSCs should be split or not in the following hypotheses to be tested. Thereafter, subsequent SDHs were proposed by collapsing particular PSCs using topology 1 as a guide.

3.12. MSC species delimitations: validation of PSCs

MSC species delimitation analyses using BPP consistently supported a SDH with eight species and, unsurprisingly, validated *D. leporii* by fully supporting its split from the remaining PSCs (node H, Fig. 8a, Supplementary Table S13). Overall, the results were robust to both the evolutionary model and subset of loci used. However, datasets OP-0.7P50 and DF-0.7P50 showed discrepancies under evolutionary model 3, which assumed large ancestral population size and deep divergence time. Results based on the optimal set indicated that the splitting event of PSC3 and PSC4 (node A, Fig. 8a, Supplementary Table S13) was not supported under such model, while DF-0.7P50 results did not support, in addition, the split between [PSC3, PSC4] and PSC5 (node B, Fig. 8a, Supplementary Table S13). Furthermore, we detected a minor convergence problem between rjMCMC algorithms when using the subset of DF-0.7P50 loci under evolutionary model 3, rendering the support of the split between PSC6 and PSC7 (node C, Fig. 8a, Supplementary Table S13) as inconclusive.

Regarding BFD* results, all datasets used (OP-bfdP50, 636 SNPs; DF-bfdP50, 306 SNPs; OVM-bfdP50, 702 SNPs) showed the strongest support for an eight species model (Table 1).

3.13. Divergence time estimates of PSCs

After filtering, the amount of retained SNPs per dataset was as follows: (i) OP-

Table 1. BFD* analysis results for competing species delimitation hypotheses (SDH). For each SDH, marginal-likelihood estimates (MLE), Bayes factors comparisons (*2lnBf*), number of species are shown. N/A indicates the best model per dataset.

			OP-bfdP50		DF-bfdP50		OVM-bfdP50	
SDH	Species	Num. species	MLE	*2lnBf*	MLE	*2lnBf*	MLE	*2lnBf*
H1	PCS1, PCS2, PCS3, PCS4, PCS5, PCS6, PCS7, PSC8	8	-6651.3436	N/A	-3293.8689	N/A	-7072.5424	N/A
H2	PSC1-2, PSC3, PSC4, PSC5, PSC6, PSC7, PSC8	7	-7126.6336	950.58	-3540.3319	492.926	-7837.1633	1529.242
H3	PSC1, PSC2, PSC3, PSC4, PSC5, PSC6-7, PSC8	7	-6747.8514	193.0156	-3329.9940	72.2502	-7143.2471	141.4094
H4	PSC1, PSC2, PSC3-4, PSC5, PSC6-7, PSC8	6	-7150.8434	998.9996	-3510.8807	434.0236	-7599.6006	1054.116
H5	PSC1, PSC2, PSC3-4, PSC5, PSC6-7-8	5	-7449.7209	1596.755	-3611.1303	634.5228	-7939.8419	1734.599
H6	PSC1, PSC2, PSC3-4, PSC5, PSC6, PSC7, PSC8	7	-7049.1235	795.5598	-3475.4909	363.2440	-7521.9262	898.7676

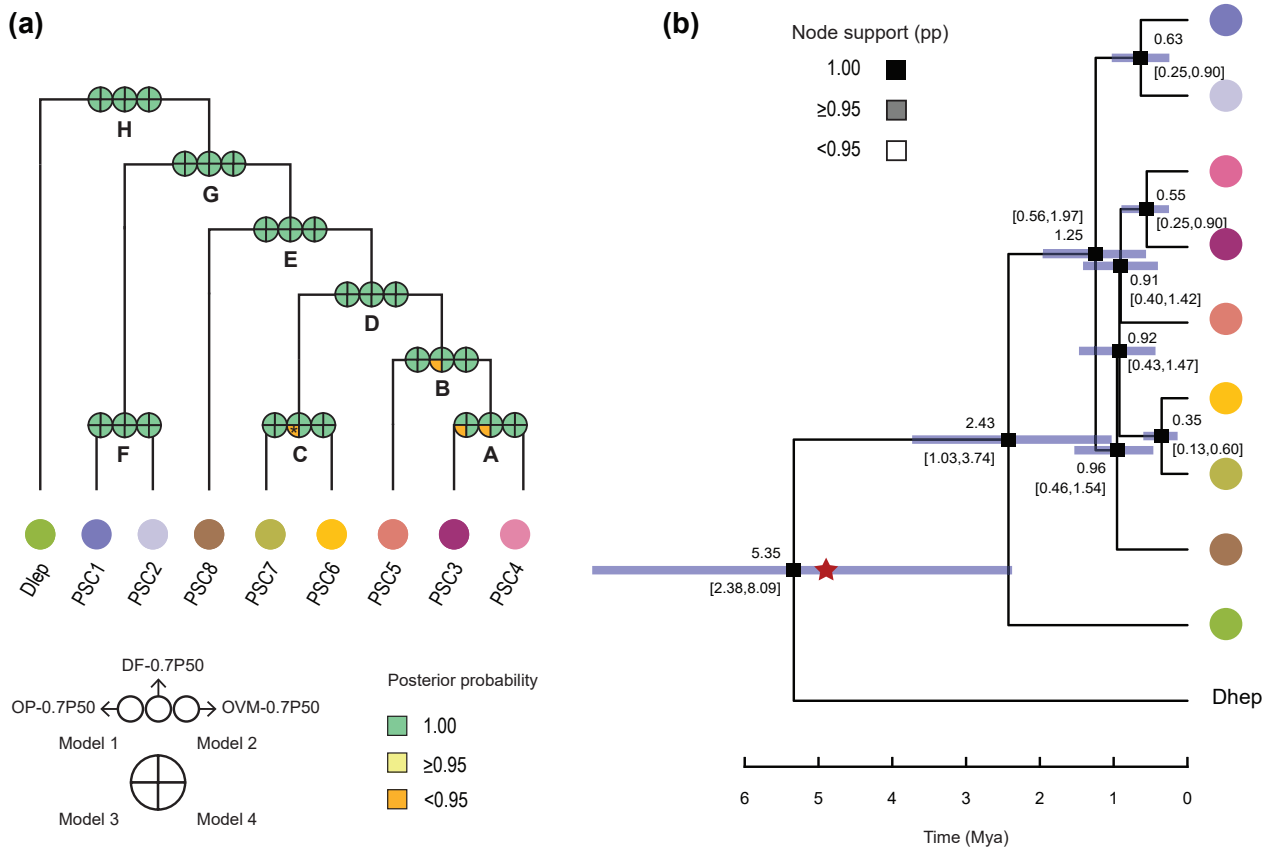


Figure 8. (a) Schematic representation summarizing the guided tree used for BPP analysis based on 500 random OVM-0.7P50 loci. Color scheme indicates the posterior probability of each node/splitting event (A-H). (b) Time-calibrated SNAPP tree of the Corso-Sardinian clade showing the phylogenetic relationships among primary species candidates (PSC) and the estimated date of each node in million years (Mya). Blue bars represent the 95% high posterior density (HPD) interval of nodal time estimates (also indicated in squared brackets). Crimson star indicates the calibrated node. (*) Split with convergence problems between algorithms A0 and A1.

0.7P50, 408 SNPs; (ii) DF-0.7P50, 208 SNPs; (iii) OVM-0.7P50, 775 SNPs. Divergence time estimates between validated PSCs suggested that candidate taxa might have diverged during the Pleistocene in a relatively short span of time, while *D. leporii* might have diverged during the Late Pleistocene or Early Pliocene (OP: 2.32 Mya, DF: 3.17 Mya, OVM: 2.43 Mya) (Fig. 8b, Supplementary Fig. S15). Overall, OP and DF results showed greater

discordance (Supplementary Fig. S16a,b) than OVM results (Supplementary Fig. S16c), as indicated by the fuzziness of the cloudograms, lack of nodal suport in certain nodes, and the rearrangement of certain PSCs, yielding novel topological results that established PSC3 and PCS4 (*D. xeropotamica*) as taxa sister to the remaining Corsican and Sardinian PSCs.

4. Discussion

4.1. RADseq dataset optimization

RADseq has been widely adopted among researchers given its utility and relative low cost of laboratory work for shallow systematics in non-model species. However, under proper conditions for taxon sampling and sequencing coverage, RADseq can be a powerful approach for large-scale phylogenomics as well (Eaton et al., 2017). It is particularly widespread for population genetic and phylogeographic studies (reviewed in Narum et al., 2013) but also for phylogenetics (e.g. Appelhans et al., 2020; Eaton and Ree, 2013; Ferrer Obiol et al., 2021) and species delimitation (e.g. Ferrer Obiol et al., 2023; Hosegood et al., 2020; Newton et al., 2020). However, all this is dependent on a good yield of orthologous loci.

One of the main factors affecting the number of orthologous loci recovered in RADseq data is the assembly parameters applied to construct loci catalogs. It is a common procedure, especially in population genomic analyses, to optimize assembly parameters to minimize the inclusion of paralogous and singleton loci, and researchers often follow the guidelines proposed by Paris et al. (2017). But also, the amount of missing data, which can greatly impact resulting phylogenetic inferences (Eaton et al., 2017). If divergence times are deep or taxon coverage parameters are set too high, data set size can rapidly dwindle at the cost of informative sites (Cariou et al., 2013; Rubin et al., 2012). Furthermore, in our case, heterogeneous ploidy composition among samples and mosaic-Meselson effect in freshwater planarians might be factors contributing to erroneous orthologous assembly and, overall, oversplitting by separating alleles of a single locus into different stacks. Thus, species with higher divergence levels (e.g. mosaic-Meselson effect) will lose more variation than those with lower divergence across datasets assembled with the same parameters (Harvey et al., 2016).

To navigate these issues, we used three sets of catalog loci, viz. OP, DF, and OVM, to assess the robustness in our results. Results showed that, before filtering, OVM catalog presented the lowest percentage of singleton loci (potential errors or rare alleles), showing comparable levels to those reported in studies on diploid organisms (e.g. Ferrer Obiol et al., 2021). However, OVM recovered ~2.5x less loci when compared with OP and DF catalogs, which was to be expected since we were applying lower similarity thresholds to assemble loci. Differences in singleton composition could be explained by the differences regarding assembly parameter n. While OP assembly parameters (m2 M1 n2) will facili-

tate oversplitting at individual level (i.e. alleles within an individual that should collapse into the same locus will result in separate loci given the low settings for *M*), especially given *Dugesia*'s mosaicism (see 'star-like pattern' in Leria et al., 2019a), the $n = 2$ can revert part of the oversplitting at population level when constructing the catalog, thus having an impact on singletons. However, DF assembly parameters (m3 M2 n1) will fail more often to find homologous loci in different samples because *n* is too stringent, thus increasing singletons spuriously. In principle, OVM assembly parameters (m3 M5 n6) should be more resilient to noise and oversplitting. This, in turn, could explain why results recovered using OVM datasets were, overall, more consistent across analyses.

4.2. Consideration for phylogenetic inference

Taxon coverage parameters are relevant (e.g. Tripp et al., 2017), and our phylogenetic results indicate so. Perhaps more than assembly parameters under certain circumstances. We can clearly see the effect of taxon coverage parameters when examining phylogenies inferred using P100 loci with broad taxon sampling (e.g. including *D. leporii* and *D. hepta*). P100 datasets showed a massive loss of polymorphic loci compared to their P50 and P75 counterparts (Supplementary Table S7), resulting in a normalization of the variation across samples and, in consequence, yielding phylogenies with shorter internodes and more ambiguous or unresolved relationships. Interestingly, phylogenies inferred with OVM-P100 loci that included all taxa (n = 63) seemed to be less affected, often showing better nodal support than their OP and DF counterparts (Supplementary Table S14), but also retaining 2~3x more loci, possibly due to a lesser prevalence of singletons. In addition, topologically OVM-P100 trees were less impacted than OP-P100 and DF-P100 trees and, even though MC was included in the sample set, topologies yielded by OVM-0.4P100 and OVM-0.6P100 datasets had great similarities with topology 1 (cf. Fig. 3b and Supplementary Fig. S4), which we argue it should be considered the correct topology (discussed below).

On the other hand, methodological approaches can affect the phylogenetic inference. RADseq datasets have low phylogenetic information per locus, which might result in poorly resolved gene trees, and in such cases, concatenation can be more accurate than summary coalescent approaches to reflect the true species tree (Mirarab et al., 2016; Molloy and Warnow, 2018). However, under high levels of ILS (Mendes and Hahn, 2018) or abundant introgression (Mallet et al., 2016) concatenation analyses may result in an average topology that differs from the true species tree. In our case, when hybrids from population MC were present in the sample set, and enough informative sites were used (P50 and P75 loci), topological results were largely similar across different datasets yet differing from the true species tree (cf. Supplementary Figs 6 and 7c). We argue that this might happen due to the inclusion of hybrids in the trees.

Concurrently, after excluding MC, phylogenetic results produced two discordant to-

pologies (Fig. 2b,c). Topology 1 recovered *D. leporii* as sister taxon to all PSCs, showed overall no signs of ambiguous or unresolved relationships, supported the monophyly of Corsican populations, which has been well-established in other studies (Benítez-Álvarez et al., 2023; Dols-Serrate et al., 2024, 2025), and is congruent with STRUCTURE and PCA results. In contrast, topology 2 recovered *D. leporii* as sister taxon to all PSCs, but showed shorter internodes, sings of nodal support discordance, and disarray of the Corsican clade following a pattern seemingly grouping long branches near the root of the trees. Neither clustering (Fig. 4; Supplementary Fig. S8) nor recent coancestry (Fig. 5; Supplementary Fig. S11) results suggested any relationship between southern Corsican and Sardinian populations that could support topology 2 as a valid hypothesis. Although it holds true that topology 2 was most prevalent in OVM datasets, OVM catalog loci have a greater ratio of SNPs per locus (Supplementary Fig. S17) which can make their loci prone to include more high heterogeneity rate sites and be affected by long branch attraction (LBA) (Bergsten, 2005). However, it is worth mentioning that when missing data was maximized (i.e. OP/DF/OVM-0.4P50), all datasets recovered the same topology (Fig. 2d; Supplementary Fig. S6), emphasizing the importance of exploring several taxon coverage filters.

Additionally, concatenation and quartet-based coalescent methods yielded incongruent topologies when comparing respective results for the same dataset (Supplementary Figs S6 and S7). OP-0.7P50 and DF-0.7P50 datasets yielded a new topology not recovered before in any phylogenetic analysis, regardless of assembly and/or taxon coverage parameters used. Such results are hardly believable, particularly given that clustering and fineRADstructure analyses have consistently discriminated *D. leporii* against any other species or population. Contrastingly, OVM-0.7P50 quartet-based results recovered topology 1, in congruence with results of phylogenetic analyses with low taxon coverage parameters and clustering analyses as opposed to the concatenation results for the same dataset (which yielded topology 2). We argue that this change in topology yield responds to the fact that we used only one random SNP per locus for the SVDQuartets analyses, possibly reducing the impact of LBA.

4.3. Corso-Sardinian *Dugesia* clade: how many species?

While species delimitation in freshwater planarians using morphology has a long tradition, the practice of accommodating multiple lines of evidences under an integrative taxonomic framework is comparatively recent (e.g. Dols-Serrate et al., 2024, 2025; Harrath et al., 2019; Leria et al., 2020; Sluys et al., 2013; Stocchino et al., 2013). Although, strictly speaking, in our present study we do not describe any new species, our results provided sufficient evidence to warrant a taxonomical update of the Corso-Sardinian *Dugesia* clade of the Western Mediterranean. Furthermore, our study supposes, to our knowledge, the first successful use of RADseq data to perform species delimitation in triclads (Platyhelminthes, Tricladida). Additionally, it also constitutes the first molecular insights of *D. leporii*.

Based on our results, we find no support for the monophyletic status of *D. mariae.* Phylogenetic, clustering, and species delimitation analyses were concordant in that regard. Phylogenetic results, irrespective of dataset assembly parameters, taxon coverage, or sample set, consistently recovered *D. mariae* as paraphyletic, in contrast with results from previous phylogenetic studies (Dols-Serrate et al., 2024, 2025) and the mtDNA phylogenetic tree used in this study to identify our samples (Supplementary Fig. S3). PCA analyses clearly discriminated the type locality of *D. mariae* (BA) from the rest and showed potential splitting of up to four genetic clusters (Supplementary Fig. S10). STRUCTURE analyses (Supplementary Fig. S8), overall, had problems discriminating genetic clusters, probably due to the divergence between some of the taxa present in the analyses or differences in sample size (Lawson et al., 2018). This, in turn, led to the inference of "optimal" values for *K* that often underrepresented the acknowledged taxonomic diversity of the pertinent sample set. We found that cluster representation capped at certain *K* values, which made better biological sense, and we decided to weight them positively over the "optimal" *K*. Thus, hierarchical STRUCTURE analyses of the Corsican subsample indicated that *D. mariae* was comprised of two allelic populations. PSC8 (AI, SP, TG, and UC) was clearly discriminated, in accordance with phylogenetic and PCA analyses, yet PSC5 (BA), PSC6 (CT1288-21 and CT1288-23), and PSC7 (CC) showed a pattern of admixture with a "ghost" population. Said pattern could be the result of (i) admixture with an extinct or unsampled population, or (ii) ILS. Both phenomena could either be caused by genetic diversity in northern Corsica that did not establish in southern populations, potentially due to population bottlenecks during a stepwise southward range expansion alongside lack of gene flow given their low vagility, or the result of fragmentations and demographic contractions of a broader ancestral population. Distinguishing between these scenarios with our sampling and amount of data, at present, is difficult. However, throughout our analyses we have seen a pattern of higher genetic structure and diversification in northern Corsica (5 out of the 8 delimited PSCs occur in that region). pattern that has also been reported for vertebrates (Bisconti et al., 2013; Salvi et al., 2017) and plants (Migliore et al., 2011), which suggests that the underlying causes could be climatic, as we discuss further below.

Our analyses also suggested that *D. xeropotamica* might constitute two separate species, viz. PSC3 (CA and CT1288-22) and PSC4 (CT1288-24 and MG1292-08). Spatial clustering, phylogenetic, and species delimitation results were largely congruent across datasets and taxon coverage parameters. It is acknowledged that MSC methods have a tendency towards mistaking structure due to population-level processes and that due to species boundaries (Chambers and Hillis, 2020; Sukumaran and Knowles, 2017). While this may hold true, especially when considering the dispersal capabilities of freshwater planarians and the potential impact of seasonal droughts on population fragmentation or bottlenecks, our results can only be fully interpreted once other reliable sources of data corroborate the split of PSC3 and PSC4 in a future study.

However, during fieldwork we were able to observe some differences, specifically between samples of CA and MG. Most notably, specimens from CA were particularly large,

measuring up to ~30 mm in length, almost twice the length of MG specimens or any Corso-Sardinian *Dugesia* in general (Dols-Serrate pers. obs.).

Similarly, all analyses were largely congruent concerning *D. benazzii*'s potential split into two species (PSC1 and PSC2). Noteworthy, previous studies (Dols-Serrate et al., 2020, 2024, 2025) had also highlighted MT (PSC2) as a distinct population, genetically. Most Sardinian populations can be quite heterogeneous in terms of karyological composition or reproductive strategies. For instance, in populations such as BU or MA, it is common to find diploid and aneuploid individuals co-occurring in the same locations, altogether with other species, such as *D. hepta*. Moreover, these populations typically exhibit regression of the copulatory apparatus when reared in the laboratory within a matter of weeks (Stocchino and Manconi, 2013). But not MT. This monospecific population, located in a small valley by the monte Albo massif in eastern Sardinia, is isolated from the rest of known *D. benazzii* populations (Stocchino, 2018), and only encompasses diploid and sexual individuals (Stocchino pers. comm.) that did not reabsorb the copulatory apparatus under lab conditions, irrespective of time passed (Dols-Serrate pers. obs.). Nevertheless, all things considered, pertinent taxonomic decisions will have to wait until consideration of other sources of evidence.

4.4. Found anew: *Dugesia leporii*

Ever since its formal description, *D. leporii* had remained elusive. Our niche modelling approach, based on *D. benazzii* s.l. distribution data (i.e. including populations of *D. benazzii* and *D. mariae* as of present days), proved to be successful. As a result, we were able to collect the first specimens of this species since 1999. While it holds true that the niche model could be improved with a broader taxon sampling, or accounting for absence in visited sites, it falls within reason to assume that species of the Corso-Sardinian clade, overall, share broadly similar niches given our present results. Such environmental niche similarities imply that present distribution patterns of the Corso-Sardinian species might be consequence of historical factors, which is expected from low vagility species (Barbato et al., 2018; Fattorini, 2009).

Our analyses consistently discriminated *D. leporii* and morphological data corroborated them. Future studies on the group should strive to sample new populations across southern Sardinia to assess *D. leporii*'s potential distribution range and unveil possible population structure, or, perhaps, uncover new taxa that might help unravel the complex evolutionary history of the Corso-Sardinian species.

4.5. Evolutionary history of the Corso-Sardinian clade

Currently, the taxonomy of the Corso-Sardinian clade encompasses six species, viz. *D. benazzii*, *D. hepta*, *D. hoidi*, *D. leporii*, *D. mariae*, and *D. xeropotamica*. Our study em-

ployed genome-wide data that partially resolved the phylogenetic tree of the Corso-Sardinian clade and elucidated potential new species boundaries. Unfortunately, it was not possible to obtain genomic data of *D. hoidi* and, therefore, its phylogenetic position shall remain unresolved for now. However, our analyses showed a pattern of substantial diversification that occurred mainly on Corsica.

It is acknowledged that microplate tectonics may have played an important role in the diversification of the Western Mediterranean *Dugesia* (Leria et al., 2022). However, it has been estimated that the diversification within the Corso-Sardinian clade occurred between 5.7 and 2.2 Mya (Solà et al., 2022), long after the settlement of the Corso-Sardinian microplate in its current position and the first separation between both islands (9 Mya; Alvarez, 1972; Alvarez et al., 1974; Rosenbaum et al., 2002). Given our time-calibrated results (Fig. 8b; Supplementary Figs S15 and S16), it falls within reason to presume that changes in paleoclimatic conditions were important for the diversification of this group. Suc (1984) suggested, based on palynological and macroflora analyses, that changes in paleoenvironmental conditions of the north-western Mediterranean that resulted on the onset of the Mediterranean climate in Europe were characterized by a progressive decrease in moisture, with the installation of summer dryness about ~3.2 Mya, and a first dry period at ~2.3 Mya. Our results align with the estimated timing of these events, which could explain the divergence of *D. leporii* right before the split of the Corsican and Sardinian subclades that occurred between 2.2 and 1.2 Mya (Fig. 8b; Supplementary Figs S15 and S16). Thus, the posterior diversification of all PSCs might have likely occurred during the Pleistocene glacial-interglacial oscillations (2.58 Mya – 12 kya).

During the Pleistocene glacial phases, extensive glaciers formed on the central mountain chain of Corsica (Kuhlemann et al., 2005) and it has been suggested that polar air invasions might have recursively impacted northern Corsica during the last glacial maximum (Kuhlemann et al., 2008). Furthermore, prevalence of Pleistocene microrefugia in northern Corsica has been proposed to explain the phylogeographic structure of certain endemic species (Bisconti et al., 2013; Salvi et al., 2017). This hypothesis aligns particularly well with our results, considering that five PSCs occur in said region (Fig. 9), characterized by an abundance of river valleys. All things considered, the paleoclimatic conditions discussed above could have caused multiple population fragmentation, contraction, and extinction events, consequently driving speciation or, at least, the genetic structure patterns found in our study in a rather short period of time. The occurrence of multiple species of *Dugesia* within a confined geographical area is not unique. Similar situations have been documented in other Mediterranean regions, such as the islands of Mallorca (Leria et al., 2022) and Corfu (Solà et al., 2013), where different species coexist in close proximity. These findings support the notion that fluvial basins function as isolated habitats for planarians (Solà et al., 2013). Consequently, processes that disrupt fluvial connections, such as glacial periods, may influence the structuring and speciation of planarian populations.

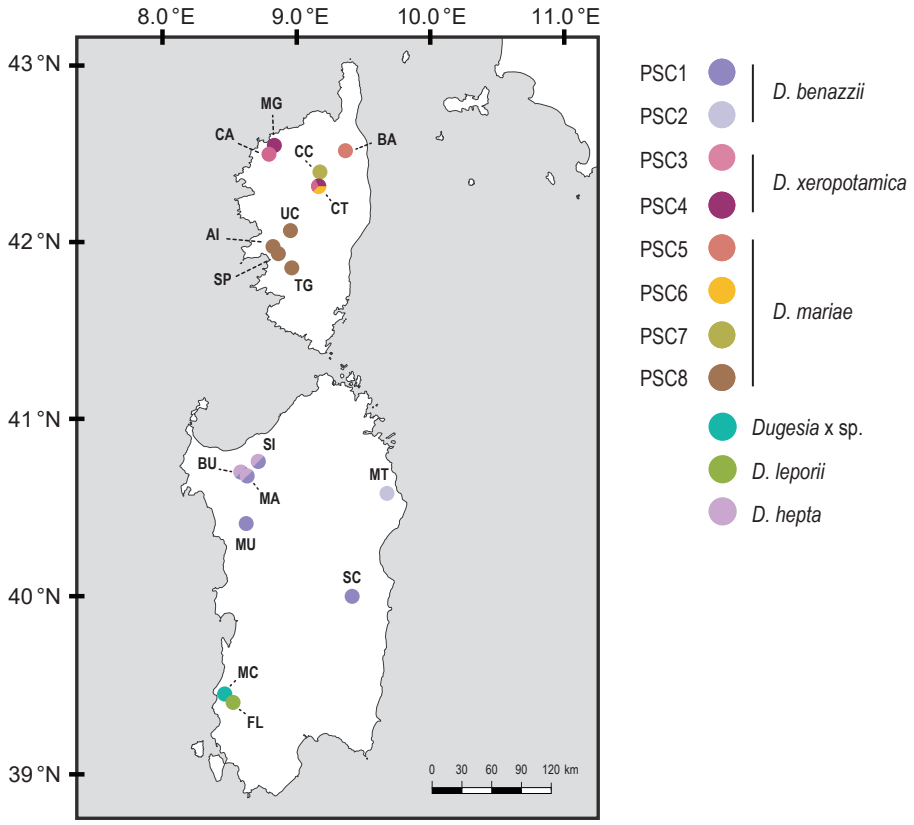


Figure 8. Map showing the distribution of all primary species candidates (PSC), *D. leporii*, *D. hepta*, and the hybrid population from MC.

4.6. On the status of interspecific hybrids

Gene flow, and ultimately hybridization, is a fundamental control on the speciation process and its impact in the formation and persistence of independently evolving lineages can be disparate (Mallet, 2007). The notion of hybridization in freshwater planarians is rather unexplored and often the term has been used to describe syngamy between different reproductive biotypes of the same species (e.g. Beukeboom et al., 1998; Beukeboom and Vrijenhoek, 1998; Pongratz et al., 1998). To our knowledge, only two studies have addressed interspecific hybridization in freshwater planarians to this date. Pala et al. (1982) proposed, based on karyological evidence, that one of *D. benazzii*'s biotypes (see Pala et al., 1982 and references therein) was a natural stabilized parthenogenetic hybrid between *D. benazzii* and *D. hepta*. Said biotype was characterized, among other things, by presenting aneuploid chromosome sets often ranging from 27 to 32 chromosomes within a same individual. However, this hypothesis was questioned (Benazzi-Lentati and Benazzi, 1985). More recently, a phylogenetic study analyzing *D. benazzii*, *D. hepta*, and the populations where both species coexist found molecular evidence compatible with interspecific hybridization (Dols-Serrate et al., 2020).

Our karyological data showed that the karyotype that Pala et al. (1982) had attributed to H×B putative hybrids (aneuploids in Supplementary Table S3) was prevalent across different *D. mariae* populations, as well as in other monospecific *D. benazzii* populations (e.g.

SC). We found that similar karyological results had been reported for Corsican populations before (Deri, 1981, 1971). Specifically, cases of hiperdiploid chromosome complements with ~30 chromosomes. It is worth mentioning that while putative H×B hybrids from BU and MA did present an aneuploid chromosome complement as in Pala et al. (1982), all H×B putative hybrids from SI were diploid (2n = 16). Moreover, our phylogenetic, clustering, and introgression analyses did not support the hybrid origin of the candidate H×B putative hybrids, indicating, by all means, that they must be considered as members of *D. benazzii*. These results suggest that our previous results for interspecific hybridization were, most likely, a case of ILS (cf. Dols-Serrate et al., 2020).

On the other hand, our analyses indicated that population MC probably had a hybrid origin. Noteworthy, all MC individuals were asexual without copulatory apparatus. The genomic contributions of both parental species to the hybrid lineage are expected to be equal in F_1 hybrids, but unequal if they derive from backcrossing of hybrids (Payseur and Rieseberg, 2016). While it holds true that our results revealed introgression signatures across all *D. benazzii* populations (Fig. 7; Supplementary Fig. S14 and Table S12), possibly discarding a recent origin of the hybridization event, it is difficult to reach categorical conclusions given the difficulty to assess the role that the potentially mixoploid chromosome complement of MC might play into all this. However, INTROGRESS and f4-ratio tests results showed patterns compatible with an old origin and backcrossing with *D. leporii*'s lineage. Similar results have been reported in pen shells (Sekino et al., 2023).

It is acknowledged that triploid and polyploid freshwater planarians of the genus *Dugesia* are capable of reproducing pseudogametically (Benazzi-Lentati, 1979; Benazzi and Benazzii-Lentati, 1976) and crossing experiments on the genus Schmidtea have demonstrated that polyploid parthenogens can fertilize haploid oocytes of sexual individuals, generating different lineages of polyploid parthenogenetic offspring (Pongratz et al., 1998). Moreover, Chinone et al. (2014) demonstrated that triploid specimens of *Dugesia ryukyuensis* Kawakatsu, 1976 presented a complex meiotic system that allowed them to produce haploid gametes, as well as diploid ones (usually, oocytes). In all, our results suggest that *D. leporii* likely hybridized with an ancestral *benazzii* lineage. If syngamy had involved diploid oocytes, F_1 triploid or polyploid hybrids could have still retained meiotic activity and backcross with *D. leporii*, potentially generating the mixoploid hybrid offspring that we have analyzed. Finally, our analyses did not find evidence of ongoing hybridization. Perhaps the putative F_1 hybrids went extinct and MC hybrids thrived thanks to fissiparous reproduction. Or maybe they are still out there, waiting to be discovered. Only time and future research will tell.

5. Conclusions and final remarks

Despite the anticipated enhancement in our understanding of the evolutionary history and diversity within the Corso-Sardinian clade that should come from using genome-wide loci rather than legacy Sanger loci, it is important to acknowledge existing

challenges. Intraspecific ploidy variation, as observed in our case study species (Deri, 1981; Pala et al., 1982), complicates most genetic analyses as most bioinformatic pipelines are designed for diploid organisms. Our approach using over-merging assembly parameters has produced consistent results, overall yielding more robust results than the optimal and default approaches. We recommend that phylogenetic studies using RADseq data for non-model non-diploid organisms should explore alternative assembly parameters to minimize the effect of paralog and singleton content, or at least not rely on low assembly parameters.

Concurrently, we highlight a pattern of high genetic diversity within the Corso-Sardinian clade, probably driven by paleoclimatic changes, that warrant a re-evaluation of the taxonomic group. We present a phylogenetic hypothesis within a paleobiogeographic context in which *D. leporii* diverged during a period of decreased moisture in the Western Mediterranean. This event likely occurred prior to the potential speciation outburst in Corsica and Sardinia during the Pleistocene. Most notably, our study provides robust evidence for the first time of natural interspecific hybridization in freshwater planarians and triclads.

CREDIT authorship contribution statement

Daniel Dols-Serrate: Conceptualization, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Longhua Guo**: Investigation, Supervision. **Leonid Kruglyak**: Funding acquisition, Supervision. **Marta Riutort**: Conceptualization, Resources, Writing – review & editing, Funding acquisition, Supervision.

Conflicts of interest

The authors declare that they have no known competing financial or personal interests that could have influenced the work presently reported in this paper.

Funding

This research was supported by the Ministerio de Economía y Competitividad (project PGC2018-093924-B-100) from Spain. This research was supported also by the Ajuts de Personal Investigador predoctoral en Formació (APIF) grant from the Universitat de Barcelona (Daniel Dols-Serrate), the Helen Hay Whitney Foundation Postdoc fellowship (Longhua Guo), and the Howard Hughes Medical Institute (Leonid Kruglyak).

Acknowledgements

We would like to express our gratitude to Gema Blasco for her dedication and exceptional maintenance of our planarian samples in the laboratory, and Dr. Giacinta A. Stocchino for preparing and examining the morphological slides that allowed to corroborate the presence of *D. leporii* among our samples. We thank Dr. Laia Leria for her invaluable support and guidance through fieldwork and sampling campaigns, and Dr. Julio Rozas for providing the computational resources needed to conduct the research presented in this paper. Furthermore, we are grateful to Vadim A. Pisarenco and Marta Olivé-Muñiz for their constant support and providing bioinformatic assistance. We would also like to thank Dr. Joan Ferrer Obiol and Guillem Izquierdo-Aranega for their thoughtful insights, which have helped us many times throughout the preparation of this paper.

References

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. Nucleic Acids Research 25, 3389-3402. https://doi.org/10.1093/nar/25.17.3389

Alvarez W., 1972. Rotation of the Corsica-Sardinia microplate. Nature Physical Science 135, 103-105. https://doi.org/10.1038/physci235103a0

Alvarez W., Cocozza T., Wezel F.C., 1974. Fragmentation of the Alpine orogènic belt by microplate dispersal. Nature 248, 309-314. https://doi.org/10.1038/248309a0

Appelhans M.S., Paetzold C., Wood K.R., Wagner W.L., 2020. RADseq resolves the phylogeny of Hawaiian *Myrsine* (Primulaceae) and provides evidence for hybridization. Journal of Systematics and Evolution 58, 823-840. https://doi.org/10.1111/jse.12668

Baird N.A., Etter P.D., Atwood T.S., Currey M.C., Shiver A.L., Lewis Z.A., Selker E.U., Cresko W.A., Johnson E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3, 1-7. https://doi.org/10.1371/journal.pone.0003376

Barbato D., Benocci A., Manganelli G., 2018. The biogeography of non-marine molluscs in the Tuscan Archipelago reveals combined effects of current eco-geographical drivers and paleogeography. Organisms Diversity and Evolution 18, 443-457. https://doi.org/10.1007/s13127-018-0378-2

Bayona-Vásquez N.J., Glenn T.C., Kieran T.J., Pierson T.W., Hoffberg S.L., Scott P.A., Bentley K.E., Finger J.W., Louha S., Troendle N., Diaz-James P., Mauricio R., Faircloth B.C., 2019. Adapterama III: quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). PeerJ 7:e7724. https://doi.org/10.7717/peerj.7724

Benazzi M., Benazzi-Lentati G., 1976. Platyhelminthes. In: John B. (ed.), *Animal Cytogenesis* Gebrueder Borntraeger, Berlin-Stuttgart, 1: 1-177.

Benazzii-Lentati G., 1979. Polyploidy and types of development in the triclad *Dugesia lugubris* s.l.: interaction between the ploidy level and the origin of amphimixis or pseudogamy. Caryologia 32, 247-263. https://doi.org/10.1080/00087114.1979.10796791

Benazzi-Lentati G., Benazzi M., 1985. On the origin and the true nature of the tetraploid asynaptic planarian called Benazzi's G biotype. Caryologia 38, 269-279. https://doi.org/10.1080/00087114.1985.10797750

Benítez-Álvarez L., Leria L., Fernández R., Mateos E., El Ouanighi Y., Bennas N., El Alami M., Yacoubi-Khebiza M., Ougougdal H.A., Riutort M., 2023. Phylotranscriptomics interrogation uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean. Molecular Phylogenetics and Evolution 178:107649. https://doi.org/10.1016/j.ympev.2022.107649

Beukeboom L.W., Sharbel T.F., Michiels N.K., 1998. Reproductive modes, ploidy distribution, and supernumerary chromosome frequencies of the flatworm *Polycelis nigra* (Platyhelminthes: Tricladida). Hydrobiologia 383, 277-285.

Beukeboom L.W., Vrijenhock R.C., 1998. Evolutionary genetics and ecology of sperm-dependent parthenogenesis. Journal of Evolutionary Biology 11, 755-782. https://doi.org/10.1007/s000360050117

Bisconti R., Canestrelli D., Salvi D., Nascetti G., 2013. A geographic mosaic of evolutionary lineages within the insular endemic newt *Euproctus montanus*. Molecular Ecology 22, 143-156. https://doi.org/10.1111/mec.12085

Blighe K., Lun A., 2023. PCAtools: everything principal component analysis. https://doi.org/10.18129/B9.bioc.PCAtools. R package version 2.14.0. Available from: https://bioconductor.org/packages/PCAtools (last accessed October 2023).

Bourgeois Y.X.C., Warren B.H., 2021. An overview of current population genomics methods for the analysis of whole-genome resequencing data in eukaryotes. Molecular Ecology 30, 6036-6071. https://doi.org/10.1111/mec.15989

Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A., RoyChoudhury A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29, 1917-1932. https://doi.org/10.1093/molbev/mss086

Camacho C., Coulouris G., Avagyan V., Ma N., Papadodoulos J., Bealer K., Madden T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421-430. https://doi.org/10.1186/1471-2105-10-421

Cariou M., Duret L., Charlat S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and Evolution 3, 846-852. https:/doi.org/10.1002/ece3.512

Challis R., Richards E., Rajan J., Cochrane G., Blaxter M., 2020. BlobToolKit – Interactive

quality assessment of genome assemblies. G3: Genes, Genomes, Genetics 10, 1361-1374. https://doi.org/10.1534/g3.119.400908

Chambers E.A., Hillis D.M., 2020. The multispecies coalescent over-splits species in the case of greographically widespread taxa. Systematic Biology 69, 184-193. https://doi.org/10.1093/sysbio/syz042

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., Lee J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4:7. https://doi.org/10.1186/s13742-015-0047-8

Chifman J., Kubatko L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317-3324. https://10.1093/bioinformatics/btu530

Chinone A., Nodono H., Matsumoto M., 2014. Triploid planarian reproduces truly bisexually with euploid gametes produced through a different meiotic system between sex. Chromosoma 123, 265-272. https://doi.org/10.1007/s00412-013-0449-2

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R., 1000 Genomes Project Analysis Group., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

Deri P., 1971. Ricerche cariologiche e di biología della riproduzione in popolazioni di *Dugesia benazzii* della Corsica. Atti della Società Toscana di Scienze Naturali 78, 55-74.

Deri P., 1981. Nuove osservazioni cariologiche e di biología riproduttiva su populazioni di *Dugesia benazzii* (Triclade paludicolo) della Corsica. Atti della Società Toscana di Scienze Naturali 88, 61-72.

Dols-Serrate D., Leria L., Aguilar J.P., Stocchino G.A., Riutort M., 2020. *Dugesia hepta* and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex? Organisms Diversity and Evolution 20, 369-386. https://doi.org/10.1007/s13127-020-00438-z

Dols-Serrate D., Stocchino G.A., Sluys R., Riutort M., 2024. Fantastic beasts and how to delimit them: an integrative approach using multispecies coalescent methods reveals two new, endemic *Dugesia* species (Platyhelminthes, Tricladida) from Corsica and Sardinia. Zoological Journal of the Linnean Society. In Press. https://doi.org/10.1093/zoolinnean/zlad135

Dols-Serrate D., Stocchino G.A., Nuin-Villabona P., Sluys R., Riutort M., 2025. New insights into the evolution and biogeography of freshwater planarians on islands in the Tyrrhenian Sea, Western Mediterranean Basin, with the integrative description of a new endemic species from Corsica (Platyhelminthes, Tricladida, *Dugesia*). Zoological Journal of the Linnean Society. In review.

Eaton D.A.R., Ree R.H., 2013. Inferring phylogeny and introgression using RADseq data:

an example from flowering plants (*Pedicularis*: Orobanchaceae). Systematic Biology 62, 689-706. https://doi.org/10.5061/dryad.bn281

Eaton D.A.R., Spriggs E.L., Park B., Donoghue M.J., 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Systematic Biology 66, 399-412. https://doi.org/10.1093/sysbio/syw092

Etter P.D., Preston J.L., Bassham S., Cresko W.A., Johnson E.A., 2011. Local *de novo* assembly of RAD pair-end contigs using short sequencing reads. PLoS ONE 6:e18561. https://doi.org/10.1371/journal.pone.0018561

Evanno G., Regnaut S., Goudet J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14, 2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fattorini S., 2009. Faunal patterns in tenebrionids (Coleoptera: Tenebrionidae) on the Tuscan Islands: the dominance of paleogeography over recent geography. European Journal of Entomology 106, 415-423.

Ferrer Obiol J., Herranz J.M., Paris J.R., Whiting J.R., Rozas J., Riutort M., González-Solís J., 2023. Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds. Molecular Phylogenetics and Evolution 179:107671. https://doi.org/10.1016/j.ympev.2022.107671

Ferrer Obiol J., James H.F., Chesser R.T., Bretagnolle V., González-Solís J., Rozas J., Riutort M., Welch A.J., 2021. Integrating sequence capture and restriction site-associated DNA sequencing to resolve recent radiations of pelagic seabirds. Systematic Biology 70, 976-996. https://doi.org/10.1093/sysbio/syaa101

Flouri R., Jiao X., Rannala B., Yang Z., 2018. Species tree inference with BPP using genomic sequences and the multispecies coalescent. Molecular Biology and Evolution 35, 2585-2593. https://doi.org/10.1093/molbev/msy147

Frajman B., Záveská E., Gamisch A., Moser T., The STEPPE Consortium, Schönswetter P., 2019. Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana* s.l. (Euphorbiaceae). Molecular Phylogenetics and Evolution 134, 238-252. https://doi.org/10.1016/j.ympev.2018.10.046

Girard C.A., 1850. A brief account of the fresh water planariae of the United States. Proceedings of the Boston Society of Natural History 3, 264-265.

Gompert Z., Buerkle C.A., 2010. INTROGRESS: a software package for mapping components of isolation in hybrids. Molecular Ecology Resources 10, 378-384. https://doi.org/10.1111/j.1755-0998.2009.02733.x

Guindon S., Dufayard J-F., Lefort V., Anisimova M., Hordijk W., Gascuel O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the per-

formance of PhyML 3.0. Systematic Biology 59, 307-321. https://doi.org/10.1093/sysbio/syq010

Harrath A.H., Sluys R., Mansour L., Folefack G.L., Aldahmash W., Alwasel S., Solà E., Riutort M., 2019. Molecular and morphological indentification of two new African species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from Cameroon. Journal of Natural History 53, 253-271. https://doi.org/10.1080/00222933.2019.1577508

Harvey M.G., Smith B.T., Glenn T.C., Faircloth B.C., Brumfield R.T., 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Systematic Biology 65, 910-924. https://doi.org/10.1093/sysbio/syw036

Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35, 518-522. https://doi.org/10.5281/zenodo.854445

Hosegood J., Humble E., Ogden R., de Bruyn M., Creer S., Stevens G.M.W., Abudaya M., Bassos-Hull K., Bonfil R., Fernando D., Foote A.D., Hipperson H., Jabado R.W., Kaden J., Moazzam M., Peel L.R., Pollett S., Ponzo A., Poortvliet M., Salah J., Senn H., Steward J.D., Wintner S., Carvalho G., 2020. Phylogenomics and species delimitation for effective conservation of manta and devil rays. Molecular Ecology 29, 4783-4796. https://doi.org/10.1111/mec.15683

Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14, 587-589. https://doi.org/10.1038/nmeth.4285

Kass R.E., Raftery A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773-795.

Katoh K., Standley D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30, 772-780. 10.1093/molbev/mst010

Kawakatsu M., Oki I., Tamura S., Sugino H., 1976. Studies on the morphology, karyology and taxonomy of the Japanese freshwater planarian *Dugesia japonica* Ichikawa and Kawakatsu, with a description of a new subspecies, *Dugesia japonica ryukyuensis* subspec. nov. Bulletin of Fuji Women`s College no. 14, ser. II, 81-126.

Kishino H., Miyata T., Hasegawa M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. Journal of Molecular Evolution 31, 151-160.

Kuhlemann J., Frisch W., Székely B., Dunkl I., Danišík M., Krumrei I., 2005. Würmian maximum glaciation in Corsica. Austrian Journal of Earth Sciences 97, 68-81.

Kuhlemann J., Rohling E.J., Krumrei I., Kubik P., Ivy-Ochs S., Kucera M., 2008. Regional synthesis of Mediterranean atmospheric circulation during the Last Glacial Maximum. Science 321, 1338-1340. https://doi.org/10.1126/science.1157638

Laetsch D.R., Blaxter M.L., 2017. BlobTools: interrogation of genomes assemblies. F1000Research 6:1287. https://doi.org/10.12688/f1000research.12232.1

Lawson D.J., van Dorp L., Falush D., 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature Communications 9:3258. http://dx.doi.org/10.1038/s41467-018-05257-7

Lázaro E.M., Sluys R., Pala M., Stocchino G.A., Baguñà J., Riutort M., 2009. Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). Molecular Phylogenetics and Evolution 52, 835-845. https://doi.org/10.1016/j.ympev.2009.04.022

Leaché A.D., Fujita M.K., Minin V.N., Bouckaert R.R., 2014. Species delimitation using genome-wide SNP data. Systematic Biology 63, 534-542. https://doi.org/10.1093/sysbio/syu018

Lepori N.G., 1951. Sulle caratteristiche morfologiche e sulla posizione sistematica della planaria di Sardegna e Corsica già ascritta a *Dugesia* (=*Euplanaria*) *gonocephala* (Dugès). Atti Società Toscana di Scienze Naturali 58, 28-47.

Leria L., Riutort M., Romero R., Ferrer X., Vila-Farré M., 2022. Microplate tectonics and environmental factors as distribution drivers in Western Mediterranean freshwater planarians. Journal of Biogeography 49, 1124-1136. https://doi.org/10.1111/jbi.14373

Leria L., Vila-Farré M., Álvarez-Presas M., Sánchez-Gracia A., Rozas J., Sluys R., Riutort M., 2020. Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhelminthes, Tricladida): extreme intraindividual genetic diversity, morphological stasis, and karyological variability. Molecular Phylogenetics and Evolution 143:106496. https://doi.org/10.1016/j.ympev.2019.05.010

Leria L., Vila-Farré M., Solà E., Riutort M., 2019. Oustanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. BMC Evolutionary Biology 19, 130-149. https://doi.org/10.1186/s12862-019-1440-1

Linke S., Lehner B., Dallaire C.O., Ariwi J., Grill G., Anand M., Beames P., Burchard-Levine V., Maxwell S., Moidu H., Tan F., Thieme M., 2019. Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Scientific Data 6, 1-15. https://doi.org/10.1038/s41597-019-0300-6

Malinsky M., Matschiner M., Svardal H., 2021. Dsuite – Fast *D*-statistics and related admixture evidence from VCF files. Molecular Ecology Resources 21, 584-595. https://doi.org/10.1111/1755-0998.13265

Malinsky M., Trucchi E., Lawson D.J., Falush D., 2018. RADpainter and fineRADstructure: population inference from RADseq data. Molecular Biology and Evolution 35, 1284-1290. https://doi.org/10.1093/molbev/msy023

Mallet J., 2007. Hybrid speciation. Nature 446, 279-283. https://doi.org/10.1038/na-

ture05706

Mallet J., Besansky N., Hahn M.W., 2016. How reticulated are species? BioEssays 38, 140-149. https://doi.org/10.1002/bies.201500149

Mendes F.K., Hahn M.W., 2018. Why concatenation fails near the anomaly zone. Systematic Biology 67, 158-169. https://doi.org/10.1093/sysbio/syx063

Migliore J., Baumel A., Juin M., Diadema K., Hugot L., Verlaque R., Médail F., 2011. Genetic diversity and structure of a Mediterranean endemic plant in Corsica (*Mercurialis corsica*, Euphorbiaceae). Population Ecology 53, 573-586. https://doi.org/10.1007/s10144-011-0266-5

Mirarab S., Bayzid M.S., Warnow T., 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Systematic Biology 65, 366-380. https://doi.org/10.1093/sysbio/syu063

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37, 1530-1534. https://doi.org/10.1093/molbev/msaa015

Molloy E.K., Warnow T., 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. Systematic Biology 67, 285-303. https://doi.org/10.1093/sysbio/syx077

Newton L.G., Starrett J., Hendrixson B.E., Derkarabetian S., Bond J.E., 2020. Integrative species delimitation reveals cryptic diversity in the southern Appalachian *Antrodiaetus unicolor* (Araneae: Antrodiaetidae) species complex. Molecular Ecology 29, 2269-2287. https://doi.org/10.1111/mec.15483

Nguyen L-T., Schmidt H.A., von Haeseler A., Minh B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32, 268-274. https://doi.org/10.1093/molbev/msu300

Peterson B.K., Weber J.N., Kay E.H., Fisher H.S., Hoekstra H.E., 2012. Double digest RADseq: an inexpensive method for *de novo* discovery and genotyping in model and non-model species. PLoS ONE 7:e37135. https://doi.org/10.1371/journal.pone.0037135

Pala M., Casu S., Lepori N.G., 1982. Stabilized natural interspecific hybrid population of the fresh wáter planarians *Dugesia gonocephala* s.l. (Turbellaria, Tricladida). Caryologia 35, 247-256. https://doi.org/10.1080/00087114.1982.10796929

Pala M., Casu S., Vacca R.A., 1981. *Dugesia hepta*, nuova specie di planaria di acqua dolce di Sardegna appartenente alla superspecie *Dugesia gonocephala* (Dugès). Bolletino della Società Sarda di Scienze Naturali 20, 97-107. https://core.ac.uk/reader/11689490 (last accessed 7 June 2023).

Pala M., Stocchino G.A., Corso G., Casu S., 2000. A new species of freshwater planarian

belonging to the genus *Dugesia* (Platyhelminthes, Tricladida) from Sardinina. Italian Journal of Zoology 67, 379-383. https://doi.org/10.1080/11250000009356343

Paris J.R., Stevens J.R., Catchen J.M., 2017. Lost in parameter space: a road map for STACKS. Methods in Ecology and Evolution 8, 1360-1373. https://doi.org/10.1111/2041-210X.12775

Payseur B.A., Riesenberg L.H., 2016. A genomic perspective on hybridization and speciation. Molecular Ecology 25, 2337-2360. https://doi.org/10.1111/mec.13557

Phillips S.J., Anderson R.P., Dudík M., Schapire R.E., Blair M.E., 2017. Opening the black box: an open-source release of Maxent. Ecography 40, 887-893. https://doi.org/10.1111/ecog.03049

Phillips S.J., Dudík M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161-175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

Pina-Martins F., Silva D.N., Fino J., Paulo O.S., 2017. *Structure_threader*: an improves method for automation and parallelization of programs STRUCTURE, FASTSTRUCTURE and *MavericK* on multicore CPU systems. Molecular Ecology Resources 17, e268-e274. https://doi.org/10.1111/1755-0998.12702

Pongratz N., Sharbel T.F., Beukeboom L.W., Michiels N.K., 1998. Allozyme variability in sexual and parthenogenetic freshwater planarians: evidence for polyphyletic origin of parthenogenetic lineages through hybridization with coexisting sexuals. Heredity 81, 38-47. https://doi.org/10.1046/j.1365-2540.1998.00359.x

Pritchard J.K., Stephens M., Donnelly P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945-959. https://doi.org/10.1093/genetics/155.2.945

Ravinet M., Faria R., Butlin R.K., Galindo J., Bierne N., Rafajlović M., Noor M.A.F., Mehlig B., Westram A.M., 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. Journal of Evolutionary Biology 30, 1450-1477. https://doi.org. 10.1111/jeb.13047

Rochette N.C., Catchen J.M., 2017. Deriving genotypes from RAD-seq short-read data using Stacks. Nature Protocols 12, 2640-2659. http://dx.doi.org/10.1038/nprot.2017.123

Rochette N.C., Rivera-Colón A.G., Catchen J.M., 2019. Stacks 2: analytical methods for pair-end sequencing improve RADseq-based population genomics. Molecular Ecology 28, 4737-4754. https://doi.org/10.1111/mec.15253

Rosenbaum G., Lister G.S., Duboz C., 2002. Reconstruction of the tectonic evolution of the western Mediterranean since the Oligocene. Journal of the Virtual Explorer 8, 107-130. https://doi.org/10.3809/jvirtex.2002.00053

Rubin B.E.R., Ree R.H., Moreau C.S., 2013. Inferring phylogenies from RAD sequence

data. PLoS ONE 7:e33394. https://doi.org/10.1371/journal.pone.0033394

Salvi D., Pinho C., Harris D.J., 2017. Digging up the roots of an insular hotspot of genetic diversity: decoupled mito-nuclear histories in the evolution of the Corsican-Sardinian endemic lizard *Podarcis tiliguerta*. BMC Evolutionary Biology 17:63. https://doi.org/10.1186/s12862-017-0899-x

Schweyen H., Rozenberg A., Leese F., 2014. Detection and removal of PCR duplicates in population genomics ddRAD studies bu addition of a degenerate base region (DBR) in sequencing adapters. Biological Bulletin 227, 146-160. https://doi.org/10.1086/BBLv227n2p146

Sekino M., Hashimoto K., Nakamichi R., Yamamoto M., Fujinami Y., Sasaki T., 2023. Introgressive hybridization in the west Pacific pen shells (genus *Atrina*): restricted interspecies gene flow within the genome. Molecular Ecology 32, 2945-2963. https://doi.org/10.1111/mec.16908

Shimodaira H., 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology 51, 492-508. https://doi.org/10.1080/10635150290069913

Sluys R., Riutort M., 2018. Planarian diversity and phylogeny. In: Rink, J. C. (ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology*, Methods in Molecular Biology, Vol. 1774. Springer Science+Business Media, New York, 1-56. https://doi.org/10.1007/978-1-4939-7802-1_1

Sluys R., Solà E., Gritzalis K., Vila-Farré M., Mateos E., Riutort M., 2013. Integrative delineation of species of Mediterranean freshwater planarians (Platyhelminthes: Tricladida: Dugesiidae). Zoological Journal of the Linnean Society 169, 532-547. https://doi.org/10.1111/zoj.12077

Solà E., Leria L., Stocchino G.A., Bagherzadeh R., Balke M., Daniels S.R., Harrath A.H., Khang T.S., Krailas D., Kumar B., Li M-H., Maghsoudlou A., Matsumoto M., Naser N., Ober B., Segev O., Thielicke M., Tong X., Zivanovic G., Manconi R., Baguñà J., Riutort M., 2022. Three dispersal routes out of Africa: a puzzling biogeographical history in freshwater planarians. Journal of Biogeography 49, 1219-1233. https://doi.org/10.1111/jbi.14371

Solà E., Sluys R., Gritzalis K., Riutort M., 2013. Fluvial basin history in the northeastern Mediterranean region underlies dispersal and speciation patterns in the genus *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae). Molecular Phylogenetics and Evolution 66, 877-888. https://doi.org/10.1016/j.ympev.2012.11.010

Stobie C.S., Oosthuizen C.J., Cunningham M.J., Bloomer P., 2018. Exploring the phylogeography of a hexaploid freshwater fish by RAD sequencing. Ecology and Evolution 8, 2326-2342. https://doi.org/10.1002/ece3.3821

Stocchino G.A., 2018. 80 years of research on planarians (Platyhelminthes, Tricladida) from Sardinia, Italy: an annotated checklist. Zootaxa 4532, 539-552. https://doi.org/10.11646/zootaxa.4532.4.5

Stocchino G.A., Corso G., Manconi R., Casu S., Pala M., 2005. Endemic freshwater planarians of Sardinia: redescription of *Dugesia hepta* (Platyhelminthes, Tricladida) with a comparison of the Mediterranean species of the genus. Journal of Natural Histology 39, 1947-1960. https://doi.org/10.1080/00222930500060025

Stocchino G.A., Dols-Serrate D., Sluys R., Riutort M., Onnis C., Manconi R., 2021. Amphibioplanidae: a new branch and family on the phylogenetic tree of the triclad flatworms (Platyhelminthes: Tricladida), represented by a species from Sardinian caves with a remarkable lifestyle. Zoological Journal of the Linnean Society 193, 1364–1391. https://doi.org/10.1093/zoolinnean/zlaa183

Stocchino G.A., Manconi R., 2013. Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). Italian Journal of Zoology 803, 319-328. ttp://dx.doi.org/10.1080/11250003.2013.822025

Stocchino G.A., Sluys R., Deri P., Manconi R., 2013. Integrative taxonomy of a new species of planarian from the Lake Ohrid basin, including an analysis of biogeographical patterns in freshwater triclads from the Ohrid region (Platyhelminthes, Tricladida, Dugesiidae). ZooKeys 313, 25-43. https://doi.org/10.3897/zookeys.313.5363

Suc J.P., 1984. Origin and evolution of the Mediterranean vegetation and climate in Europe. Nature 307, 429-432.

Sukumaran J, Knowles L.L. (2017). Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences 114, 1607-1612. https://doi.org/10.1073/pnas.1607921114

Swofford D., 2003. PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Tonzo V., Papadopoulou A., Ortego J., 2019. Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. Molecular Ecology 28, 3869-3886. https://doi.org/10.1111/mec.15189

Tripp E.A., Tsai Y-H. E., Zhuang Y., Dexter K.G., 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. Ecology and Evolution 7, 7920-7936. https://doi.org/10.1002/ece3.3274

Winsor L., Sluys R., 2018. Basic histological techniques for planarians. In: Rink, J. C. (ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology*, Methods in Molecular Biology, Vol. 1774. Springer Science+Business Media, New York, 285-351. https://doi.org/10.1007/978-1-4939-7802-1_9

Yang Z., 2015. The BPP program for species tree estimation and species delimitation. Current Zoology 61, 854-865. https://doi.org/10.1093/czoolo/61.5.854

Yang X., Rannala B., 2010. Bayesian species delimitation using multilocus sequence data. Genetics 164, 1645-1656. https://doi.org/10.1073/pnas.0913022107

Supplementary Material: Figures

Mixed, not stirred: Genomic analyses of *Dugesia* (Platyhelminthes, Tricladida) shed light on the evolutionary history and diversification of the Corso-Sardinian clade and reveal the first case of interspecific hybridization in freshwater flatworms

Daniel Dols-Serrate[1,2], Longhua Guo[3,4], Leonid Kruglyak[5,6] & Marta Riutort[1,2]

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[3] Geriatrics Center and Institute of Gerontology, University of Michigan, Ann Arbor, Michigan, USA

[4] Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, Michigan, United States

[5] Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

[6] Department of Human Genetics, Department of Biological Chemistry, University of California, Los Angeles, California, USA

ORCIDs:
0000-0003-0820-7324 Daniel Dols-Serrate

0000-0001-9690-9750 Longhua Guo

0000-0002-8065-3057 Leonid Kruglyak

0000-0002-2134-7674 Marta Riutort

Corresponding author: D. Dols-Serrate danieldols@ub.edu/bcn27101993@gmail.com

Table of Contents

Supplementary Fig. S1 – Map of localities visited in 2019-2021 sampling campaign **239**

Supplementary Fig. S2 – Example of chromosome complement of a mixoploid individual **240**

Supplementary Fig. S3 – Mitochondrial gene tree used for species identification **241**

Supplementary Fig. S4 – Phylogenetic trees including all samples **242**

Supplementary Fig. S5 – Phylogenetic trees of Corsica and summary of topological results ... **246**

Supplementary Fig. S6 – Phylogenetic trees without hybrids from MC **251**

Supplementary Fig. S7 – SVDQuartets phylogenetic trees **255**

Supplementary Fig. S8 – STRUCTURE analysis results **257**

Supplementary Fig. S9 – PCA analysis results for all samples and all ingroup samples **260**

Supplementary Fig. S10 – PCA analysis results of Corsican and Sardinian subsample sets **262**

Supplementary Fig. S11 – fineRADstructure analysis results **263**

Supplementary Fig. S12 – PCA results for candidate parental-hybrid trios **266**

Supplementary Fig. S13 – Summary of INTROGRESS and ancestry-painting analyses results for candidate parental-hybrid trios **267**

Supplementary Fig. S14 – ABBA/BABA test results **268**

Supplementary Fig. S15 – Time-calibrated SNAPP trees **269**

Supplementary Fig. S16 – Cloudograms of SNAPP trees **270**

Supplementary Fig. S17 – Summary of polymorphic loci prevalence in loci catalogs **271**

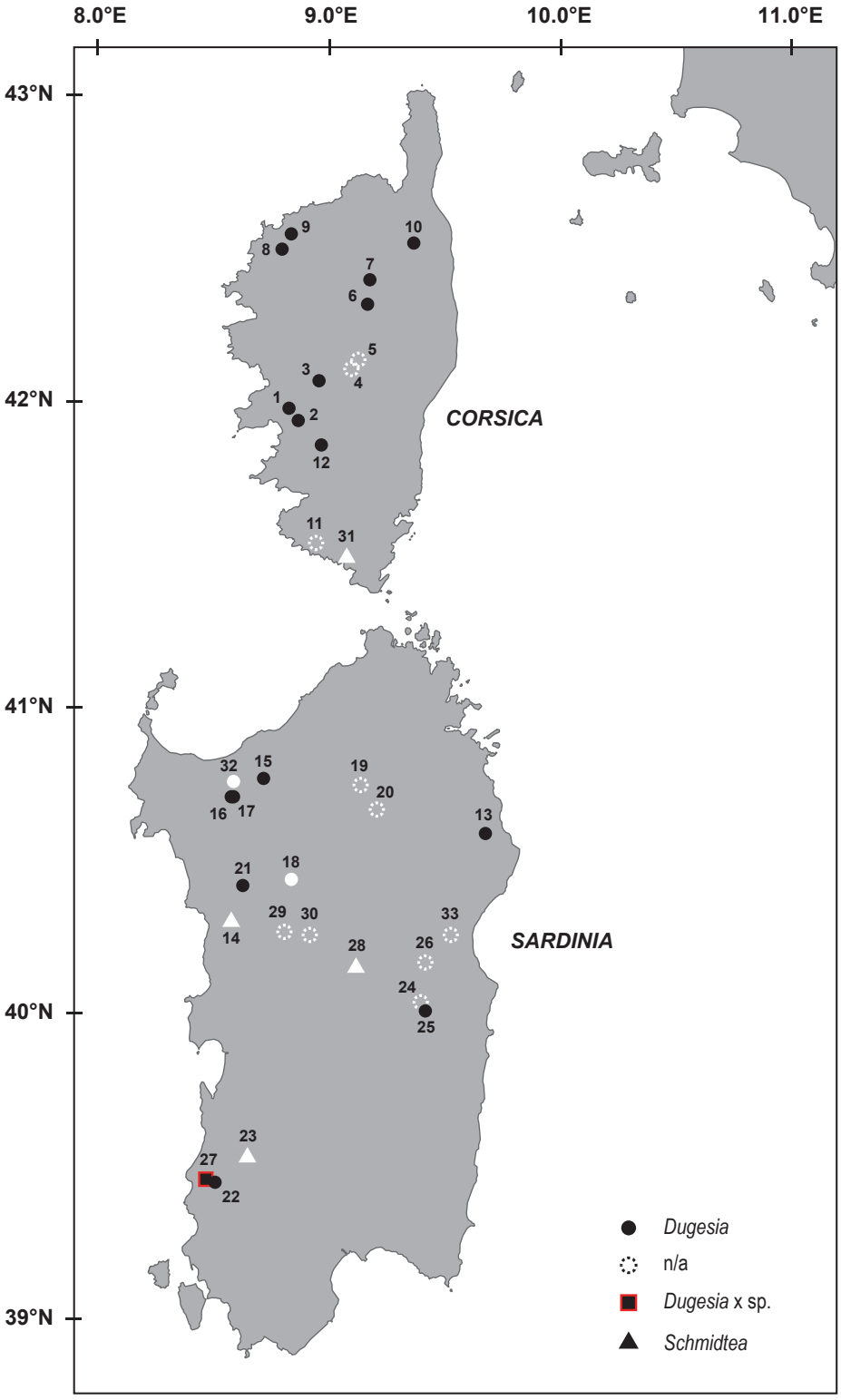


Fig. S1. Map of Corsica and Sardinia showing all localities visited between 2019 and 2021 (see Supplemental Information Table S1). Localities featured in the present study are presented in black. Solid circles indicate presence of *Dugesia* specimens. Triangles indicate presence of *Schmidtea* specimens. Distribution of hybrids between *D. leporii* and *D. benazzii* reported in this study are indicated with a black-red square. n/a indicates absence of freshwater flatworms in the visited locality.

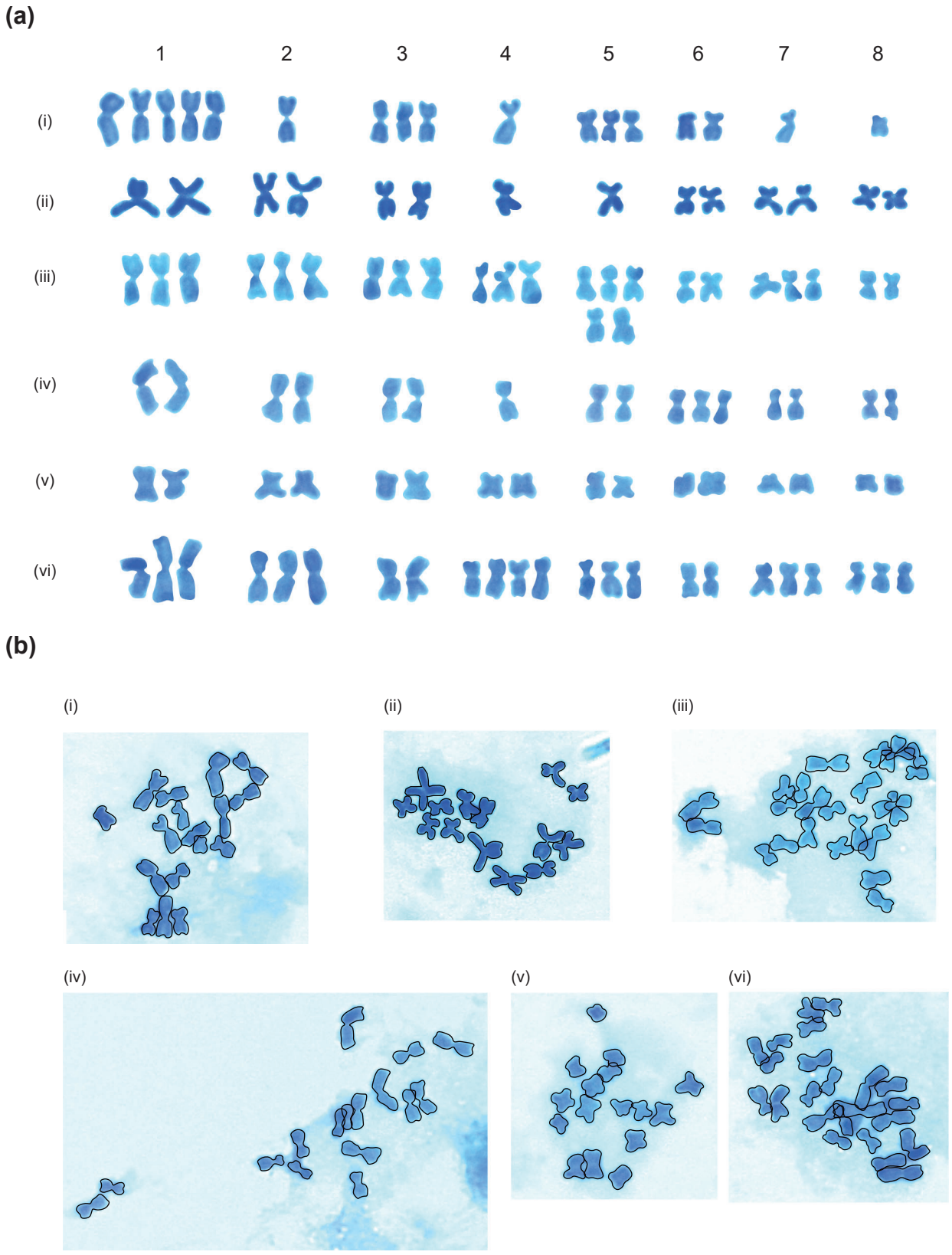


Fig. S2. Example of chromosome complements from mixoploid individual MC1397-08, a natural hybrid between *D. leporii* and *D. benazzii*. (a) Chromosome sets arranged in pairs, triplets or more. Western digits (1-8) indicate the 8 types of chrosomosomes present, the haploid complement, differing in size and centromere position. Roman numbers (i-vi) indicate the pertinent metaphase plate of each arrangement shown in (b).

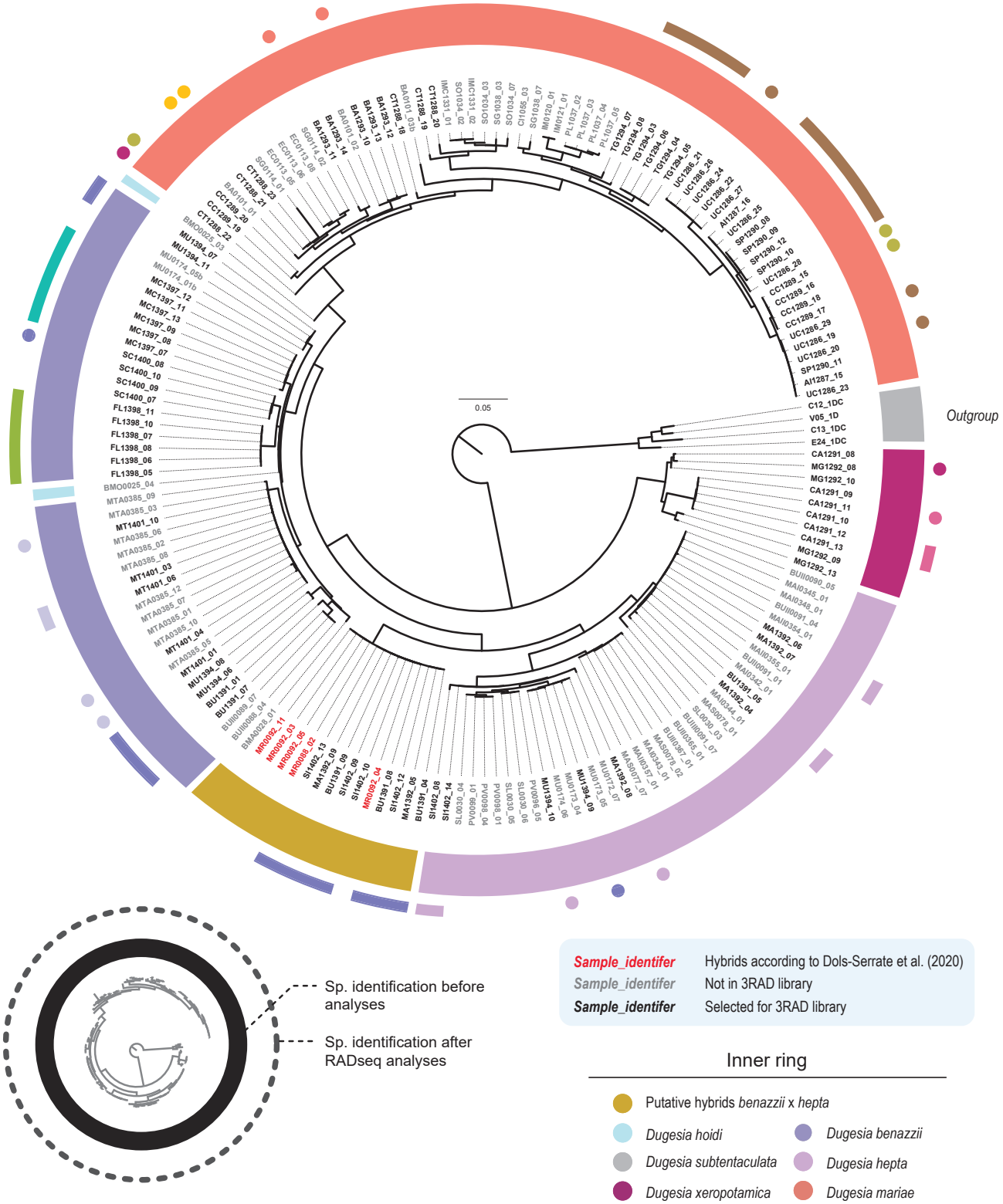
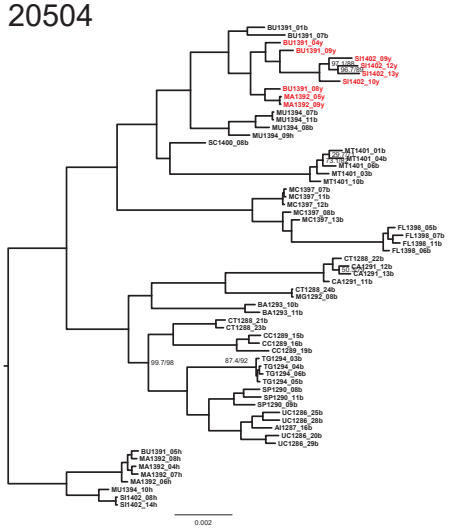


Fig. S3. Phylogenetic tree based on concatenated mitochondrial data (COI + NADH1-tRNAW-COII) used for species identification. Inner colored ring shows species identification prior to the present study analyses. Outer colored ring shows resulting species delineation into primary species candidates (PSC) after analyses. RADseq samples that yielded depths of coverage <10x were not analyzed, hence the lack of affiliation to a PSC.

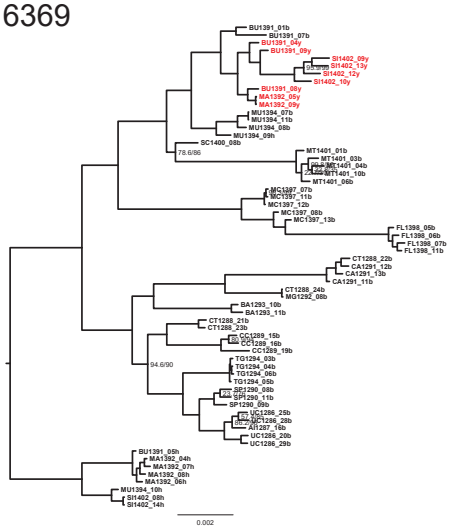
Fig. S4. Phylogenetic trees inferred with IQ-Tree 2.0 of (a) optimal, (b) default, and (c) over-merged loci datasets using nine different combinations of taxon coverage parameters using all samples with a depth of coverage equal to or greater than 10x. Nodal support values are only given for nodes with less than 100/100 (SH-aLRT test/UFboot) support. Putative *hepta* x *benazzii* hybrids identified according to Dols-Serrate et al. (2020) criterion (see Supplementary Fig. S3) are indicated in red. Names of the employed datasets and their pertinent number of polymorphic loci are also shown over each inferred tree. (For a better interpretation of this supplementary figure, we recommend the reader to examine it in a non-physical format) (pp 243-245).

(a)

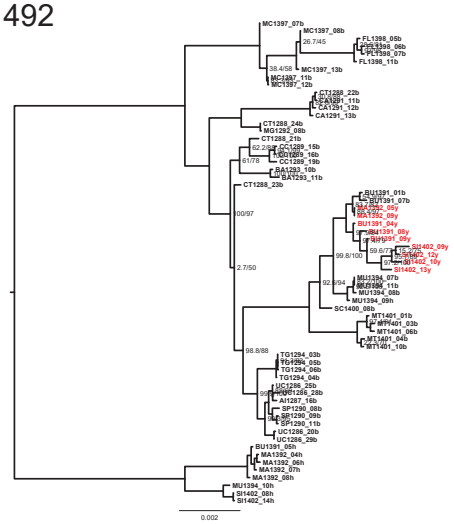
OP-0.4P50
20504



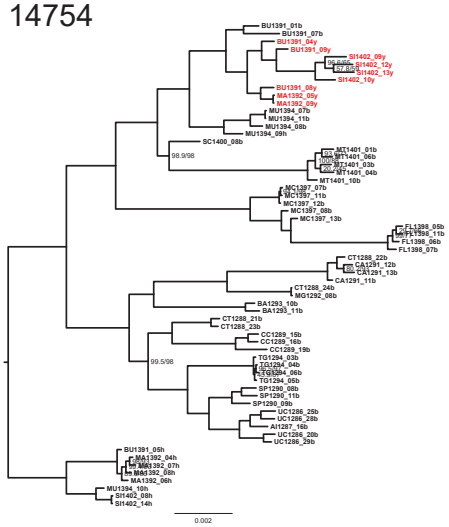
OP-0.4P75
6369



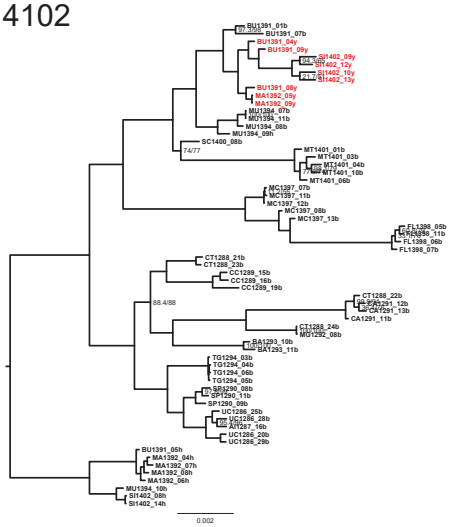
OP-0.4P100
492



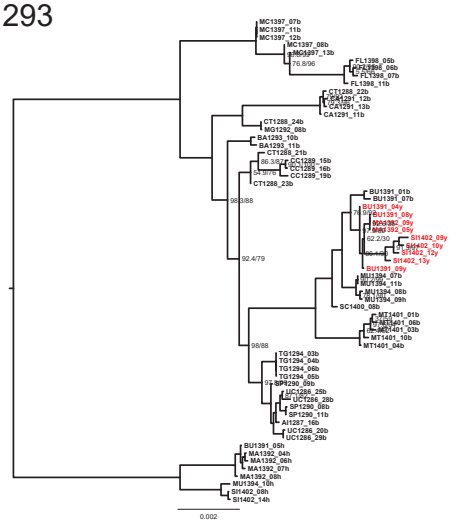
OP-0.6P50
14754



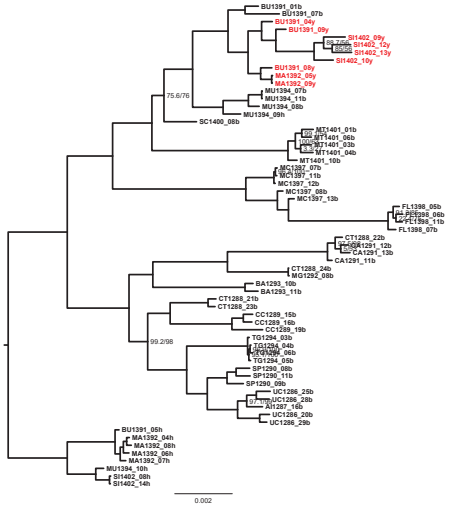
OP-0.6P75
4102



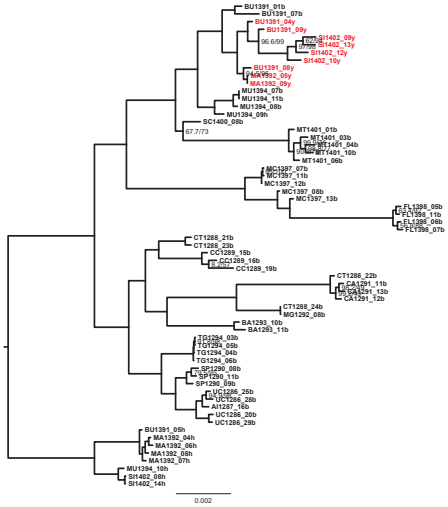
OP-0.6P100
293



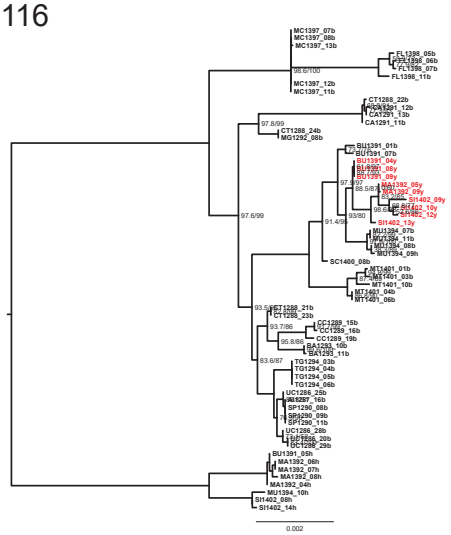
OP-0.7P50
11390



OP-0.7P75
2653

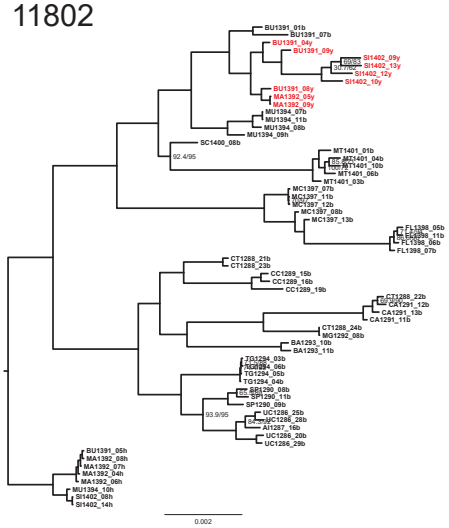


OP-0.7P100
116

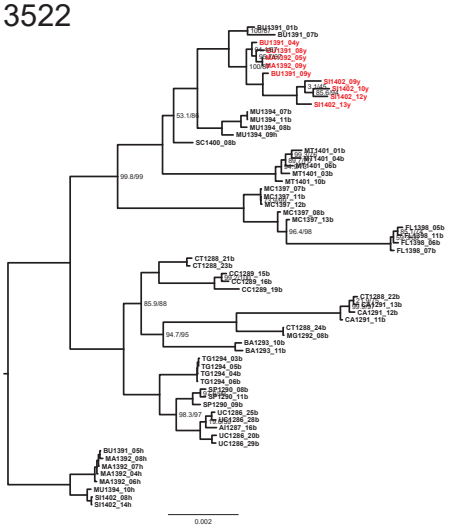


(b)

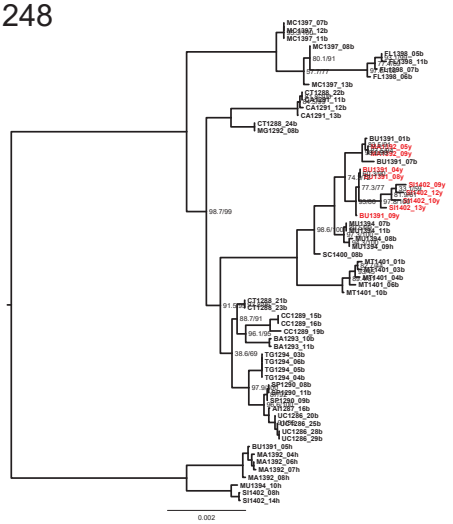
DF-0.4P50
11802



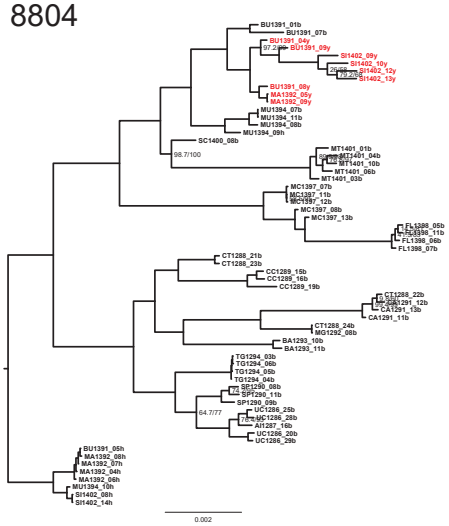
DF-0.4P75
3522



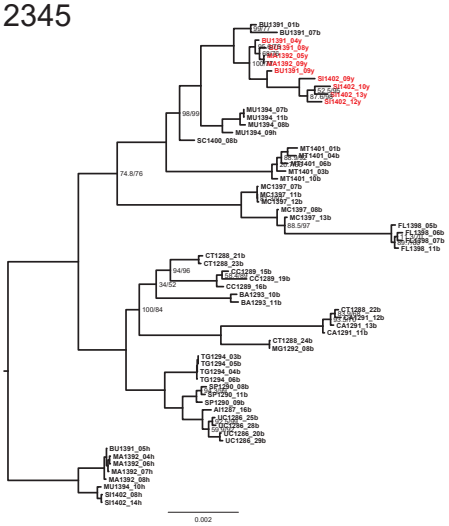
DF-0.4P100
248



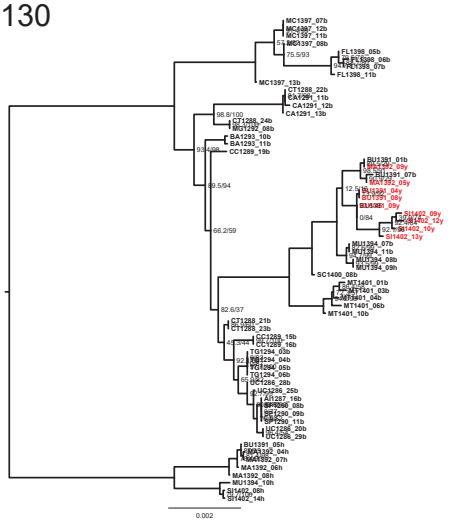
DF-0.6P50
8804



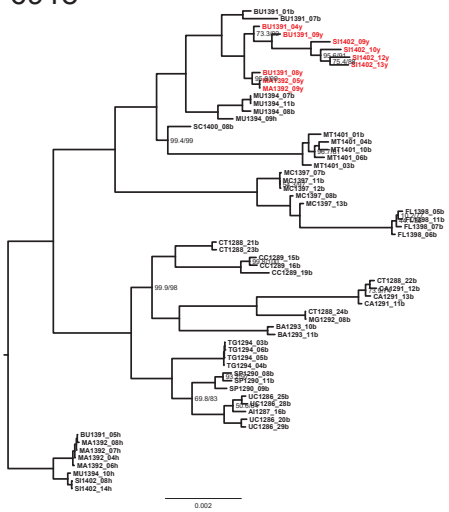
DF-0.6P75
2345



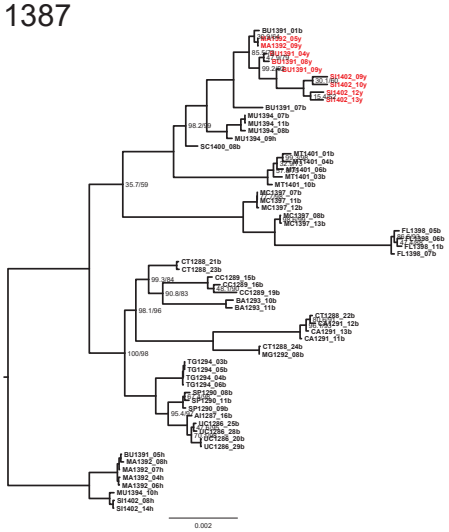
DF-0.6P100
130



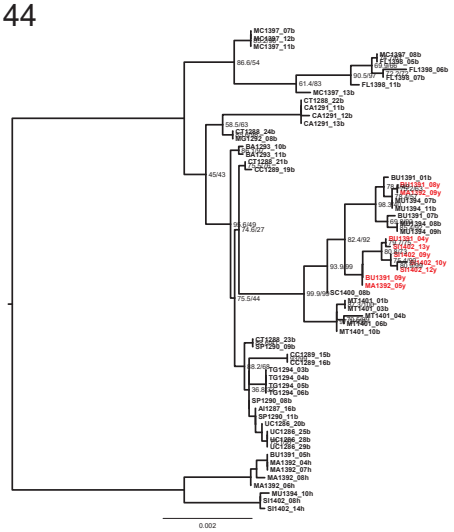
DF-0.7P50
6918



DF-0.7P75
1387

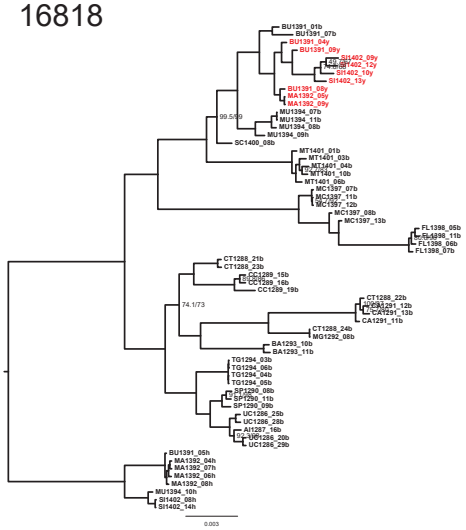


DF-0.7P100
44

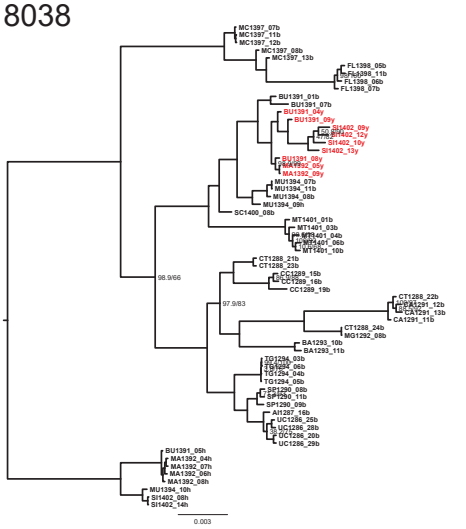


(c)

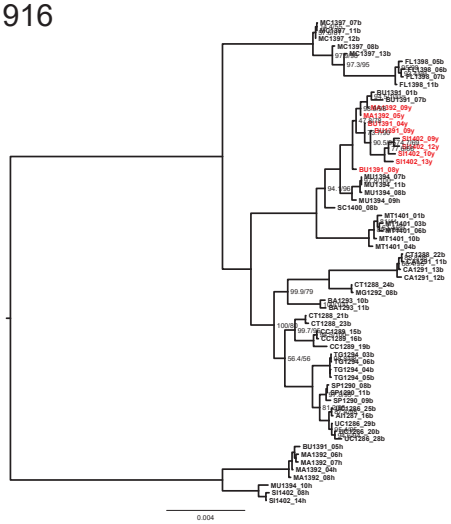
OVM-0.4P50
16818



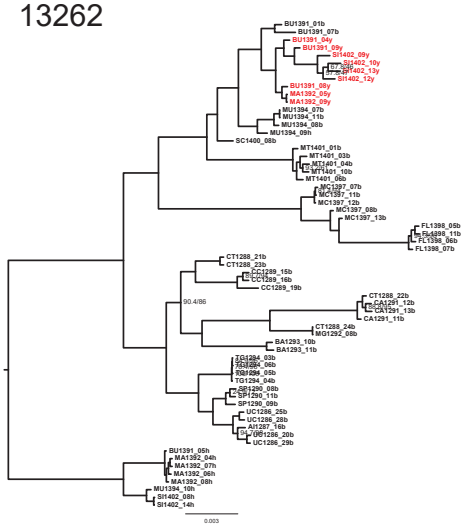
OVM-0.4P75
8038



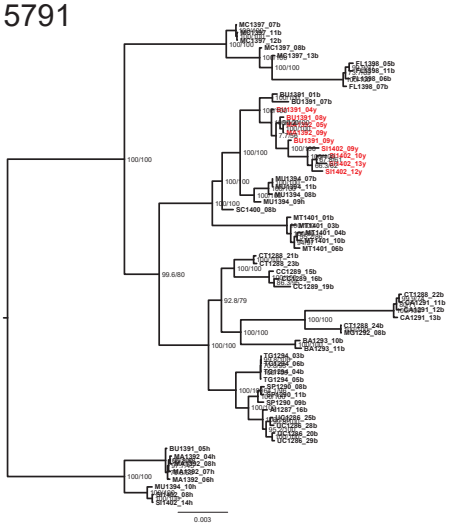
OVM-0.4P100
916



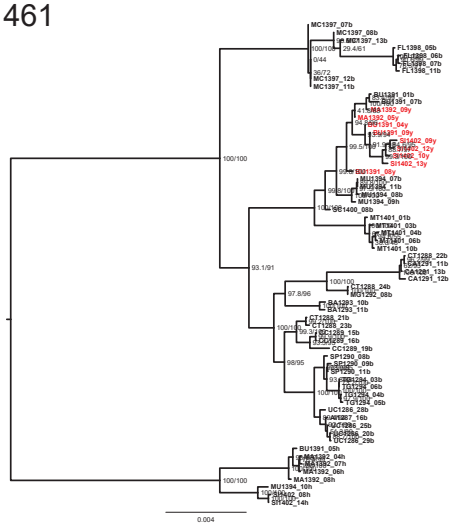
OVM-0.6P50
13262



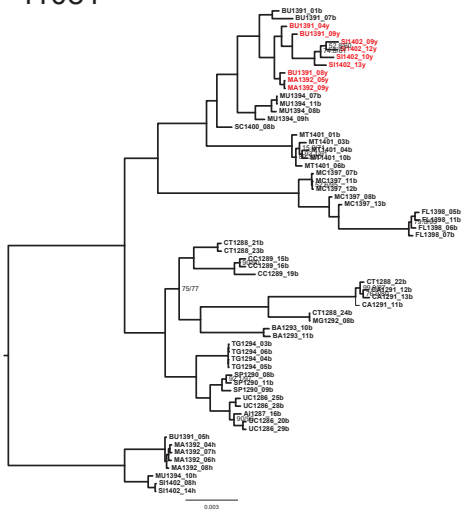
OVM-0.6P75
5791



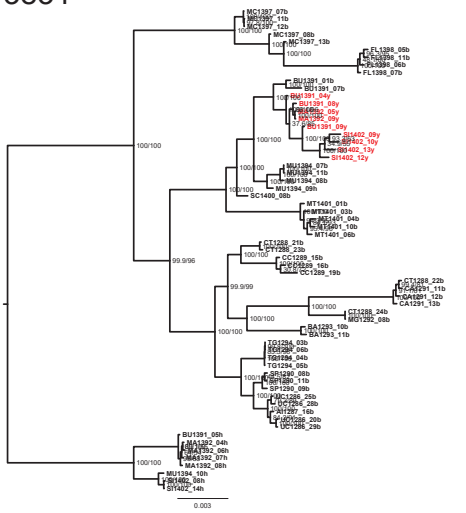
OVM-0.6P100
461



OVM-0.7P50
11081



OVM-0.7P75
3831



OVM-0.7P100
127

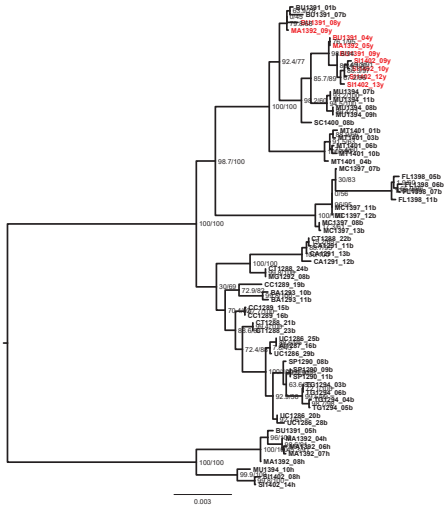
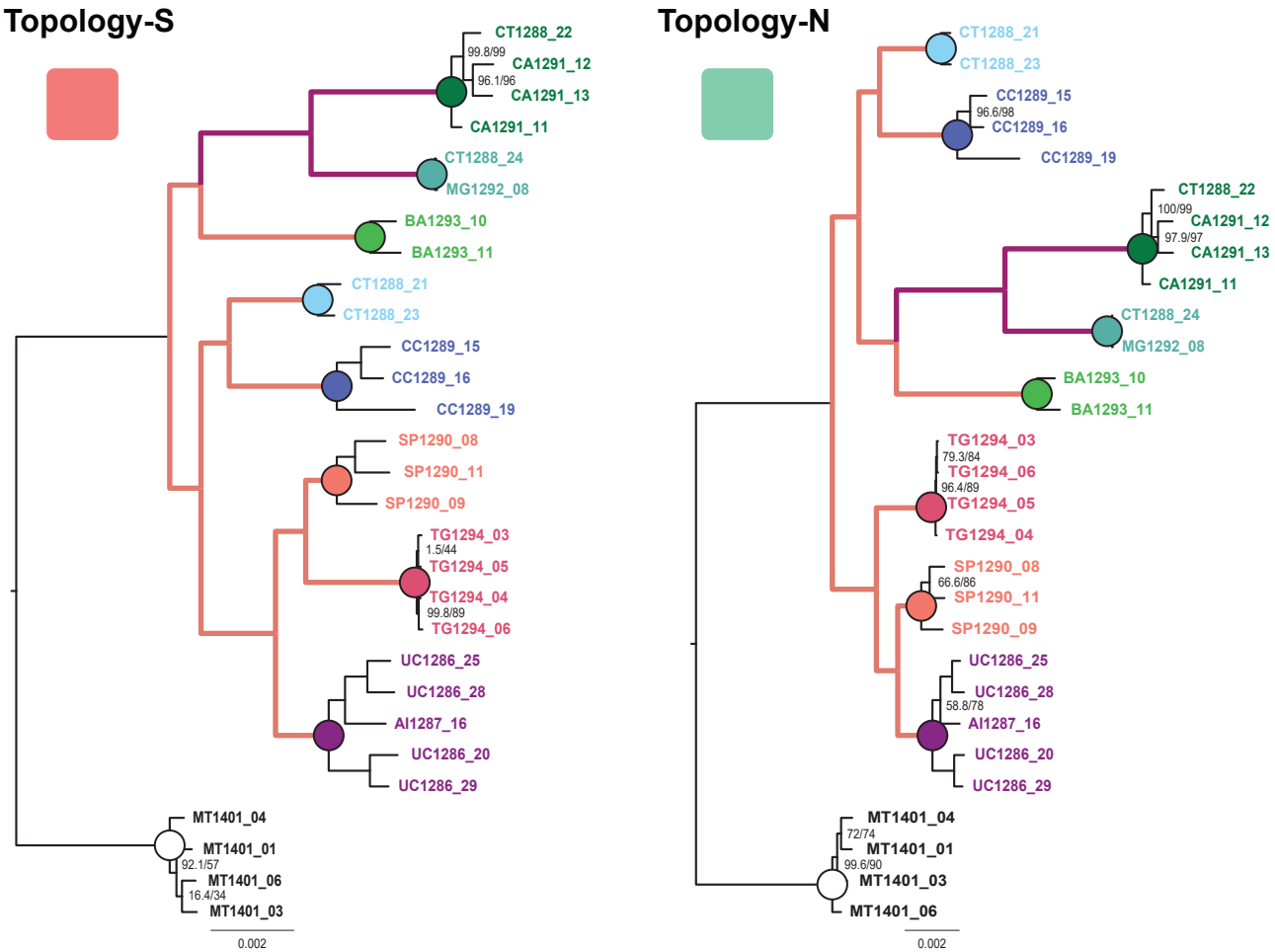


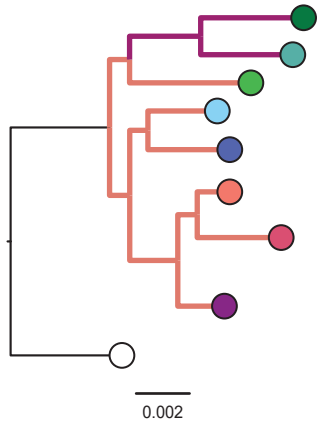
Fig. S5. (a) Phylogenetic trees inferred with IQ-Tree 2.0 of all Corsican samples, using sardinian population MT as outgroup. Color coded nodes and tips represent all subclades of interest and their relative phylogenetic position according to topology-S (left, based on OP-0.4CP50: cyan and blue northern subclades sister to southern subclades, viz. orange, magenta, and violet) or topology-N (right, based on OVM-0.4CP50: cyan and blue subclades sister to northern subclades, viz. light green, teal, and dark green). Branches are colored by taxon. Below both trees are represented the topologies yielded by each datasets using nine different taxon coverage parameters. Red: topology-S. Green: topology-N. Node supports are shown for nodes with less than 100/100 (SH-aLRT/UFboot) support. Number of polymorphic loci of each dataset are shown below its name. (b) Topological results for OP datasets. (c) Topological results for DF datasetets. (d) Topological results for OVM datasets (pp 247-250).

(a)

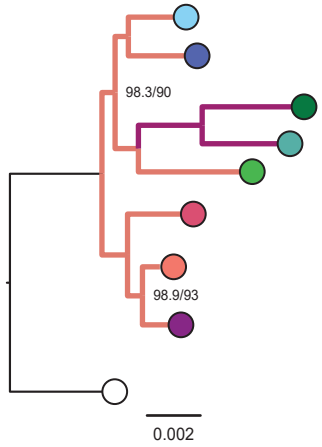

Topology-S
CT1288_22
99.8/99
CA1291_12
96.1/96
CA1291_13
CA1291_11
CT1288_24
MG1292_08
BA1293_10
BA1293_11
CT1288_21
CT1288_23
CC1289_15
CC1289_16
CC1289_19
SP1290_08
SP1290_11
SP1290_09
TG1294_03
1.5/44
TG1294_05
TG1294_04
99.8/89
TG1294_06
UC1286_25
UC1286_28
AI1287_16
UC1286_20
UC1286_29
MT1401_04
MT1401_01
92.1/57
MT1401_06
16.4/34
MT1401_03
0.002
Topology-N
CT1288_21
CT1288_23
CC1289_15
96.6/98
CC1289_16
CC1289_19
CT1288_22
100/99
CA1291_12
97.9/97
CA1291_13
CA1291_11
CT1288_24
MG1292_08
BA1293_10
BA1293_11
TG1294_03
79.3/84
TG1294_06
96.4/89
TG1294_05
TG1294_04
SP1290_08
66.6/86
SP1290_11
SP1290_09
UC1286_25
UC1286_28
58.8/78
AI1287_16
UC1286_20
UC1286_29
MT1401_04
72/74
MT1401_01
99.6/90
MT1401_03
MT1401_06
0.002
Dugesia mariae
Dugesia xeropotamica
OP
P50
P75
P100
0.4
0.6
0.7
DF
P50
P75
P100
0.4
0.6
0.7
OVM
P50
P75
P100
0.4
0.6
0.7

(b)

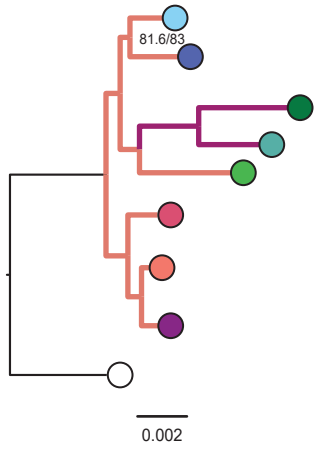
OP-0.4P50
26754



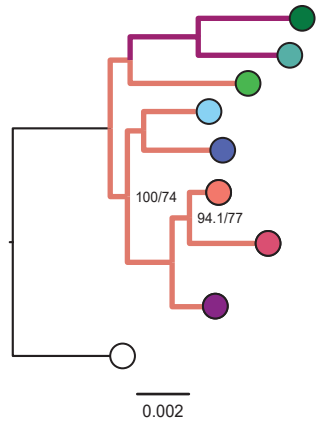
OP-0.4P75
13654



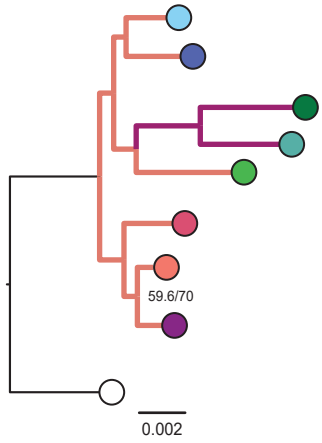
OP-0.4P100
2360



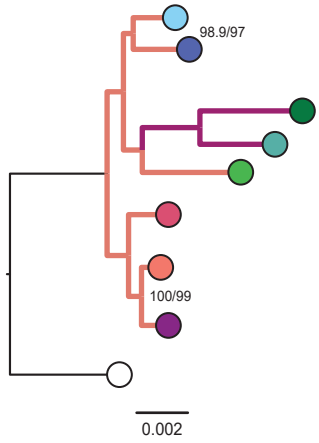
OP-0.6P50
20049



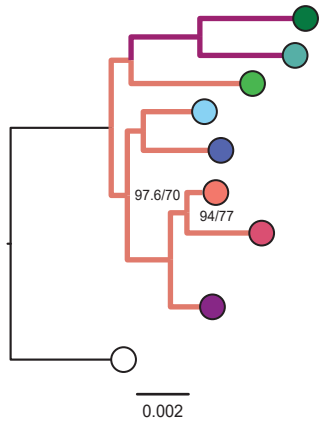
OP-0.6P75
9290



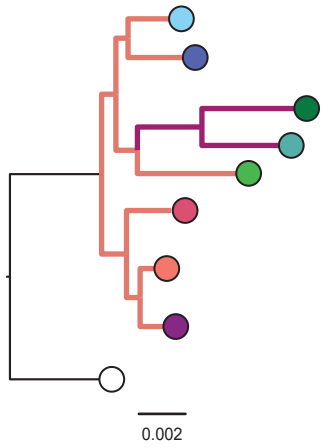
OP-0.6P100
1344



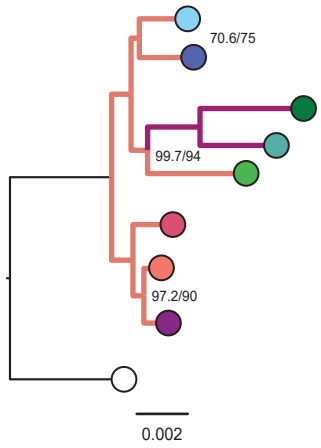
OP-0.7P50
14649



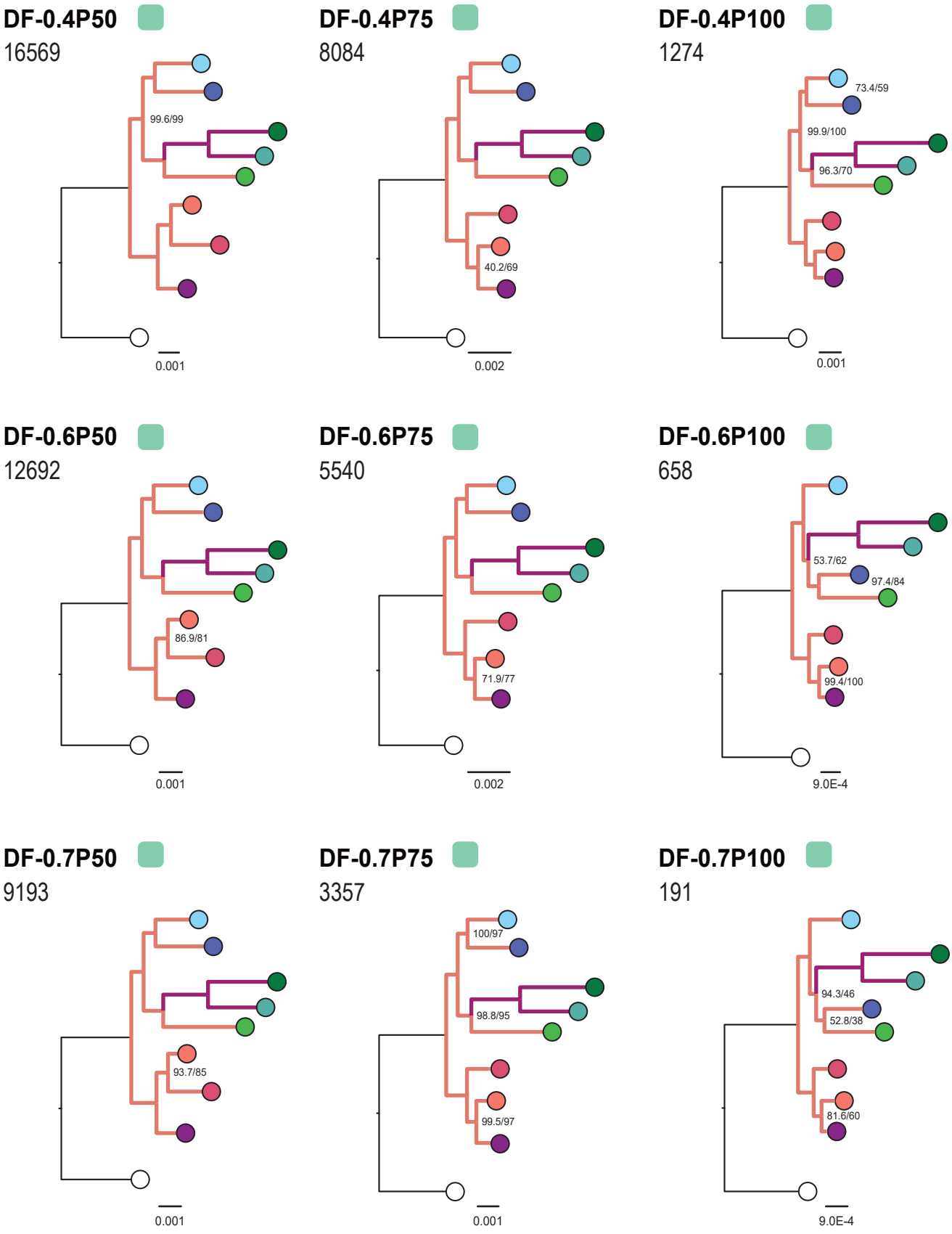
OP-0.7P75
6121



OP-0.7P100
512



(c)


DF-0.4P50
16569
99.6/99
0.001
DF-0.4P75
8084
40.2/69
0.002
DF-0.4P100
1274
73.4/59
99.9/100
96.3/70
0.001
DF-0.6P50
12692
86.9/81
0.001
DF-0.6P75
5540
71.9/77
0.002
DF-0.6P100
658
53.7/62
97.4/84
99.4/100
9.0E-4
DF-0.7P50
9193
93.7/85
0.001
DF-0.7P75
3357
100/97
98.8/95
99.5/97
0.001
DF-0.7P100
191
94.3/46
52.8/38
81.6/60
9.0E-4

(d)

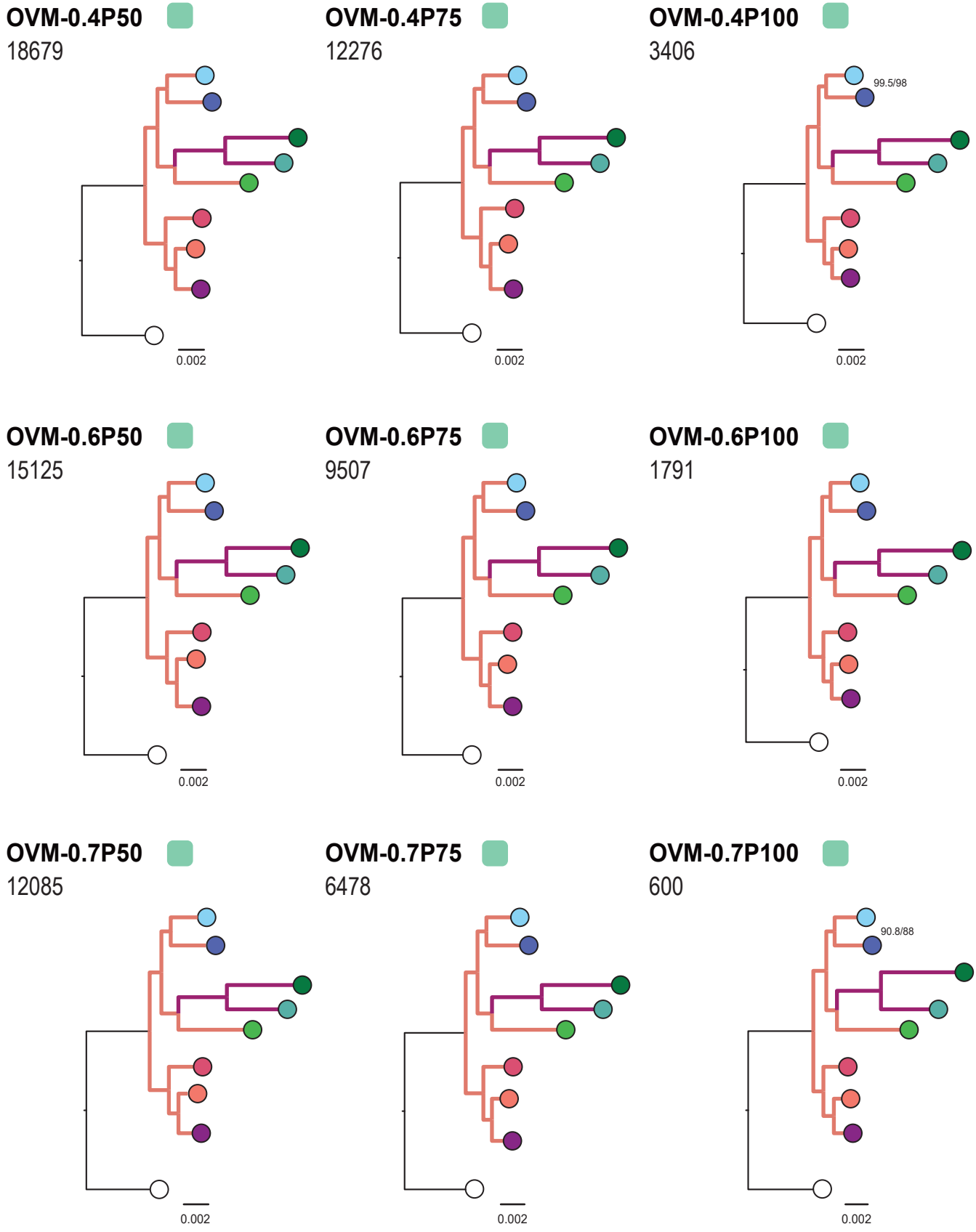
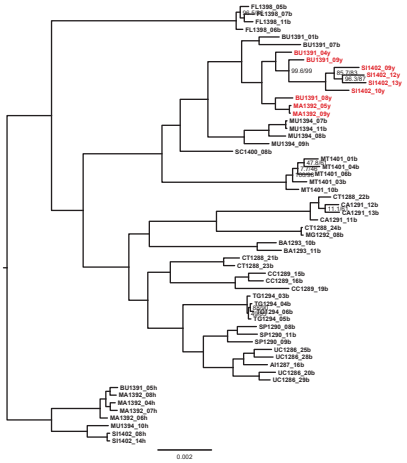

OVM-0.4P50
18679
0.002
OVM-0.4P75
12276
0.002
OVM-0.4P100
3406
99.5/98
0.002
OVM-0.6P50
15125
0.002
OVM-0.6P75
9507
0.002
OVM-0.6P100
1791
0.002
OVM-0.7P50
12085
0.002
OVM-0.7P75
6478
0.002
OVM-0.7P100
600
90.8/88
0.002

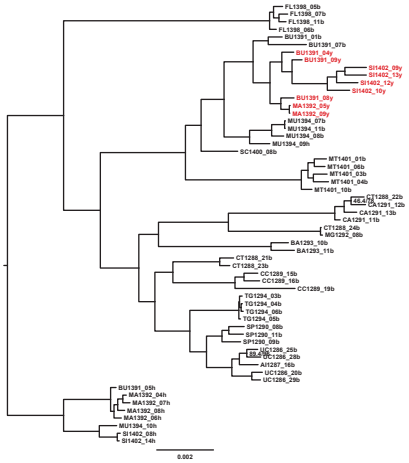
Fig. S6. Phylogenetic trees inferred with IQ-Tree 2.0 of (a) optimal, (b) default, and (c) over-merged loci datasets using nine different combinations of taxon coverage parameters using all samples with a depth of coverage equal to or greater than 10x, but excluding samples from population MC. Nodal support values are shown for nodes with less than 100/100 (SH-aLRT test/UFboot) support. Red tips: putative *hepta* x *benazzii* hybrids identified according to Dols-Serrate et al. (2020) criterion (see Supplemental Information Fig. S2) . Names of the employed datasets and their pertinent number of polymorphic loci are also shown over each inferred tree. (For a better interpretation of this supplemental figure, we recommend the reader to examine it in a non-physical format) (pp 252-254).

(a)

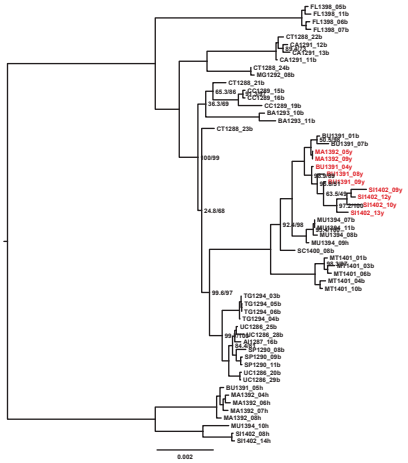
OP-0.4P50
20504



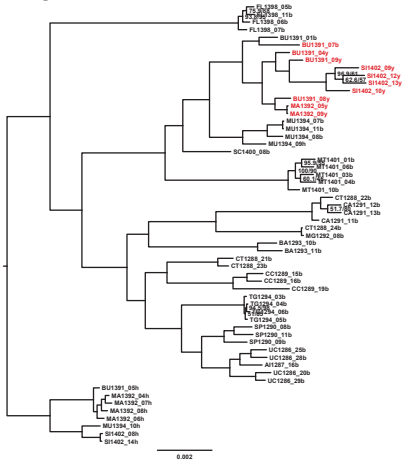
OP-0.4P75
6369



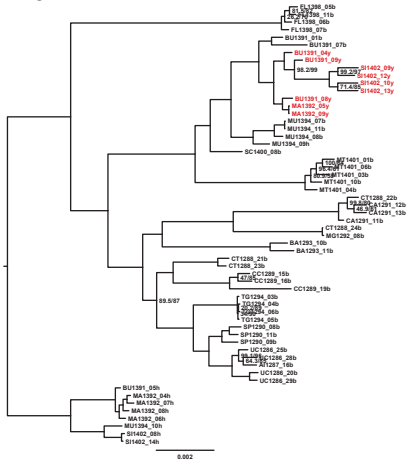
OP-0.4P100
492



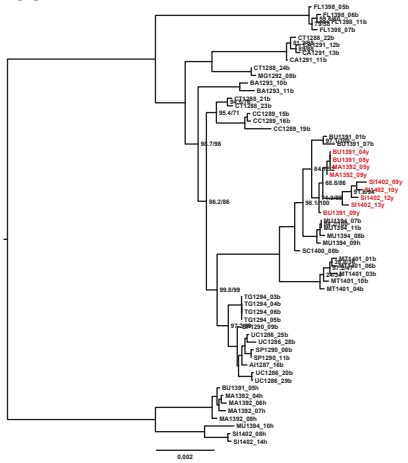
OP-0.6P50
14754



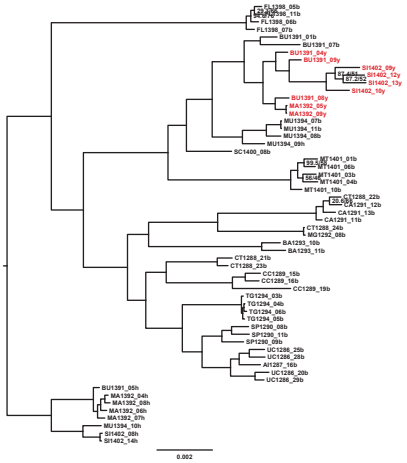
OP-0.6P75
4102



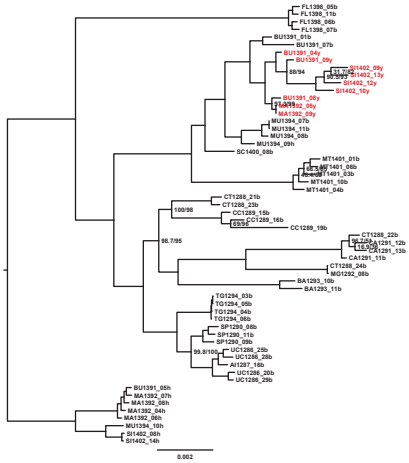
OP-0.6P100
293



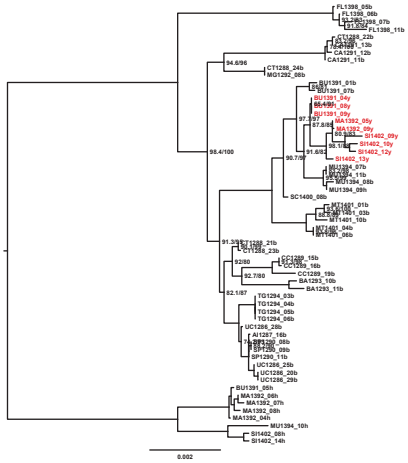
OP-0.7P50
11390



OP-0.7P75
2653

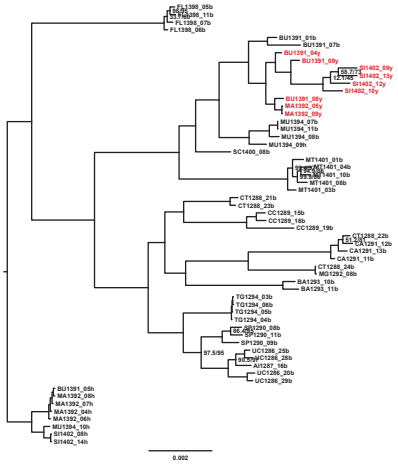


OP-0.7P100
116

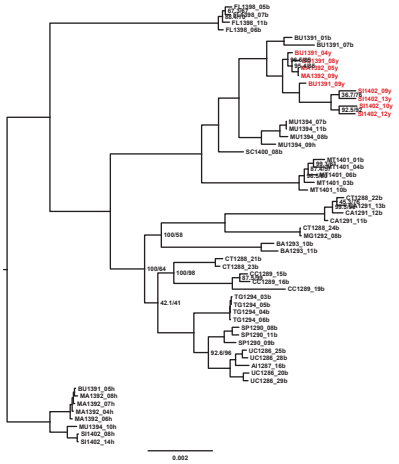


(b)

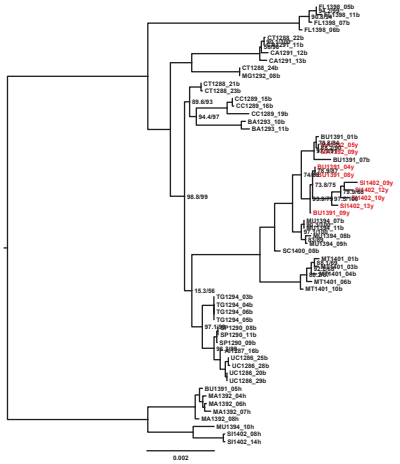
DF-0.4P50
11802



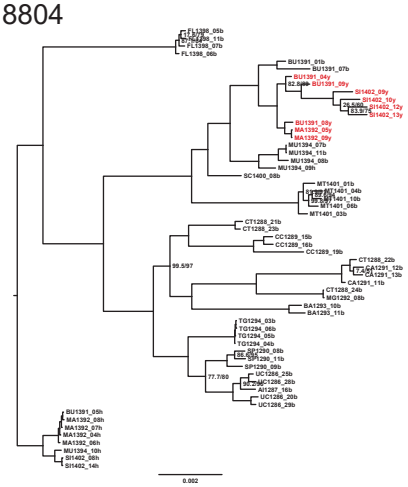
DF-0.4P75
3522



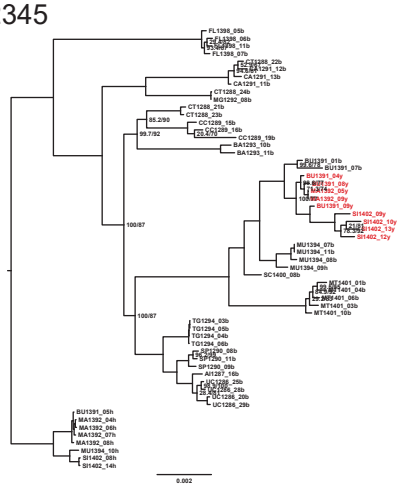
DF-0.4P100
248



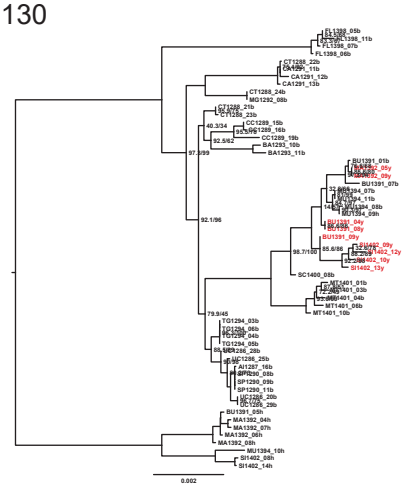
DF-0.6P50
8804



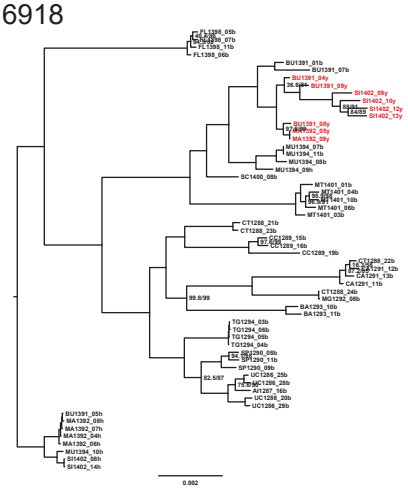
DF-0.6P75
2345



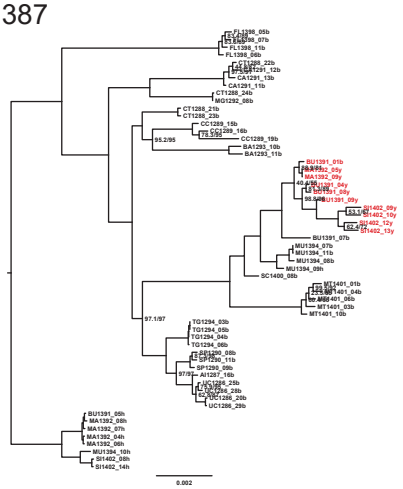
DF-0.6P100
130



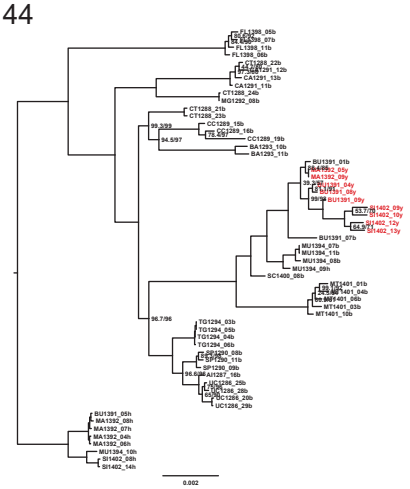
DF-0.7P50
6918



DF-0.7P75
1387

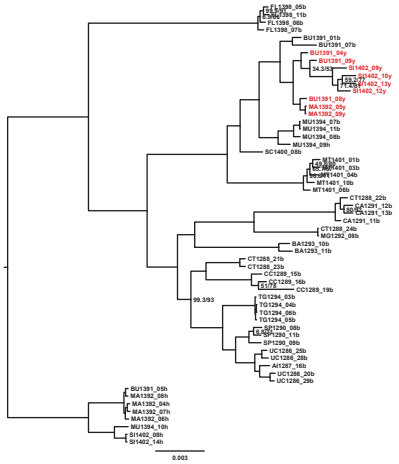


DF-0.7P100
44

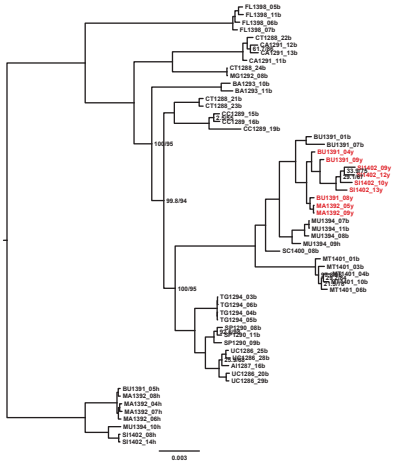


(c)

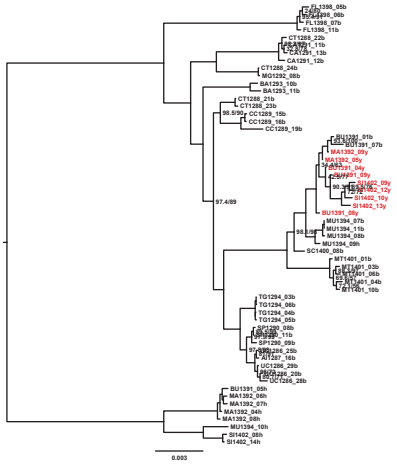
OVM-0.4P50
16818



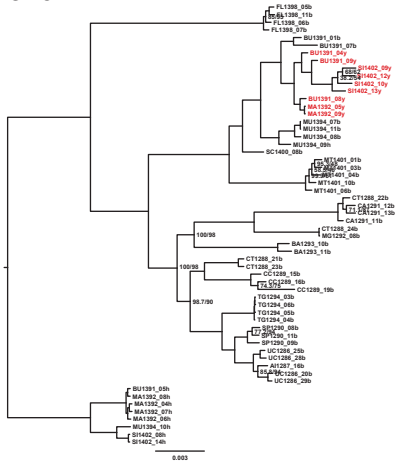
OVM-0.4P75
8038



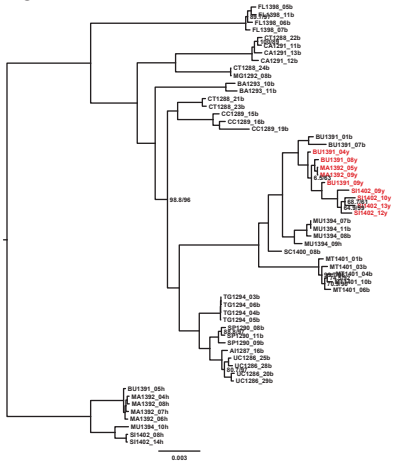
OVM-0.4P100
916



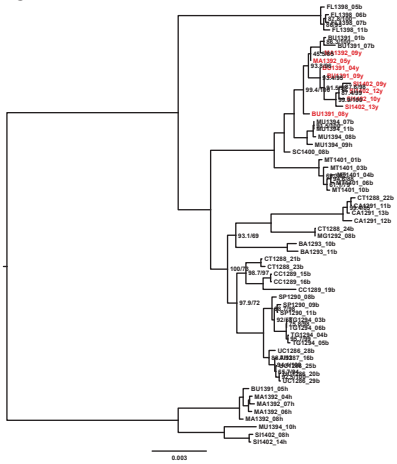
OVM-0.6P50
13262



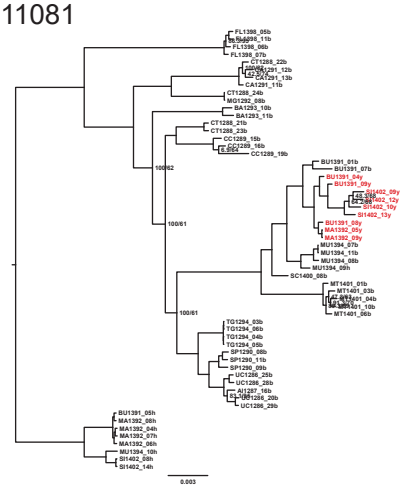
OVM-0.6P75
5791



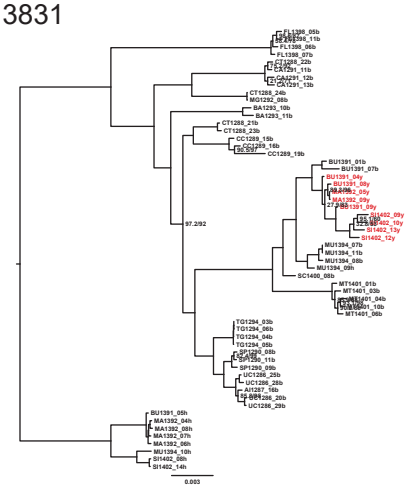
OVM-0.6P100
461



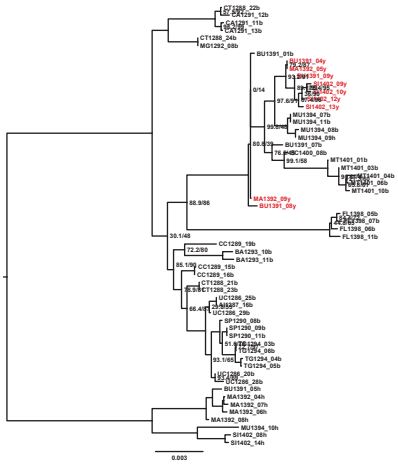
OVM-0.7P50
11081



OVM-0.7P75
3831



OVM-0.7P100
127



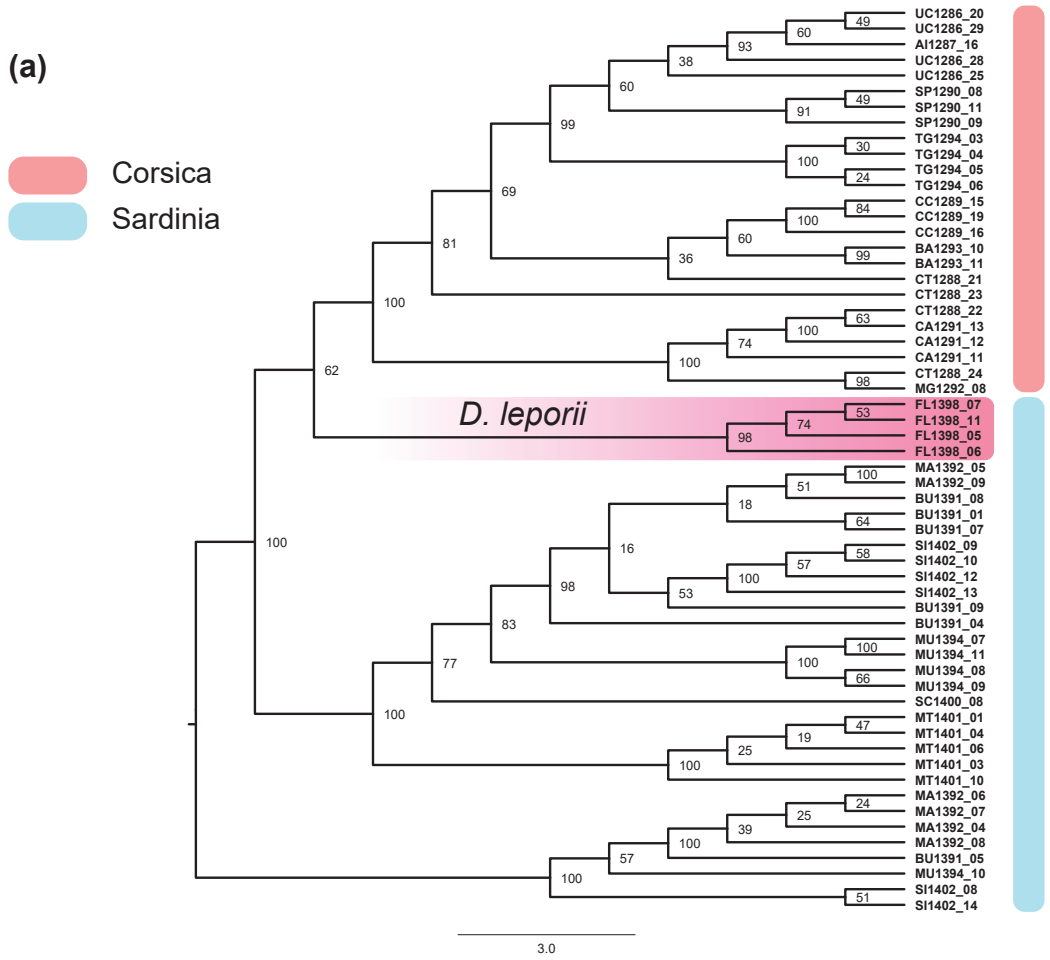
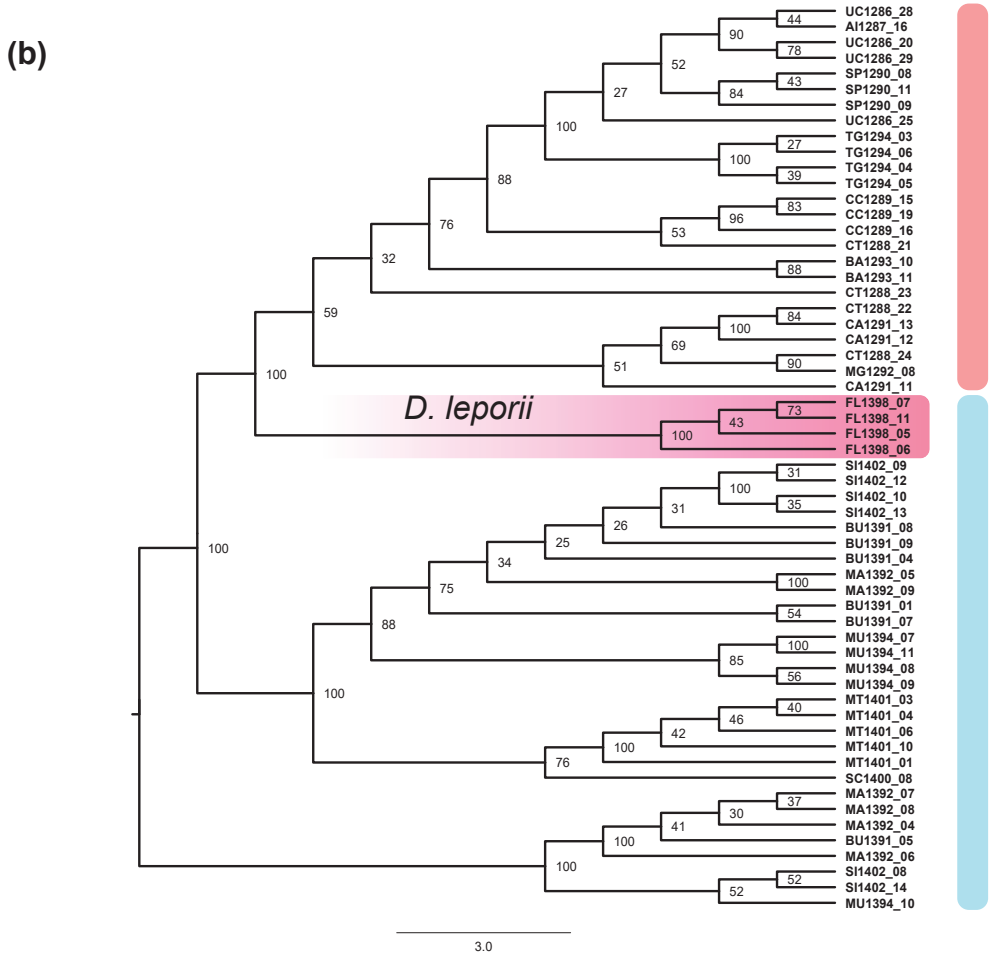


Fig. S7. Results for SVDQuartets phylogenetic inference using (a) OP-0.7P50, (b) DF-0.7P50, and (c) OVM-0.7P50 datasets (pp 255-256).


(b)
UC1286_28
AI1287_16
UC1286_20
UC1286_29
SP1290_08
SP1290_11
SP1290_09
UC1286_25
TG1294_03
TG1294_06
TG1294_04
TG1294_05
CC1289_15
CC1289_19
CC1289_16
CT1288_21
BA1293_10
BA1293_11
CT1288_23
CT1288_22
CA1291_13
CA1291_12
CT1288_24
MG1292_08
CA1291_11
D. leporii
FL1398_07
FL1398_11
FL1398_05
FL1398_06
SI1402_09
SI1402_12
SI1402_10
SI1402_13
BU1391_08
BU1391_09
BU1391_04
MA1392_05
MA1392_09
BU1391_01
BU1391_07
MU1394_07
MU1394_11
MU1394_08
MU1394_09
MT1401_03
MT1401_04
MT1401_06
MT1401_10
MT1401_01
SC1400_08
MA1392_07
MA1392_08
MA1392_04
BU1391_05
MA1392_06
SI1402_08
SI1402_14
MU1394_10
3.0

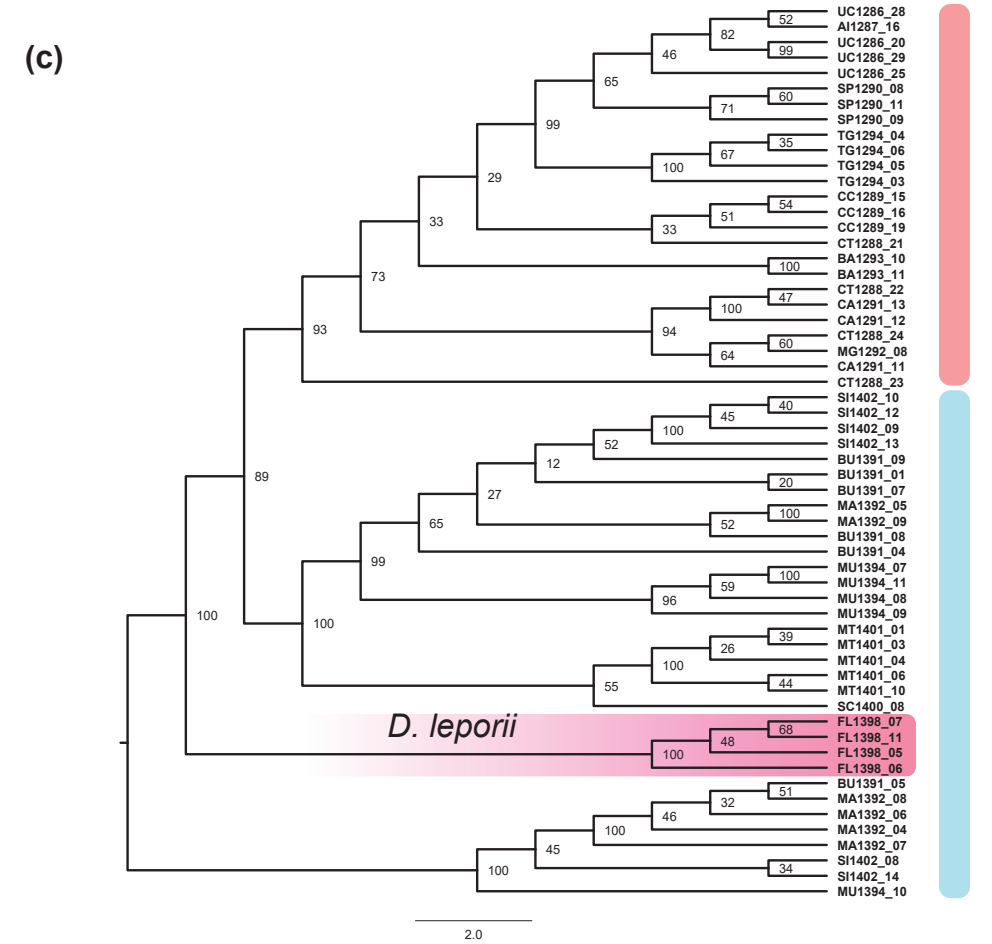
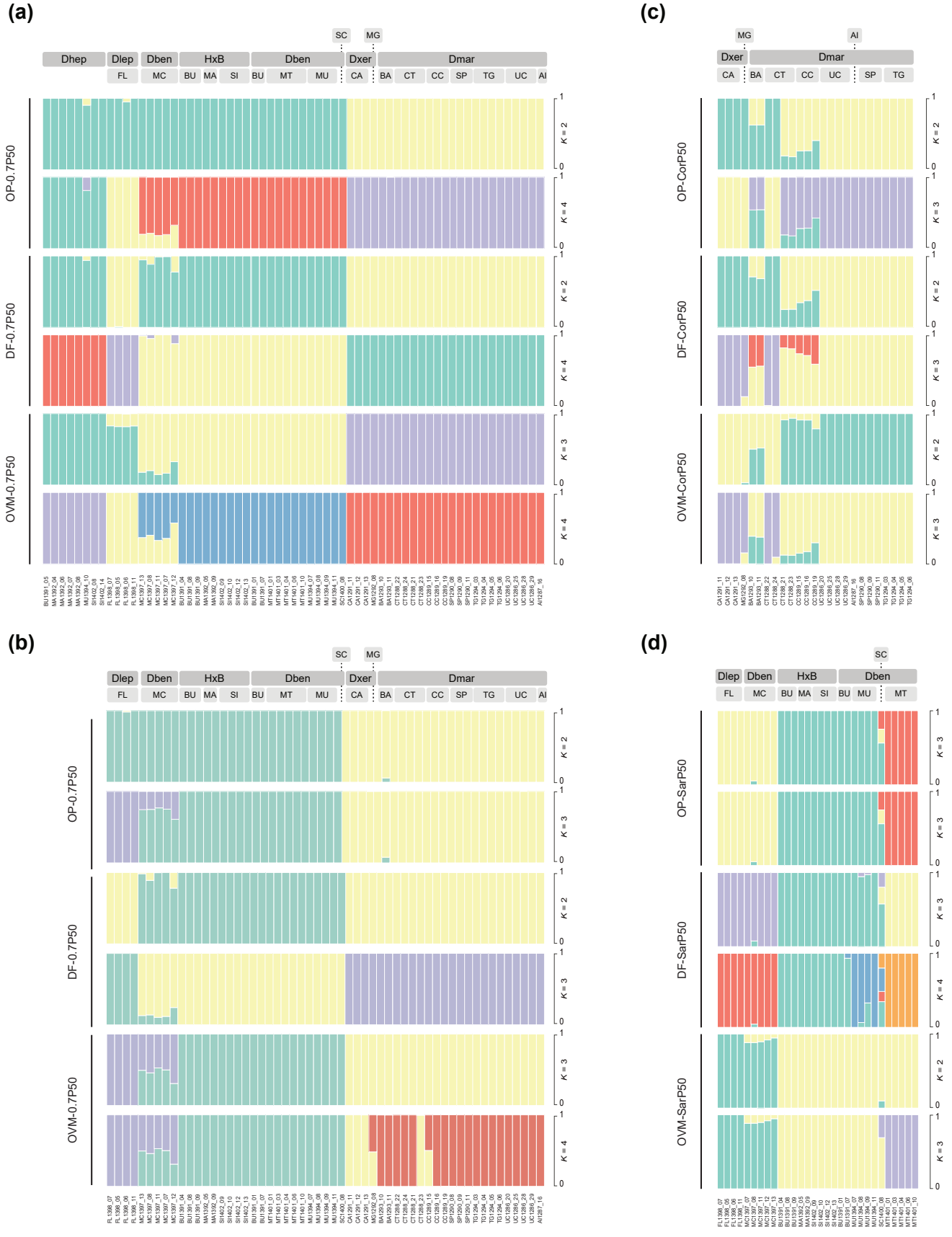

(c)
UC1286_28
AI1287_16
UC1286_20
UC1286_29
UC1286_25
SP1290_08
SP1290_11
SP1290_09
TG1294_04
TG1294_06
TG1294_05
TG1294_03
CC1289_15
CC1289_16
CC1289_19
CT1288_21
BA1293_10
BA1293_11
CT1288_22
CA1291_13
CA1291_12
CT1288_24
MG1292_08
CA1291_11
CT1288_23
SI1402_10
SI1402_12
SI1402_09
SI1402_13
BU1391_09
BU1391_01
BU1391_07
MA1392_05
MA1392_09
BU1391_08
BU1391_04
MU1394_07
MU1394_11
MU1394_08
MU1394_09
MT1401_01
MT1401_03
MT1401_04
MT1401_06
MT1401_10
SC1400_08
D. leporii
FL1398_07
FL1398_11
FL1398_05
FL1398_06
BU1391_05
MA1392_08
MA1392_06
MA1392_04
MA1392_07
SI1402_08
SI1402_14
MU1394_10
2.0

Fig. S8. STRUCTURE analysis results for OP, DF, and OVM datasets using different sets of taxa. Facets above the plots represent: *D. hepta* (Dhep), *D. benazzii* (Dben), *D. leporii* (Dlep), *D. mariae* (Dmar), *D. xeropotamica* (Dxer), the putative hybrids of *benazzii* x *hepta* (HxB), the candidate *leporii* x *benazzii* hybrids (LxB), and acronyms of the populations (as in Supplemental Information Table S3). (a) Results for all samples (n = 63). (b) Results excluding *D. hepta*. (c) Hierarchical results for Corsican populations. (d) Hierarchical results for Sardinian populations, excluding *D. hepta*. (e-f) Hierarchical results for the parental-hybrid trio of *D. benazzii*, *D. leporii*, and LxB. (g-h) Hierarchical results for the parental-hybrid trio of *D. benazzii*, *D. hepta*, and HxB (pp 258-259).


(a)
SC
MG
Dhep
Dlep
Dben
HxB
Dben
Dxer
Dmar
FL
MC
BU
MA
SI
BU
MT
MU
CA
BA
CT
CC
SP
TG
UC
AI
OP-0.7P50
DF-0.7P50
OVM-0.7P50
K = 2
K = 4
K = 3
(c)
MG
AI
Dxer
Dmar
CA
BA
CT
CC
UC
SP
TG
OP-CorP50
DF-CorP50
OVM-CorP50
K = 2
K = 3
(b)
SC
MG
Dlep
Dben
HxB
Dben
Dxer
Dmar
FL
MC
BU
MA
SI
BU
MT
MU
CA
BA
CT
CC
SP
TG
UC
AI
OP-0.7P50
DF-0.7P50
OVM-0.7P50
K = 2
K = 3
K = 4
(d)
SC
Dlep
Dben
HxB
Dben
FL
MC
BU
MA
SI
BU
MU
MT
OP-SarP50
DF-SarP50
OVM-SarP50
K = 3
K = 4
K = 2

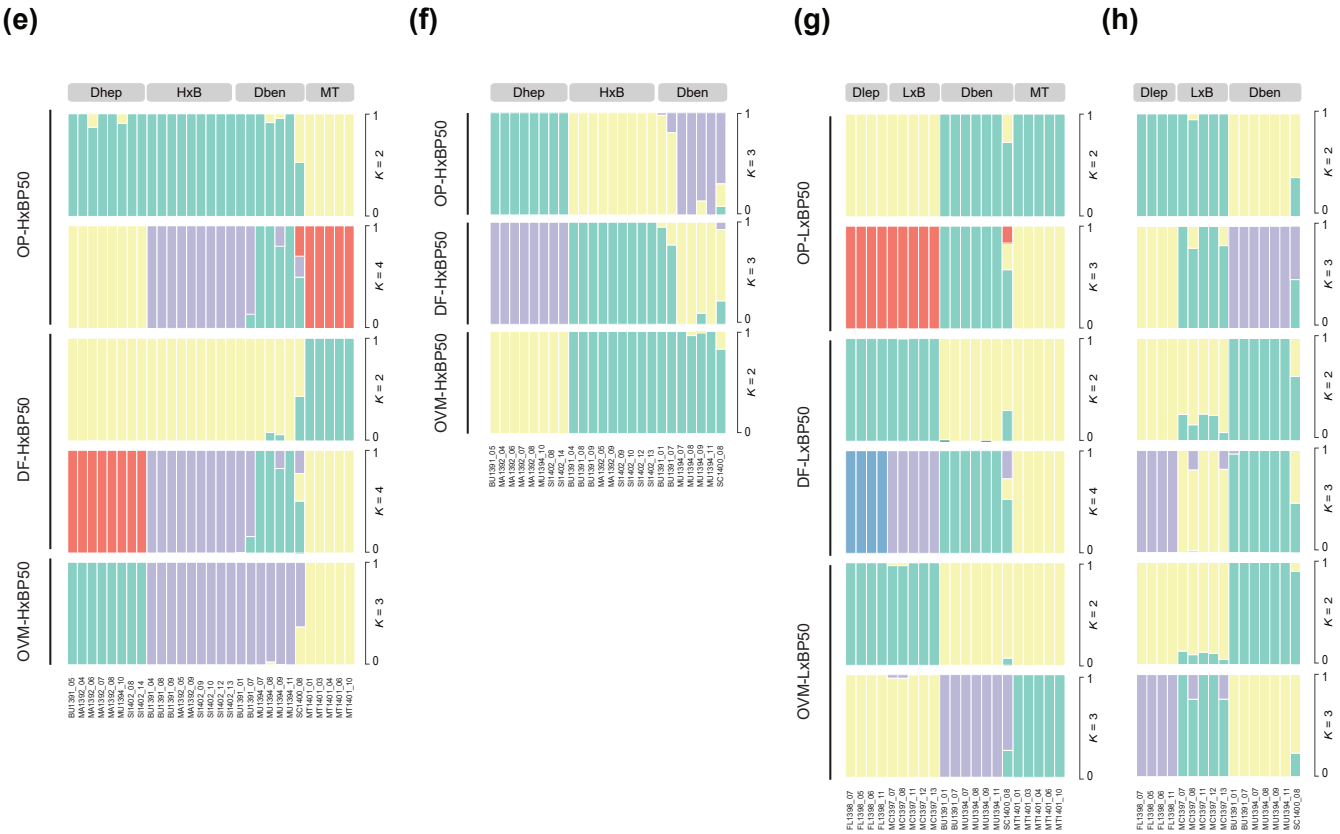
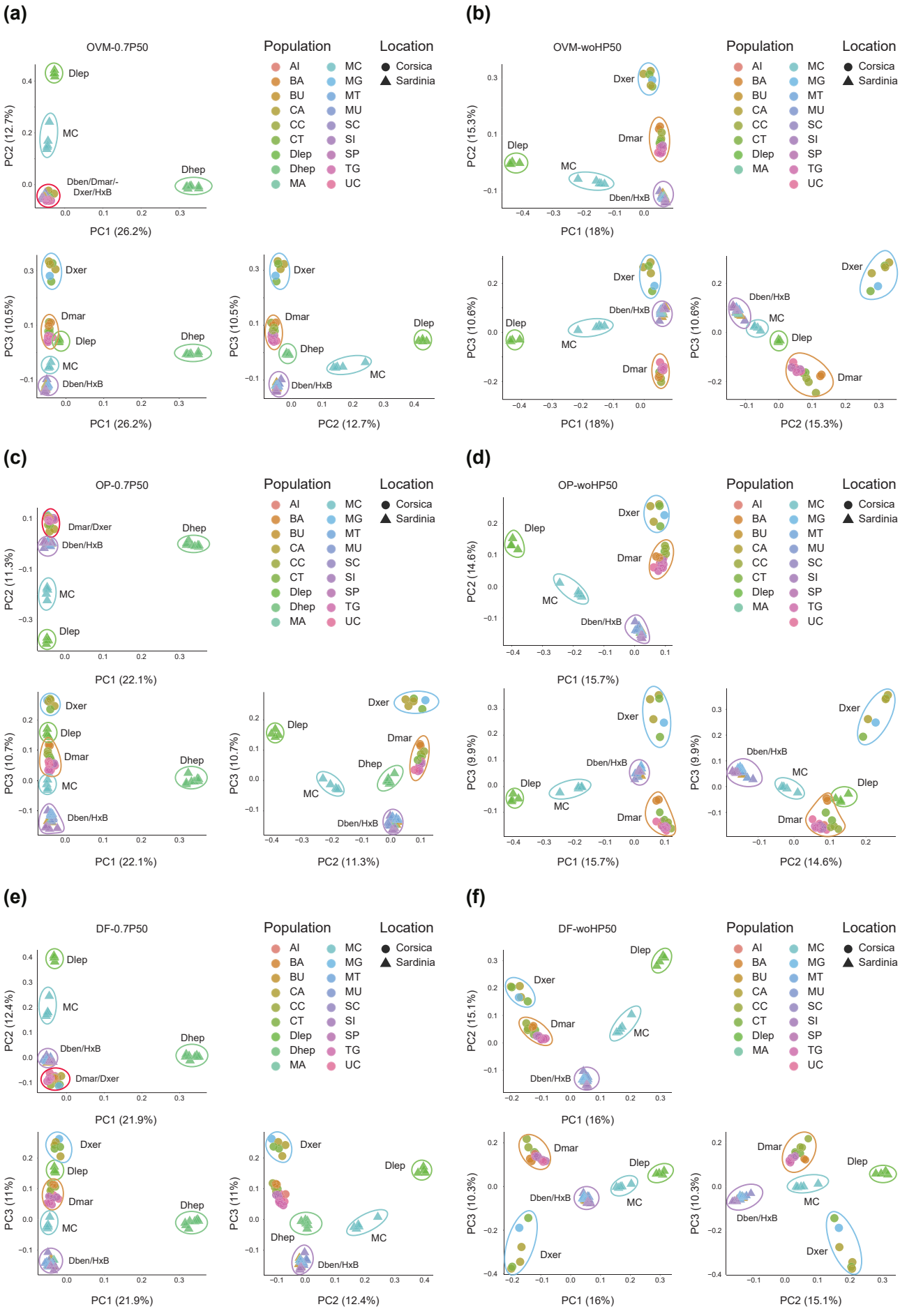

(e)
(f)
(g)
(h)
Dhep
HxB
Dben
MT
Dlep
LxB
OP-HxBP50
DF-HxBP50
OVM-HxBP50
OP-LxBP50
DF-LxBP50
OVM-LxBP50
K = 2
K = 3
K = 4

Fig. S9. Principal Component Analyses (PCA) showing PC1 to PC3 combinations for optimal (OP), default (DF), and over-merged (OVM) datasets. (a, c, e) All samples included. (b, d, f) Analyses performed without *D. hepta* (Dhep). Dben: *D. beanzzii*. Dlep: *D. leporii.* Dmar: *D. mariae.* Dxer: *D.xeropotamica.* HxB: putative *hepta* x *benazzii* hybrids. Samples are color coded by population.


(a)
OVM-0.7P50
Population
Location
Corsica
Sardinia
Dlep
MC
Dhep
Dben/Dmar/-Dxer/HxB
PC1 (26.2%)
PC2 (12.7%)
Dxer
Dmar
Dben/HxB
PC3 (10.5%)
(b)
OVM-woHP50
PC1 (18%)
PC2 (15.3%)
PC3 (10.6%)
(c)
OP-0.7P50
Dmar/Dxer
PC1 (22.1%)
PC2 (11.3%)
PC3 (10.7%)
(d)
OP-woHP50
PC1 (15.7%)
PC2 (14.6%)
PC3 (9.9%)
(e)
DF-0.7P50
Dmar/Dxer
PC1 (21.9%)
PC2 (12.4%)
PC3 (11%)
(f)
DF-woHP50
PC1 (16%)
PC2 (15.1%)
PC3 (10.3%)
AI BA BU CA CC CT Dlep Dhep MA
MC MG MT MU SC SI SP TG UC

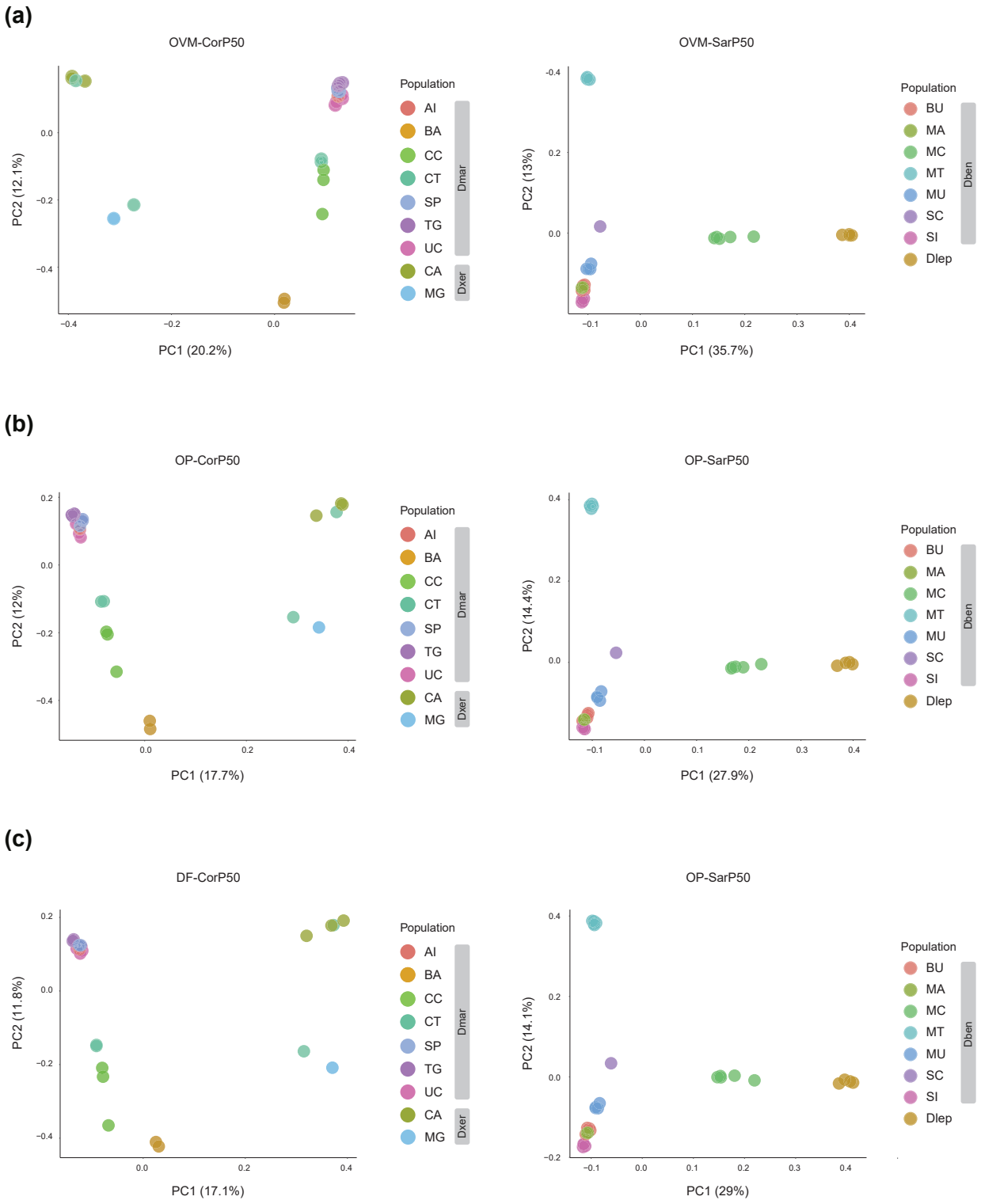
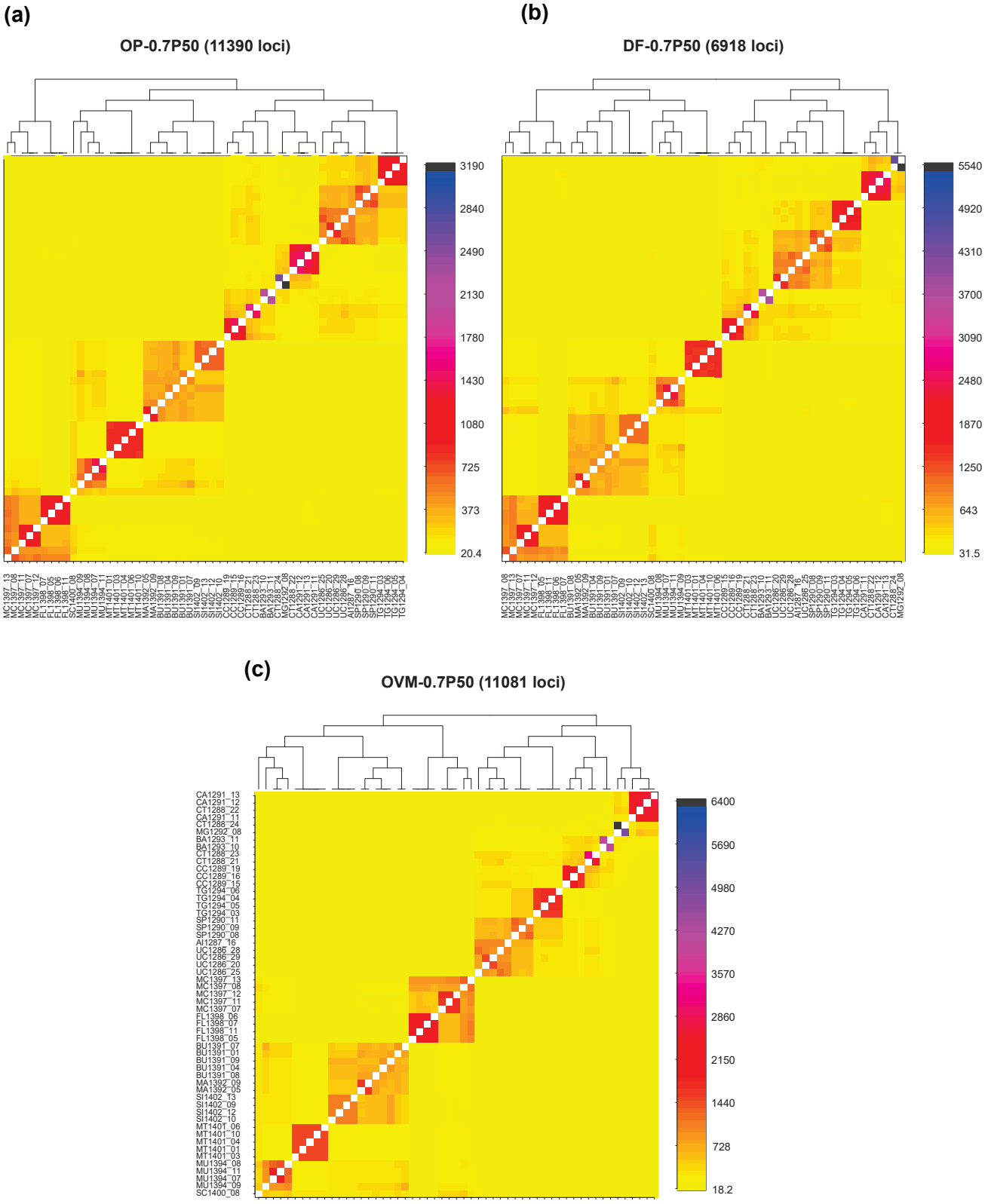
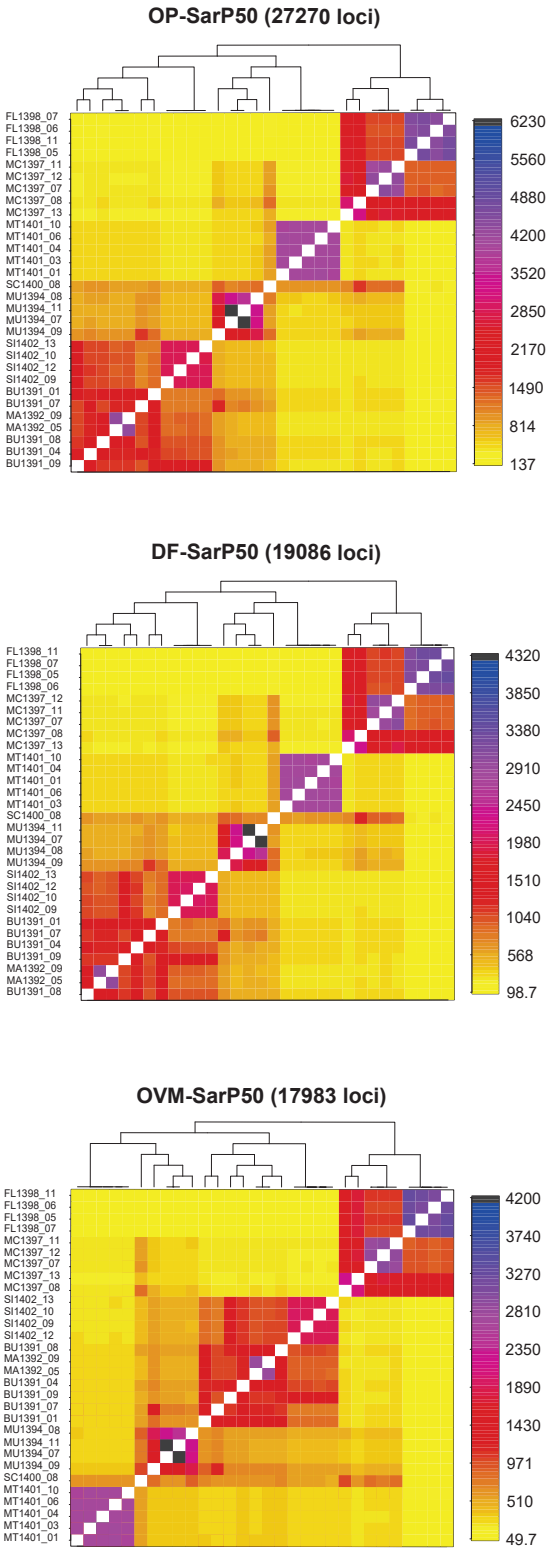


Fig. S10. Principal Component Analyses (PCA) showing PC1 and PC2 of (a) over-merged (OVM), (b) optimal (OP), and (c) default (DF) datasets. Left side plots show genetic clustering of Corsican samples. Right side plots show genetic clustering of Sardinian samples. *Dugesia hepta* not included. Dben: *D. benazzii*. Dlep: *D. leporii.* Dmar: *D. mariae*. Dxer: *D. xeropotamica.*

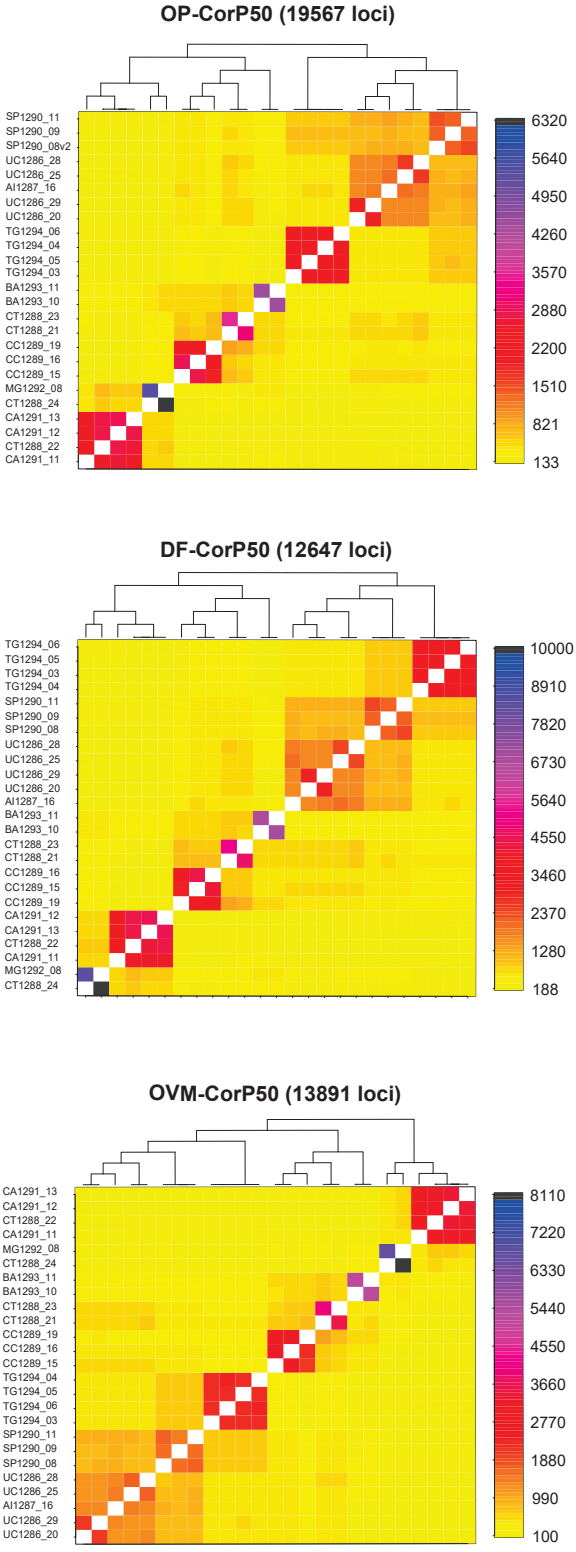
Fig. S11. fineRADstructure heatmaps of coancestry coefficients among all ingroup samples (a-c) inferred using OP, DF, and OVM loci. (d) In a column, from top to bottom: coancestry coefficients of all Sardinian samples based on optimal (OP-SarP50), default (DF-SarP50), and over-merged (OVM-SarP50) loci. (e) In a column, from top to bottom: coancestry patterns of all Corsican samples based on optimal (OP-CorP50), default (DF-CorP50), and over-merged (OVM-CorP50) datasets. Atop each heatmap there is a fineRADstructure clustering dendrogram based on the matrix of coancestry coefficients. Branch support is shown for branches with posterior probabilities <1. Coancestry coefficients are colour coded from low (yellow) to high (blue-black) corresponding to the values in the legend. Only one random SNP per loci (indicated in brackets) was used (pp 264-265).


(a)
OP-0.7P50 (11390 loci)
3190
2840
2490
2130
1780
1430
1080
725
373
20.4
(b)
DF-0.7P50 (6918 loci)
5540
4920
4310
3700
3090
2480
1870
1250
643
31.5
(c)
OVM-0.7P50 (11081 loci)
6400
5690
4980
4270
3570
2860
2150
1440
728
18.2

(d)


OP-SarP50 (27270 loci)
FL1398_07
FL1398_06
FL1398_11
FL1398_05
MC1397_11
MC1397_12
MC1397_07
MC1397_08
MC1397_13
MT1401_10
MT1401_06
MT1401_04
MT1401_03
MT1401_01
SC1400_08
MU1394_08
MU1394_11
MU1394_07
MU1394_09
SI1402_13
SI1402_10
SI1402_12
SI1402_09
BU1391_01
BU1391_07
MA1392_09
MA1392_05
BU1391_08
BU1391_04
BU1391_09
6230
5560
4880
4200
3520
2850
2170
1490
814
137
DF-SarP50 (19086 loci)
FL1398_11
FL1398_07
FL1398_05
FL1398_06
MC1397_12
MC1397_11
MC1397_07
MC1397_08
MC1397_13
MT1401_10
MT1401_04
MT1401_01
MT1401_06
MT1401_03
SC1400_08
MU1394_11
MU1394_07
MU1394_08
MU1394_09
SI1402_13
SI1402_12
SI1402_10
SI1402_09
BU1391_01
BU1391_07
BU1391_04
BU1391_09
MA1392_09
MA1392_05
BU1391_08
4320
3850
3380
2910
2450
1980
1510
1040
568
98.7
OVM-SarP50 (17983 loci)
FL1398_11
FL1398_06
FL1398_05
FL1398_07
MC1397_11
MC1397_12
MC1397_07
MC1397_13
MC1397_08
SI1402_13
SI1402_10
SI1402_09
SI1402_12
BU1391_08
MA1392_09
MA1392_05
BU1391_04
BU1391_09
BU1391_07
BU1391_01
MU1394_08
MU1394_11
MU1394_07
MU1394_09
SC1400_08
MT1401_10
MT1401_06
MT1401_04
MT1401_03
MT1401_01
4200
3740
3270
2810
2350
1890
1430
971
510
49.7

(e)


OP-CorP50 (19567 loci)
SP1290_11
SP1290_09
SP1290_08v2
UC1286_28
UC1286_25
AI1287_16
UC1286_29
UC1286_20
TG1294_06
TG1294_04
TG1294_05
TG1294_03
BA1293_11
BA1293_10
CT1288_23
CT1288_21
CC1289_19
CC1289_16
CC1289_15
MG1292_08
CT1288_24
CA1291_13
CA1291_12
CT1288_22
CA1291_11
6320
5640
4950
4260
3570
2880
2200
1510
821
133
DF-CorP50 (12647 loci)
TG1294_06
TG1294_05
TG1294_03
TG1294_04
SP1290_11
SP1290_09
SP1290_08
UC1286_28
UC1286_25
UC1286_29
UC1286_20
AI1287_16
BA1293_11
BA1293_10
CT1288_23
CT1288_21
CC1289_16
CC1289_15
CC1289_19
CA1291_12
CA1291_13
CT1288_22
CA1291_11
MG1292_08
CT1288_24
10000
8910
7820
6730
5640
4550
3460
2370
1280
188
OVM-CorP50 (13891 loci)
CA1291_13
CA1291_12
CT1288_22
CA1291_11
MG1292_08
CT1288_24
BA1293_11
BA1293_10
CT1288_23
CT1288_21
CC1289_19
CC1289_16
CC1289_15
TG1294_04
TG1294_05
TG1294_06
TG1294_03
SP1290_11
SP1290_09
SP1290_08
UC1286_28
UC1286_25
AI1287_16
UC1286_29
UC1286_20
8110
7220
6330
5440
4550
3660
2770
1880
990
100

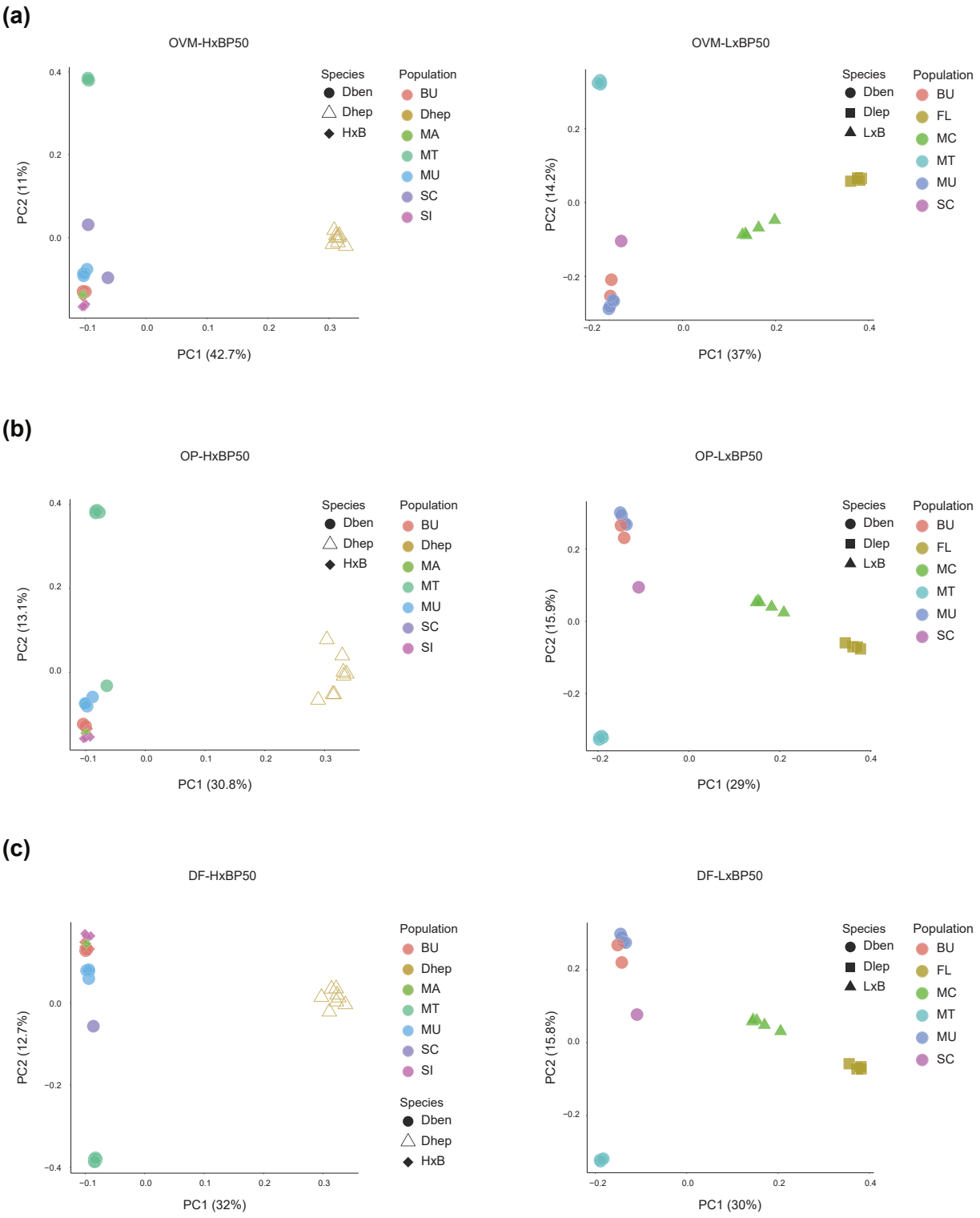


Fig. S12. Principal Component Analyses (PCA) showing PC1 and PC2 of (a) over-merged (OVM), (b) optimal (OP), and (c) default (DF) datasets of candidate parental-hybrid trios. Left side plots show clustering patterns of HxB parental-hybrid trio. Right side plots show clustering patterns of LxB parental-hybrid trio. HxB: putative *hepta* x *benazzii* hybrid. LxB: *leporii* x *benazzii* hybrid from MC. Dben: *D. benazzii.* Dhep: *D. hepta*. Dlep: *D. leporii.*

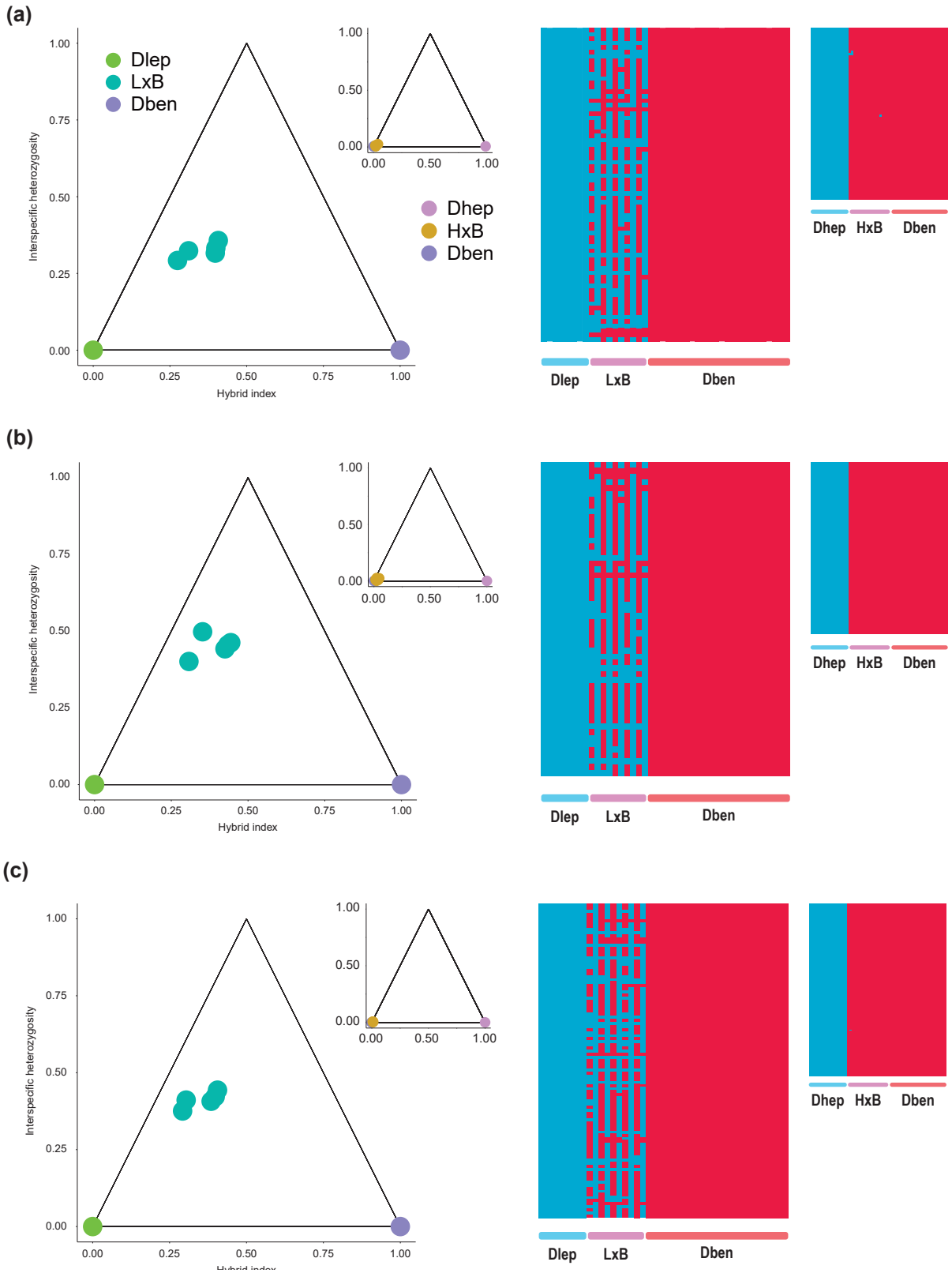


Fig. S13. From left to right, results for the INTROGRESS analyses of the candidate parental-hybrid trios *D. leporii*, *D. benazzii*, and LxB hybrids; *D. hepta*, *D. benazzii*, and HxB putative hybrids; and pertinent ancestry-painting plots. (a) Results for OP-LxBP50 and OP-HxBP50; (b) DF-LxBP50 and DF-HxBP50; (c) OVM-LxBP50 and OVM-HxBP50 datasets. Dben: *D. benazzii*; Dhep: *D. hepta*; Dlep: *D. leporii*; LxB: *leporii* x *benazzii* hybrids; HxB: *hepta* x *benazzii* putative hybrids.

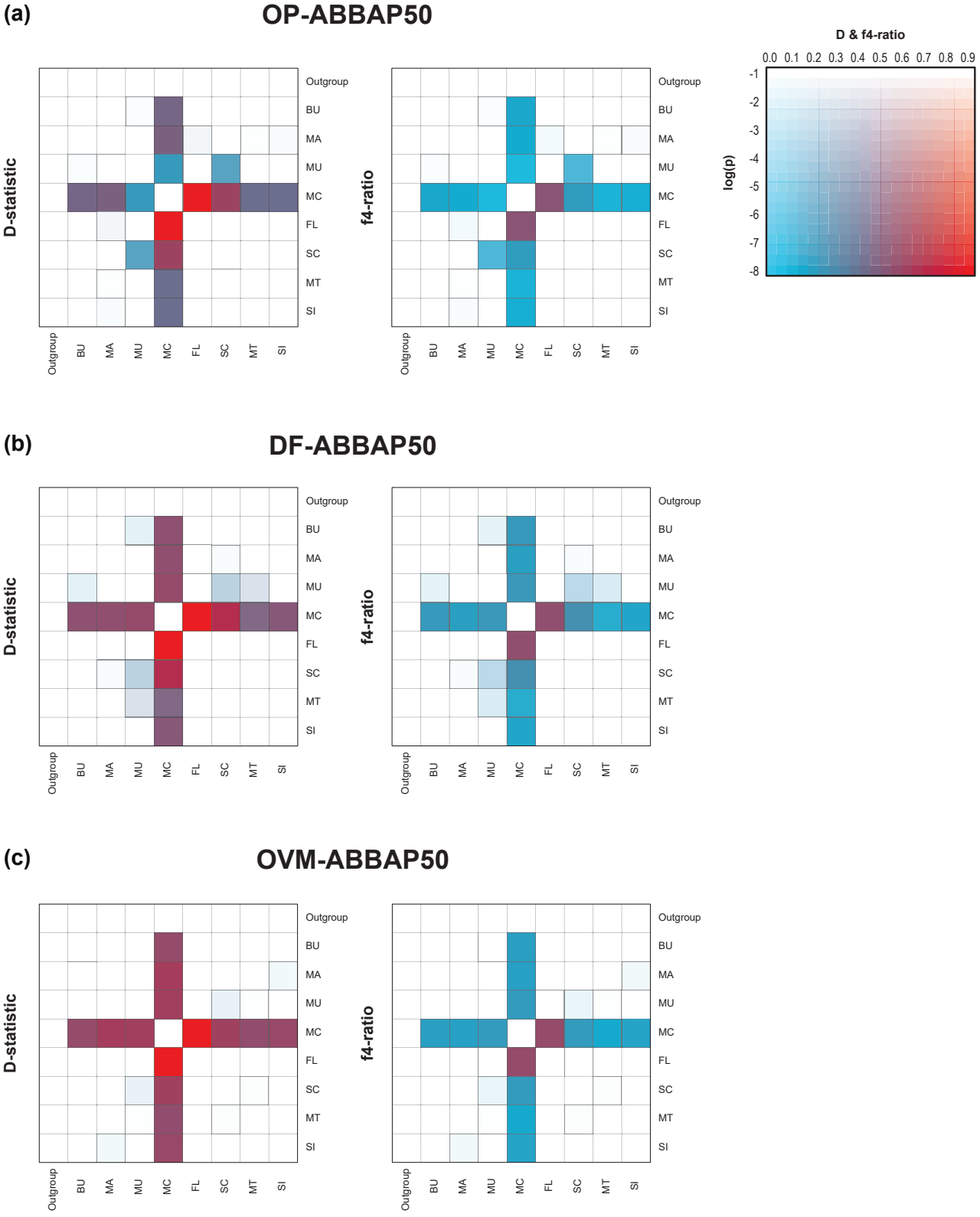


Fig. S14. Results for the ABBA/BABA tests based on (a) optimal, (b) default, and (c) over-merged loci sets. Heatmap cells: blue hues indicate low values for *D*-statistics and f4-ratio tests. Conversely, red hues indicate high values. Color intensity indicates varying degree of p-value significance after applying the Benjamini-Hochberg correction for multiple testing.

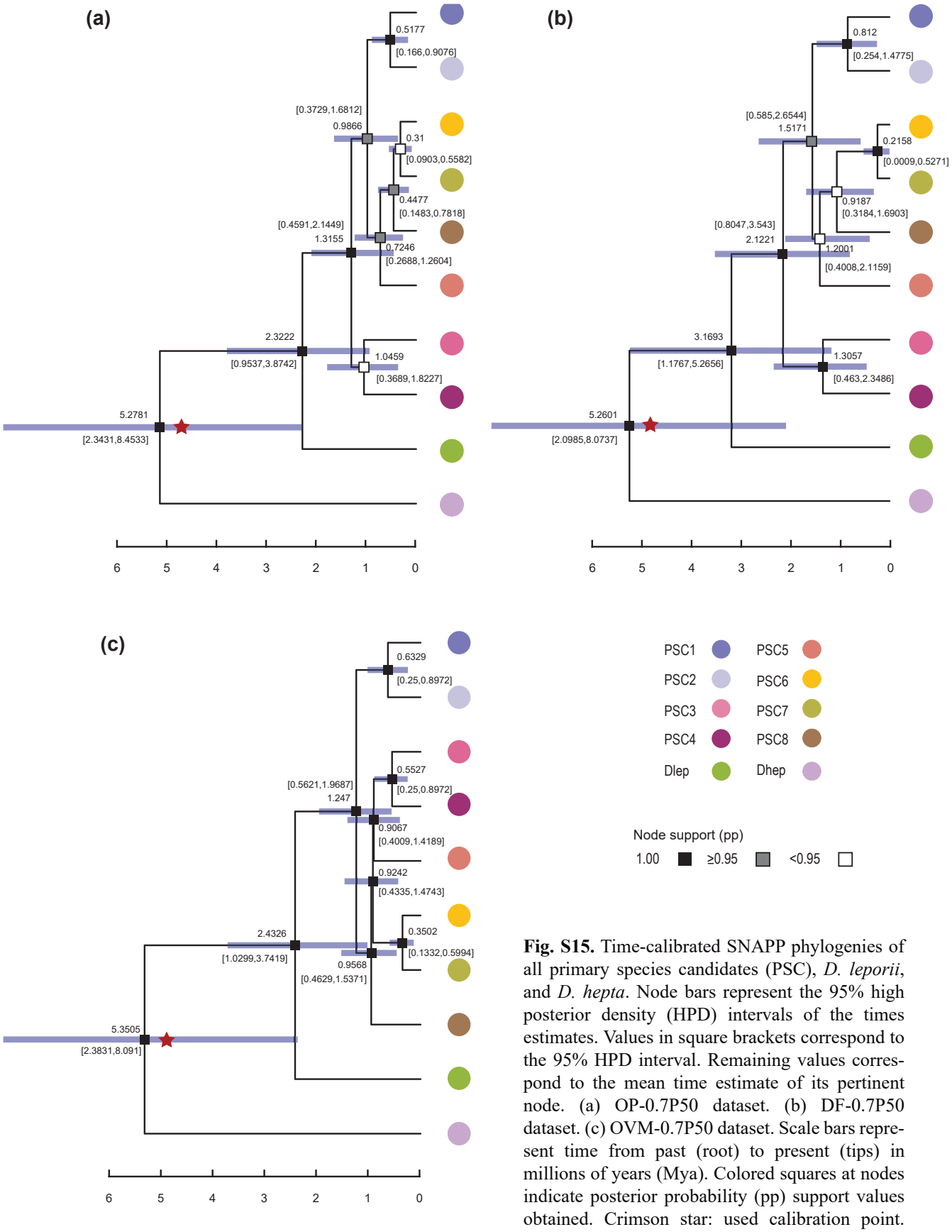


Fig. S15. Time-calibrated SNAPP phylogenies of all primary species candidates (PSC), *D. leporii*, and *D. hepta*. Node bars represent the 95% high posterior density (HPD) intervals of the times estimates. Values in square brackets correspond to the 95% HPD interval. Remaining values correspond to the mean time estimate of its pertinent node. (a) OP-0.7P50 dataset. (b) DF-0.7P50 dataset. (c) OVM-0.7P50 dataset. Scale bars represent time from past (root) to present (tips) in millions of years (Mya). Colored squares at nodes indicate posterior probability (pp) support values obtained. Crimson star: used calibration point. Dhep: *D. hepta*. Dlep: *D. leporii*.

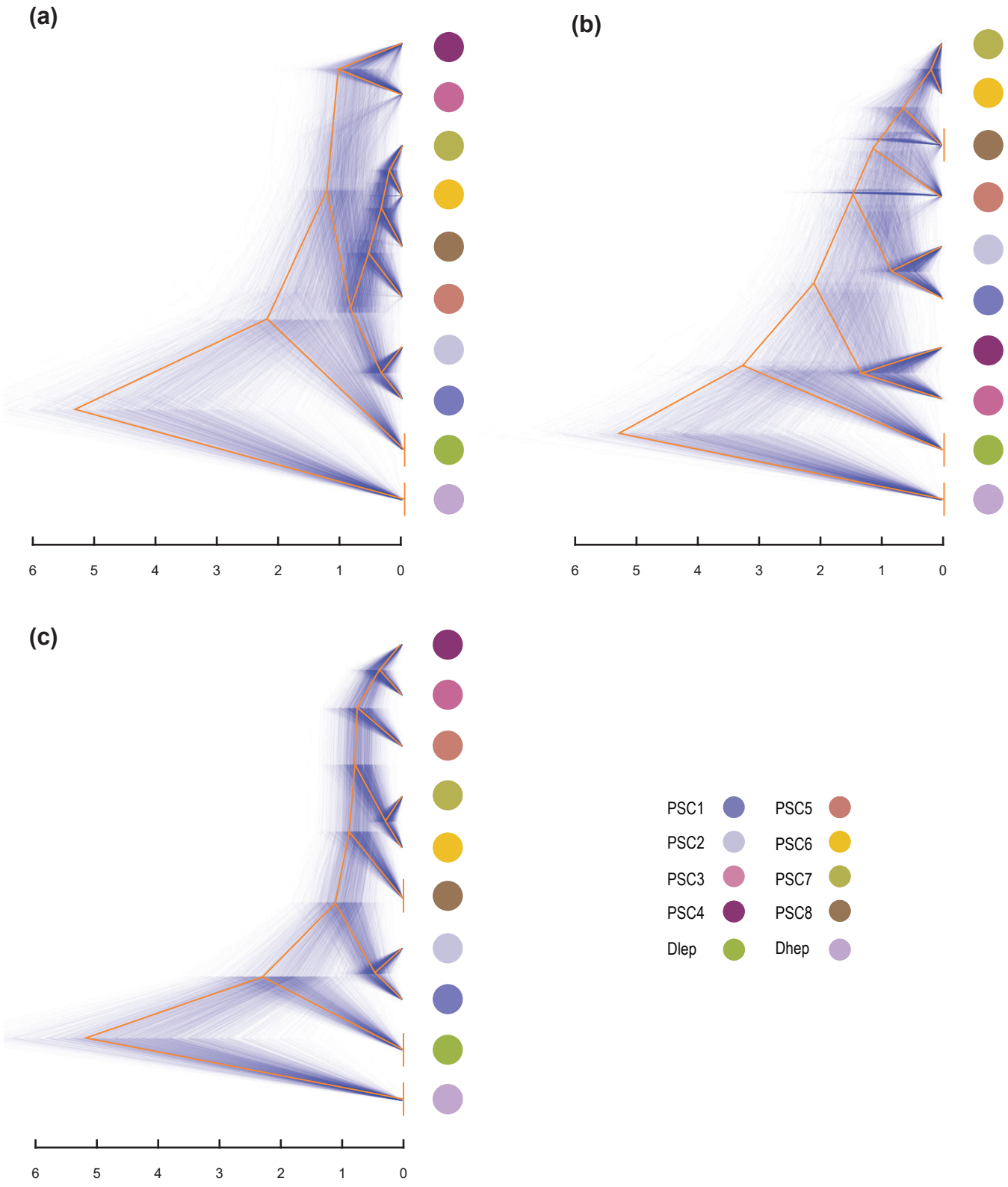


Fig. S16. Cloudograms of SNAPP trees from the posterior tree distribution of (a) OP-0.7P50, (b) DF-0.7P50, and (c) OVM-0.7P50 datasets showing topological and branch length variation. Summary trees of each analyses are shown in orange, embedded within their pertinent cloudogram. Scale bars represent time from past (root) to present (tips) in millions of years (Mya). PSC: primary species candidate. Dhep: *D. hepta.* Dlep: *D. leporii.*

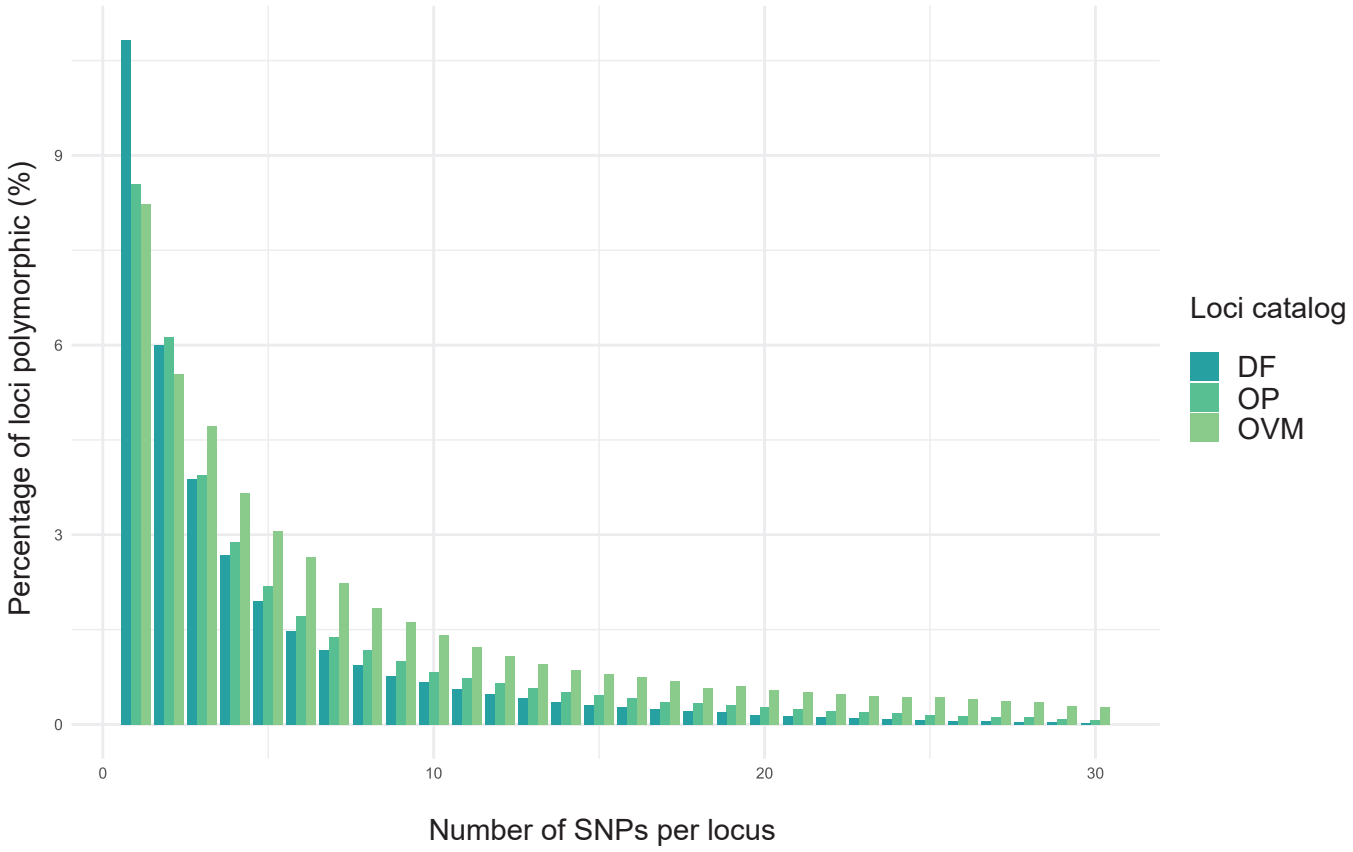


Fig. S17. Percentage of polymorphic loci for each number of SNPs per locus. Data shown for each catalog loci assembled using differing assembly parameters. OP: optimal set (m2 M1 n2), DF: default set (m3 M2 n1), OVM: over-merged set (m3 M5 n6).

Supplementary Material: Tables

Mixed, not stirred: Genomic analyses of *Dugesia* (Platyhelminthes, Tricladida) shed light on the evolutionary history and diversification of the Corso-Sardinian clade and reveal the first case of interspecific hybridization in freshwater flatworms

Daniel Dols-Serrate[1,2], Longhua Guo[3,4], Leonid Kruglyak[5,6] & Marta Riutort[1,2]

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[3] Geriatrics Center and Institute of Gerontology, University of Michigan, Ann Arbor, Michigan, USA

[4] Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, Michigan, United States

[5] Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

[6] Department of Human Genetics, Department of Biological Chemistry, University of California, Los Angeles, California, USA

ORCIDs:

0000-0003-0820-7324 Daniel Dols-Serrate

0000-0001-9690-9750 Longhua Guo

0000-0002-8065-3057 Leonid Kruglyak

0000-0002-2134-7674 Marta Riutort

Corresponding author: D. Dols-Serrate danieldols@ub.edu/bcn27101993@gmail.com

Table of Contents

Supplementary Table S1 – List of localities visited in 2019-2021 sampling campaign..... **275**

Supplementary Table S2 – Sampling coordinates used to train the niche model **278**

Supplementary Table S3 – Complete list of individuals used in this study **279**

Supplementary Table S4 – List of all individuals used to infer a mitochondrial gene tree **283**

Supplementary Table S5 – PCR details per reaction **290**

Supplementary Table S6 – List of specific primers used for amplification and sequencing in the present study **291**

Supplementary Table S7 – List of all datasets assembled **292**

Supplementary Table S8 – Individuals included in each subsample set used **295**

Supplementary Table S9 – AUC values of the environmental niche model **297**

Supplementary Table S10 – Relative contribution of the hydro-environmental variables to the environmental niche model **298**

Supplementary Table S11 – Topology test results **299**

Supplementary Table S12 – ABBA/BABA test results **302**

Supplementary Table S13 – BPP results **303**

Supplementary Table S14 – Mean nodal support and standard deviation per dataset of phylogenetic trees inferred with all populations **304**

Supplementary Table S1. List of localities sampled across Corsica and Sardinia between 2019 and 2021 for the present study.

Population number and Locality	River	Abbrev.	Coordinates	Dugesiidae species	Sample code	Number of samples	Collectors	Date
1 – Aiaccio, Corsica	La Gravona	AI	41°58'25.0"N, 8°49'53.4"E	*Dugesia mariae*	AI1287	54	D. Dols-Serrate & L. Leria	26/11/2019
2 – Setti Polli, Corsica	Le Prunelli	SP	41°55'43.9"N, 8°52'15.6"E	*Dugesia mariae*	SP1290	13	D. Dols-Serrate & L. Leria	26/11/2019
3 – U Crucoli, Corsica	Ruisseau du Crucoli	UC	42°03'31.9"N, 8°57'54.5"E	*Dugesia mariae*	UC1286	50	D. Dols-Serrate & L. Leria	26/11/2019
4 – Riu Golu, Corsica	La Gravona	-	42°05'38.2"N, 9°06'06.1"E	-	-	-	D. Dols-Serrate & L. Leria	26/11/2019
5 – Terminale 20, Corsica	Ruisseau de Fulminato	-	42°07'12.4"N, 9°08'00.1"E	-	-	-	D. Dols-Serrate & L. Leria	26/11/2019
6 – Corte, Corsica	Le Tevignano	CT	42°18'18.3"N, 9°09'55.9"E	*Dugesia mariae*	CT1288	73	D. Dols-Serrate & L. Leria	26/11/2019
7 – Camping Campita, Corsica	Le Golo	CC	42°23'34.2"N, 9°10'34.8"E	*Dugesia mariae*	CC1289	29	D. Dols-Serrate & L. Leria	26/11/2019
8 – Calvi, Corsica	Ruisseau de Lioli	CA	42°29'10.8"N, 8°48'09.2"E	*Dugesia xeropotamica*	CA1291	20	D. Dols-Serrate & L. Leria	27/11/2019
	‘’	‘’	“	*Dugesia xeropotamica*	CA1405	66	D. Dols-Serrate & L. Leria	28/05/2021
9 – Montegrosso, Corsica	Seccu	MG	42°32'17.4"N, 8°50'38.5"E	*Dugesia xeropotamica* and *Dugesia sicula*	MG1292	32	D. Dols-Serrate & L. Leria	27/11/2019
10 – Barchetta, Corsica	Ruisseau de Casaconi	BA	42°30'24.2"N, 9°22'27.3"E	*Dugesia mariae*	BA1293	27	D. Dols-Serrate & L. Leria	27/11/2019

11 – Cioccia, Corsica	Ruisseau de la Cioccia	-	41°29'58.6"N 8°59'54.1"E	-	-	-	D. Dols-Serrate & L. Leria	28/11/2019
12 – Turghja, Corsica	Torrent de Marcuggio	TG	41°51'09.1"N, 8°58'00.1"E	*Dugesia mariae*	TG1294	44	D. Dols-Serrate & L. Leria	28/11/2019
13 – Monte Albo, Sardinia	Riedu su Manganu	MT	40°34'30"N, 9°40'31"E	*Dugesia benazzii*	MT1401	8	D. Dols-Serrate & L. Leria	24/05/2021
14 – Badu de Mola, Sardinia	Sa Laccheddu	-	40°18'11"N, 8°34'57"E	*Schmidtea polychroa*	MR1395	42	D. Dols-Serrate & L. Leria	23/05/2021
15 – Silis, Sardinia	Silis	SI	40°45'24"N, 8°43'28"E	*Dugesia benazzii* and *Dugesia hepta*	SI1402	76	D. Dols-Serrate & L. Leria	25/05/2021
16 – Bunnari, Sardinia	Bunnari	BU	40°42'05"N, 8°35'30"E	*Dugesia benazzii* and *Dugesia hepta*	BU1391	42	D. Dols-Serrate & L. Leria	22/05/2021
17 – Mascari, Sardinia	Mascari	MA	40°41'57"N, 8°35'03"E	*Dugesia benazzii* and *Dugesia hepta*	MA1392	32	D. Dols-Serrate & L. Leria	22/05/2021
18 – Santa Lucia, Sardinia	Santa Lucia	-	40°25'53"N, 8°50'09"E	*Dugesia* sp.	MR1393	1	D. Dols-Serrate & L. Leria	22/05/2021
19 – Otti, Sardinia	Bizzolu	-	40°43'55"N, 9°08'36"E	-	-	-	D. Dols-Serrate & L. Leria	25/05/2021
20 – Canale, Sardinia	Canale Luca 'e Sozzu	-	40°39'16.8"N, 9°12'49.5"E	-	-	-	D. Dols-Serrate & L. Leria	25/05/2021
21 – Mulinu, Sardinia	Mulina	MU	40°24'25"N, 8°37'35"E	*Dugesia benazzii* and *Dugesia hepta*	MU1394	44	D. Dols-Serrate & L. Leria	22/05/2021
22 – Fluminimaggiore, Sardinia	Bau Porcus	FL	39°26'11.4"N, 8°30'35.1"E	*Dugesia leporii*	FL1398	31	D. Dols-Serrate & L. Leria	23/05/2021
23 – Mulinu Savio, Sardinia	Bellu	-	39°31'32.5"N, 8°38'48.1"E	*Schmidtea* sp.	MR1396	21	D. Dols-Serrate & L. Leria	23/05/2021

24 – Tomba dei Giganti, Sardinia	Piricanas	-	40°00'56.8"N, 9°23'55.9"E	-	-	-	D. Dols-Serrate & L. Leria	24/05/2021
25 – Sa Carcaredda, Sardinia	Bacu Silacaccero	SC	40°00'04.4"N, 9°25'16.0"E	*Dugesia benazzii*	SC1400	23	D. Dols-Serrate & L. Leria	24/05/2021
26 – Conchedda, Sardinia	Biddo	-	40°08'43.1"N, 9°12'01.5"E	-	-	-	D. Dols-Serrate & L. Leria	24/05/2021
27 – Monte Cidro, Sardinia	Mannu di Fluminimaggiore	MC	39°27'03.5"N, 8°28'27.3"E	*Dugesia* sp.	MC1397	49	D. Dols-Serrate & L. Leria	23/05/2021
28 – Olzai, Sardinia	Taloro	-	40°09'12.7"N, 9°06'59.1"E	*Schmidtea* sp.	-	15	D. Dols-Serrate & L. Leria	24/05/2021
29 – Birori, Sardinia	S'Adde	-	40°14'54.5"N, 8°48'26.9"E	-	-	-	D. Dols-Serrate & L. Leria	22/05/2021
30 – SP6, Sardinia	Murtazzolu	-	40°14'35.9"N, 8°55'28.5"E	-	-	-	D. Dols-Serrate & L. Leria	22/05/2021
31 – Pianottoli, Corsica	Ruisseau de Canella	-	41°29'13.5"N, 9°04'38.6"E	*Schmidtea* sp.	-	30	D. Dols-Serrate & L. Leria	28/05/2021
32 – Gabaru, Sardinia	Ottava	-	40°45'16.9"N, 8°35'35.8"E	*Dugesia hepta* and *Dugesia benazzii*	-	29	D. Dols-Serrate & L. Leria	25/05/2021
33 – Campo Dorgali, Sardinia	Truncu 'e Murto	-	40°14'17.7"N, 9°31'31.7"E	-	-	-	D. Dols-Serrate & L. Leria	26/05/2021

Supplementary Table S2. List of sampling sites coordinates extracted from Dols-Serrate et al. (2020) to train and infer the environmental niche model of *D. benazzii* s.l.

Longitude	Latitude
8.981389	40.62833
9.116389	40.78111
8.5825	40.30306
8.591667	40.70139
8.596111	40.70472
9.511667	40.74083
9.584167	40.86917
8.584167	40.69917
8.589722	40.7
8.626389	40.40694
9.143333	40.73194
8.835833	40.43139
8.594167	40.82972
9.675278	40.575
9.659722	40.13139

Supplementary Table S3. Complete list of specimens used for the present study and pertinent Gen Bank accession numbers.

Locality	Acronym			Coordinates			Population number
Individual	**Taxonomic identification before analyses**	**Taxonomic identification after analyses**	**Karyotype**	**COI**	**NADH1-tRNAW-COII**	**Collectors***	**Date of collection**
Bunnari, Sardinia	**BU**			**40°42'05"N, 8°35'30"E**			**16**
BU1391-01	*D. benazzii*	*D. benazzii* (PSC1)	-	PP254880	PP259355	DS, L	22/05/2021
BU1391-04	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8[a]	PP254881	PP259356	DS, L	22/05/2021
BU1391-05	*D. hepta*	*D. hepta*	n = 7	PP254882	PP259357	DS, L	22/05/2021
BU1391-07	*D. benazzii*	*D. benazzii* (PSC1)	n = 8[t]	PP254883	PP259358	DS, L	22/05/2021
BU1391-08	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8[a]	PP254884	PP259359	DS, L	22/05/2021
BU1391-09	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8[a]	PP254885	PP259360	DS, L	22/05/2021
Mascari, Sardinia	**MA**			**40°41'60"N, 8°35'23"E**			**17**
MA1392-04	*D. hepta*	*D. hepta*	n = 7	PP254897	PP259367	DS, L	22/05/2021
MA1392-05	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8[a]	PP254898	PP259368	DS, L	22/05/2021
MA1392-06	*D. hepta*	*D. hepta*	n = 7	PP254899	PP259369	DS, L	22/05/2021
MA1392-07	*D. hepta*	*D. hepta*	n = 7	PP254900	PP259370	DS, L	22/05/2021
MA1392-08	*D. hepta*	*D. hepta*	n = 7	PP254901	PP259371	DS, L	22/05/2021
MA1392-09	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8[a]	PP254902	PP259372	DS, L	22/05/2021
Mulinu, Sardinia	**MU**			**40°24'25"N, 8°37'35"E**			**21**
MU1394-06	*D. benazzii*	-	mixoploid	PP254915	PP259384	DS, L	22/05/2021
MU1394-07	*D. benazzii*	*D. benazzii* (PSC1)	mixoploid	PP254916	-	DS, L	22/05/2021
MU1394-08	*D. benazzii*	*D. benazzii* (PSC1)	n = 8[t]	PP254917	PP259385	DS, L	22/05/2021
MU1394-09**	*D. hepta*	*D. benazzii* (PSC1)	mixoploid	PP254918	PP259386	DS, L	22/05/2021
MU1394-10	*D. hepta*	*D. hepta*	n = 7	PP254919	PP259387	DS, L	22/05/2021
MU1394-11	*D. benazzii*	*D. benazzii* (PSC1)	mixoploid	PP254920	PP259388	DS, L	22/05/2021
Monte Cidro, Sardinia	**MC**			**39°27'03.5"N, 8°28'27.3"E**			**27**
MC1397-07	*D. benazzii*	*Dugesia* × *sp.*	mixoploid	PP254903	PP259373	DS, L	23/05/2021
MC1397-08	*D. benazzii*	*Dugesia* × *sp.*	mixoploid	PP254904	PP259374	DS, L	23/05/2021
MC1397-09	*D. benazzii*	-	mixoploid	PP254905	PP259375	DS, L	23/05/2021

MC1397-11	*D. benazzii*	*Dugesia* × *sp.*	mixoploid	PP254906	PP259376	DS, L	23/05/2021
MC1397-12	*D. benazzii*	*Dugesia* × *sp.*	mixoploid	PP254907	PP259377	DS, L	23/05/2021
MC1397-13	*D. benazzii*	*Dugesia* × *sp.*	mixoploid	PP254908	PP259378	DS, L	23/05/2021
Fluminimaggiore, Sardinia	**FL**		**39°26'11.4"N, 8°30'35.1"E**				**22**
FL1398-05	*D. benazzii*	*D. leporii*	n = 8	PP254891	PP259361	DS, L	23/05/2021
FL1398-06	*D. benazzii*	*D. leporii*	n = 8	PP254892	PP259362	DS, L	23/05/2021
FL1398-07	*D. benazzii*	*D. leporii*	n = 8	PP254893	PP259363	DS, L	23/05/2021
FL1398-08	*D. benazzii*	*D. leporii*	n = 8	PP254894	PP259364	DS, L	23/05/2021
FL1398-10	*D. benazzii*	*D. leporii*	n = 8	PP254895	PP259365	DS, L	23/05/2021
FL1398-11	*D. benazzii*	*D. leporii*	n = 8	PP254896	PP259366	DS, L	23/05/2021
Sa Carcaredda, Sardinia	**SC**		**40°00'04.4"N, 9°25'16.0"E**				**25**
SC1400-07	*D. benazzii*	-	n = 8[a]	PP254921	PP259389	DS, L	24/05/2021
SC1400-08	*D. benazzii*	*D. benazzii* (PSC1)	n = 8[a]	PP254922	PP259390	DS, L	24/05/2021
SC1400-09	*D. benazzii*	-	n = 8[a]	PP254923	PP259391	DS, L	24/05/2021
SC1400-10	*D. benazzii*	-	n = 8[a]	PP254924	PP259392	DS, L	24/05/2021
Monte Albo, Sardinia	**MT**		**40°34'30"N, 9°40'31"E**				**13**
MT1401-01	*D. benazzii*	*Dugesia* sp (PSC2)	n = 8*	PP254910	PP259379	DS, L	24/05/2021
MT1401-03	*D. benazzii*	*Dugesia* sp (PSC2)	n = 8*	PP254911	PP259380	DS, L	24/05/2021
MT1401-04	*D. benazzii*	*Dugesia* sp (PSC2)	n = 8*	PP254912	PP259381	DS, L	24/05/2021
MT1401-06	*D. benazzii*	*Dugesia* sp (PSC2)	n = 8*	PP254913	PP259382	DS, L	24/05/2021
MT1401-10	*D. benazzii*	*Dugesia* sp (PSC2)	n = 8*	PP254914	PP259383	DS, L	24/05/2021
Silis, Sardinia	**SI**		**40°45'24.0"N 8°43'28.0"E**				**15**
SI1402-08	*D. hepta*	*D. hepta*	n = 7	PP254925	PP259393	DS, L	25/05/2021
SI1402-09	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8	PP254926	PP259394	DS, L	25/05/2021
SI1402-10	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8	PP254927	PP259395	DS, L	25/05/2021
SI1402-12	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8	PP254928	PP259396	DS, L	25/05/2021
SI1402-13	*Dugesia* × sp.	*D. benazzii* (PSC1)	n = 8	PP254929	PP259397	DS, L	25/05/2021
SI1402-14	*D. hepta*	*D. hepta*	n = 7	PP254930	PP259398	DS, L	25/05/2021

U Crocolli, Corsica	**UC**			**42°03'31.9"N, 8°57'54.5"E**			**3**
UC1286-20	*D. mariae*	*Dugesia* sp (PSC8)	-	OR650821	OR640529	DS, L	26/11/2019
UC1286-25	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR650826	OR640534	DS, L	26/11/2019
UC1286-26	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR560827	-	DS, L	26/11/2019
UC1286-27	*D. mariae*	*Dugesia* sp (PSC8)	-	OR650828	OR640535	DS, L	26/11/2019
UC1286-28	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	PP254931	-	DS, L	26/11/2019
UC1286-29	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR650829	-	DS, L	26/11/2019
Aiaccio, Corsica	**AI**			**41°58'25.0"N, 8°49'53.4"E**			**1**
AI1287-13	*D. mariae*	-	n = 8[a]	-	-	DS, L	26/11/2019
AI1287-16	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR650785	OR640491	DS, L	26/11/2019
AI1287-17	*D. mariae*	-	n = 8[a]	-	-	DS, L	26/11/2019
AI1287-22	*D. mariae*	-	-	-	-	DS, L	26/11/2019
Corte, Corsica	**CT**			**42°18'18.3"N, 9°09'55.9"E**			**6**
CT1288-18	*D. mariae*	-	-	PP254890	-	DS, L	26/11/2019
CT1288-19	*D. mariae*	-	-	OR650800	OR640505	DS, L	26/11/2019
CT1288-21	*D. mariae*	*Dugesia* sp (PSC6)	n = 8[a]	OR650802	OR640507	DS, L	26/11/2019
CT1288-22	*D. mariae*	*Dugesia* sp (PSC3)	-	OR650803	OR640508	DS, L	26/11/2019
CT1288-23	*D. mariae*	*Dugesia* sp (PSC6)	22**	OR650804	OR640509	DS, L	26/11/2019
CT1288-24	*D. mariae*	*D. xeropotamica* (PSC4)	n = 8	-	-	DS, L	26/11/2019
Camping Campita, Corsica	**CC**			**42°23'34.2"N, 9°10'34.8"E**			**7**
CC1289-15	*D. mariae*	*Dugesia* sp (PSC7)	n = 8[a]	OR650796	OR640501	DS, L	26/11/2019
CC1289-16	*D. mariae*	*Dugesia* sp (PSC7)	n = 8[a]	PP254888	-	DS, L	26/11/2019
CC1289-18	*D. mariae*	-	n = 8[t]	OR650798	OR640503	DS, L	26/11/2019
CC1289-19	*D. mariae*	*Dugesia* sp (PSC7)	-	OR650799	OR640504	DS, L	26/11/2019
CC1289-20	*D. mariae*	-	-	PP254889	-	DS, L	26/11/2019
Setti Polli, Corsica	**SP**			**41°55'43.9"N, 8°52'15.6"E**			**2**
SP1290-08	*D. mariae*	*Dugesia* sp (PSC8)	n = 8	OR650809	OR640517	DS, L	26/11/2019
SP1290-09	*D. mariae*	*Dugesia* sp (PSC8)	n = 8	OR650810	OR640518	DS, L	26/11/2019
SP1290-11	*D. mariae*	*Dugesia* sp (PSC8)	n = 8	OR650812	OR640520	DS, L	26/11/2019
SP1290-12	*D. mariae*	*Dugesia* sp (PSC8)	n = 8	OR650813	OR640521	DS, L	26/11/2019

Calvi, Corsica	**CA**		**42°29'10.8"N, 8°48'09.2"E**				**8**
CA1291-08	*D. xeropotamica*	-	n = 8	OR650792	OR640497	DS, L	27/11/2019
CA1291-09	*D. xeropotamica*	-	n = 8	OR650793	OR640498	DS, L	27/11/2019
CA1291-10	*D. xeropotamica*	-	n = 8	PP254886	-	DS, L	27/11/2019
CA1291-11	*D. xeropotamica*	*Dugesia* sp (PSC3)	n = 8	OR650794	OR640499	DS, L	27/11/2019
CA1291-12	*D. xeropotamica*	*Dugesia* sp (PSC3)	n = 8	PP254887	-	DS, L	27/11/2019
CA1291-13	*D. xeropotamica*	*Dugesia* sp (PSC3)	n = 8	OR650795	OR640500	DS, L	27/11/2019
Montegrosso, Corsica	**MG**		**42°32'17.4"N, 8°50'38.5"E**				**9**
MG1292-08	*D. xeropotamica*	*D. xeropotamica* (PSC4)	n = 8	OR650805	OR640511	DS, L	27/11/2019
MG1292-09	*D. xeropotamica*	-	n = 8	OR650806	OR640512	DS, L	27/11/2019
MG1292-10	*D. xeropotamica*	-	-	PP254909	-	DS, L	27/11/2019
MG1292-11	*D. xeropotamica*	-	-	-	-	DS, L	27/11/2019
MG1292-12	*D. xeropotamica*	-	-	-	-	DS, L	27/11/2019
MG1292-13	*D. xeropotamica*	-	n = 8	OR650807	OR640513	DS, L	27/11/2019
Barchetta, Corsica	**BA**		**42°30'24.2"N, 9°22'27.3"E**				**10**
BA1293-10	*D. mariae*	*D. mariae* (PSC5)	n = 8[a]	OR650786	OR640492	DS, L	27/11/2019
BA1293-11	*D. mariae*	*D. mariae* (PSC5)	n = 8	OR650787	OR640493	DS, L	27/11/2019
Turghja, Corsica	**TG**		**41°51'09.1"N, 8°58'00.1"E**				**12**
TG1294-03	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR650814	OR640522	DS, L	28/11/2019
TG1294-04	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR650815	OR640523	DS, L	28/11/2019
TG1294-05	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR650816	OR640524	DS, L	28/11/2019
TG1294-06	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR650817	OR640425	DS, L	28/11/2019
TG1294-07	*D. mariae*	*Dugesia* sp (PSC8)	-	OR650818	OR640426	DS, L	28/11/2019
TG1294-08	*D. mariae*	*Dugesia* sp (PSC8)	n = 8[a]	OR650819	OR640527	DS, L	28/11/2019

***Collectors:** (DS) D. Dols-Serrate, (L) L. Leria. [a] **Aneuploidy** [t] **Triploidy**

** Cross contaminated sample during legacy Sanger sequencing

Supplementary Table S4. List of individuals featured in the mitochondrial gene tree employed for species identification with pertinent Gen Bank accession numbers.

Locality		**Coordinates**	
Individual	**Taxonomic identification based on the mitochondrial gene tree**	**COI**	**NADH1-tRNAW-COII**
Badde Manna, Sardinia		**40°46'52"N, 9°06'59"E**	
BMA0028-01	*D. benazzii*	MN525640	-
Badu de Mola, Sardinia		**40°18'11"N, 8°34'57"E**	
BMO0025-03	*D. hoidi*	MN525637	OP720240
BMO0025-04	*D. hoidi*	OR650791	-
Bunnari, Sardinia		**40°42'05"N, 8°35'30"E**	
BU1391-01	*D. benazzii*	PP254880	PP259355
BU1391-04	*Dugesia* × sp.	PP254881	PP259356
BU1391-05	*D. hepta*	PP254882	PP259357
BU1391-07	*D. benazzii*	PP254883	PP259358
BU1391-08	*Dugesia* × sp.	PP254884	PP259359
BU1391-09	*Dugesia* × sp.	PP254885	PP259360
MR0088-02	*Dugesia* × sp.	MN525658	-
BUII0088-04	*D. benazzii*	MK712640	-
BUII0089-07	*D. benazzii*	MN525665	-
BUII0365-01	*D. hepta*	MN525738	-
BUII0367-01	*D. hepta*	MN525739	-
Bunnari, Sardinia		**40°42'17"N, 8°35'46"E**	
BUIII0090-05	*D. hepta*	MN525671	-
BUIII0091-01	*D. hepta*	MN525674	OP720245
BUIII0091-04	*D. hepta*	MN525677	OP720246
BUIII0091-07	*D. hepta*	MN525680	-
Mascari, Sardinia		**40°41'60"N, 8°35'23"E**	
MA1392-04	*D. hepta*	PP254897	PP259367
MA1392-05	*Dugesia* × sp.	PP254898	PP259368

MA1392-06	*D. hepta*	PP254899	PP259369
MA1392-07	*D. hepta*	PP254900	PP259370
MA1392-08	*D. hepta*	PP254901	PP259371
MA1392-09	*Dugesia* × sp.	PP254902	PP259372
MAII0354-01	*D. hepta*	MN525735	OR640510
MAII0355-01	*D. hepta*	MN525736	OP720281
MAII0357-01	*D. hepta*	MN525737	-
Mascari, Sardinia		**40°41'57"N, 8°35'03"E**	
MAS0077-07	*Dugesia hepta*	MN525649	-
MAS0078-01	*Dugesia hepta*	MN525650	-
MAS0078-02	*Dugesia hepta*	MN525651	-
MAI0342-01	*D. hepta*	MK385923	OP720277
MAI0343-01	*D. hepta*	MN525730	OP720278
MAI0344-01	*D. hepta*	MK385924	OP720279
MAI0345-01	*D. hepta*	MN525731	OP720280
MAI0348-01	*D. hepta*	MN525732	-
Mulinu, Sardinia		**40°24'25"N, 8°37'35"E**	
MU1394-06	*D. benazzii*	PP254915	PP259384
MU1394-07	*D. benazzii*	PP254916	-
MU1394-08	*D. benazzii*	PP254917	PP259385
MU1394-09	*D. hepta*	PP254918	PP259386
MU1394-10	*D. hepta*	PP254919	PP259387
MU1394-11	*D. benazzii*	PP254920	PP259388
MU0172-07	*D. hepta*	MK712639	-
MU0173-04	*D. hepta*	MN525722	OP720273
MU0173-05	*D. hepta*	MN525723	OR640514
MU0174-01	*D. benazzii*	MN525725	OP720274
MU0174-05	*D. benazzii*	MN252726	OP720275
MU0174-06	*D. hepta*	MN525727	OP720276
Monte Cidro, Sardinia		**39°27'03.5"N, 8°28'27.3"E**	
MC1397-07	*D. benazzii*	PP254903	PP259373

MC1397-08	*D. benazzii*	PP254904	PP259374
MC1397-09	*D. benazzii*	PP254905	PP259375
MC1397-11	*D. benazzii*	PP254906	PP259376
MC1397-12	*D. benazzii*	PP254907	PP259377
MC1397-13	*D. benazzii*	PP254908	PP259378
Fluminimaggiore, Sardinia		**39°26'11.4"N, 8°30'35.1"E**	
FL1398-05	*D. benazzii*	PP254891	PP259361
FL1398-06	*D. benazzii*	PP254892	PP259362
FL1398-07	*D. benazzii*	PP254893	PP259363
FL1398-08	*D. benazzii*	PP254894	PP259364
FL1398-10	*D. benazzii*	PP254895	PP259365
FL1398-11	*D. benazzii*	PP254896	PP259366
Sa Carcaredda, Sardinia		**40°00'04.4"N, 9°25'16.0"E**	
SC1400-07	*D. benazzii*	PP254921	PP259389
SC1400-08	*D. benazzii*	PP254922	PP259390
SC1400-09	*D. benazzii*	PP254923	PP259391
SC1400-10	*D. benazzii*	PP254924	PP259392
Santa Lucia, Sardinia		**40°25'53"N, 8°50'09"E**	
SL0030-03	*D. hepta*	MN525641	OP720242
SL0030-04	*D. hepta*	MN525642	-
SL0030-05	*D. hepta*	MK712638	OP720243
SL0030-06	*D. hepta*	MN525643	OR640516
Monte Albo, Sardinia		**40°34'30"N, 9°40'31"E**	
MT1401-01	*D. benazzii*	PP254910	PP259379
MT1401-03	*D. benazzii*	PP254911	PP259380
MT1401-04	*D. benazzii*	PP254912	PP259381
MT1401-06	*D. benazzii*	PP254913	PP259382
MT1401-10	*D. benazzii*	PP254914	PP259383
MTA0385-01	*D. benazzii*	MK385925	OP720282
MTA0385-02	*D. benazzii*	MK385926	OP720283

MTA0385-03	*D. benazzii*	MN525740	OP720284
MTA0385-05	*D. benazzii*	MN525742	OP720286
MTA0385-06	*D. benazzii*	MN525743	OP720287
MTA0385-07	*D. benazzii*	MN525744	-
MTA0385-08	*D. benazzii*	MN525745	OP720288
MTA0385-09	*D. benazzii*	MN525746	OP720289
MTA0385-10	*D. benazzii*	MN525747	OP720290
MTA0385-12	*D. benazzii*	MN525749	OP720292
Silis, Sardinia		**40°49'47"N, 8°35'39"E**	
MR0092-03	*Dugesia* × sp.	MK712643	-
MR0092-04	*Dugesia* × sp.	MN525683	-
MR0092-05	*Dugesia* × sp.	MN525684	-
MR0092-11	*Dugesia* × sp.	MN525690	-
Ponte Valenti, Sardinia		**40°28'09"N, 8°48'21"E**	
PV0096-05	*D. hepta*	MN525697	-
PV0098-01	*D. hepta*	MN525699	OP720247
PV0098-04	*D. hepta*	MN525700	OP720249
PV0099-01	*D. hepta*	MN525701	OP720250
Silis, Sardinia		**40°45'24.0"N 8°43'28.0"E**	
SI1402-08	*D. hepta*	PP254925	PP259393
SI1402-09	*Dugesia* × sp.	PP254926	PP259394
SI1402-10	*Dugesia* × sp.	PP254927	PP259395
SI1402-12	*Dugesia* × sp.	PP254928	PP259396
SI1402-13	*Dugesia* × sp.	PP254929	PP259397
SI1402-14	*D. hepta*	PP254930	PP259398
U Crocolli, Corsica		**42°03'31.9"N, 8°57'54.5"E**	
UC1286-19	*D. mariae*	OR650820	OR640528
UC1286-20	*D. mariae*	OR650821	OR640529
UC1286-21	*D. mariae*	OR650822	OR640530
UC1286-22	*D. mariae*	OR650823	OR640531
UC1286-23	*D. mariae*	OR650824	OR640532

UC1286-24	*D. mariae*	OR650825	OR640533
UC1286-25	*D. mariae*	OR650826	OR640534
UC1286-26	*D. mariae*	OR560827	-
UC1286-27	*D. mariae*	OR650828	OR640535
UC1286-28	*D. mariae*	PP254931	-
UC1286-29	*D. mariae*	OR650829	-
Aiaccio, Corsica		**41°58'25.0"N, 8°49'53.4"E**	
AI1287-15	*D. mariae*	OR650784	OR640490
AI1287-16	*D. mariae*	OR650785	OR640491
Corte, Corsica		**42°18'18.3"N, 9°09'55.9"E**	
CT1288-18	*D. mariae*	PP254890	-
CT1288-19	*D. mariae*	OR650800	OR640505
CT1288-21	*D. mariae*	OR650802	OR640507
CT1288-22	*D. mariae*	OR650803	OR640508
CT1288-23	*D. mariae*	OR650804	OR640509
Camping Campita, Corsica		**42°23'34.2"N, 9°10'34.8"E**	
CC1289-15	*D. mariae*	OR650796	OR640501
CC1289-16	*D. mariae*	PP254888	-
CC1289-18	*D. mariae*	OR650798	OR640503
CC1289-19	*D. mariae*	OR650799	OR640504
CC1289-20	*D. mariae*	PP254889	-
Setti Polli, Corsica		**41°55'43.9"N, 8°52'15.6"E**	
SP1290-08	*D. mariae*	OR650809	OR640517
SP1290-09	*D. mariae*	OR650810	OR640518
SP1290-10	*D. mariae*	OR650811	OR640519
SP1290-11	*D. mariae*	OR650812	OR640520
SP1290-12	*D. mariae*	OR650813	OR640521
Calvi, Corsica		**42°29'10.8"N, 8°48'09.2"E**	
CA1291-08	*D. xeropotamica*	OR650792	OR640497
CA1291-09	*D. xeropotamica*	OR650793	OR640498
CA1291-10	*D. xeropotamica*	PP254886	-

CA1291-11	*D. xeropotamica*	OR650794	OR640499
CA1291-12	*D. xeropotamica*	PP254887	-
CA1291-13	*D. xeropotamica*	OR650795	OR640500
Montegrosso, Corsica		**42°32'17.4"N, 8°50'38.5"E**	
MG1292-08	*D. xeropotamica*	OR650805	OR640511
MG1292-09	*D. xeropotamica*	OR650806	OR640512
MG1292-10	*D. xeropotamica*	PP254909	-
MG1292-13	*D. xeropotamica*	OR650807	OR640513
Barchetta, Corsica		**42°30'24.2"N, 9°22'27.3"E**	
BA0101-01	*D. mariae*	MN525702	OP720251
BA0101-02	*D. mariae*	MN525703	OP720252
BA0101-03	*D. mariae*	OP561860	OP720253
BA1293-10	*D. mariae*	OR650786	OR640492
BA1293-11	*D. mariae*	OR650787	OR640493
BA1293-12	*D. mariae*	OR650788	OR640494
BA1293-13	*D. mariae*	OR650789	OR640495
BA1293-14	*D. mariae*	OR650790	OR640496
Turghja, Corsica		**41°51'09.1"N, 8°58'00.1"E**	
TG1294-03	*D. mariae*	OR650814	OR640522
TG1294-04	*D. mariae*	OR650815	OR640523
TG1294-05	*D. mariae*	OR650816	OR640524
TG1294-06	*D. mariae*	OR650817	OR640425
TG1294-07	*D. mariae*	OR650818	OR640426
TG1294-08	*D. mariae*	OR650819	OR640527
Ecole Communale, Corsica		**42°26.648'N, 9°26.386'E**	
EC0113-05	*D. mariae*	OP561872	OP720266
EC0113-06	*D. mariae*	OP561873	OP720267
EC0113-08	*D. mariae*	OP561875	OP720269
La Solenzara, Corsica		**41°50'39.6"N, 9°20'34"E**	
SO1034-02	*D. mariae*	MN525751	OP720294
SO1034-03	*D. mariae*	MN525752	OP720295

SO1034-07	*D. mariae*	OP561879	OP720299
Cioccia, Corsica		**41°32'41.7"N, 9°01'24.1"E**	
CI1055-03	*D. mariae*	MN525765	OP720307
Pero Longo, Corsica		**41°32'27.6"N, 8°58'07.2"E**	
PL1037-02	*D. mariae*	MN525756	OP720303
PL1037-03	*D. mariae*	MN525757	OP720304
PL1037-04	*D. mariae*	MN525758	OP720305
PL1037-05	*D. mariae*	MN525759	OP720306
Molara, Sardinia		**40°52'04.2"N, 9°43'36.6"E**	
IM0120-01	*D. mariae*	PP087950	PP100177
IM0121-01	*D. mariae*	PP087951	PP100178
Montecristo. Italy		**42°20'26.0"N, 10°19'05.8"E**	
IMC1331-01	*D. mariae*	PP087950	PP100179
IMC1331-02	*D. mariae*	PP087951	PP100180
Trelles, Spain		**43°28'33.6"N, 6°44'00.9"W**	**Outgroup**
C12.1DC	*D. subtentaculata*	MK712596	MK712667 + MK713188 + MK713061
Luarca, Spain		**43°32'30.8"N 6°32'07.4"W**	**Outgroup**
C13.1DC	*D. subtentaculata*	MK712612	MK712669 + MK713190 + MK713063
Monte de Vento, Portugal		**39°17'20.2"N, 7°23'01.8"W**	**Outgroup**
E24.1DC	*D. subtentaculata*	MK712611	MK712682 + MK713200 + MK713074
Peralejos de las Truchas, Spain		**40°35'35.5"N, 1°55'32.2"W**	**Outgroup**
V05.1D	*D. subtentaculata*	MK712608	MK712704 + MK713222 + MK713096

Supplementary Table S5. The following PCR specifics are per reaction, with a final PCR volume of 25 µL. The concentrations listed in the table represent the concentrations of the aliquots that were used in the PCR preparation.

Reagent	Volume (µl)	Concentration
Promega 5x Green GoTaq Flexi Buffer	5	-
$MgCl_2$	2.5	25 mM
dNTPs	1.5	0.5 mM
Forward primer	1	10 µM
Reverse primer	1	10 µM
Taq polymerase	0.15	1 µL
Milli Q autoclaved water	12.85	-
Sample DNA extract	1	-

Supplementary Table S6. List of specific primers used for amplification and sequencing in the present study.

Primers		Marker(s)	Sequence 5'-3'	Annealing Temp. (ºC)	Source
SamCF	F[1]	COI	GCTAATAATTTGAGTTTTTG	51	Dols-Serrate et al., 2020
TarlyCR	R[2]	COI	CATTTTAAAACAACATTACC	51	Dols-Serrate et al., 2020
Cupi2	F	NADH1-tRNAW-COII	ATTTGRCTYCGTTCTTC	43	Leria et al., 2020
coIIP	R	NADH1-tRNAW-COII	CCACAAACCTCAGAAC	43	Leria et al., 2020

[1] Forward [2] Reverse

Supplementary Table S7. Complete list of all datasets assambled for the present study, with indications on the taxon coverage parameters employed and the number of loci and polymorphic loci sampled for each analysis. Cells in green denote usage of pertinent dataset for a given analysis. Green cells with numbers indicate number of retained loci (BPP column) or SNPs (Introgress/Ancestry-painting plot/SNAPP_time_estimate/BFD* columns) after further filtering (see Material and Methods). Sample sets are highlighted in bright orange. For a comprehensive list of individuals selected for each sample set, see Supplementary Table S8.

Loci/sample set	Dataset	Num. individuals (n)	Loci present in at least *i* % of individuals within a population (-r)	Loci present in at least *j* % of populations (-p)	Number of loci	Number of polymorphic loci	Phylogenetic trees	Phylogenetic trees w/o MC	SVDQuartets	Topology test	Structure	PCA	fineRAD structure	Introgress	Ancestry painting plot	ABBA/BABA	SNAPP calibrated tree	BPP	BFD*
All populations																			
Optimal set	OP-0.4P50	63	40	50	20643	20504													
	OP-0.6P50	63	60	50	14936	14754													
	OP-0.7P50	63	70	50	11653	11390											**408**	**500**	
	OP-0.4P75	63	40	75	6464	6369													
	OP-0.6P75	63	60	75	4224	4102													
	OP-0.7P75	63	70	75	2797	2653													
	OP-0.4P100	63	40	100	521	492													
	OP-0.6P100	63	60	100	313	293													
	OP-0.7P100	63	70	100	129	116													
Default set	DF-0.4P50	63	40	50	11888	11802													
	DF-0.6P50	63	60	50	8904	8804													
	DF-0.7P50	63	70	50	7052	6918											**208**	**500**	
	DF-0.4P75	63	40	75	3581	3522													
	DF-0.6P75	63	60	75	2406	2345													
	DF-0.7P75	63	70	75	1461	1387													
	DF-0.4P100	63	40	100	254	248													
	DF-0.6P100	63	60	100	137	130													
	DF-0.7P100	63	70	100	49	44													
Over merged set	OVM-0.4P50	63	40	50	16884	16818													
	OVM-0.6P50	63	60	50	13339	13262													
	OVM-0.7P50	63	70	50	11208	11081											**775**	**500**	

| | | |

OVM-CorP50	25	70	50	14294	13891													
Exclusively Sardinian populations w/o *D. hepta*																		
OP-SarP50	30	70	50	28451	27270													
DF-SarP50	30	70	50	19936	19086													
OVM-SarP50	30	70	50	18619	17983													
All populations w/o *D. hepta*																		
OP-woHP50	55	70	50	11488	11180													
DF-woHP50	55	70	50	6994	6846													
OVM-woHP50	55	70	50	10987	10823													
H×B parental-hybrid trio w/o MC																		
OP-HxBP50	29	70	50	35284	33472								**754**	**71**				
DF-HxBP50	29	70	50	25656	24273								**290**	**31**				
OVM-HxBP50	29	70	50	22052	21170								**1478**	**214**				
L×B parental-hybrid trio w/o H×B putative hybrids																		
OP-LxBP50	21	70	50	29185	27875								**2154**	**75**				
DF-LxBP50	21	70	50	20287	19336								**1189**	**54**				
OVM-LxBP50	21	70	50	19580	18894								**2692**	**101**				
Sardinian populations + UC/AI outgroup																		
OP-ABBAP50	35	70	50	24226	23448													
DF-ABBAP50	35	70	50	16282	15724													
OVM-ABBAP50	35	70	50	16905	16470													
BFD dataset w/o *D. hepta*, *D. leporii* & MC hybrids																		
OP-bfdP50	46	70	50	11218	10959													**636**
DF-bfdP50	46	70	50	6604	6460													**306**
OVM-bfdP50	46	70	50	10523	10364													**702**

Supplementary Table S8. Lists of individuals included in each sample set for the analyses of this study (see Supplementary Table S7).

All populations	Downsample for SNAPP time tree	Corsican populations + MT outgroup	Exclusively Corsican populations	Exclusively Sardinian populations_w /o *D. hepta*	All populations w/o Dhepta	H×B parental-hybrid trio w/o MC	L×B parental-hybrid trio w/o H×B putative hybrids	Sardinian populations + UC/AI outgroup	BFD dataset w/o *D. hepta*, *D. leporii* & MC hybrids
UC1286_20	AI1287_16	UC1286_20	UC1286_20	BU1391_01	UC1286_20	BU1391_01	BU1391_01	UC1286_20	UC1286_20
UC1286_25	CT1288_21	UC1286_25	UC1286_25	BU1391_04	UC1286_25	BU1391_04	BU1391_07	UC1286_25	UC1286_25
UC1286_28	CT1288_22	UC1286_28	UC1286_28	BU1391_07	UC1286_28	BU1391_05	MU1394_07	UC1286_28	UC1286_28
UC1286_29	CT1288_23	UC1286_29	UC1286_29	BU1391_08	UC1286_29	BU1391_07	MU1394_08	UC1286_29	UC1286_29
AI1287_16	CT1288_24	AI1287_16	AI1287_16	BU1391_09	AI1287_16	BU1391_08	MU1394_09	AI1287_16	AI1287_16
CT1288_21	CC1289_15	CT1288_21	CT1288_21	MA1392_05	CT1288_21	BU1391_09	MU1394_11	BU1391_01	CT1288_21
CT1288_22	CC1289_16	CT1288_22	CT1288_22	MA1392_09	CT1288_22	MA1392_04	MC1397_07	BU1391_04	CT1288_22
CT1288_23	CC1289_19v2	CT1288_23	CT1288_23	MU1394_07	CT1288_23	MA1392_05	MC1397_08	BU1391_07	CT1288_23
CT1288_24	SP1290_11	CT1288_24	CT1288_24	MU1394_08	CT1288_24	MA1392_06	MC1397_11	BU1391_08	CT1288_24
CC1289_15	CA1291_12	CC1289_15	CC1289_15	MU1394_09	CC1289_15	MA1392_07	MC1397_12	BU1391_09	CC1289_15
CC1289_16	CA1291_13	CC1289_16	CC1289_16	MU1394_11	CC1289_16	MA1392_08	MC1397_13	MA1392_05	CC1289_16
CC1289_19v2	MG1292_08	CC1289_19v2	CC1289_19v2	MC1397_07	CC1289_19v2	MA1392_09	FL1398_05	MA1392_09	CC1289_19v2
SP1290_08v2	BA1293_10v2	SP1290_08v2	SP1290_08v2	MC1397_08	SP1290_08v2	MU1394_07	FL1398_06	MU1394_07	SP1290_08v2
SP1290_09	BA1293_11	SP1290_09	SP1290_09	MC1397_11	SP1290_09	MU1394_08	FL1398_07	MU1394_08	SP1290_09
SP1290_11	TG1294_06	SP1290_11	SP1290_11	MC1397_12	SP1290_11	MU1394_09	FL1398_11	MU1394_09	SP1290_11
CA1291_11	BU1391_01	CA1291_11	CA1291_11	MC1397_13	CA1291_11	MU1394_10	SC1400_08	MU1394_11	CA1291_11
CA1291_12	MA1392_07	CA1291_12	CA1291_12	FL1398_05	CA1291_12	MU1394_11	MT1401_01	MC1397_07	CA1291_12
CA1291_13	MU1394_11	CA1291_13	CA1291_13	FL1398_06	CA1291_13	SC1400_08	MT1401_03	MC1397_08	CA1291_13
MG1292_08	FL1398_06	MG1292_08	MG1292_08	FL1398_07	MG1292_08	MT1401_01	MT1401_04	MC1397_11	MG1292_08
BA1293_10v2	FL1398_07	BA1293_10v2	BA1293_10v2	FL1398_11	BA1293_10v2	MT1401_03	MT1401_06	MC1397_12	BA1293_10v2
BA1293_11	FL1398_11	BA1293_11	BA1293_11	SC1400_08	BA1293_11	MT1401_04	MT1401_10	MC1397_13	BA1293_11
TG1294_03	SC1400_08	TG1294_03	TG1294_03	MT1401_01	TG1294_03	MT1401_06		FL1398_05	TG1294_03
TG1294_04	MT1401_03	TG1294_04	TG1294_04	MT1401_03	TG1294_04	MT1401_10		FL1398_06	TG1294_04
TG1294_05	MT1401_06	TG1294_05	TG1294_05	MT1401_04	TG1294_05	SI1402_08		FL1398_07	TG1294_05
TG1294_06	MT1401_10	TG1294_06	TG1294_06	MT1401_06	TG1294_06	SI1402_09		FL1398_11	TG1294_06
BU1391_01	SI1402_08	MT1401_04		MT1401_10	BU1391_01	SI1402_10		SC1400_08	BU1391_01
BU1391_04	SI1402_14	MT1401_01		SI1402_09	BU1391_04	SI1402_12		MT1401_01	BU1391_04
BU1391_05		MT1401_06		SI1402_10	BU1391_07	SI1402_13		MT1401_03	BU1391_07
BU1391_07		MT1401_03		SI1402_12	BU1391_08	SI1402_14		MT1401_04	BU1391_08
BU1391_08		UC1286_20		SI1402_13	BU1391_09			MT1401_06	BU1391_09
BU1391_09					MA1392_05			MT1401_10	MA1392_05
MA1392_04					MA1392_09			SI1402_09	MA1392_09
MA1392_05					MU1394_07			SI1402_10	MU1394_07
MA1392_06					MU1394_08			SI1402_12	MU1394_08
MA1392_07					MU1394_09			SI1402_13	MU1394_09
MA1392_08					MU1394_11				MU1394_11
MA1392_09					MC1397_07				SC1400_08
MU1394_07					MC1397_08				MT1401_01
MU1394_08					MC1397_11				MT1401_03

MU1394_09					MC1397_12				MT1401_04
MU1394_10					MC1397_13				MT1401_06
MU1394_11					FL1398_05				MT1401_10
MC1397_07					FL1398_06				SI1402_09
MC1397_08					FL1398_07				SI1402_10
MC1397_11					FL1398_11				SI1402_12
MC1397_12					SC1400_08				SI1402_13
MC1397_13					MT1401_01				UC1286_20
FL1398_05					MT1401_03				
FL1398_06					MT1401_04				
FL1398_07					MT1401_06				
FL1398_11					MT1401_10				
SC1400_08					SI1402_09				
MT1401_01					SI1402_10				
MT1401_03					SI1402_12				
MT1401_04					SI1402_13				
MT1401_06									
MT1401_10									
SI1402_08									
SI1402_09									
SI1402_10									
SI1402_12									
SI1402_13									
SI1402_14									

Supplementary Table S9. Obtained AUC values for the environmental niche models of *D. benazzii* s.l. from Sardinia.

Species	AUC training ± Standard deviation	AUC testing ± Standard deviation
D. benazzii s.l.	0.724 ± 0.0279	0.701 ± 0.180

Supplementary Table S10. Relative contribution of the hydro-environmental variables to the distribution model of *D. benazzii* s.l.

Variable	Percent contribution
Actual evapotranspiration	77.8
Natural discharge	21.9
Air temperature	0.2
Land cover classes	0.1
Precipitation	0

Supplementary Table S11. Topology test results. Topology-N (N) and topology-S (S) are ranked as 1 according to the unconstrained state topology of the dataset analyzed. deltaL: logL difference from the maximal logL in the set. bp-RELL: bootstrap proportion using RELL method. p-KH: p-value of one sided Kishino-Hasegawa test. p-SH: p-value of Shimodaira-Hasegawa test. p-WKH: p-value of weighted KH test. p-WSH: p-value of weighted SH test. p-AU: p-value of approximately unbiased (AU) test. Tests highlighted in bolt denote significant exclusion of the alternate topology (rank 2). (* Probable artefactual result)

Dataset	Topology	Rank	logL	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	p-AU
OP-0.4CP50	S	1	-10549965.70	0.000	0.9992	0.9998	1.0000	0.9998	0.9998	0.9997
	N	2	-10550262.17	296.466	**0.0008**	**0.0002**	**0.0000**	**0.0002**	**0.0002**	**0.0003**
OP-0.6CP50	S	1	-8016520.67	0.000	0.9121	0.9106	0.9106	0.9106	0.9119	0.9081
	N	2	-8016626.62	105.948	0.0879	0.0894	0.0894	0.0894	0.0881	0.0919
OP-0.7CP50	S	1	-5933780.17	0.000	0.7987	0.8036	0.8036	0.8036	0.8036	0.8110
	N	2	-5933842.62	62.446	0.2013	0.1964	0.1964	0.1964	0.1964	0.1890
OP-0.4CP75	N	1	-6348537.44	0.000	0.9403	0.9364	0.9364	0.9364	0.9364	0.9327
	S	2	-6348666.55	112.513	0.0597	0.0636	0.0636	0.0636	0.0636	0.0673
OP-0.6CP75	N	1	-3769435.25	0.000	0.9975	0.9976	0.9976	0.9976	0.9976	0.9974
	S	2	-3769768.07	332.822	**0.0025**	**0.0024**	**0.0024**	**0.0024**	**0.0024**	**0.0026**
OP-0.7CP75	N	1	-2493309.18	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-2493584.43	275.253	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
OP-0.4CP100	N	1	-929667.18	0.000	0.9996	0.9992	0.9992	0.9992	0.9992	0.9998
	S	2	-929824.09	153.910	**0.0004**	**0.0008**	**0.0008**	**0.0008**	**0.0008**	**0.0002**
OP-0.6CP100	N	1	-528440.10	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-528567.97	127.871	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
OP-0.7CP100	N	1	-201708.80	0.000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998
	S	2	-201794.64	85.845	**0.0000**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0002**
DF-0.4CP50	N	1	-6348537.44	0.000	0.9824	0.9825	0.9825	0.9825	0.9825	0.9795
	S	2	-6348666.55	129.115	**0.0176**	**0.0175**	**0.0175**	**0.0175**	**0.0175**	**0.0205**

DF-0.6CP50	N	1	-4925615.12	0.000	0.9905	0.9909	0.9909	0.9909	0.9909	0.9895
	S	2	-4925747.08	131.966	**0.0095**	**0.0091**	**0.0091**	**0.0091**	**0.0091**	**0.0105**
DF-0.7CP50	N	1	-3597735.26	0.000	0.9968	0.9957	0.9957	0.9957	0.9957	0.9952
	S	2	-3597873.95	138.691	**0.0032**	**0.0043**	**0.0043**	**0.0043**	**0.0043**	**0.0048**
DF-0.4CP75	N	1	-3157125.43	0.000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999
	S	2	-3157332.97	207.535	**0.0000**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**
DF-0.6CP75	N	1	-2181206.89	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-2181420.73	213.837	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
DF-0.7CP75	N	1	-1323976.70	0.000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998
	S	2	-1324127.37	150.668	**0.0000**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0002**
DF-0.4CP100	N	1	-490374.57	0.000	0.9987	0.9981	0.9981	0.9981	0.9981	0.9990
	S	2	-490455.63	81.061	**0.0013**	**0.0019**	**0.0019**	**0.0019**	**0.0019**	**0.0010**
DF-0.6CP100	N	1	-250351.20	0.000	0.9881	0.9890	0.9890	0.9890	0.9890	0.9896
	S	2	-250437.35	86.151	**0.0119**	**0.0110**	**0.0110**	**0.0110**	**0.0110**	**0.0104**
DF-0.7CP100	N	1	-72813.01	0.000	0.9617	0.9599	0.9599	0.9599	0.9599	0.9636
	S	2	-72849.45	36.443	**0.0383**	**0.0401**	**0.0401**	**0.0401**	**0.0401**	**0.0364**
OVM-0.4CP50	N	1	-7682758.80	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-7683755.72	994.589	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
OVM-0.6CP50	N	1	-6331603.49	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-6332632.41	928.917	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
OVM-0.7CP50	N	1	-5123771.92	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-5124615.88	843.960	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
OVM-0.4CP75	N	1	-5165917.37	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-5166911.96	994.589	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
OVM-0.6CP75	N	1	-4049803.26	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9364
	S	2	-4050689.77	886.506	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	0.0635*
OVM-0.7CP75	N	1	-2783978.91	0.000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
	S	2	-2784701.40	722.488	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0005**

OVM-0.4CP100	N	1	-1442810.17	0.000	1.0000	1.0000	1.0000	1.0000	0.9999
	S	2	-1443281.89	471.717	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0001**
OVM-0.6CP100	N	1	-761171.682	0.000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-7615441.387	369.705	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
OVM-0.7CP100	N	1	-262764.95	0.000	1.0000	1.0000	1.0000	1.0000	1.0000
	S	2	-263132.03	167.074	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**

Supplementary Table S12. Analyses of introgression using *D*-statistics and f4-ratio tests based on OP, DF, and OVM-ABBAP50 datasets.

P_1	P_2	P_3	*D*	*Z*-score	p-value	f4-ratio	ABBA sites	BABA sites	Dataset
FL	MC	BU	0.449303	7.60145	2.93E-14	0.154397	109.615	41.6508	
FL	MC	MA	0.470489	6.72281	1.78E-11	0.135854	91.7881	33.0521	
FL	MC	MT	0.413805	6.06836	1.29E-09	0.103638	98.7898	40.9604	
FL	MC	MU	0.475504	7.4464	9.59E-14	0.159929	112.54	40.0043	
FL	MC	SC	0.600318	10.1148	2.30E-16	0.204707	102.091	25.4974	
FL	MC	SI	0.412711	7.23969	4.50E-13	0.118472	104.709	43.5296	OP-ABBAP50
BU	MC	FL	0.863332	45.2247	2.30E-16	0.522232	567.868	41.6508	
MA	MC	FL	0.875531	50.0197	2.30E-16	0.507448	498.036	33.0521	
MT	MC	FL	0.847353	27.6089	2.30E-16	0.506204	495.708	40.9604	
MU	MC	FL	0.861995	37.7826	2.30E-16	0.511782	539.746	40.0043	
SC	MC	FL	0.900516	47.036	2.30E-16	0.524301	487.095	25.4974	
SI	MC	FL	0.849017	37.6052	2.30E-16	0.516362	533.087	43.5296	
FL	MC	BU	0.541178	9.16521	2.30E-16	0.214918	68.9399	20.524	
FL	MC	MA	0.550795	9.5306	2.30E-16	0.192421	57.5993	16.6843	
FL	MC	MT	0.461576	5.69687	1.22E-08	0.140003	53.408	19.6748	
FL	MC	MU	0.574493	9.47374	2.30E-16	0.223293	72.9679	19.7196	
FL	MC	SC	0.688809	13.3296	2.30E-16	0.265653	57.8881	10.6668	
FL	MC	SI	0.512896	9.48396	2.30E-16	0.161215	62.2913	20.0558	DF-ABBAP50
BU	MC	FL	0.877946	42.9654	2.30E-16	0.537437	315.786	20.524	
MA	MC	FL	0.885346	44.9822	2.30E-16	0.52747	274.354	16.6843	
MT	MC	FL	0.871728	39.0089	2.30E-16	0.530682	287.091	19.6748	
MU	MC	FL	0.879427	37.7676	2.30E-16	0.537685	307.378	19.7196	
SC	MC	FL	0.928958	57.0193	2.30E-16	0.550898	289.632	10.6668	
SI	MC	FL	0.873643	40.8742	2.30E-16	0.527363	297.392	20.0558	
FL	MC	BU	0.571007	15.044	2.30E-16	0.191839	137.662	37.5914	
FL	MC	MA	0.633632	16.3281	2.30E-16	0.186679	118.383	26.5492	
FL	MC	MT	0.560746	10.9656	2.30E-16	0.145325	129.462	36.4355	
FL	MC	MU	0.621113	16.0679	2.30E-16	0.20967	148.695	34.7531	
FL	MC	SC	0.613234	9.62589	2.30E-16	0.210864	114.504	27.4519	
FL	MC	SI	0.582121	13.7698	2.30E-16	0.169365	134.793	35.6023	OVM-
BU	MC	FL	0.915669	106.294	2.30E-16	0.571725	853.926	37.5914	ABBAP50
MA	MC	FL	0.932626	104.759	2.30E-16	0.566847	761.559	26.5492	
MT	MC	FL	0.912538	103.691	2.30E-16	0.566754	796.735	36.4355	
MU	MC	FL	0.920012	115.983	2.30E-16	0.566721	834.203	34.7531	
SC	MC	FL	0.930678	116.856	2.30E-16	0.562799	764.557	27.4519	
SI	MC	FL	0.916342	101.189	2.30E-16	0.564477	815.538	35.6023	

Supplementary Table S13. Results of guided Bayesian species delimitation in BPP (model A10) based on datasets OP, DF, & OVM-0.7P50, using four different evolutionary and demographic models. Posterior probabilities are shown for each node (A-H, in Fig. 8a). A0: algorithm 0; A1: algorithm 1. Non-supported splits are highlighted in bold.

Dataset & nodes	Model 1 $\theta_1 \sim$ IG(3, 0.002) $\tau_1 \sim$ IG(3, 0.002)		Model 2 $\theta_2 \sim$ IG(3, 0.1) $\tau_2 \sim$ IG(3, 0.002)		Model 3 $\theta_3 \sim$ IG(3, 0.1) $\tau_3 \sim$ IG(3, 0.1)		Model 4 $\theta_4 \sim$ IG(3, 0.002) $\tau_4 \sim$ IG(3, 0.1)	
	A0	A1	A0	A1	A0	A1	A0	A1
OP-0.7P50								
A	1.00	1.00	1.00	1.00	**0.00**	**0.00**	1.00	1.00
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
D	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
E	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
G	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
H	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DF-0.7P50								
A	1.00	1.00	1.00	1.00	**0.00**	**0.00**	1.00	1.00
B	1.00	1.00	1.00	1.00	**0.00**	**0.00**	1.00	1.00
C	1.00	1.00	1.00	1.00	1.00	**0.00**	1.00	1.00
D	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
E	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
G	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
H	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
OVM-0.7P50								
A	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
B	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
D	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
E	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
F	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
G	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
H	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Supplementary Table S14. Mean nodal support and standard deviation per dataset of phylogenetic trees inferred with all populations.

Datasets	SH-aLRT	stdev	Ufboot	stdev
OVM-0.4P50	92,58	18,02	94,68	10,33
OVM-0.6P50	92,82	17,63	94,33	14,25
OVM-0.7P50	92,62	17,90	95,05	10,16
OVM-0.4P75	92,51	20,77	96,05	8,18
OVM-0.6P75	93,60	15,77	93,68	13,45
OVM-0.7P75	92,21	17,45	94,07	11,91
OVM-0.4P100	90,06	16,57	91,37	14,96
OVM-0.6P100	87,97	21,00	93,42	13,70
OVM-0.7P100	78,43	29,34	86,42	16,59
Global OVM	**90,31**	**19,38**	**93,23**	**12,62**
DF-0.4P50	94,22	12,90	96,78	7,60
DF-0.6P50	91,13	20,85	96,67	9,22
DF-0.7P50	91,57	18,45	95,52	8,90
DF-0.4P75	90,60	22,32	93,02	12,69
DF-0.6P75	90,09	19,42	92,60	12,30
DF-0.7P75	85,96	23,23	92,72	12,33
DF-0.4P100	83,38	24,53	88,52	19,92
DF-0.6P100	74,77	33,00	82,27	21,90
DF-0.7P100	58,33	38,23	76,63	22,25
Global DF	**84,45**	**23,66**	**90,52**	**14,12**
OP-0.4P50	93,72	12,90	97,00	7,60
OP-0.6P50	94,98	20,85	95,38	9,22
OP-0.7P50	93,33	18,45	92,63	8,90
OP-0.4P75	90,35	12,90	93,60	7,60
OP-0.6P75	90,45	20,85	92,82	9,22
OP-0.7P75	92,24	18,45	92,23	8,90
OP-0.4P100	80,12	24,53	88,75	19,92
OP-0.6P100	80,56	33,00	86,25	21,90
OP-0.7P100	71,18	38,23	82,18	22,25
Global OP	**87,44**	**22,24**	**91,21**	**12,83**

Chapter IV

The first chromosome-scale assemblies of *Dugesia* shed light on the genome size evolution in flatworms

Dols-Serrate, D., Tenaguillo, I., Pisarenco, V.A., Olivé-Muñiz, M., Rozas, J., Riutort, M.

In preparation

The first chromosome-scale assemblies of *Dugesia* shed light on the genome size evolution in flatworms

Daniel Dols-Serrate[1,2], Ignacio Tenaguillo[1,2], Vadim A. Pisarenco[1,2], Marta Olivé-Muñiz[1,2], Julio Rozas[1,2] & Marta Riutort[1,2]

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

ORCIDs:
0000-0003-0820-7324 Daniel Dols-Serrate
0000-0001-8011-5286 Ignacio Tenaguillo
0000-0002-8085-7674 Marta Olivé-Muñiz
0000-0002-4968-4090 Vadim A. Pisarenco
0000-0002-6839-9148 Julio Rozas
0000-0002-2134-7674 Marta Riutort

Corresponding author: M. Riutort mriutort@ub.edu

Abstract

Genomes are an indispensable source of information for understanding the processes underpinning life on Earth. Knowing their sequences not only empowers scientific discovery for the focal species, but also related taxa. Planarian flatworms (Platyhelminthes, Tricladida) constitute a diverse taxonomic group comprising hundreds to thousands of species distributed globally that exhibit a broad spectrum of phenotypic diversity and have colonized terrestrial, marine, and freshwater environments. However, this wealth of diversity contrasts with the almost complete lack of genomic resources for this taxonomic group. Here, we present the chromosome-scale genome assemblies of *Dugesia etrusca*, *Dugesia hepta* and *Dugesia liguriensis*, which constitute the firsts of their kind for the longstanding genus of freshwater planarians, *Dugesia*. These new assemblies represent a significant improvement in contiguity, exceeding 200 times the level achieved by the single available genome of the genus up to the present day, which is that of the model species *Dugesia japonica*. We performed macrosynteny comparative analyses among our chromosome-scale genomes and that of *Schmidtea mediterranea*. We found a substantial lack of synteny conservation between both genera. Furthermore, our analyses revealed an expansion of DNA transposable elements in *Dugesia* not present in Schmidtea, and discarded the hypothesis of a whole-genome duplication as the underlying cause for the differences in genome size between both genera. In addition, our analyses shed light on the chromosomal rearrangements that originated the uncommon chromosomal complement of *D. hepta* (n = 7).

Keywords: *Dugesia* flatworms; PacBio; genome assemblies; chromosomal rearrangements; genome size evolution; *Schmidtea mediterranea*.

1. Introduction

Evolution acts on genomic changes, subtly weaving the continous and complex process that is speciation. The advent of next-generation sequencing (NGS) has greatly contributed to the advance of comparative genomics and the understanding of how genomes relate to phenotype, evolve, and are regulated and structured (Giani et al., 2020; Kim et al., 2013; Rowley and Corces, 2018; Satam et al., 2023; Slatko et al., 2018). Chromosome-scale genome sequences are necessary for elucidating the evolutionary history of genome structural variation, chromosomal rearrangements and the consequences they may have on reproductive isolation and speciation (Kawakami et al., 2011; King, 1987; Lucek et al., 2023; Mackintosh et al., 2023; Mérot, 2020; Simakov et al., 2022) and, fortunately, such resources have become increasingly available and very useful for non-model organisms. However, research efforts are not equally distributed among taxa and certain taxonomic groups remain under-sampled and mostly unexplored from a genomic standpoint (Hotaling et al., 2021). Such is the case, surprisingly, for spiders (Escuer et al., 2022), aquatic insects (Hotaling et al., 2020), cold-water corals (MacLeod et al., 2023) or, as subjects of this study, planarians.

Planarian flatworms (Platyhelminthes, Tricladida) are a species-rich taxon encompassing hundreds, if not thousands, of species that exist worldwide and represent a rich phenotypic diversity that can be found in terrestrial, marine, and freshwater environments (Sluys and Riutort, 2018). Flatworms have been extensively studied in ecology, (Vila-Farré and Rink, 2018; and references therein), autophagy dynamics (González-Estévez, 2008; and references therein), and biogeography (e.g., Sluys et al., 1998; Solà et al., 2022) to name a few disciplines. Furthermore, planarians exhibit highly diverse reproductive strategies (Beukeboom et al., 1998; Chinone et al., 2014; Sluys and Riutort, 2018; Stocchino and Manconi, 2013; Vila-Farré and Rink, 2018), even among conspecific populations, wherein not only multiple biotypes can coexist (e.g., Benazzi-Lentati, 1979; Vacca et al., 1993), but also intercross (e.g., Pongratz et al., 1998). However, planarians are mostly studied for their unmatched regenerative abilities (Ivankovic et al., 2019; Reddien, 2018; Wagner et al., 2011) and their pluripotent stem cells (neoblasts), which represent 20-30% of the parenchymatic cells (González-Estévez and Saló, 2010). Anecdotally, this regenerative prowess has earned planarians the reputation of being '*immortal under the edge of the knife*' (Dalyell, 1814) but, most notably, has established them as model systems in regeneration and stem cell dynamics. It would be reasonable to expect that such a diverse taxonomic group would possess a wealth of genomic resources, especially considering how the model species *Schmidtea mediterranea* Benazzi, Baguñà, Ballester, Puccinelli & Del Papa, 1975, was among the early cohort of Sanger-sequenced genomes (Robb et al., 2015, 2007). However, only three genome assemblies of triclads are currently available for the planarian research community: *S. mediterranea* (Grohme et al., 2018; Guo et al., 2022), *Schmidtea polychroa* (Schmidt, 1861) (Póti et al., 2024), and *Dugesia japonica* Ichikawa & Kawakatsu, 1964 (Tian et al., 2022). Thus, it is desirable to obtain new genomic resources of the taxon.

The longstanding genus *Dugesia* Girard, 1850 is a species-rich genus of freshwater planarians which encompasses approximately 110 species, roughly half of the known species of the family Dugesiidae Ball, 1974, where *Schmidtea* belongs as well (only 4 spp.). Through a series of studies focused on *Dugesia* of the Western Mediterranean region, the potential of these taxa as model systems for evolutionary biology and phylogeography has been highlighted (Leria et al., 2019, 2022; Benítez-Álvarez et al., 2023; Dols-Serrate et al. 2025). Within this region, a species with a rather peculiar chromosomal set can be found. *Dugesia hepta* Pala, Casu & Vacca, 1981, Sardinian endemic and member of the Corso-Sardinian clade, possesses a haploid complement of 7 chromosomes (n = 7), unique among all *Dugesia* species in the Wester Palearctic region (Stocchino et al., 2005) where 8 chromosomes is the most common complement. All sister taxa to *D. hepta* within the Corso-Sardinian clade share the aforementioned haploid complement, making it an interesting case for studying how chromosomal rearrangements might interplay with speciation.

Here, we present the first three high-quality chromosome-scale genome assemblies for the species *D. hepta*, *Dugesia etrusca* Benazzi, 1946, and *Dugesia liguriensis* de Vries, 1988 using Pacific Bioscience (PacBio) long reads and proximity ligation libraries (Omni-C libraries), and provide a first comparative macrosynteny perspective with the model species S. mediterranea. These new genomic resources will benefit the planarian research community and contribute to the developing understanding of the structure, organization, and genome evolution in planarians.

2. Materials and methods

2.1. Samples

All animals used for these analyses were collected from wild populations and maintained in laboratory in 1× Montjuïc water (1.6 mM NaCl, 1.0 mM $CaCl_2$, 1.0 mM $MgSO_4$, 0.1 mM $MgCl_2$, 0.1 mM KCl, and 1.2 mM $NaHCO_3$ in DI H_2O, pH 6.9–8.1) at a constant temperature of 14 ºC. The *D. hepta* samples were collected at 40.756564; 8604074 near the city of Sassari, Sardinia (Italy). The *D. etrusca* samples were collected at 43.497778, 10.625833, near the town of Rivalto (Italy). The *D. liguriensis* samples were collected at 43.823611, 6.581944, near the village of La Garde (France). Animals were starved for a week prior to DNA/RNA extractions. Animals destined to DNA extractions were frozen with liquid nitrogen and stored at -80 ºC before being sent to Dovetail Genomics (Scotts Valley, CA).

2.2. Flow cytometry

We estimated the genome size of *D. etrusca*, *D. hepta* and *D. liguriensis* by means of flow cytometry readings. Samples were prepared by incubating living individuals in a solution of 2% N-acetyl-L-cysteine at pH 7 to remove the mucus and minimize the formation of cell aggregates. Animals were rinsed using a mixture of distilled water and tap

water (1:1) for 1 minute and placed in separated Petri dishes. Thereafter, samples were minced using a sterilized stainless steel razor and incubated in 1 ml Galbraith's isolation buffer (Galbraith et al., 1983) supplemented with 10 µl of RNase A (1 mg/µl) for 15 minutes. After incubation, the cells were retrieved and separated by gently pipetting the liquid contents through a nylon mesh (pore size, 75 µm) into an Eppendorf tube. Further, the cell suspension was stained for 15 minutes with 30 µl of propidium iodide (stock, 1 mg/ml). The quantity of DNA was measured with a Gallios Flow Cytometer at the Unitat de Citometria dels Centres Científics i Tecnològics de la UB (CCiT, UB). Finally, genome size was determined by comparing the readings of fluorescence of our samples against those of males of *Blatella germanica*, with a known genome size of 2.025 Gb.

2.2. DNA/RNA extractions, sequencing and genome assembly

RNA extractions were performed using Trizol (Thermo Fisher Scientific, USA) following the manufacturer's instructions. Total RNA quantification and integrity number were assessed with Qubit and Bioanalyzer in the Centres Científics i Tecnològics, Universitat de Barcelona (CCiT, UB). Truseq stranded and ribo-zero libraries were constructed in Macrogen Inc., (Macrogen Europe, Madrid) to obtain Illumina paired-end reads.

DNA extraction, sequencing and genome assembly were performed at Dovetail Genomics (Scotts Valley, CA). High-quality genomic DNA of each species was extracted from pooled samples following the DNA extraction protocol detailed in Zhang and Alvarado (2018). Genomic sequence data was obtained using a single library per species. We used PacBio CLR (*D. hepta*, *D. etrusca*) and PacBio HiFi (*D. liguriensis*) libraries. Next, Omni-C libraries (one library per each species) were build, followed by Illumina 150-bp paired-end (HiSeq X platform) sequencing to perform the chromosome-level scaffolding using the HiRise software (Putnam et al., 2016). We determined the completeness of the new assemblies by applying the pipeline of Benchmarking Universal Single Copy Orthologues (BUSCO; v.5.4.2; Seppey et al., 2019), searching the metazoa (odb10; 954 genes), arthropoda (odb10; 1013 genes) and eukaryota (odb10; 255 genes) data sets.

2.3. Contaminants sequence removal

Mitochondrial genome sequences were removed from the assembled scaffolds using BLAST+ *blastn* v.2.12.0 (Altschul et al., 1997; Camacho et al., 2009) searches against the mitogenomes of *D. japonica* (GenBank accession number AB618487) and *Dugesia ryukyuensis* Kawakatsu, 1976 (AB618487). Any hit of >1kb was removed and replaced by Ns in the pertinent scaffold. We filtered out potential contaminant sequences and/or errors present in the scaffolds using different criteria. We mapped our genome assemblies against the nucleotide database of the NCBI using BLAST+ *blastn* and Blobtools v.3.4.0, as implemented in BlobToolKit (Laetsch and Blaxter, 2017; Challis et al., 2020) and discarded any scaffold with bacterial, virus or fungal hits. Thereafter, we calculated the mean depth

of coverage for the remaining scaffolds and discarded those that fell outside the 95% confidence interval and did not exhibit BUSCO hits.

2.4. Identification of repetitive elements

We analysed repetitive regions of the genome assemblies using a combination of de novo transposable elements (TE) identification with RepeatModeler v2.0.3 (Smit and Humbley, 2008-2015; RepeatModeler) and DeepTE (Haidong et al., 2020), and a similarity-based search using RepearMasker v.4.1.2 (Smit et al., 2013-2015; RepeatMasker). We built a database of repetitive elements (RE) of each species. We modelled RE using RepeatModeler and parsed its output to classify RE into identified and unknown elements. Thereafter, we used DeepTE to classify the unknown RE using the metazoans model (-sp M). Lastly, all classified RE were concatenated into a single database and used to soft-mask (i.e., RE sequences in lowercase) the repetitive sequences of each genome assembly using the –xsmall option of RepeatMasker.

2.5. Structural and functional genome annotation

Structural annotation of the soft-masked genome assemblies was performed with BRAKER2 (Brůna et al., 2021; Hoff et al., 2019; Stanke et al., 2006, 2008) using RNAseq data of each species, which had been previously mapped to their respective assemblies using STAR v.2.7.10b (Dobin et al., 2013). To functionally annotate gene models, we used BLASTP v.2.2 (*E*-value = 10^{-3}) searches against Swiss-Prot. Furthermore, we also conducted searches against the eggNOG database using eggNOG-mapper v.2.1.9 (Huerta-Cepas et al., 2019; Buchfink et al., 2021; Cantalapiedra et al., 2021) and searched the predicted peptides for specific protein domain signatures within InterProScan v.5.56.89 (Jones et al., 2014). All functional evidence was integrated. Concurrently, we detected the transfer RNA genes (tRNAs) encoded in the genomic sequences using the tRNAscan-SE v2.0.7 software (Chan et al., 2021).

2.6. Detection of whole-genome duplication

We constructed the K_s distribution of *D. etrusca*, *D. hepta*, and *D. liguriensis* to evaluate the possibility of a whole-genome duplication (WGD) event as the underlying cause for the conspicuous difference of genome size between *Dugesia* and *S. mediterranea*. First, using wgd v2 (Chen and Zwaenepoel, 2023), we inferred the paralogous gene families of *D. hepta*, *D. etrusca* and *D. liguriensis* using the coding sequences of each assembly (wgd dmd). Thereafter, the K_s distribution was inferred (wgd ksd) and represented (wgd viz). Furthermore, we used BLASTP v.2.2 (*E*-value = 10^{-10} and top 5 matches) to search all potential homologous gene pairs of protein sequences within each species and identified the collinear gene pairs with MCScanX (Wang et al., 2012).

2.7. Intron and CDS lengths

We estimated the distribution of intron and CDS lengths of *Dugesia* and *S. mediterranea* using the largest scaffolds or pseudochromosomes (*D. etrusca*; *D. hepta*, 7 scaffolds and *D. liguriensis*, 8 scaffolds; *S. mediterranea*, 4 scaffolds, Guo et al. (2022) assembly). We inferred orthologous groups among species using Broccoli v.1.2 (Derelle et al., 2020) and selected all single copy orthologues (1:1:1:1). We visualized intron and CDS length distributions using *ggplot2* and compared the results for each *Dugesia* species against *S. mediterranea*.

2.8. Macrosynteny analysis

To explore the possible chromosomal rearrangement occurred in *D. hepta*, we used the R package GENESPACE v.1.2.3 (Lovell et al., 2022), which uses a combination of OrthoFinder (Emms and Kelly, 2019) and a reimplementation of the MCScanX algorithm, to infer gene-ordered based syntenic blocks and visualized them using the built-in riparian plot function.

3. Results and discussion

3.1. New *Dugesia* genomes: assembly and annotations

Here we present three new chromosome-scale genome assemblies for the freshwater flatworms *D. etrusca*, *D. hepta* and *D. liguriensis* with an assembly size of 2.96 Gb, 2.09 Gb and 2.14 Gb respectively (Fig. 1, Table 1). Despite having a large number of scaffolds, in special *D. etrusca*'s assembly (Detr_v1), the N50 and L50 values (Table 1) indicate a high contiguity of the assemblies, mostly due to the largest scaffolds. The N50 values are similar to those reported for *Schmidtea* genomes (Ivankovic et al., 2023 preprint) and greater than the values reported for *D. japonica* (cf. Tian et al., 2022). The largest scaffolds, or pseudochromosomes, represent 82.7% (Dhep_v1, 7 pseudochromosomes), 74.3% (Dlig_v1, 8 pseudochromosomes) and 52.5% (Detr_v1, 8 pseudochromosomes) of total genome assembly size that (Fig. 2), overall, match the known karyology of each species (*D. hepta*, Pala et al., 1982; *D. etrusca* and *D. liguriensis*, Lázaro et al., 2009). While it is true that the results for Detr_v1, at first sight, are not impressive, it is worth mentioning that estimates of *D. etrusca*'s genome size based on flow cytometry results indicate a relative genome size of approximately 2 Gb, rather than ~3 Gb (Table 1). The pseudochromosomes of *D. etrusca* account for 1.55 Gb, which, would represent approximately 75% of the estimated genome size, a similar fraction to Dlig_v1's. The strong compositional bias (>70% A/T) and the abundance of repetitive elements, as well as the known mosaicism in *Dugesia* (see mosaic-Meselson effect in Leria et al., 2019) might explain the highly fragmented assembly . Most notably, all three assemblies share a similar high fraction of repeat content (Table 2) that substantially exceed reported values for *D. japonica* (~65%, Tian et al., 2022) and *S. mediterranea* (~62%, Grohme et al., 2018).

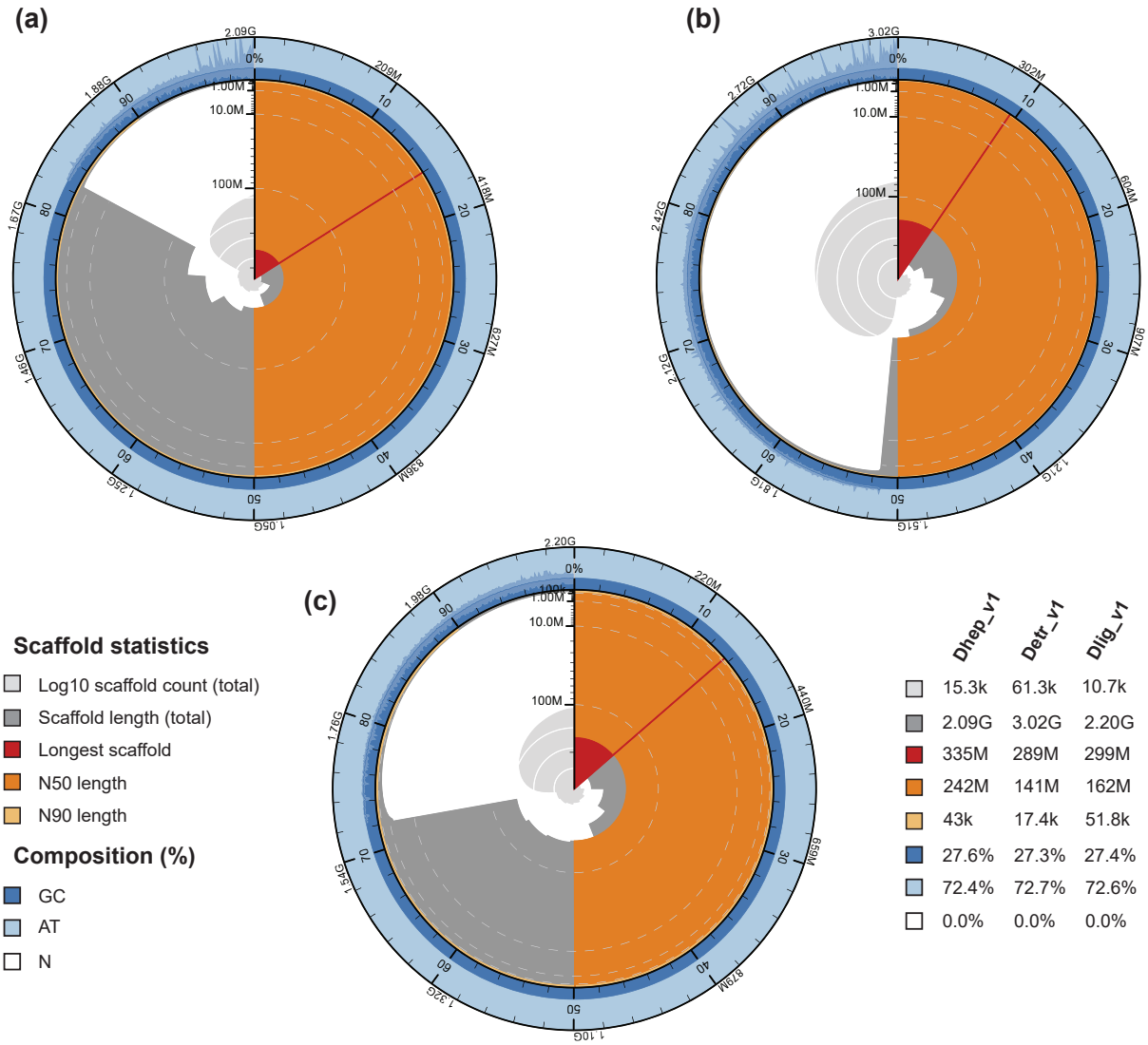


Figure 1. Summary of genome assembly statistics for (a) *D. hepta* (Dhep_v1), (b) *D. etrusca* (Detr_v1), and (c) *D. liguriensis* (Dlig_v1).

The structural annotation (Table 2) revealed that *D. hepta* encodes 35,345 protein-coding genes (37,773 transcripts), *D. liguriensis* encodes 45,367 protein-coding genes (47,772 transcripts), and *D. etrusca* encodes 53,829 protein-coding genes (56,813 transcripts). These values substantially exceed the reported protein-coding genes in *D. japonica* (12,031 genes). However, results for *D. hepta* are similar to those reported in *S. mediterranea* (~31,000 genes in Grohme et al., 2018; 34,976 genes modelled using Guo et al., 2022 assembly). Sequence similarity-based searches showed that the majority of protein-coding genes had hits against the surveyed protein databases for functional annotation (67.9%, 79.4% and 77.1% for *D. etrusca*, *D. hepta* and *D. liguriensis* respectively). Furthermore, the BUSCO search results against the Arthropoda, Metazoa, and Eukaryota data sets generally improved when using the structurally annotated assemblies (Fig. 3, Table 2), except for Dlig_v1. Similar results had been previously reported in planarians (Ivankovic et al., preprint, 2023), suggesting that BUSCO searches are sub-optimal for planarian genomes.

Table 1. Genome assembly statistics after contaminants removal.

	Detr_v1	Dhep_v1	Dlig_v1
Flow cytometry (Gb)	2.02	2.05	2.01
Assembly size (bp)	2,961,733,428	2,090,052,221	2,136,556,669
GC content (%)	27.22	27.64	27.34
Number of scaffolds	59,495	15,220	8,935
Longest scaffold (Mb)	288.8	335.3	299.4
Sequence in chromosomes (%)	52.50	82.73	74.26
N50 (Mb)	141.2	242.1	162.2
L50	8	4	5
Complete BUSCOs			
Metazoa data set	553 (58.0%)	616 (64.6%)	659 (69.1%)
Arthropoda data set	599 (59.1%)	591 (58.3%)	636 (62.7%)
Eukaryota data set	189 (74.1%)	180 (70.6%)	199 (78.0%)

Moreover, our results add to the accumulating evidence that planarians would benefit from having a group-specific BUSCO set (e.g., Benítez-Álvarez et al., 2023; Ivankovic et al., preprint, 2023; Tian et al., 2022) as indicated by the high number of missing BUSCOs (Supplementary Table S1).

All things considered, the new chromosome-scale genome assemblies of *D. hepta*, *D. etrusca* and *D. liguriensis* represent a new genomic resource for the planarian research community and may offer an opportunity to deepen the knowledge on this taxonomic group. In terms of contiguity, these assemblies imply an improvement of between 215 and 370 times (based on N50 values) compared to the only publicly available *Dugesia* chromosome-scale assembly. This improvement is also reflected in the high number of annotated genes (Table 2).

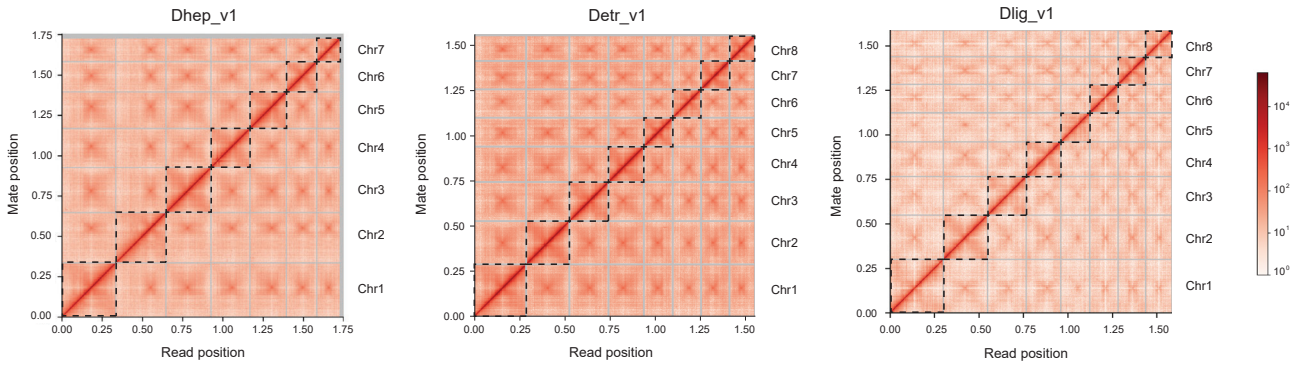


Figure 2. Long-range contact heatmap of paired Omni-C reads depicting the major scaffolds of pseudochromosomes (highlighted in dotted squares) of Dhep_v1, Detr_v1, and Dlig_v1 assemblies. Axes show the mapping position of the first and second read in a read pair (in Gb), respectively, grouped into bins, which are coloured based on the number of read pairs in that bin.

3.1. Possible causes of genome size difference between *Schmidtea* and *Dugesia*

3.1.1. Whole-genome duplication

Little is known about the evolution of genome sizes in freshwater planarian. Re-

Table 2. Genome annotation statistics.

	Detr_v1	Dhep_v1	Dlig_v1
Repetitive regions (%)	78.32	78.68	78.91
Transposable & repetitive elements (bp)	2,318,718,339	1,643,800,512	1,685,580,944
Number of elements	5,922,150	4,067,511	4,794,532
Genome annotation			
Protein-coding genes	53,829	35,435	45,367
Number of transcripts	56,813	37,773	47,772
Supported genes	27,240	18,588	21,268
Functionally annotated genes	39,543	28,146	34,991
Number of tRNA	1336	6106	19549
Number of tRNA after filtering	842	3247	16080
Complete BUSCOs			
Metazoa data set	699 (73.3%)	689 (72.2%)	539 (56.5%)
Arthropoda data set	718 (70.9%)	701 (69.2%)	561 (55.3%)
Eukaryota data set	74.1 (82.8%)	205 (80.4%)	170 (66.7%)

cently, a study by Ivankovic et al. (preprint 2023) reported genome size estimates of the four known species within the genus *Schmidtea*, revealing substantial differences between a clade of small genome sizes (*S. mediterranea* and *S. polychroa*; 0.82~0.78 Gb respectively) and a clade of relatively large genome sizes (*Schmidtea lugubris* Schmidt, 1861 and *Schmidtea nova* Benazzi, 1982; 1.5~1.2 Gb respectively). It was proposed that a large portion of said increase could be attributed to a TE expansion and an increase of the protein-coding genes due transposon insertions that would have expanded intron regions. Such hypothesis would reasonably explain why despite the massive occurrence of syntenic block rearrangements within *Schmidtea* (Ivankovic et al., preprint 2023) the haploid complement (n = 4) is mostly well preserved between both clades (Leria et al., 2018).

Dugesia represents one of the closest relative taxa to *Schmidtea* and had been considered as its sister taxon until the discovery of the genus *Recurva* Sluys, 2013. Many *Dugesia*

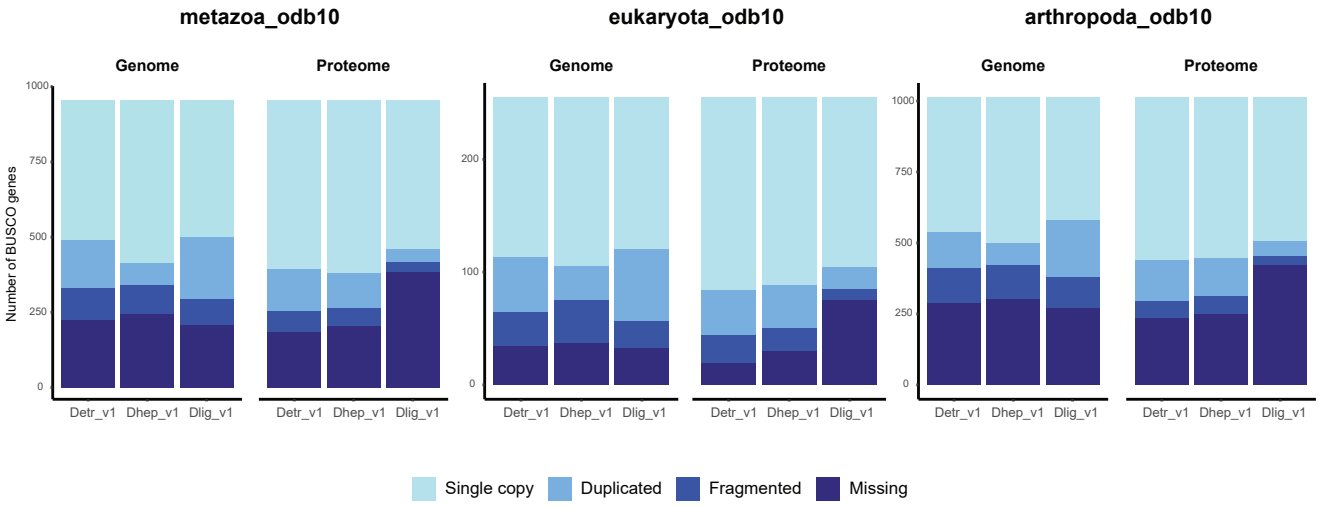


Figure 3. BUSCO assessment of the genome assemblies before (Genome) and after (Proteome) functional annotations.

species share a haploid complement number of n = 8 and some genome size estimates of Western Mediterranean species suggest that ~2 Gb is a common size (e.g., *D. benazzii* 2.05 Gb, *D. leporii* 2.38 Gb, *D. xeropotamica* 2.09 Gb, Dols-Serrate at al., unpublished data). In view of the aforementioned points, it fell within reason to contemplate a WGD event as the underlying cause for the differences observed between *Schmidtea* and *Dugesia*. The availability of the three new chromosome-scale genome assemblies provided the first opportunity to explore this scenario.

Our results for the K_s distribution (Fig. 4, Supplementary Fig. S1), however, indicate that most retained gene duplicates are recent and do not exhibit the expected signal of a WGD (Tiley et al., 2018). The distribution of K_s can serve as a proxy for the distribution of the divergence times of all duplication events that have left a trace in the genome of interest. A distribution pattern of exponential decay, as present in our case, is expected under the assumption of a constant gene duplication rate and a constant rate of loss of duplicated copies (Lynch & Conery, 2000). However, a WGD event entails the duplication of all gene copies and a massive loss or divergence of said copies in a relatively short period of time, which would leave a peak signature of a large number of duplication events with similar divergence times. Furthermore, WGDs are expected to leave large blocks of intragenomic collinearity (Vanneste et al., 2014). Our results show that only 1.04% (*D. hepta*), 1.40% (*D. etrusca*) and 2.19% (*D. liguriensis*) of genes are collinear, supporting the K_s distribution results (e.g., *Arabidopsis thaliana*, an autopolyploid, has approximately ~27% collinear genes).

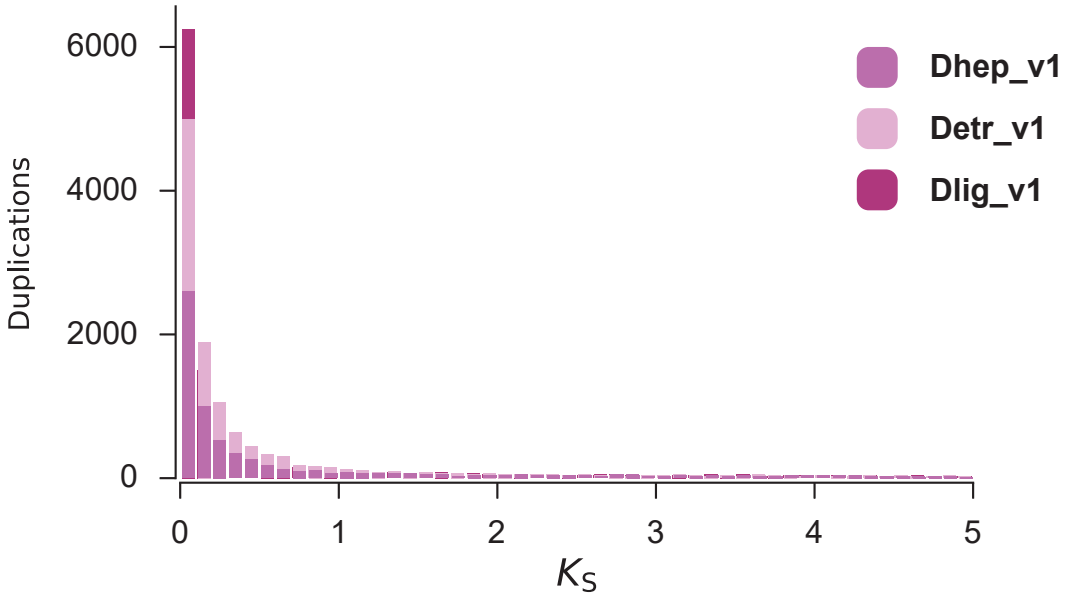


Figure 4. Overlayed paranome K_s distributions of the three genome assemblies showing a pattern of exponential decay.

3.1.2. Genome architecture of non-coding regions

To explore other genomic features that could account for the differences in genome size between both genera, we plotted the density distribution of intron and CDS lengths

(Fig. 5, Table 3) of single-copy orthologues (SO). Homology-search results based on Broccoli v.1.2 identified 3065 orthogroups (1:1:1:1) shared among *D. hepta*, *D. etrusca*, *D. liguriensis* and *S. mediterranea*, 864 of which were exon-equivalent (same number of exons in all species).

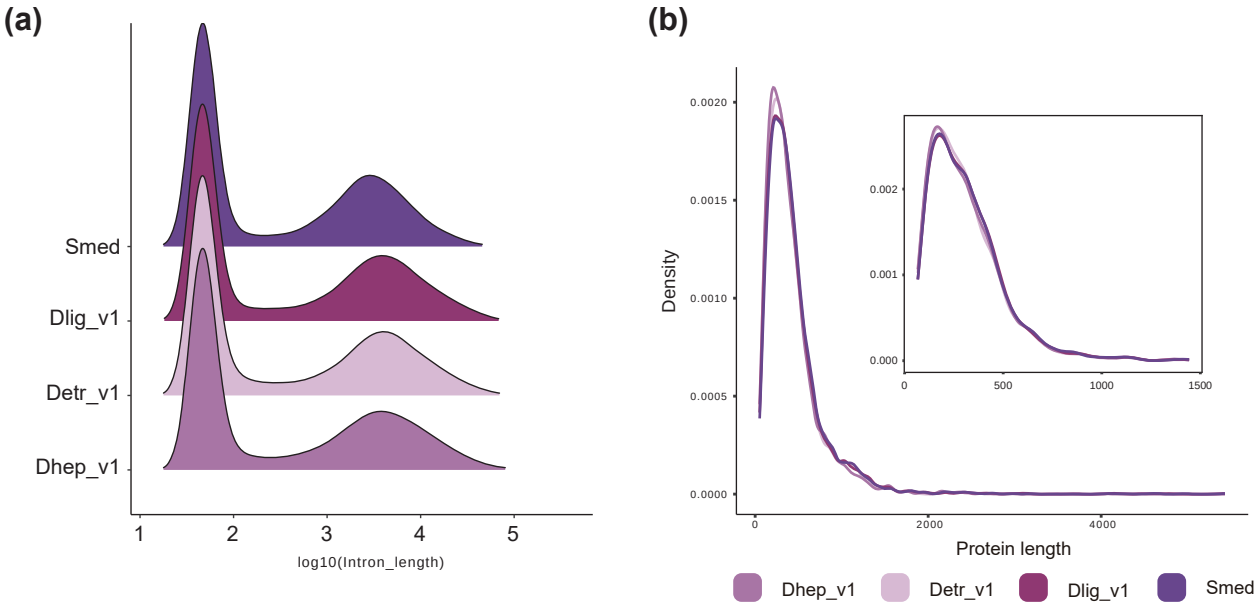


Figure 5. (a) Ridgeline plots showing the density distribution of intron lengths of all introns comprised in the 3065 single-copy orthogroups shared between *D. hepta* (Dhep_v1), *D. etrusca* (Detr_v1), *D. liguriensis* (Dlig_v1), and *S. mediterranea* (Smed). (b) Density distribution of CDS lengths of all 3065 single-copy orthologues. Inset shows the same distribution for exon-equivalent orthologues, 864 in total.

Based on the ridgeline plot results (Fig. 5a, Supplementary Fig. S2), we set a threshold value of 2.5 for the logarithmic intron length (316 bp) to classify introns in two classes: small-sized (≤ 316 bp) and large-sized (> 316 bp). Differences in small-sized introns were meaningless among *Dugesia* species and *S. mediterranea*, although the latter presented a lesser number of small-sized introns. Furthermore, large-sized introns were larger and more abundant in *Dugesia* than *S. mediterranea* (Table 3), a difference that can also be observed by the displacement of the large-sized introns peak in *Dugesia*'s ridgelines (Fig. 5a). On the other hand, results for the CDS density distribution were largely similar among the studied species (Fig. 5b, Table 3), indicating that the changes that led to the increase of genome size likely occurred in non-coding regions.

Intrigued by these findings, we examined the repeat content of the four genomes (Fig. 6, Supplementary Fig. S3). Transposable element annotations revealed that a large portion of the repetitive content in *Dugesia* was assigned to DNA elements (50-65%), whereas in *S. mediterranea* said type of elements contributed to less than 40% of the repeat content (Supplementary Fig. S3a). In absolute terms, the increase in genome size could be attributable to an expansion of transposable elements (TE), most particularly, DNA/*Maverick*, DNA/*DDE* and, as reported in *Schmidtea*, LTR/*Gypsy* elements (Fig. 6a,b). When comparing the relative composition of the repeat content we find that the proportion of DNA/

Table 3. Summary length (bp) statistics of introns and protein-coding sequences (CDS) of single-copy orthologues

All introns	**Detr_v1**	**Dhep_v1**	**Dlig_v1**	**Smed**
Number of introns	15,509	14,901	14,768	13,016
Shortest intron	33	38	38	34
Median lenght	101	97	160	104
Mean length	3335	3467	3412	2364
Longest intron	133,684	114,161	102,489	168,052
Short introns (≤316 bp)				
Number of introns	8431	8070	7729	6959
Shortest intron	33	38	38	34
Median lenght	47	47	47	47
Mean length	59.55	58.81	58.92	57.76
Longest intron	316	316	316	316
Large introns (>316 bp)				
Number of introns	7078	6831	7039	6057
Shortest intron	318	317	317	317
Median lenght	4076	3974	4047	2988
Mean length	7236	7492	7095	5013
Longest intron	133,684	114,161	102,489	168,052
All CDS				
Number of CDS	3065	3065	3065	3065
Shortest CDS	52	67	67	67
Median length	342	335	355	361
Mean length	425.4	417.4	444.9	450.8
Longest CDS	4462	4405	5417	5431
Exon-equivalent CDS				
Number of CDS	864	864	864	864
Shortest CDS	71	69	71	67
Median length	268.5	269.0	277.5	276.0
Mean length	304.7	305.5	312.7	312.9
Longest CDS	1441	1387	1441	1443

DDE elements is similar between *Schmidtea* and *Dugesia*. However, in *Dugesia* there is a clear expansion, most notably in *D. etrusca*, of DNA/*Maverick* TEs, which are the largest TEs known in eukaryotes, 15 ~ 40 kb (Haapa-Paananen et al., 2014; Bellas and Sommaruga, 2021). Comparisons with the S. mediterranea genome also revealed that *Dugesia* and Schmidtea share a highly similar LTR composition, with > 90% of LTR TE being *Gypsy* elements, yet significant differences are found when examining TEs of the LINE and SINE orders (Fig. 6d,e). In *Dugesia* there is a clear expansion of LINE/ RTE elements, the smallest type of non-LTR retrotransposons (Župunski et al., 2001), as well as an expansion of the SINE/*5S* and SINE/*7SL* subfamilies, which are not present in *Schmidtea* (Supplementary Fig. S3e). Interestingly, we found that more than 60% of the SINE elements in *D. hepta* are SINE/*5S* elements, which is almost 4 times more than the prevalence of this type of elements in *D. etrusca* or *D. liguriensis*.

All things considered, our results support the hypothesis that changes in genome size between *Schmidtea* and *Dugesia* were likely driven by an expansion of TEs in the latter rather than a WGD event.

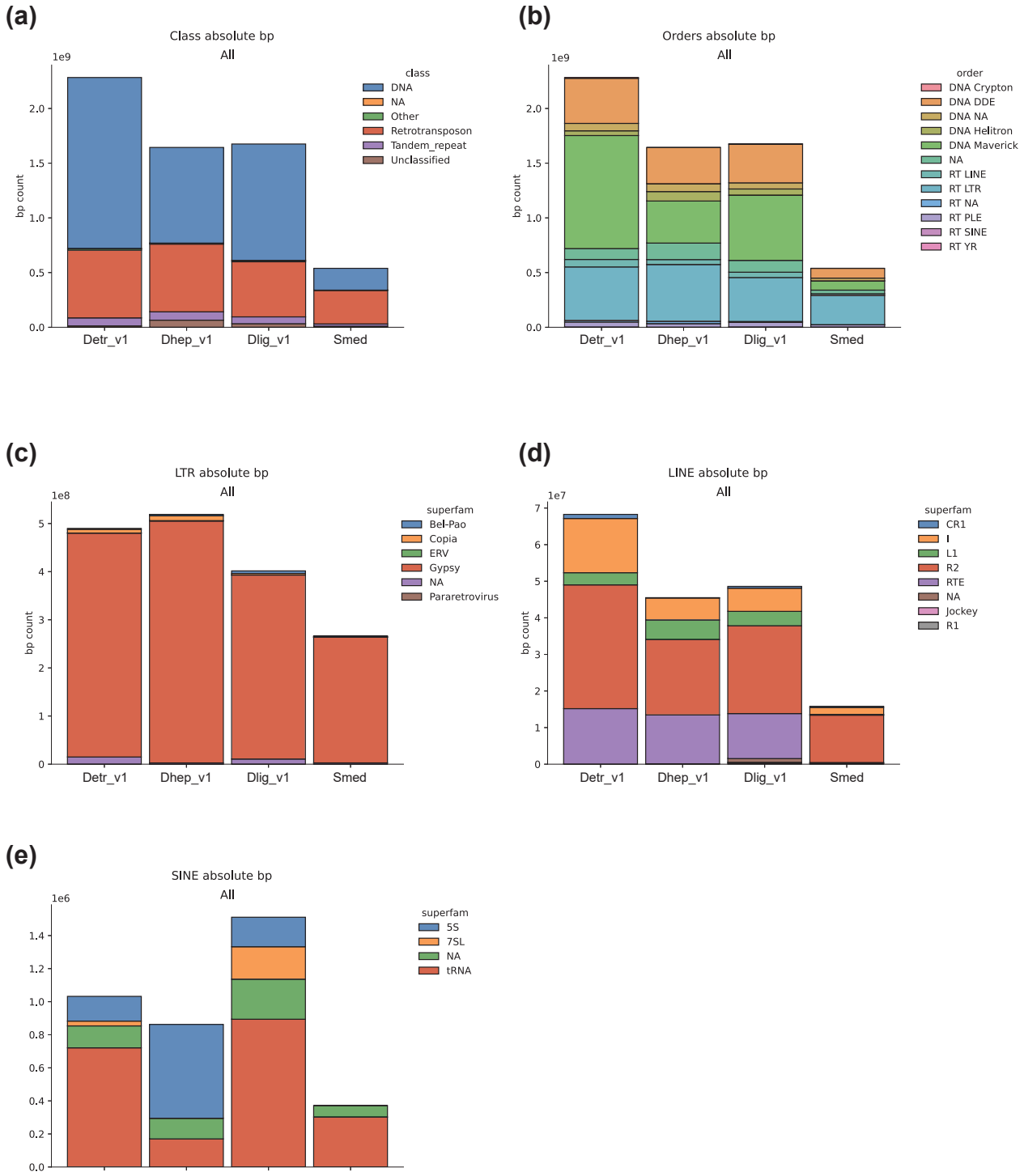


Figure 6. Stacked histograms showing the total amount (bp) of repeat content of each genome assembly classified into (a) classes of repetitive elements and (b) orders within said classes. (c) Close up look at the proportion of LTR retrotransposon elements. (d) Close up looks at the proportion of LINE retrotransposon elements and (e) SINE retrotransposon elements.

3.2. Macrosynteny and the chromosomal rearrangements

We assessed the synteny between the four genomes in an attempt to elucidate the origin of the peculiar haploid complement of *D. hepta* (n = 7) and to establish a first mac-

rosynteny comparison between *Dugesia* and *Schmidtea*. Furthermore, it has been shown that structural genome rearrangements are frequent and surprisingly common in *Schmidtea*, including inter-chromosomal translocations (Ivankovic, preprint 2023), which sparked in us an interest to check if such rearrangements were also common in *Dugesia*.

Our results indicate that macrosynteny is well preserved between *D. etrusca* and *D. liguriensis*, showing no apparent signs of major translocations or inversions (Fig. 7) even though both species shared they last common ancestor approximately 9.5 Mya (Solà et al., 2022). In *D. hepta*, as expected, chromosomal rearrangements can be observed. Chromosome 2 shares a major syntenic block with *D. etrusca*'s and *D. liguriensis*' chromosome 7, and minor ones with their chromosomes 5 and 6 (Fig. 7a). Furthermore, *D. hepta*'s chromosome 6 is comprised of two syntenic blocks associated to *D. etrusca*'s and *D. liguriensis*' chromosomes 5 and 6.

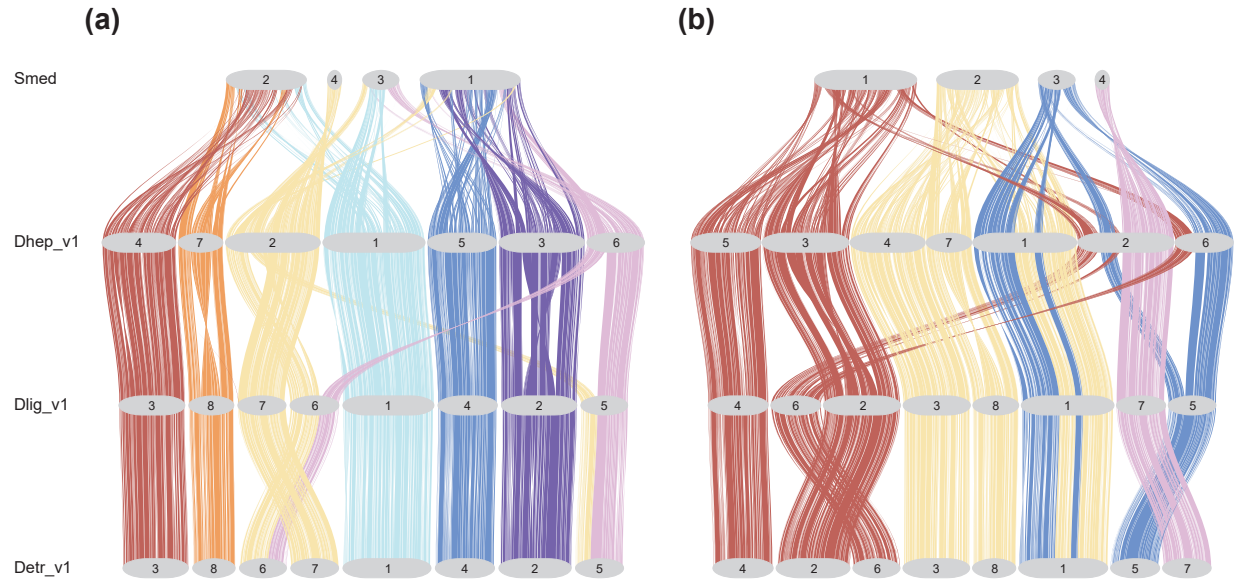


Figure 7. GENESPACE synteny order results comparing four genome assemblies. Ribbons are coloured using (a) *D. hepta*'s chromosomes as reference and (b) *S. mediterranea*.

Overall, two major fusion events can be observed in the chromosome complement of *D. hepta*. Considering the generally conserved synteny between *D. hepta* and *D. etrusca*, as well as *D. liguriensis*, across the rest of chromosomes, and the rarity of *D. hepta*'s karyotype among *Dugesia* species of the Western Mediterranean, it is reasonable to assume that the last common ancestor with *D. hepta* (~30 Mya; Solà et al., 2022) likely had a haploid complement of n = 8. Similarly, the last common ancestor of the Corso-Sardinian clade (~5.7 Mya) might have had the same complement and a similar karyotype to that of *D. etrusca* and *D. liguriensis*. The present *D. hepta* karyotype might be the result of a fission of the ancestral chromosomes 5 and 6. Fragments of the ancestral chromosomes 5 and 6 would have fused with the ancestral chromosome 7, resulting in *D. hepta*'s current chromosome 2, and the remaining fragments of the fissioned ancestral chromosomes would have merged into the current chromosome 6. Furthermore, two pericentric inversions can

be observed in *D. hepta*'s chromosomes 3 and 7.

It is acknowledged that TEs play an important role in the structural rearrangement and evolution of genomes (reviwed in Chénais et al., 2012). *Dugesia hepta* showed, comparatively, a striking expansion of SINE/*5S* elements. Although SINE retrotransposons seem to represent a minimal fraction of the assembly (< 1.4 Mb) it would be interesting to assess whether or not this expansion was recent or if there is any particular enrichment of these elements near the synteny breakpoints of *D. hepta*'s chromosomes 2 and 6 that could suggest their involvement in the observed chromosomal rearrangements.

Contrastingly, macrosynteny is highly altered between *Schmidtea* and *Dugesia* (Fig. 7b). It is estimated that both genera might have diverged more than 155 Mya (Solà et al., 2022), yet surprisingly our preliminary results suggest that *S. mediterranea*'s chromosome 4 is seemingly preserved in a single syntenic block in *Dugesia* (chromosome 2 in Dhep_v1; chromosome 7 in Detr_v1 & Dlig_v1) despite being a chromosome with a marked density of transposons.

4. Conclusions

The chromosome-scale assemblies of *D. etrusca*, *D. hepta* and *D. liguriensis* constitute the first high-quality and contiguous genomes of their kind for the genus *Dugesia* of freshwater planarians. These new assemblies will contribute to advance towards a more comprehensive genomic representation of free-living planarians in science and alleviate their meagre representation in the Tree of Life. Furthermore, our results have shown that macrosynteny is well-preserved in *Dugesia* while refuting the hypothesis that a WGD could be the underlying cause of the substantial difference in genome size between *Dugesia* and *S. mediterranea*. Moreover, our results suggest that this difference is likely due to an expansion of TEs, most notably of DNA/*Maverick* elements, the largest type reported in eukaryotes.

Concurrently, our results offered a first view on the chromosomal rearrangements that shaped *D. hepta*'s particular karyotype and highlighted an expansion of SINE retrotransposons in this species, suggesting their putative involvement in the rearrangement events. However, further research will be necessary to elucidate the role of this type of TEs in the conformation of *D. hepta*'s chromosome portrait.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein data base search programs. *Nucleic Acids Research*, 25(17), 3389-3402. https://doi.org/10.1093/nar/25.17.3389

Ball, I. R. (1974). A contribution to the phylogeny and biogeography of freshwater triclads (Platyhelminthes: Turbellaria). In: Riser NW, Morse MP (eds), *Biology of the Turbellaria*. New York: McGraw–Hill; 339–401

Benazzi, M. (1946). Sopra una nuova planaria d'acqua dolce. *Archivio Zoologico Italiano*, 31, 93-102.

Benazzi, M. (1982). Speciation events as evidenced in Turbellaria. In C. Barigozzi (Ed.), *Mechanisms of speciation*. New York, NY: Alan R. Liss; 307-344.

Benazzi, M., Baguñà, J., Ballester, R., Puccinelli, I., & Del Papa, R. (1975). Further contributions to the taxonomy of the "*Dugesia lugubris-polychroa* group" with description of *Dugesia mediterranea* n. sp. (Tricladida, Paludicola). *Bolletino di Zoologia*, 42(1), 81-89. https://doi.org/10.1080/11250007509430132

Benazzii-Lentati, G. (1979). Polyploidy and types of development in the triclad *Dugesia lugubris* s.l.: interaction between the ploidy level and the origin of amphimixis or pseudogamy. *Caryologia*, 32(2), 247-263. https://doi.org/10.1080/00087114.1979.10796791

Benítez-Álvarez, L., Leria, L., Fernández, R., Mateos, E., El Ouanighi, Y., Bennas, N., El Alami, M., Yacoubi-Khebiza, M., Ougougdal. H. A., & Riutort, M. (2023). Phylotranscriptomics interrogation uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean. *Molecular Phylogenetics and Evolution*, 178:107649. https://doi.org/10.1016/j.ympev.2022.107649

Beukeboom, L.W., Sharbel, T.F., & Michiels N.K. (1998). Reproductive modes, ploidy distribution, and supernumerary chromosome frequencies of the flatworm *Polycelis nigra* (Platyhelminthes: Tricladida). *Hydrobiologia*, 383(3), 277-285.

Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark- EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1), 1– 11. https://doi.org/10.1093/nargab/lqaa108

Buchfink, B., Reuter, K., & Drost, H-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, 18(4), 366-368. http://dx.doi.org/10.1038/s41592-021-01101-x

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadodoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421-430. https://doi.org/10.1186/1471-2105-10-421

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepa, 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

Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit – Inter-

active quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics*, 10(4), 1361-1374. https://doi.org/10.1534/g3.119.400908

Chan, P. P., Lin, B. Y., Mak, A. J., & Lowe, T. M. (2021). tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research*, 49(16), 9077-9096. https://doi.org/10.1093/nar/gkab688

Chénais, B., Caruso, A., Hiard, S., & Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. *Gene*, 509(1), 7-15. http://dx.doi.org/10.1016/j.gene.2012.07.042

Chinone, A., Nodono, H., & Matsumoto M. (2014). Triploid planarian reproduces truly bisexually with euploid gametes produced through a different meiotic system between sex. *Chromosoma*, 123(3), 265-272. https://doi.org/10.1007/s00412-013-0449-2

Dalyell, J. G. (1814). Observations on some interesting phenomena in animal physiology, exhibited by several species of planariae: Illustrated by coloured figures of living animals. Edinburgh: Archibald Constable & Co.

Derelle, R., Philippe, H., & Colbourne, J. K. (2020). Broccoli: Combining phylogenetic and network analyses for orthology assignment. *Molecular Biology and Evolution*, 37(11), 3389-3396. https://doi.org/10.1093/molbev/msaa159

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21. https://doi.org/10.1093/bioinformatics/bts635

Dols-Serrate, D., Guo, L., Kruglyak, L., & Riutort, M. (2025). Knot expected: Phylogenomics of *Dugesia* (Platyhelminthes, Tricladida) shed light on the evolutionary history and diversification of the Corso-Sardinian clade and uncover the first case of interspecific hybridization in freshwater planarians. In process.

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20:238. https://doi.org/10.1186/s13059-019-1832-y

Escuer, P., Pisarenco, V. A., Fernández-Rui, A. A., Vizueta, J., Sánchez-Herrero, J. F., Arnedo, M. A., Sánchez-Gracia, A., & Rozas, J. (2022). The chromosome-scale assembly of the Canary Islands endemic spider *Dysdera silvatica* (Arachnida, Araneae) sheds light on the origin and genome structure of chemoreceptor gene families in chelicerates. *Molecular Ecology Resources*, 22(1), 375-390. https://doi.org/10.1111/1755-0998.13471

Giani, A. M., Gallo, G. R., Gianfranceschi L., & Formenti G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, 18, 9-19. https://doi.org/10.1016/j.csbj.2019.11.002

Girard, C. A. (1850). A brief account of the fresh water planariae of the United States. *Proceedings of the Boston Society of Natural History*, 3, 264-265.

González-Estévez, C. (2008). Autophagy in freshwater planarians. *Methods in Enzymology*,

451(8), 439-465. http://dx.doi.org/10.1016/S0076-6879(08)03227-8

González-Estévez, C., & Saló, E. (2010). Autophagy and apoptosis in planarians. *Apoptosis*, 15(3), 279-292. https://doi.org/10.1007/s10495-009-0445-4

Grohme, M. A., Schloissnig, S., Rozanski, A., Pippel, M., Young, G. R., Winkler, S., Brandl, H., Henry, I., Dahl, A., Powell, S., Hiller, M., Myers, E., & Rink, J. C. (2018). The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature*, 554(7690), 56-61. https://doi.org/10.1038/nature25473

Guo, L., Bloom, J. S., Dols-Serrate, D., Boocock, J., Ben-David, E., Shubert, O. T., Kozuma, K., Ho, K., Warda, E., Chui, C., Wei, Y., Leighton, D., Vergara, T. L., Riutort, M., Alvarado, A. S., & Kruglyak, L. (2022). Island-specific evolution of a sex-primed autosome in a sexual planarian. *Nature*, 606(7913), 329-334. https://doi.org/10.1038/s41586-022-04757-3

Hengchi, C., & Zwaenepoel, A. (2023). Inference of ancient polyploidy from genomic data. In: Van de Peer, Y. (eds) *Polyploidy. Methods in Molecular Biology*, vol 2545. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-2561-3_1

Hotaling, S., Kelley, J. L., & Frandsen, P. B. (2020). Aquatic insects are dramatically underrepresented in genomic research. *Insects*, 11(9), 601-608. https://doi.org/10.3390/insects11090601

Hotaling, S., Kelley, J. L., & Frandsen, P. B. (2021). Toward a genome sequence for every animal: where are we now? *Proceedings of the National Academy of Sciences of the United States of America*, 118(52), e2109019118. https://doi.org/10.1073/pnas.2109019118

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-genome annotation with BRAKER. *Methods in Molecular Biology*, 1962, 65– 95. https://doi.org/10.1007/978-1-4939-9173-0_5

Ichikawa, A., & Kawakatsu, M. (1964). A new freshwater planarian, *Dugesia japonica*, commonly but erroneously known as *Dugesia gonocephala* (Dugès). *Annotationes Zoologicae Japoneses*, 37, 185-194.

Ivankovic, M., Brand, J. N., Pandolfini, L., Brown, T., Pippel, M., Rozanski, A., Schubert, T., Grohme, M. A., Winkler, S., Robledillo, L., Zhang, M., Codino, A., Gustincich, S., Vila-Farré, M., Zhang, S., Papatonis, A., Marques, A., & Rink, J. C. (2023). A comparative analysis of planarian genomes reveals regulatory conservation in the face of rapid structural divergence. *BioRxiv*, preprint. https://doi.org/10.1101/2023.12.22.572568

Ivankovic, M., Haneckova, R., Thommen, A., Grohme, M. A., Vila-Farré, M., Werner, S., & Rink, J. C. (2019). Model systems for regeneration: planarians. *Development*, 146(17), dev167684. https://doi.org/10.1242/dev.167684

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

Kawakami, T., Butlin, R. K., & Cooper, S. J. B. (2011). Chromosomal speciation revisited: Modes of Diversification in Australian morabine grasshoppers (*Vandiemenella, viatica* species group). *Insects*, 2, 49-61. https://doi.org/10.3390/insects2010049

Kawakatsu, M., Oki, I., Tamura, S., & Sugino, H. (1976). Studies on the morphology, karyology and taxonomy of the Japanese freshwater planarian *Dugesia japonica* Ichikawa and Kawakatsu, with a description of a new subspecies, *Dugesia japonica ryukyensis* subspec. nov. *Bulletin of Fuji Women's College*, no. 14, ser. II**:**81–126.

Kim, J., Larkin, D. M., Cai, Q., Asan, Zhang, Y., Ge, R-L., Auvil, L., Capitanu, B., Zhang, G., Lewin, H. A., & Ma, J. (2013). Reference-assisted chromosome assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 110(5), 1785-1790. https://doi.org/10.1073/pnas.1220349110

King, M. (1987). Chromosomal rearrangements, speciation and the theoretical approach. *Heredity*, 59, 1-6. https://doi.org/10.1038/hdy.1987.90

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: interrogation of genomes assemblies. *F1000Research*, 6:1287. https://doi.org/10.12688/f1000research.12232.1

Lázaro, E. M., Sluys, R., Pala, M., Stocchino, G. A., Baguñà, J., & Riutort, M. (2009). Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution*, 52(3), 835-845. https://doi.org/10.1016/j.ympev.2009.04.022

Leria, L., Riutort, M., Romero, R., Ferrer, X., & Vila-Farré, M. (2022). Microplate tectonics and environmental factors as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography*, 49(6), 1124-1136. https://doi.org/10.1111/jbi.14373

Leria, L., Sluys, R., & Riutort, M. (2018). Diversification and biogeographic history of the Western Palearctic freshwater flatworm genus *Schmidtea* (Tricladida: Dugesiidae), with a redescription of *Schmidtea nova*. *Journal of Zoological Systematics and Evolutionary Research*, 56(3), 335-351. https://doi.org/10.1111/jzs.12214

Leria, L., Vila-Farré, M., Solà, E., & Riutort, M. (2019). Oustanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. *BMC Evolutionary Biology*, 19(1), 130-149. https://doi.org/10.1186/s12862-019-1440-1

Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M., & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, 11:e78526. https://doi.org/10.7554/eLife.78526

Lucek, K., Giménez, M. D., Joron, M., Rafajlović, M., Searle, J. B., Walden, N., Westram, A. M., & Faria, R. (2023). The impact of chromosomal rearrangements in speciation: From micro- to macroevolution. *Cold Spring Harbor Perspectives in Biology*, 15(11), a041447. https://doi.org/10.1101/cshperspect.a041447

Lynch, M. (2007).

Lynch, M., & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, 290(5494), 1151-1155. https://doi.org/10.1126/science.290.5494.1151

Mackintosh, A., Vila, R., Laetsch, D. R., Hayward, A., Martin, S. H., & Lohse, K. (2023). Chromosome fissions and fusions act as barriers to gene flow between *Brenthis* fritillary butterflies. *Molecular Biology and Evolution*, 40(3): msad043 https://doi.org/10.1093/molbev/msad043

MacLeod, K. L., Paris, J. R., Jenkins, T. L., & Stevens, J. R. (2023). The first genome of the cold-water octocoral, the pink sea fan, *Eunicella verrucosa*. *Genome Biology and Evolution*, 15(6), 1-7. https://doi.org/10.1093/gbe/evad083

Mérot, C. (2020). Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Molecular Ecology*, 29(14), 2513-2516. https://doi.org/10.1111/mec.15500

Pala, M., Casu, S., & Vacca, R. A. (1981). *Dugesia hepta*, nuova specie di planaria di acqua dolce di Sardegna appartenente alla superspecie *Dugesia gonocephala* (Dugès). *Bolletino della Società Sarda di Scienze Naturali*, 20, 97-107. https://core.ac.uk/reader/11689490

Pala, M., Stocchino, G. A., Corso, G., & Casu S. (2000). A new species of freshwater planarian belonging to the genus *Dugesia* (Platyhelminthes, Tricladida) from Sardinina. *Italian Journal of Zoology*, 67, 379-383. https://doi.org/10.1080/11250000009356343

Pongratz, N., Sharbel, T. F., Beukeboom, L.W., & Michiels, N. K. (1998). Allozyme variability in sexual and parthenogenetic freshwater planarians: evidence for polyphyletic origin of parthenogenetic lineages through hybridization with coexisting sexuals. *Heredity*, 81, 38-47. https://doi.org/10.1046/j.1365-2540.1998.00359.x

Póti, Á., Szütz, D., & Vermezovic, J. (2024). Mutational profile of the regenerative process and *de novo* genome assembly of the planarian *Schmidtea polychroa*. *Nucleic Acids Research*,

Reddien, P. W. (2018). The cellular and molecular basis for planarian regeneration. *Cell*, 175(2), 327-345. https://doi.org/10.1016/j.cell.2018.09.021

Rowley, M. J., & Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nature Reviews Genetics*, 19(2), 789-800. https://doi.org/10.1038/s41576-018-0060-8

Robb, S. M. C., Gotting, K., Ross, E., & Alvarado, A. S. (2015). SmedGC 2.0: The *Schmidtea mediterranea* genome database. *Genesis*, 53(8), 535-546. https://doi.org/10.1002/dvg.22872

Robb, S. M. C., Ross, E., & Alvarado, A. S. (2007). SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Research*, 36, 599-605. https://doi.org/10.1093/nar/gkm684

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., & Malonia, S. K. (2023). Next-generation sequencing Technology: Current trends and advancements. *Biology*, 12(7), 997 https://doi.org/10.3390/biology12070997

Schmidt, O. (1861). Die Dendrocoelen Strudelwürmer aus den Umgebungen von Graz.

Zeitschrift für wissenschaftliche Zoologie, 11, 89-94.

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. In Kollmar, M. (2019). *BUSCO: Assessing genome assembly and annotation completeness BT - Gene prediction: Methods and protocols* (Vol. 1962, pp. 227– 245). https://doi.org/10.1007/978-1-4939-9173-0_14

Simakov, O., Bredeson, J., Berkoff, K., Marletaz, F., Mitros, T., Schults, D. T., O'Connell, B. L., Dear, P., Martínez, D. E., Steele, R. E., Green, R. E., David, C. N., & Rokhsar, D. S. (2022). Deeply conserved synteny and the evolution of metazoan chromosomes. *Science Advances*, 8(5), eabi5884. https://doi.org/10.1126/sciadv.abi5884

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of the next generation sequencing technologies. *Current Protocols in Molecular Biology*, 122(1), e59. https://doi.org/10.1002/cpmb.59

Sluys, R., Kawakatsu, M., & Winsor, L. (1998). The genus *Dugesia* in Australia, with its phylogenetic analysis and historical biogeography (Platyhelminthes, Tricladida, Dugesiidae). *Zoologica Scripta*, 27(4), 273-289. https://doi.org/10.1111/j.1463-6409.1998.tb00461.x

Sluys, R., & Riutort, M. (2018). Planarian diversity and phylogeny. In: Rink, J. C. (ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology*, Methods in Molecular Biology, Vol. 1774. Springer Science+Business Media, New York, 1-56. https://doi.org/10.1007/978-1-4939-7802-1_1

Smit, A. F., & Hubley, R. (2008- 2015). *RepeatModeler-1.0*. Retrieved from http://www.repeatmasker.org

Smit, A. F., Hubley, R., & Green, P. (2013- 2015). *RepeatMasker-4.0*. http://www.repeatmasker.org

Solà, E., Leria, L., Stocchino, G. A., Bagherzadeh, R., Balke, M., Daniels, S. R., Harrath, A. H., Khang, T. S., Krailas, D., Kumar, B., Li, M-H., Maghsoudlou, A., Matsumoto, M., Naser, N., Ober, B., Segev, O., Thielicke, M., Tong, X., Zivanovic, G., Manconi, R., Baguñà, J., & Riutort, M. (2022). Three dispersal routes out of Africa: a puzzling biogeographical history in freshwater planarians. *Journal of Biogeography*, 49(7), 1219-1233. https://doi.org/10.1111/jbi.14371

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, 24(5), 637–644. https://doi.org/10.1093/bioinformatics/btn013

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, 7, 1–11. https://doi.org/10.1186/1471-2105-7-62

Stocchino, G. A., Corso, G., Manconi, R., Casu, S., & Pala, M. (2005). Endemic freshwater planarians of Sardinia: redescription of *Dugesia hepta* (Platyhelminthes, Tricladida) with a comparison of the Mediterranean species of the genus. *Journal of Natural Histology*, 39(22), 1947-

1960. https://doi.org/10.1080/00222930500060025

Stocchino, G. A., & Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology*, 803(3), 319-328. ttp://dx.doi.org/10.1080/11250003.2013.822025

Tian, Q., Guo, Q., Guo, Y., Luo, L., Kristiansen, K., Han, Z., Fang, H., & Zhang, S. (2022). Whole-genome sequence of the planarian *Dugesia japonica* combining Illumina and PacBio data. *Genomics*, 114(2):110293. https://doi.org/10.1016/j.ygeno.2022.110293

Tiley, G. P., Barker, M. S., & Burleigh, J. G. (2018). *Genome Biology and Evolution*, 10(11), 2882-2898. https://doi.org/10.1093/gbe/evy200

Vacca, R. A., Casu, S., & Pala, M. (1993). Popolamento planariologico dei fiumi del Nord Sardegna. II. I cariotipi dei tricladi d'acqua dolce rinvenuti nel bacino idrografico del fiume Coghinas. *Bolletino della Società Sarda di Scienze Naturali*, 29, 59-73.

Vanneste, K., Baele, G., Maere, S., & Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of succesful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Research*, 24(8), 1334-1347. https://doi.org/10.1101/gr.168997.113

Vila-Farré, M., & Rink, J. C. (2018). The ecology of freshwater planarians. In: Rink, J. C. (ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology*, Methods in Molecular Biology, Vol. 1774. Springer Science+Business Media, New York, 173-205. https://doi.org/10.1007/978-1-4939-7802-1_1

de Vries, E. J. (1988). Further contributions to the taxonomy and biogeography of the subgenus *Dugesia* (Platyhelminthes: Tricladida: Paludicola) in the Mediterranean region and Middle East. *Israel Journal of Zoology*, 35(3), 109-136. https://doi.org/10.1080/00212210.1988.10688607

Wagner, D. E., Wang, I. E., & Reddien, P. W. (2011). Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science*, 332(6031), 811-816. https://doi.org/10.1126/science.1203983

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene syntenty and collinearity. *Nucleic Acids Research*, 40(7), e79. https://doi.org/10.1093/nar/gkr1293

Zhang, S., & Alvarado, A. S. (2018). Planarian high molecular weight DNA isolation by spooling. In: Rink, J. C. (ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology*, Methods in Molecular Biology, Vol. 1774. Springer Science+Business Media, New York, 277-284. https://doi.org/10.1007/978-1-4939-7802-1_1

Župunski, V., Gubenšek, F., & Kordiš, D. (2001). Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons. *Molecular Biology and Evolution*, 18(10), 1849-1863. https://doi.org/10.1093/oxfordjournals.molbev.a003727

Supplementary Material: Figures & Tables

The first chromosome-scale assemblies of *Dugesia* shed light on the genome size evolution in flatworms

Daniel Dols-Serrate[1,2], Ignacio Tenaguillo[1,2], Vadim A. Pisarenco[1,2], Julio Rozas[1,2] & Marta Riutort[1,2]

[1] Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

[2] Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028, Barcelona, Catalonia, Spain

ORCIDs:

0000-0003-0820-7324 Daniel Dols-Serrate

0000-0001-8011-5286 Ignacio Tenaguillo

0000-0002-8085-7674 Marta Olivé-Muñiz

0000-0002-4968-4090 Vadim A. Pisarenco

0000-0002-6839-9148 Julio Rozas

0000-0002-2134-7674 Marta Riutort

Corresponding author: D. Dols-Serrate danieldols@ub.edu/bcn27101993@gmail.com

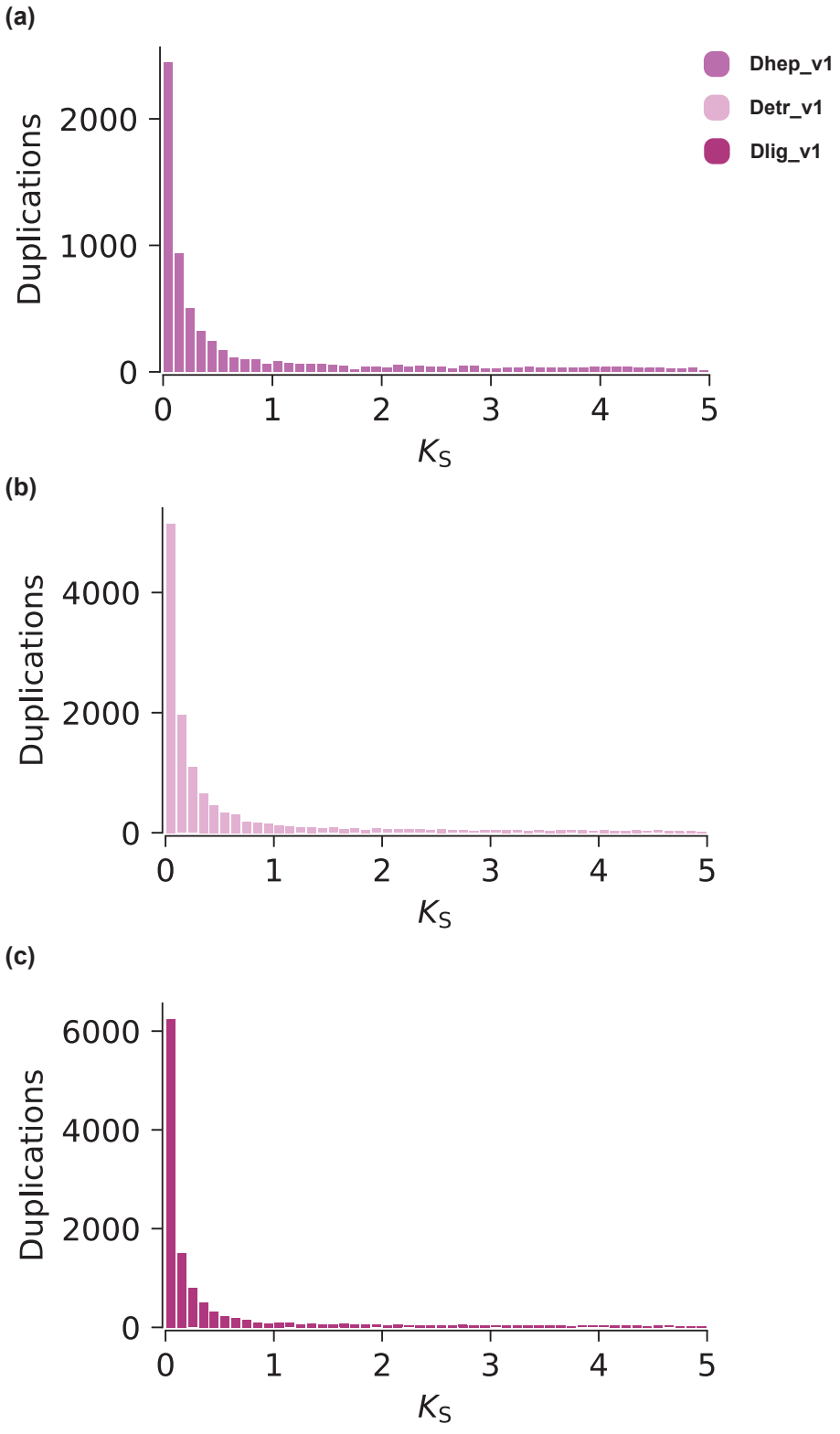


Figure S1. Separate paranome Ks distributions of the three genome assemblies showing a pattern of exponential decay. (a) Dhep_v1, (b) Detr_v1 and (c) Dlig_v1.

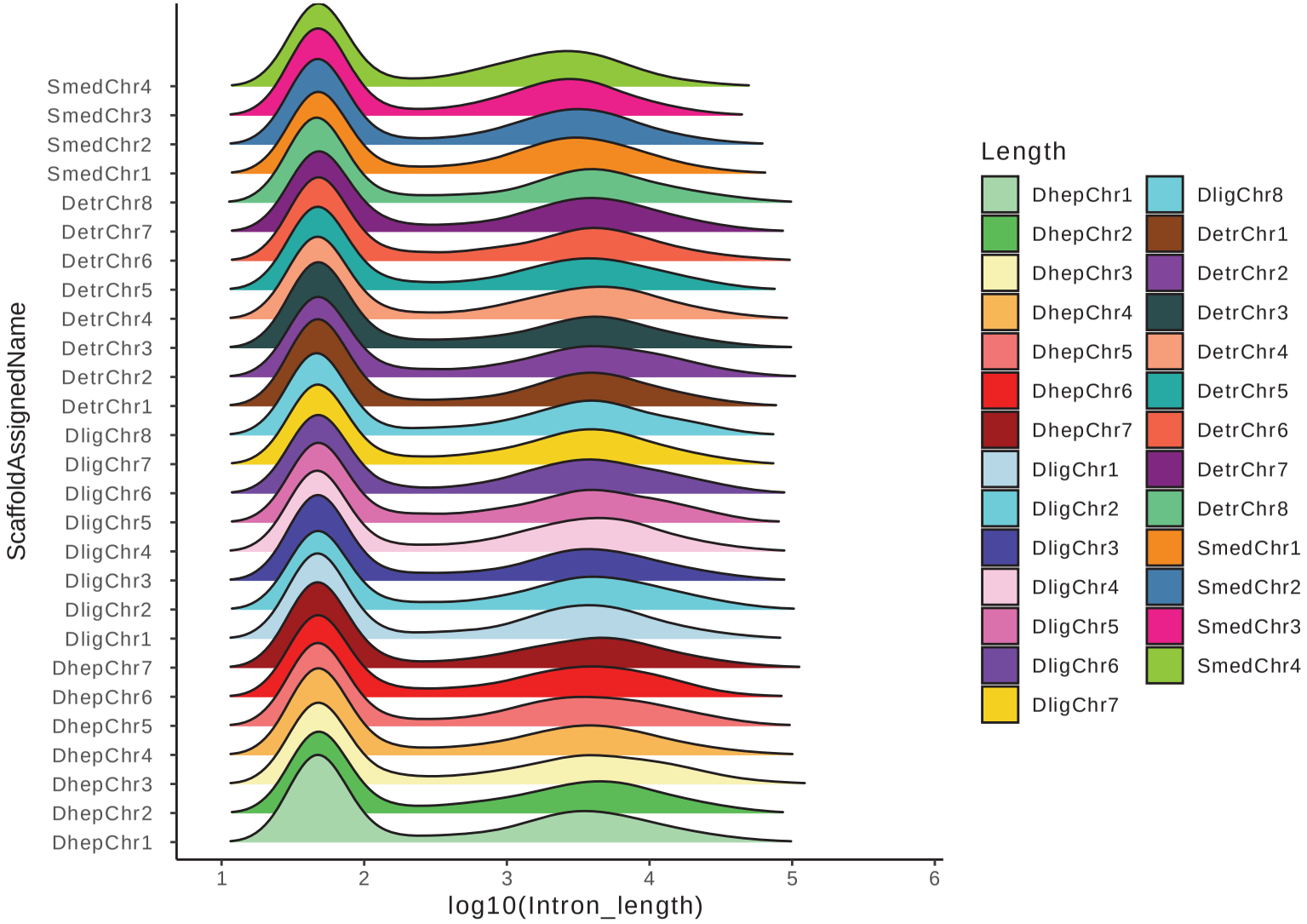


Figure S2. Ridgeline plots showing the density distribution of intron lengths of all introns comprised in the 3065 single-copy orthogroups shared between *S. mediterranea* (Smed), *D. hepta* (Dhep), *D. etrusca* (Detr) and *D. liguriensis* (Dlig) separated by chromosomes.

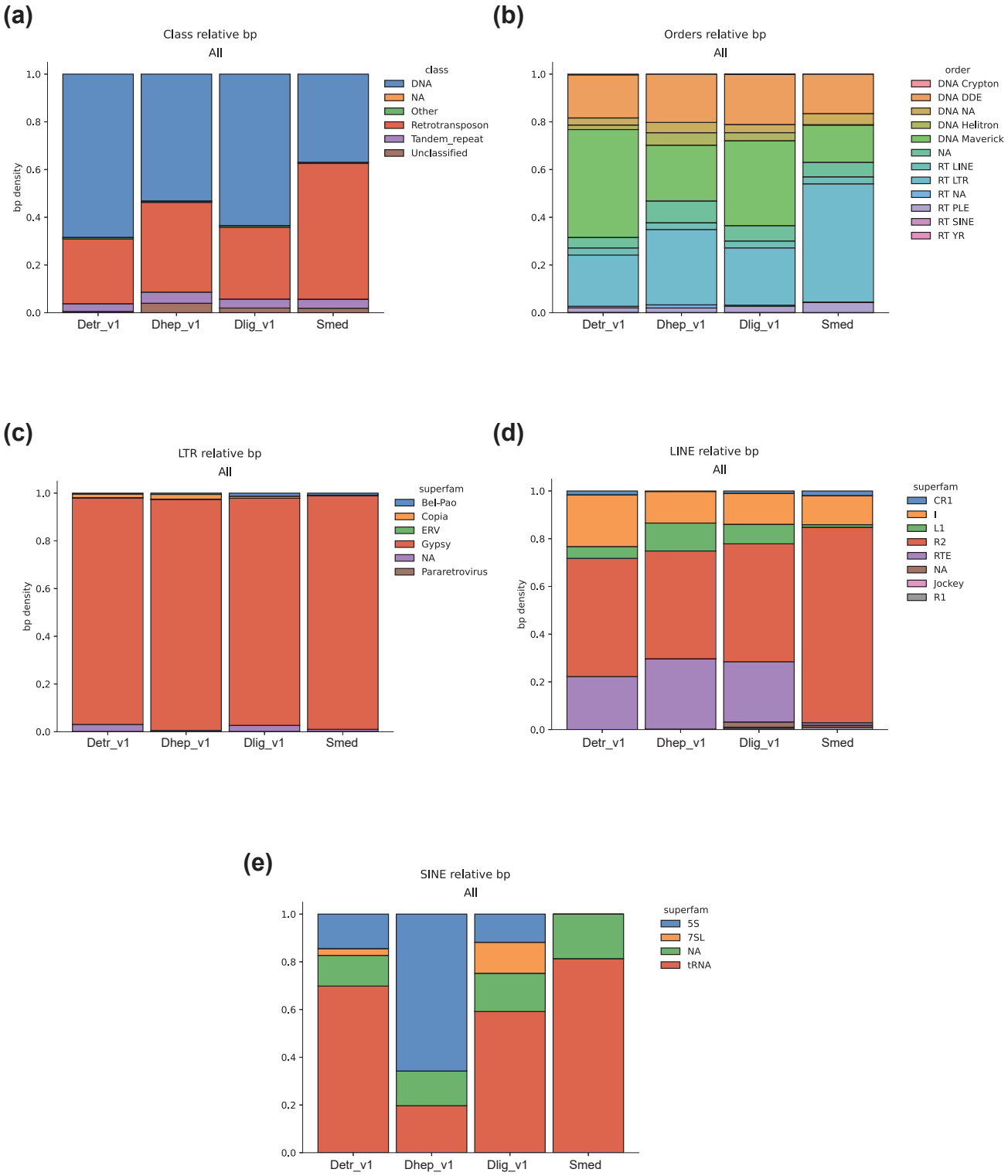


Figure S3. Comparison of the relative contributions (%) of each class (a) and order (b) of repetitive elements to the repeat content of each genome assembly. (c) Relative superfamilies contributions of repetitive elements to the LTR retrotransposon elements. Relative superfamilies contributions of repetitive elements to the (d) LINE retrotransposon and (e) SINE retrotransposon elements.

Supplementary Table S1. BUSCO analysis results.

	metazoa_odb10 (954 genes)			arthropoda_odb10 (1013 genes)			eukaryota_odb10 (255 genes)		
	Detr_v1	**Dhep_v1**	**Dlig_v1**	**Detr_v1**	**Dhep_v1**	**Dlig_v1**	**Detr_v1**	**Dhep_v1**	**Dlig_v1**
Before decontamination									
Single	463 (48.5%)	540 (56.6%)	453 (47.5%)	474 (59.5%)	513 (50.6%)	432 (42.6%)	142 (55.7%)	150 (58.8%)	135 (52.9%)
Duplicated	161 (16.9%)	76 (8.0%)	206 (21.6%)	129 (12.7%)	78 (7.7%)	204 (20.1%)	48 (18.8%)	30 (11.8%)	64 (25.1%)
Fragmented	106 (11.1%)	94 (9.8%)	86 (9.0%)	123 (12.1%)	121 (11.9%)	106 (10.5%)	31 (12.2%)	38 (14.9%)	24 (9.4%)
Missing	224 (23.5%)	244 (25.6%)	209 (21.9%)	287 (28.4%)	301 (29.7%)	271 (26.8%)	34 (13.3%)	37 (14.5%)	32 (12.6%)
After decontamination									
Single	464 (48.6%)	540 (56.6%)	460 (48.2%)	473 (46.7%)	513 (50.6%)	433 (42.7%)	143 (56.1%)	150 (58.8%)	138 (54.1%)
Duplicated	158 (16.6%)	76 (8.0%)	199 (20.9%)	126 (12.4%)	78 (7.7%)	203 (20.0%)	46 (18.0%)	30 (11.8%)	61 (23.9%)
Fragmented	106 (11.1%	94 (9.8%)	86 (9.0%)	124 (12.2%)	121 (11.9%)	106 (10.5%)	32 (12.5%)	38 (14.9%)	24 (9.4%)
Missing	226 (23.7%)	244 (25.6%)	209 (21.9%)	290 (28.7%)	301 (29.7%)	271 (26.8%)	34 (13.4%)	37 (14.5%)	32 (12.6%)
Proteome									
Single	562 (58.9%)	573 (60.6%)	495 (51.9%)	575 (56.8%)	565 (55.8%)	508 (50.1%)	171 (67.1%)	166 (65.1%)	151 (59.2%)
Duplicated	137 (14.4%)	116 (12.1%)	44 (4.6%)	143 (14.1%)	136 (13.4%)	53 (5.2%)	40 (15.7%)	39 (15.3%)	19 (7.5%)
Fragmented	72 (7.5%)	62 (6.5%)	32 (3.4%)	61 (6.0%)	64 (6.3%)	30 (3.0%)	25 (9.8%)	20 (7.8%)	10 (3.9%)
Missing	183 (19.2%)	203 (21.3%)	383 (40.1%)	234 (23.1%)	248 (24.5%)	422 (41.7%)	19 (7.4%)	30 (11.8%)	75 (29.4%)

General discussion

In this thesis, I aimed to unravel some of the processes that have contributed to the diversification of Tricladida, with a particular focus on a fairly small taxon known as the Corso-Sardinian clade. To this end, I undertook dedicated sampling efforts across Corsica and Sardinia to study the endemic populations of *Dugesia* on the islands. This led me to perform phylogenetic, population, karyological, and, with the invaluable help of my colleagues, morphological analyses that have proved instrumental to identify new members of the Corso-Sardinian clade while uncovering hidden patterns of diversity and reconstruct the biogeographical history of this taxonomic group.

1. On the use of molecular markers to identify species of the Corso-Sardinian clade

It is undeniable that the use of molecular markers has sprung up our understanding of evolutionary relationships of flatworms ever since the first molecular phylogeny of Platyhelminthes came through (Carranza et al., 1997). Since the advent of the first-generation DNA sequencing techniques, a plethora of molecular markers have been developed. Many differ in their informativeness, which can be used to facilitate taxonomic identification across different levels. The phylogenetic information preserved within their sequences may function as individualized *barcodes*, enabling the precise assignment of a given sample into its pertinent species. Among these markers, the mitochondrial gene cytochrome *c* oxidase subunit I (*COI*) stands out as an excellent barcode for metazoan species. Its rapid evolutionary rate empowers precise species-level assignment and even demarcation of phylogeographic clusters within a species. So much so that international DNA barcoding projects such as the Barcode of Life Data System (BOLD; Ratnasingham and Hebert, 2007) revolve around this marker. Even though *COI* is one of the most used barcodes (Valentini et al., 2008), freshwater planarians of the genus *Dugesia* have other relevant

markers at their disposal such as *18S* and *28S* rRNAs for higher-level taxonomy analyses, or the internal transcribed spacer-1 (*ITS-1*) for shallow taxonomic levels.

More recently, several novel nuclear markers developed by Leria and collaborators (2019), known as *Dunuc* (shortening of "*Dugesia* nuclear"), have successfully been used to help reconstruct the biogeographical history of *Dugesia* (Leria et al., 2022; Solà et al., 2022) as well as to delimit cryptic species while resolving evolutionary relationships within congenerics of the Western Mediterranean (Leria et al., 2020). Thus, displaying a wide range of taxonomic informativeness.

Throughout this thesis, several of the above-mentioned markers have been routinely used to either identify samples, describe new species, or infer the intra- and interspecific phylogenetic relationships of the Corso-Sardinian species. However, after this thesis, it has become apparent that not all markers are equally fit for performing phylogenetic analyses on this taxonomic group. Thus, I will discuss the potential uses and limitations of the loci used in a phylogenetic context (Chapters I and II).

ITS-1 is a commonly used marker that has been featured in several works studying the interspecific relationships of *Dugesia* species of the Mediterranean, often in concatenation with *COI* or other molecular markers (Baguñà et al., 1999; Lázaro et al., 2009; Solà et al., 2013; Sluys et al., 2013; Leria et al., 2020). However, as far as the Corso-Sardinian clade is concerned, its resolving phylogenetic power is exceedingly limited for studying this taxonomic group. Overall, *ITS-1* is only valid for assuredly discriminating individuals of the Corso-Sardinian clade from the rest of Mediterranean *Dugesia* without needing to perform karyological or morphological analyses. In Chapter I, it was established that all populations of *D. benazzii* s.l. (sensu lato), whether from Corsica or Sardinia, shared the same *ITS-1* sequence, differing only in a few sites from that of *D. hepta*. Thereafter, in Chapter II.1, *D. benazzii* s.l. was divided into three species, viz. *D. benazzii* s.s. (sensu stricto), *D. hoidi* and *D. mariae*, all of which exhibited the same *ITS-1* sequence. Additionally, *ITS-1* was sporadically used in Chapter III to ensure that individuals of the allochthonous species *D. sicula* were not included in the analyses. We recovered the *ITS-1* sequence of several individuals of *D. xeropotamica*, which proved to be identical to that shared by *D. benazzii*, *D. hoidi*, and *D. mariae* (not shown).

It is noteworthy that *ITS-1* was instrumental to establish our first interspecific hybridization hypothesis between *D. benazzii* and *D. hepta* (Chapter I) because it misled our attention to a clade (*mixed* clade based on mitochondrial results; Chapter I, Fig. 1) that seemingly comprised specimens of both species. Evidently, in light of Chapter III results, we now know that that hypothesis was in fact erroneous, and that the patterns that we had observed in Chapter I were likely a byproduct of incomplete lineage sorting (ILS). Most importantly, Chapter III results consequently indicated the existence of a population of *D. benazzii* that has retained, at least as far as we know, two distinct *ITS-1* sequences that were previously considered diagnostic for distinguishing between *D. hepta* and *D. benazzii* s.l. Thus, meaning that *ITS-1* is no longer a reliable marker to unmistakably identify between *D. hepta* and other Corso-Sardinian taxa. Regarding *D. leporii*, it is unclear

whether *ITS-1* could be used to discriminate it as, supposedly, the only existing sequences of this species are those presented in Chapter III, and we did not search specifically for *ITS-1* among our RAD loci. Interestingly, in a molecular study by Lázaro and collaborators (2009), an individual seemingly collected in the type locality of *D. leporii* was included. Although it had been identified as *D. benazzii* in a previous study (Baguñà et al., 1999), its *ITS-1* seemed to differ a bit from that of other Sardinian *D. benazzii* featured in Lázaro and collaborators' study (2009).

Two nuclear markers were used in Chapters I and II, viz. *Dunuc10* (Chapter II) and *Dunuc12* (Chapters I and II). These markers by themselves have shown low phylogenetic resolving power and I would advise anyone interested in studying the Corso-Sardinian clade against exclusively relying on these two markers. While it is true that they possess more phylogenetic information than *ITS-1*, *Dunuc10* and *Dunuc12* essentially share the same practical role: to discern Corso-Sardinian species from other Mediterranean *Dugesia* species. Furthermore, *Dunuc12* suffers from the same uncertainties *ITS-1* does: if there are individuals that present both *benazzii*-like and *hepta*-like *Dunuc12* haplotypes (Chapter I, Fig. 5), but they are not hybrids (Chapter III, Fig. 6, Supplementary Fig. S12), then this marker cannot be used as a reliable barcode. However, these nuclear markers have proved useful when combined with mitochondrial data, as they generally aid in resolving internal nodes that were challenging to resolve solely with mitochondrial data and, in particular, they enabled the recovery of *D. hoidi* as a monophyletic species (Chapter II.2, Supplementary Fig. S1). Concurrently, *Dunuc10* and *Dunuc12* may be used as a more restrictive dataset in molecular species delimitation analyses to leverage the potential over-splitting of multispecies coalescent (MSC) approaches, like BPP (Sukumaran and Knowles, 2017), due to the prevalence of population structure in the mitochondrial data sets, as we did in Chapter II.1.

With respect to the mitochondrial markers, four genes were used in this thesis, viz. a fragment of *COI* (Chapters I, II and III), and the NADH dehydrogenase subunit 1, the tryptophan tRNA, and the cytochrome *c* oxidase subunit II, all three in a single sequence fragment (*NADH1*-tRNAW-*COII*) (Chapters II and III). Contrary to the *ITS-1* and nuclear markers, the mitochondrial markers are more reliable and have shown greater phylogenetic resolving power in general. In Chapter I, it was possible to challenge the notion of *D. benazzii* s.l. as a single species thanks to *COI*. In Chapter II, the mitochondrial markers (especially when combined with *Dunuc10* and *Dunuc12*) were instrumental to recognize and delineate *D. hoidi*, *D. mariae* and *D. xeropotamica*. However, Chapter III results indicate that the potential of *COI* and *NADH1*-tRNAW-*COII* as barcodes only goes to certain extends (cf. Chapter III, Supplementary Fig. S3). Although it is true that in Chapter III no definitive taxonomic decisions affecting the delineated molecular primary species candidates (PSC) were reached due to the lack of morphological corroboration (i.e., no new species were formally described), it is noteworthy that *D. leporii* could not be distinguished from *D. benazzii* using mitochondrial data (Chapter III, Supplementary Fig. S3). Furthermore, the elevated genetic structure within Corsica (6 PSCs) cannot be appreciated only with the mitochondrial genes employed so far. Overall, a reasonable explanation for the

problems encountered with the phylogenetic resolution of most markers discussed so far could be the rapid diversification that occurred within the Corso-Sardinian clade, recounted in Chapter III (Chapter III, Fig. 8, Supplementary Fig. S15). Many of the species (*D. benazzii*, *D. mariae*, *D. xeropotamica* and possibly *D. hoidi*) and PSCs emerged during or after the Mid-Pleistocene (~1.25 Mya). Moreover, our estimates suggest most PSCs might have started to diverge less than a million years ago (Chapter III, Supplementary Fig. S15). Which is quite recent evolutionarily speaking. Given the higher evolutionary rate of mitochondrial genes, it falls within reason to expect that phylogeographic patterns could be better detected with *COI* and *NADH1*-tRNAW-*COII* than *Dunuc10, Dunuc12* or *ITS-1*.

Considering all of the above-mentioned, if any PSC from Chapter III, particularly the ones from Corsica, were to be elevated to the species status in future studies, we could find ourselves in the following predicament: how can we confidently identify species of the Corso-Sardinian clade with just molecular data? Well, certainly not using any of the aforementioned markers by themselves. However, concatenation of multiple mitochondrial and nuclear markers followed by phylogenetic tree inference or the use of coalescent based methods as detailed in Chapter II can be of help. In Chapter II.2, this approach helped recover *D. hoidi* as monophyletic and could be useful to easily identify *D. leporii* as well, but it has not been tried since we did not have the opportunity to sequence their *Dunuc10* or *Dunuc12* markers. In addition, the development of novel molecular markers could be of aid as well, following a similar approach to that used in a previous study (Leria et al., 2019) but this time using *Dugesia* genome assemblies (e.g., Tian et al., 2022 or the genomes presented in Chapter IV) instead of using that of *S. mediterranea*. At present, though, the best way to molecularly identify any Corso-Sardinian species is to use genome-wide data, such as RADseq loci, which are particularly fitting for shallow systematics (Harvey et al., 2016) like the case at present.

With the advent of high-throughput sequencing and, hopefully, the development of novel genomic resources for triclad planarians (e.g., Benítez-Álvarez et al., 2023a; Ivanković et al., 2023; Tian et al., 2022; the genomes presented in Chapter IV), it might be possible to embrace a new era of population genomics for Tricladida in the future. Ensuring rapid and precise molecular identification of species should be a pivotal point of this new era, and the development of SNP genotyping resources, as have been developed in other taxonomic groups (e.g., Jackson et al., 2022; Silva-Junior et al., 2015; Wenne et al., 2016), could be of great use.

2. Hurdles of species delimitation in *Dugesia*

Gathering evidence for identifying or delimiting species of this genus is by no means a small feat. Species have been and are still being described on the basis of the inner morphology, being the copulatory apparatus the most notable source of diagnostic traits. Both external and inner morphology have hardly changed in the almost 130-160 million years of history of this genus (Solà et al., 2022). Most characters seem to be of non-adaptative nature, suggesting a stochastic accumulation of phenotypic differences probably caused by the isolation of populations in isolated freshwater bodies, due to either vicariance or dispersion (Solà, 2014). Their general morphological homogeneity has let researchers to consider *Dugesia* a morphological static taxon, which in turn does not leave much wiggle room for characterizing morphological diagnostic traits. Furthermore, the formal description of a species requires meticulous preparation and examination of morphological material, which demands significant expertise, patience, and a substantial time investment. However, all of these efforts can be easily thwarted if specimen fixation or sectioning procedures fail, which may happen for different unfortunate reasons, or if only asexual or immature specimens were collected during fieldwork, thus adding resilience to failure to the list of demands.

The increasing use of molecular data in freshwater flatworm systematics has certainly help in shedding light into some of the most contentious issues that morphology could

not easily address on its own. Molecular data has led to the proposal and description of new species, in revealing cryptic species or highlighting previously overlooked diversity in already established species (Benítez-Álvarez, 2023b; Leria et al., 2020; Solà, 2014; Solà et al., 2013; Wang et al., 2021 and Chapters I, II, and III of the present thesis). Likewise, molecular data has helped with the assignment of asexual forms to species (Lázaro et al., 2009; Leria et al., 2019). Yet, while the enhancement of phylogenetic resolution offers advantages, the potential risk of confusing population structure with species boundaries should be taken into account (Chambers and Hillis, 2020; Frankham et al., 2012; Sukumaran and Knowles, 2017), lest we contribute to taxonomic inflation. This potential hurdle might be of relevance in *Dugesia*.

Studies have shown that freshwater planarians of this genus not only accumulate numerous mutations, some of which even non-synonymous (Nishimura et al., 2015), but they may present highly divergent alleles within a same individual (Leria et al., 2019). Ultimately, this may have an impact on molecular delimitations performed without an *a priori* hypothesis (i.e., species discovery methods, Carsten et al., 2013). In Chapter II.1, we proposed PSCs by establishing a primary species hypothesis (PSH) based on comparing three different discovery methods, viz. ABGD (Puillandre et al., 2012), GMYC (Pons et al., 2006), and mPTP (Kapli et al., 2017). GMYC, but most notably ABGD, were prone to overestimate the number of species within *D. benazzii* s.l. with the available sampling and loci at that time (i.e., before the discovery of *D. xeropotamica* in Chapter II.2) perhaps due to the reasons detailed above. Similar results were reported for *D. subtentaculata* (Leria et al., 2020), in which delimitations performed using ABGD and GMYC often delineated numerous spurious species candidates, many of them singletons. But most interestingly, over-splitting even occurred when using cloned PCR products of a given molecular marker (i.e., a given individual could exhibit alleles so differentiated that discovery methods delineated *species* within it) (Leria et al., 2020). Although one could hastily conclude that molecular data alone contributes to the overestimation of *Dugesia* species, the truth is that over-splitting can also occur with morphological data. We can find an example of such situation in the recent synonymy of *Dugesia biblica* Benazzi & Banchetti, 1972 with *D. sicula* (Solà et al., 2015), in which molecular data was instrumental for the re-evaluation of their taxonomic status.

Indeed, based on all that has been discussed in this section, one could reasonably wonder if the delimitations established in Chapter II are reliable, especially after having pointed out the intrinsic limitations that the molecular markers employed in this thesis possess. However, throughout this thesis, we have always adopted an integrative approach when delimiting species because we were conscious of the perils of solely relying on either morphology or molecular evidence, thus ensuring robust taxonomic decisions. Hence, why we did not describe any new species in Chapter III despite uncovering highly divergent clades. Consequently, all delimited species in this thesis present molecular and morphological diagnosis. Furthermore, when morphology was inconclusive, we resorted to additional sources of evidence as in the case of *D. hoidi* using karyological data, which has been shown to be a relevant source for testing molecular clades (Leria et al., 2018, 2020).

While it holds true that the combination of morphological and molecular data has become a customary tandem for reaching reliable taxonomic conclusions in most phylogenetic studies, it would be of major interest to explore other sources of data to attain more robustness under certain scenarios like, for example, when studying recently diverged candidate species (Chapter III). Interestingly, ecological factors have never been taken into serious consideration for delimiting freshwater planarians. Approaches based on hypervolumes (Blonder, 2018; and references therein) to describe the fundamental niche of species or populations have been used in integrative studies to confidently delimit species in other organisms (Koch et al., 2016; Mammola et al., 2018; Peretti et al., 2022). Although, niche modelling in freshwater planarians has been used in several studies (Benítez-Álvarez et al., 2023c; Lázaro et al., 2013; Leria et al., 2022; Solà et al., 2022; and Chapter III), it was never used for taxonomic purposes. Nevertheless, up-to-date molecular, morphological and karyological data appear to be valuable sources for delimiting *Dugesia* species, and integrative approaches should always be favored whenever possible, as they are likely to provide the most accurate answers.

To conclude this section, I think it is important to acknowledge that few are the occasions where one can have morphological, karyological and molecular data from a same given specimen. In general, despite variations in regenerative prowess among different taxa, *Dugesia* have great regeneration capabilities (Vila-Farré et al., 2023), which means that, in principle, one could genotype, karyotype, and then analyze morphologically any given individual by just severing a small tissue fragment (often the tip of the tail or the head section) and then wait for the animal to regenerate back. Since *Dugesia* are not large creatures, tissue availability is likely to be a limiting factor. Consequently, is it is possible that not all of the aforementioned analyses can be carried out simultaneously. Although freshwater planarians are not the most complicated creatures to rear in a laboratory, not all species are equally resilient to environmental changes and casualties may accumulate the longer they spend time in suboptimal rearing conditions. Not necessarily due to one being careless during sample manipulation, but perhaps owing to unaccounted biotic or abiotic factors. For example, we experienced great losses of *D. xeropotamica* (>80%) during their first few days settling in the laboratory after field collection for reasons we could not determine.

Given the complexity and dedication that the analysis of morphological traits of any individual entails (Sluys and Riutort, 2018), it is simply not realistic to expect such type of detailed information to be available for all samples in a phylogenetic study, as it can be attained in other taxonomic groups. Furthermore, whole specimens tend to be used for establishing the holotype, *de facto* rendering them unavailable for molecular analysis unless a fragment of the animal was sequenced beforehand. Concurrently, anyone who has dabbled with karyological analysis will know that obtaining clear and precise karyotypes in a single try is not the most common of outcomes. Moreover, karyotyping may compromise the survival of the studied specimens, which can be problematic if a given specimens wants to be reused for other analyses. Additionally, the quality or quantity of the yield from a single individual (or tissue fragment) may not be enough to perform certain analyses, as it

happened to us while performing the extractions for the RADseq analysis in Chapter III. Meaning that sometimes one may be forced to postpone other analyses to ensure that at least genotyping can be performed.

I believe the following words by Alexander Pope (1688-1744) succinctly encapsulate the experience of working with freshwater planarians: "*Like following life through creatures you dissect, you lose it in the moment you detect*".

3. Novel genomic resources for the planarian research community

In order to attain robust evolutionary results from genomic data, it is essential to have a high-quality reference assembly of the studied organisms or at least a closely related species. The proliferation of reference genomes has been remarkable over the past decade, owing to advancements in sequencing and assembly technologies alongside a reduction in associated costs (Giani et al., 2020; Heather and Chain, 2016). Notably, numerous international collaborative initiatives are dedicated to promoting genome sequencing across various organismal groups. For instance, the Earth BioGenome Project (EBP) (Lewis et al., 2018) endeavors to sequence, catalogue, and characterize all eukaryotic genomes on Earth within a decade, starting in 2018. Additionally, affiliated or complementary projects, such as ik5 (Sequencing Five Thousand Arthropod Genomes) (Evans at al., 2013) and B10K (The Bird 10,000 Genomes) (Zhang, 2015), focus on collecting genome biodiversity within more localized regions of the Tree of Life. Interestingly, a domestic example may be found in the Catalan Biogenome Project (CBP) (https://www.scb.cat/biogenoma), which aims to establish a detailed catalog of the genomic diversity of eukaryote species from all Catalan speaking territories.

In Chapter IV, we presented the first chromosome-scale genomes of *Dugesia*, namely of three Mediterranean species: *D. etrusca*, *D. hepta* and *D. liguriensis*. These genomes represent a valuable resource for the planarian research community, not only for studying

the intricacies of the genome structure of these triclads, but because it will enable the establishment of the first comparative genomic analyses between *Dugesia* and the most characteristic model genus within Dugesiidae (and Tricladida), *Schmidtea*. Furthermore, these genomes will likely pave the ground for future comparative and population genomic analyses in *Dugesia*.

Indeed, we were able to compare for the first time *S. mediterranea* against our genomes and our results highlighted some interesting results. It appears that the difference of genome size in *Dugesia* compared to *Schmidtea* was not the byproduct of an event of whole genome duplication, but rather of a long process mediated by the activity of transposable elements (TE). In fact, we were able to discover that *Dugesia* exhibit an apparent expansion of DNA/*Maverick* TEs (Chapter IV, Fig. 6), which are among the largest TEs known. Owing to the lack of genome size estimates for other dugesiid genera in general, such as *Girardia* Ball, 1974 or *Recurva* Sluys, 2013, it is difficult to ascertain whether a gain in genome size occurred in *Dugesia* or a reduction in *Schmidtea*. However, if we compare the available genome size estimates for Dugesiidae at the Animal Genome Size Database (Gregory, 2024; http://www.genomesize.com, last accessed in March 2024) with the estimates presented in Chapter IV and those reported in other studies (Gregory et al., 2000; Ivanković et al., 2023), it could be suggested that a genome expansion occurred in *Dugesia* and possibly a genome reduction in the lineage of *S. mediterranea*.

Concurrently, we observed that macrosynteny in Mediterranean *Dugesia* appears to be well preserved in general (Chapter III, Fig. 7), as opposed to results shown for *Schmidtea* (Ivanković et al., 2023). No relevant structural differences were found between *D. etrusca* and *D. liguriensis*, whose lineages diverged approximately 9.5 Mya (Solà et al., 2022). Moreover, macrosynteny analysis with *D. hepta* showed that two major fusion events were responsible for its chromosome complement (n = 7), and highlighted two possible pericentric inversions, which are a phenomenon that has been reported in other *Dugesia* species (De Vries, 1986a; Ribas, 1990; Tamura et al., 1991, 1998). Barring the aforementioned rearrangements, in general the macrosynteny between *D. hepta* and its congenerics was conserved despite the considerable divergence time between their respective lineages (~30 Mya, Solà et al., 2022). It is worth mentioning that there are no dedicated dating estimates of divergence times in *Schmidtea*, therefore it is unclear if their pervasive chromosomal rearrangements are characteristic of the genus or just a common phenomenon in Dugesiidae among lineages that diverged a long (long) time ago. It would be most interesting to sequence chromosome-scale genomes of Malagasy *Dugesia* (or other ancient clades) to assess whether synteny is only preserved at 'shallow' time scales or not in *Dugesia*.

4. Karyological plasticity and its role in speciation in planarians

4.1. Ploidies in the Corso-Sardinian clade

Throughout Chapters I, II and III we collected karyological information from numerous populations. Although we did not perform an analysis for the ancestral state reconstruction of the ploidy level within the Corso-Sardinian clade, it falls within reason to assume that most likely the last common ancestor of the group had a haploid complement of n = 8 and was probably diploid. This assumption is based on the phylogenetic results shown in Chapter III (Chapter III, Fig. 8b, Supplementary Fig. S15), the available karyological knowledge on Western European *Dugesia*, and the macrosynteny results in Chapter IV. Polyploidization events appear to have occurred several times in the Corso-Sardinian clade, although at present it is unclear if said events triggered or were involved in speciation processes, provided that some delimited PSCs encompass diploid as well as triploid or aneuploid individuals (e.g., Chapter III: PSC1, PSC5, PSC8). Moreover, the existence of sexually reproducing polyploid populations is known in other freshwater planarians of the genus *Dugesia* (Chinone et al., 2014; Leria et al., 2020), and as far as we know, differences in ploidy level not necessarily preclude viable crossing, as it has been demonstrated in *Schmidtea* (Pongratz et al., 1998).

4.2. On the aneuploidy and intraindividual karyotype variability

Aneuploids were of particular interest for this thesis. Pala and collaborators (1982)

hypothesized that Sardinian aneuploids were in fact natural hybrids between *D. hepta* and *D. benazzii*. Had we had robust karyological data on the putative hybrid mitochondrial clade from Chapter I, the mystery of the interspecific *hepta* × *benazzii* hybrids could have been solved earlier. Yet, in Chapter III, we were able to prove that not only aneuploids were not hybrids, but that it was a common karyotype in Corsica as well. However, the most striking characteristic of these aneuploid individuals was in fact the high *intraindividual* karyotype variability (i.e., a given individual may present differing chromosomal numbers). Indeed, it is an oddity, but Pala and collaborators (1982) reported the same results. Reports on karyotype variability in other *Dugesia* species, such as *D. subtentaculata*, exist, but the described pattern was intraspecific (Leria et al., 2020).

Considering that neoblasts are the only type of cells with mitotic activity in planarians, this pattern of karyological variability could potentially be explained by a mis-segregation of chromosomes. Pala and collaborators (1982) contemplated this possibility and proposed a hypothesis in which first syngamy of interspecific gametes occurred, followed by polyploidization of the hybrid embryos, probably due to endomitosis. Thereon, this process would result in the generation of polyploid neoblasts that would mis-segregate during mitosis with tri- or tetrapolar spindles. Barring the initial hybridization, which we refuted in Chapter III, the rest of the process offers a reasonable explanation to the variability observed. Interestingly, it has been observed in *S. mediterranea* that some genes that are responsible of the spindle assembly checkpoint (SAC) are lacking, which is the cellular mechanism that looks after the correct segregation of chromosomes (Grohme et al., 2018).

Finally, intraindividual karyotype variability has been reported in vertebrates (Civitelli et al., 1989; Diniz and Bertollo, 2006) and invertebrates (Pavlíček et al., 2016; and references therein) as well, but most notably in cancer cells (Nicholson and Cimini, 2013; Orr and Compton, 2013). Recently, is has been observed that *S. mediterranea* exhibits tight associations of regulatory elements with their target genes (Ivanković et al., 2023). On the other hand, high GC content has been shown to prevent double strand breaks of DNA in prokaryotes (Weissman et al., 2019) but it is positively correlated with higher mutation rates (Vinogradov and Anatskaya, 2017). Considering all the aforementioned factors, along with the fact that planarians have fairly low GC content (Ivanković et al., 2023; Solà et al., 2015; Chapter IV, Table 2), one could suggest that freshwater planarians might be prone to chromosomal rearrangements but are naturally resistant to their consequences and the incidence of non-synonymous mutations that may occur due to their constant regeneration and cell turnover. Interestingly, *Hydra* (phylum Cnidaria), whose regenerative capabilities are on par with those of freshwater planarians, also exhibits low levels of GC content (Chapman et al., 2010). All things considered, one could jokingly say that the Corso-Sardinian aneuploids are akin to cross-eyed, free-living tumors. Nevertheless, it is fascinating how they are morphologically and physiologically functional, even capable of performing successful meiosis and producing progeny (Pala et al., 1982). Which begs the question: could these aneuploids be a new potential model for studying cancer dynamics?

4.3. Remarks on ploidy and reproductive strategy: a key to success?

Although remarks on the reproductive strategy were barely mentioned in Chapter III (because it was out of our scope), I must note that we found that exclusively diploid populations/species were sexual, viz. *D. leporii*, *D. xeropotamica* and PSC2 (individuals from monte Albo massif), and did not show signs of reabsorption of the copulatory apparatus. Contrastingly, polyploid and mixed populations (diploids and polyploids altogether) combined sexual and fissiparous reproduction, and displayed reabsorption of the copulatory apparatus within weeks of rearing in the laboratory. Similar cases of diverse reproductive strategies seemingly linked to ploidy level have been documented in other *Dugesia*, such as *D. sicula* (Lázaro et al., 2009) or *D. subtentaculata* (Leria et al., 2020).

Indeed, freshwater planarians may reproduce sexually via cross-fertilization or asexually by fissiparity or parthenogenesis (Pongratz et al., 2003; Stocchino and Manconi, 2013) and it appears that often polyploids favor the latter strategies, while diploids tend to reproduce sexually. The consequences of shifts in ploidy and reproductive strategy are complex and variable depending on the species. Although polyploidization has often been considered an evolutionary 'dead end' in the long run (Van de Peer at al., 2017), it seems freshwater planarians have thrived in it, as suggested by the maintenance of ancient fissiparous lineages in some species (Benítez-Álvarez et al., 2023a, *in prep.*). Notably, several *Dugesia* species, for which only natural sexual (and diploid) populations have been reported, exhibit restricted ranges of distribution, e.g., *Dugesia aurea* Leria, 2020, *Dugesia corbata* Leria, 2020, *D. hepta*, *D. hoidi, D. leporii*, *Dugesia parasagitta* Sluys & Solà, 2013, *Dugesia vilafarrei* Leria, 2020, and *D. xeropotamica* to name a few. While asexually reproducing (and often polyploid) *Dugesia* appear to have broader distribution ranges, e.g. *D. etrusca, D. sicula,* or *D. subtentaculata.* In fact, asexuality and polyploidy have been proposed as dominant factors accounting for the colonizing success of *D. sicula* (Lázaro and Riutort, 2013), but also in other freshwater genera, such in the case of the species *Girardia tigrina* (Girard, 1850) (Benítez-Álvarez et al., 2023c).

While it is true that polyploidy can have detrimental effects on fertility and fitness owing to genomic instability, mitotic and meiotic abnormalities (Van de Peer et al., 2017; and references therein), among other things for which freshwater planarians seem to be well prepared, it can also be advantageous under certain circumstances. Polyploidy may enable higher tissue regeneration rates owing to an enhanced production of RNA and proteins (e.g., Krois et al., 2013), contribute to environmental stress resilience (e.g., Ceccarelli et al., 2006; Chao et al., 2013; Pienta et al., 2022), optimize the mutation rate per individual in asexual taxa (Manning and Dickinson, 1986), or mitigate the ravages of Muller's ratchet (i.e., accumulation of deleterious mutations in the absence of recombination, as expected in asexual organisms) (Maciver, 2016). Interestingly, while most of the aforementioned points have not been dedicatedly studied in planarians, Muller's ratchet was, and no signs of it have been found in polyploid *D. subtentaculata* (Leria et al., 2019). Being polyploid ensures multiple copies of any given gene, which might help species scape the ratchet's fate through rapid reversal of spontaneous deleterious mutations by virtue of gene con-

version, a type of homologous recombination in which one sequence is replaced by another copying the homologous sequence from the donor sequence (Maciver, 2016). For instance, it has been shown that gene duplication, coupled with gene conversion between duplicated pairs, can be effective at maintaining the functionality of duplicates in human Y chromosome, a nonrecombining chromosome (Connallon and Clark, 2010). Bearing in mind the genomic characteristics of freshwater planarians that I have exposed earlier (e.g., low GC content, tight associations of regulatory elements with their target genes, polyploidy and its discussed causes), it falls within reason to contemplate shifts of ploidy and reproductive strategy as relevant causes that explain the evolutive success of freshwater planarians, who have silently colonized freshwater environments all across the globe and survived all major extinction events.

Hypothetically, polyploidization could have ensued in response to the harsh environmental conditions during Pleistocene glaciations and the probable size reduction of natural populations. Such conditions would likely have had detrimental consequences at a genomic level, and subsequent ploidy shifts to mitigate the potential negative outcomes could have contributed to the presumed rapid radiation in Corsica (Chapter III). While the extent to which polyploids establish through niche differentiation and whether polyploidy promotes ecological shifts remains insufficiently investigated (Parisod and Broennimann, 2016), polyploids are often associated with a higher tolerance and broader range of ecological and environmental conditions compared to diploids. Notably, it has been proposed that polyploids are 20% more likely to be invasive than closely related diploids, at least in plants (Pandit et al., 2011). Given the recent origin of most PSCs (Chapter III, Supplementary Fig. S15), it may even be possible that those candidate species comprising polyploid and diploid populations could be in the early stages towards a new speciation process. Or, perhaps, this pattern only means that polyploids and diploids coexist because the formers can also produce diploid offspring, as it has been shown to occur with *D. ryukyuensis* (Chinone et al., 2014; Kobayashi et al., 2008).

5. Chromosomal rearrangements and their role in *D. hepta*'s speciation

With the availability of high-quality chromosome-scale genomes, chromosomal rearrangements (CRs) can be detected in a set of related taxa by mapping reads, or pseudochromosome scaffolds, to a reference genome. Thus, inversions, duplications, deletions, fusions and translocations can be inferred based on the resulting mapping patterns (Lucek et al., 2023).

In Chapter IV, we were able to visualize for the first time the CRs involved in the formation of *D. hepta*'s karyotype (Chapter IV, Fig. 7). Bearing in mind that we were only able to observe patterns on a macrosyntenic level, we were able to detect apparent translocations and pericentric inversions in most chromosomes, but most notably in chromosomes 2, 3, 6 and 7. It is acknowledged that TEs may play an important role in the emergence of CRs (reviewed in Chénais et al., 2012), thus, we looked into the repeat content of our assemblies. Interestingly, we found that *D. hepta* exhibited relative expansions of LTR/*Gypsy* and SINE/*5S* retrotransposons, being the latter remarkably conspicuous. While we have not been able to explicitly analyze the distribution of these elements in *D. hepta*'s assembly or give them a temporal context (i.e., discern whether they constitute a recent expansion or not), we hypothesized that SINE/*5S* retrotransposons might be behind the CRs in *D. hepta*. This proposal rises from the striking differences in *D. hepta*'s SINEs composition when compared to those of *D. etrusca* and *D. liguriensis*. However, how could

have the CRs influenced, or promoted, *D. hepta*'s speciation?

CRs have been known since the early stages of genetics and their role in speciation has been lengthily studied and discussed (Faria and Navarro, 2010; Lucek et al., 2023), yet most theoretical and empirical studies have focused on a microevolutionary level (i.e., taxon pairs). The outcomes of CRs are diverse (Faria and Navarro, 2010), and they will not always contribute to reinforcing the reproductive isolation between two populations or lineages, especially when small CRs occur (Lucek et al., 2019). However, it falls within reason to expect that in the case of *D. hepta*, CRs must have played a significant role in speciation, most notably due to the resulting difference in chromosome number from the rest of its Corso-Sardinian congenerics. Changes of this nature can lead to reproductive isolation owing to heterokaryotypic individuals experiencing reduced fertility (or overall reduced fitness) caused by meiotic segregation problems. Resulting, in our particular case, in the reinforcement of the reproductive isolation between a lineage with a haploid complement of n = 7 and that with a haploid complement of n = 8.

While it is true that the change in chromosome number is a conspicuous difference, the detected inversions could have also played an important role in the divergence of *D. hepta*. In fact, they may promote speciation by delaying gene flow under uniform selection of alleles between parapatric populations (Navarro and Barton, 2003). Considering how *D. hepta* and *D. benazzii* often coexist within the same rivers, this model could adjust well to our case of study. However, it is difficult to assess the role that the inversions detected in Chapter IV might have played in the speciation of *D. hepta*. These inversions could have been concomitant with the translocations that gave rise to *D. hepta*'s karyotype or could have predated the speciation event, potentially representing a common feature among species of the Corso-Sardinian clade. Alternatively, they could have postdated the speciation event and, in turn, had nothing to do with it. All things considered, additional genome assemblies of other Corso-Sardinian species will be required to clarify the state of affairs with regards to the pericentric inversions and their involvement in the origin of *D. hepta*.

6. A new perspective on the putative hybrids

The topic of interspecific hybridization has been treated and discussed in Chapters I and III. In Chapter I, we acknowledged the possibility that potential ongoing hybridization could involve *D. hepta* and *D. benazzii*, two species with completely different chromosome complements. Truth be told, while the results we presented in Chapter I could be interpreted as evidence of the claimed natural interspecific hybridization first postulated by Pala and collaborators (1982), its interpretation was not straightforward, and other possibilities such as ILS, were considered. In Chapter III, owing to the use of genomic data (RADseq) and an improved geographical sampling, we were able to refute our hypothesis presented in Chapter I while coincidentally uncovering a seemingly genuine case of natural interspecific hybridization between *D. benazzii* and the elusive *D. leporii*.

It has been estimated that at least 10% of animal species and 25% of plant species have partaken in hybridization with at least one other species (Mallet, 2005, 2007). However, in the era of high-throughput sequencing, it is likely that the extends and impact of hybridization in the Tree of Life are larger than previously anticipated (Payseur and Risenberg, 2016; Taylor and Larson, 2019). Most importantly, it appears that the outcomes of hybridization are more variable than was expected and the geographical context of hybridization along with the initial outcome of it (the extent and time period of hybridization) are important (Taylor and Larson, 2019).

Indeed, the advances in sequencing technologies have enabled the sequencing of whole genomes or reduced-representations of them (i.e., RADseq) at a low cost. As a result, scientists have gained access to unprecedented levels of genomic resolution, which have proved instrumental to detect small genomic blocks of hybrid origin that, otherwise, would have likely been overlooked with traditional panels of a handful of markers. Thus, in recent years a growing number of studies have documented cases of ancient hybridization between lineages (Taylor and Larson, 2019; references therein). Although the terminology 'ancient hybridization' evokes the relative passing of a substantial, yet often undetermined, amount of time between the event itself and the detection of it in the present time, it is often used when there is no evidence of contemporary hybridization between taxa, either because the extant lineages are allopatric or one of the involved lineages went extinct (Palkopoulou et al., 2018; Sankararaman et al., 2014). Technically, the results discussed in Chapter III could be presented as a case of ancient hybridization in freshwater planarians, as suggested by the observed patterns of introgression (Chapter III, Figs 6 and 7) along with the fact that neither *D. leporii* or *D. benazzii* were found in the hybrids' sampling site. Yet, beyond formal labels, it is worth mentioning that, as far as we know, our findings suppose the first documented case of interspecific hybridization in Dugesiidae and Tricladida.

Even though we could not prove the hybridization hypothesis between *D. hepta* and *D. benazzii*, our results are a testament that natural hybridization is possible in freshwater planarians. Considering the numerous genomic, karyological, and reproductive particularities that freshwater planarians display, I believe our results should not deter researchers from investigating or finding signs of hybridization between *D. hepta* and other members of the Corso-Sardinian clade in future studies. As a matter of fact, it is known that in the Silis River (Sardinia, Italy), wherein *D. hepta* and *D. benazzii* coexist, several differing biotypes of *D. benazzii* can be found. However, there is one biotype in particular whose rarity has puzzled researchers. Vacca and collaborators (1988), while conducting a karyological survey of populations of freshwater planarians inhabiting the aforementioned river, documented sexual individuals that exhibited a chromosomic portrait of 22 chromosomes in total, which accounted for approximately 10% of the studied individuals. According to the authors, it was not possible to determine the haploid complement. In Chapter I, we proposed the possibility that these individuals could be hybrids, as syngamy between a diploid *D. hepta* gamete (2n = 14) and a haploid *D. benazzii* gamete (n = 8) would result in the reported chromosomal number documented by Vacca and collaborators (1988). Unfortunately, we have not encountered any animals bearing this particular karyotype throughout this thesis.

7. Accounts on the *leporii* × *benazzii* hybrids: a case of hybrid speciation?

Hybrid speciation refers to a process of speciation where hybridization plays a pivotal role in the development of reproductive barriers between a hybrid lineage and its parent lineages. There are two major types of hybrid speciation: allopolyploid and homoploid. The former type of hybrid speciation is defined as the formation of a hybrid lineage with a concomitant whole-genome duplication, which results in immediate incompatibility between the hybrid lineage and their parental species. Contrastingly, homoploid hybrid speciation does not entail a change in ploidy level or a whole-genome duplication, but the combination of parental genomes must also lead to traits that will partially or completely reproductively isolate the hybrid evolutionary lineage from their parent lineages. Furthermore, certain criteria have to be satisfied to support a case of homoploid hybrid speciation, namely: (i) genomes have to exhibit signatures of hybridization, (ii) hybrids are reproductively isolated from parental forms, and (iii) isolating mechanisms were triggered by hybridization. However, which of the two types best reflects the case of the hybrids discussed in Chapter III?

To start with, we know too little of these hybrids. Our results suggest that they are not first-generation (F_1) hybrids and probably are a result of backcrossing with one of the parental lineages, most likely *D. leporii*'s. Moreover, our karyological results suggested that hybrids possess a potentially mixoploid karyotype of 2n = 16 and 3n = 24 (Chapter III,

Supplementary Fig. S2). Hybrids did not exhibit a gonopore or any signs of reproductive apparatus development, neither when collected nor while they were reared in the laboratory. Consequently, it is safe to assume that they likely represent an asexually reproducing population. However, it is worth noting that recently several mixoploid species with adult specimens exhibiting developed reproductive apparatus, yet seemingly infertile, have been described in China, viz. *Dugesia adunca* Chen & Sluys, 2022, *Dugesia majuscula* Chen & Dong, 2021, *Dugesia semiglobosa* Chen & Dong, 2021, and *Dugesia tumida* Chen & Sluys, 2022.

It is not clear if a whole-genome duplication could have occurred at some point in the hybrid lineage, and if it had, it might not have been an impediment to reproduce sexually with other polyploids or diploids (Chinone et al., 2014; Kobayashi et al., 2008; Pongratz et al., 1998). Although the available evidence on the reproductive peculiarities of freshwater planarians does not concern interspecific crossings, anything seems to be possible when speaking of these organisms. In all, it is likely that F_1 hybrids were not allopolyploid, but not impossible. With regards to the hybrids from Chapter III, their apparently mixoploid chromosomic portrait complicates things. Although these hybrids exhibit clear genomic signatures of hybridization, they are reproductively isolated from their parental lineages owing to being asexual, and mixoploidy (probably a consequence of the hybridization between F_1 or F_n hybrids and *D. leporii*) could have triggered the reproductive isolation (as suggested by the infertility reports on mixoploid *Dugesia*), neither the allopolyploid nor the homoploid hybrid speciation concepts seem to properly fit the case at present. And yet, they can proliferate and reproduce by virtue of fissiparity, forming, as far as we know, a population like many other asexual species do.

Most interestingly, a comprehensive origin of mixoploidy in animal systems was recently proposed (Destouni and Vermeesch, 2017; Masset et al., 2021) invoking a first non-canonical cleavage of the zygote. This hypothesis has been recently demonstrated in bovine zygotes (De Coster et al., 2022), and it appears that polyspermic fertilization is the main driver of this process coined *heterogoneic division.* Thereafter, uneven spindle apposition leads to multipolar zygotic division and the segregation of parental genomes into different blastomeres, which will subsequently proceed with their developmental propagation. This, however, has only been proved in a homospecific system. But what if it occurred in an interspecific context? What if the offspring could perpetuate themselves through asexual reproduction as the hybrids of Chapter III? Could these hybrids truly be considered a hybrid species? Well, certainly they are not only reproductively isolated from the parental lineages, but also possess a novel karyotype, which can be used as a source of evidence to discriminate this new hybrid lineage from their parental species. Although the hybrids do not form a monophyletic clade in Chapter III (Chapter III, Fig. 2a), it has been shown that the inclusion of ancient asexual populations might alter phylogenetic inferences (Benítez-Álvarez, 2022; Benítez-Álvarez et al., 2023a).

All things considered, one could wonder if coining the term *heterogoneic hybrid speciation* would adequately describe the origin of these hybrids.

Conclusions

1. The use of molecular data, be it legacy Sanger loci or restriction site-associated DNA sequencing loci, has proved instrumental to elucidate previously overlooked diversity within the Corso-Sardinian clade.

2. Integrating information from genomic, morphological and karyological sources has been shown to be a suitable and reliable approach for conducting taxonomic analyses in freshwater planarians and should be encouraged in all studies of taxonomic nature.

3. The obtention of the first molecular phylogenies focused on the taxa of the Corso-Sardinian clade corroborated that *D. hepta* is not an aberrant chromosomal form of *D. benazzii* s.l.

4. *D. benazzii* s.l. harbored a complex of species distributed across Corsica and Sardinia. Consequently, three new members of the Corso-Sardinian clade were described, viz. *D. hoidi, D. mariae* and *D. xeropotamica*. Additionally, we provided a re-description of *D. benazzii* and proposed multiple species candidates, the status of which will be addressed in a forthcoming study.

5. While the phylogeny of the Corso-Sardinian clade was not fully resolved, this thesis has significantly expanded our understanding of this group and has made strides towards achieving a resolved phylogeny.

6. RADseq studies working on non-diploid organisms using a *de novo* assembly pipeline should contemplate the use of over-merging assembly parameters if optimal parameters are set too low.

7. Divergence time estimation analyses indicate that rapid diversification of the Corso-Sardinian clade occurred between the Late Pliocene and the Pleistocene, suggesting that climatic shifts, rather than microplate tectonics, were important drivers of speciation in this taxonomic group.

8. Due to the recent origin of the members of this clade, molecular barcoding cannot rely solely on the separate use of the molecular Sanger markers employed in this thesis. Concatenation of mitochondrial and nuclear markers, or the use of RAD loci, are recommended approaches to properly identify the species described in this thesis.

9. The use of niche modelling tools to guide sampling campaigns has proven successful, resulting in the discovery of specimens of *D. leporii* for the first time since 1999 and novel populations, among which a natural population of interspecific hybrids.

10. Genetic evidence demonstrates for the first time the existence of interspecific hybridization. The hybridization event likely involved *D. leporii* and the ancestor of *D.*

benazzii. Furthermore, the hybrids are likely to be the byproduct of backcrossing with *D. leporii*.

11. Hybrids could represent a case of hybrid speciation in which the established definitions of allopolyploid and homoploid hybrid speciation do not seem to apply.

12. The first chromosome-scale genome assemblies of *Dugesia* species allow to conclude that macrosynteny is well preserved in *Dugesia*, in contrast with patterns reported in *Schmidtea*.

13. Differences of genome size between *Dugesia* and *Schmidtea* are not explained by a whole-genome duplication event.

14. Changes in genome size in *Dugesia* with respect to *Schmidtea* occurred in non-coding regions and involved an expansion of DNA transposable elements of the largest kind found in eukaryotes.

15. The chromosome complement of *D. hepta* emerged from a chromosomal rearrangement involving at least two major fusion events. We found evidence indicating that an expansion of SINE retrotransposons could have intervened in the rearrangement events.

References

Abbott, R.J., Barton, N.H., Good, J.M. (2016). Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, 25(11), 2325-2332. https://doi.org/10.1111/mec.13685

Agapow, P-M. (2005). Species: demarcation and diversity. In: Purvis, A., Gittleman, J.L. (eds.), *Phylogeny and Conservation*. Cambridge: Cambridge University Press, 57-75.

Alvarez, W. (1972). Rotation of the Corsica-Sardinia microplate. *Nature Physical Science*, 235(58), 103-105. https://doi.org/10.1038/physci235103a0

Alvarez, W., Cocozza, T., Wezel, F.C. (1974). Fragmentation of the Alpine orogenic belt by microplate dispersal. *Nature*, 248(5546), 309-314. https://doi.org/10.1038/248309a0

Álvarez-Presas, M., Baguñà, J., Riutort, M. (2008). Molecular phylogeny of land and freshwater planarians (Tricladida, Platyhelminthes): from freshwater to land and back. *Molecular Phylogenetics and Evolution*, 47(2), 555-568. https://doi.org/10.1016/j.ympev.2008.01.032

Álvarez-Presas, M., Riutort, M. (2014). Planarian (Platyhelminthes, Tricladida) diversity and molecular markers: a new view of an old group. *Diversity*, 6(2), 323-338. https://doi.org/10.3390/d6020323

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., Hohenlohe, P.A. (2016). Harnessing

the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81-92. http://dx.doi.org/10.1038/nrg.2015.28

Baguñà, J. (2012). The planarian neoblast: the rambling history of its origin and some current black boxes. *International Journal of Developmental Biology*, 56(1-3), 19-37. https://doi.org/10.1387/ijdb.113463jb

Baguñà, J., Carranza, S., Pala, M., Ribera, C., Giribet, G., Arnedo, M.A., Ribas, M., Riutort, M. (1999). From morphology and karyology to molecules. New methods for taxonomical identification of asexual populations of freshwater planarians. A tribute to Professor Mario Benazzi. *Italian Journal of Zoology*, 66, 204-214. https://doi.org/10.1080/11250009909356258

Baguñà, J., Carranza, S., Paps, J., Ruiz-Trillo, I., Riutort, M. (2001). Molecular taxonomy and phylogeny of the Tricladida. In: Littlewood, D.T.J., Bray, R.A. (eds.), *Interrelationships of the Platyhelminthes*. London & New York: Taylor & Francis, 49-56.

Baird, J., McDowell, S.D.R., Fairweather, I., Murchie, A.K. (2005). Reproductive structures of *Arthurdendyus triangulatus* (Dendy): seasonality and the effect of starvation. *Pedobiologia*, 49(5), 435-442. https://doi.org/10.1016/j.pedobi.2005.05.003

Ball, I.R. (1974). A contribution to the phylogeny and biogeography of the freshwater triclads (Platyhelminthes, Turbellaria). In: *Biology of the Turbellaria*. New York: McGraw-Hill, 339-401.

Ball, I.R. (1975). Nature and formulation of biogeographic hypotheses. *Systematic Biology*, 24(4), 407-430. https://doi.org/10.1093/sysbio/24.4.407

Ball, I.R., Gourbault, N., Kenk, R. (1981). Planarians (Turbellaria) of temporary waters in Eastern North America. *Life Science Contributions of the Royal Ontario Museum*, 127:27.

Barbato, D., Benocci, A., Manganelli, G. (2018). The biogeography of non-marine molluscs in the Tuscan Archipelago reveals combined effects of current eco-geographical drivers and paleogeography. *Organisms Diversity and Evolution*, 18(4), 443-457. https://doi.org/10.1007/s13127-018-0378-2

Belton, J-M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*, 58(3), 268-276. http://dx.doi.org/10.1016/j.ymeth.2012.05.001

Benazzi, M. (1936). Razze fisiologiche di "*Euplanaria gonocephala*" differenziate dalla diversa attitudine alla scissiparità. *Rendiconti Reale Academia Nazionale del Lincei*, 23, 361-365.

Benazzi, M. (1938a). Tricladi Paludicole della Sardegna. *Archivio Zoologico Italiano*, 25, 85-94.

Benazzi, M. (1938b). Ricerche sulla riproduzione delle planarie tricladi Paludicole con particolare riguardo alla moltiplicazione asessuale. *Atti della Reale Academia Nazionale dei Lincei*, 7, 31-38.

Benazzi, M. (1945). Mutazione genomica in una razza di *Dugesia* (*Euplanaria*) *gonocephala* (Dugès). *Bolletino della Società Italiana di Biologia Sperimentale*, 20.

Benazzi, M. (1950). Ginogenesi in tricladi di acqua dolce. *Chromosoma*, 3, 474-482.

Benazzi, M. (1955). Appunti sulla distribuzione dei Tricladi in Italia. *Bolletino di Zoologia*, 22(2), 149-164. https://doi.org/10.1080/11250005509439195

Benazzi, M., Deri, P. (1980). Histo-cytological study of fissiparous planarian testicles (Tricladida, Paludicola). *Italian Journal of Zoology*, 14(3), 151-163. https://doi.org/10.1080/00269786.1980.10736353

Benazzi-Lentati, G. (1979). Polyploidy and types of development in the triclad *Dugesia lugubris* s.l.: interaction between ploidy level and the origin of amphimixis or pseudogamy. *Caryologia*, 32(2), 247-263. https://doi.org/10.1080/00087114.1979.10796791

Benazzi-Lentati, G., Piegaia, M.I. (1953). Sulle eccezioni alla riproduzione pseudogametica in biotipi di *Dugesia benazzii* (Tricladida, Paludicola). *Caryologia*, 5, 192-216.

Benítez-Álvarez, L. (2022). *Understanding diversification and dispersion patterns in planarians (Tricladida, Platyhelminthes): from one to hundreds of genes*. Doctoral Thesis, Universitat de Barcelona. http://hdl.handle.net/10803/675996

Benítez-Álvarez, L., Leal-Zanchet, A.M., Oceguera-Figueroa, A., Lopes Ferreira, R., de Medeiros Bento, D., Braccini, J., Sluys, R., Riutort, M. (2020). Phylogeny and biogeography of the Cavernicola (Platyhelminthes: Tricladida): relicts of an epigean group sheltering in caves? *Molecular Phylogenetics and Evolution*, 145, 106709. https://doi.org/10.1016/j.ympev.2019.106709

Benítez-Álvarez, L., Leria, L., Fernández, R., Mateos, E., El Ouanighi, Y., Bennas, N., El Alami, M., Yacoubi-Khebiza, M., Ayt Ougougdal, H., Riutort, M. (2023a). Phylotranscriptomics interrogation uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean. *Molecular Phylogenetics and Evolution*, 178, 107649. https://doi.org/10.1016/j.ympev.2022.107649

Benítez-Álvarez, L., Leria, L., Sluys, R., Leal-Zanchet, A.M., Riurtort, M. (2023c). Niche modelling and molecular phylogenetics unravel the colonisation biology of three species of the freshwater planarian genus *Girardia* (Platyhelminthes, Tricladida). *Hydrobiologia*, 850(14), 3125-3142. https://doi.org/10.1007/s10750-023-05239-x

Benítez-Álvarez, L., Sluys, R., Leal-Zanchet, A.M., Leria, L., Riurtort, M. (2023b). First molecular phylogeny of the freshwater planarian genus *Girardia* (Platyhelminthes: Tricladida) unveils hidden taxonomic diversity and initiates resolution of its historical biogeography. *Zoological Journal of the Linnean Society*, 197(1), 1-19. https://doi.org/10.1093/zoolinnean/zlac065

Benítez-Álvarez, L., Tenaguillo, I., Leria, L., Fernández, R., Riutort, M. (*in prep.*). A new asexual scandal: ancient planarian fissiparous lineages evolving independently brings spe-

cies concept on the line.

Berninger, J. (1911). Über die Einwirkung des Hungers auf Planarien. *Zoologische Jahrbücher, Abteilung für allgemeine Zoologie und Physiologie der Tiere*, 30, 181-216.

Beukeboom, L.W., Sharbel, T.F., Michiels, N.K. (1998). Reproductive modes, ploidy distribution, and supernumerary chromosome frequencies of the flatworm *Polycelis nigra* (Platyhelminthes: Tricladida). *Hydrobiologia*, 383, 277-285. https://doi.org/10.1023/a:1003460132521

Beukeboom, L.W., Vrijenhock, R.C. (1998). Evolutionary genetics and ecology of sperm-dependent parthenogenesis. *Journal of Evolutionary Biology*, 11(6), 755-782. http://doi.wiley.com/10.1046/j.1420-9101.1998.11060755.x

Bhardwaj, G., Ko, K.D., Hong, Y., Zhang, Z., Ho, N.L., Chintapalli, S.V., Kline, L.A., Gotlin, M., Hartranft, D.N., Patterson, M.E., Dave, F., Smith, E.J., Holmes, E.C., Patterson, R.L., van Rossum, D.B. (2012). PHYRN: a robust method for phylogenetic analysis of highly divergent sequences. *PLoS ONE*, 7(4), e34261. https://doi.org/10.1371/journal.pone.0034261

Bidegaray-Batista, L., Arnedo, M.A. (2011). Gone with the plate: the opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. *BMC Evolutionary Biology*, 11(1), 317. https://doi.org/10.1186/1471-2148-11-317

Bisconti, R., Canestrelli, D., Salvi, D., Nascetti, G. (2013). A geographic mosaic of evolutionary lineages within the insular endemic newt *Euproctus montanus*. *Molecular Ecology*, 22(1), 143-156. https://doi.org/10.1111/mec.12085

Blonder, B. (2018). Hypervolume concepts in niche- and trait-based ecology. *Ecography*, 41(9), 1441-1455. https://doi.org/10.1111/ecog.03187

Boag, B., Neilson, R., Scrimgeour, C.M. (2006). The effect of starvation on the planarian *Arthurdendyus triangulatus* (Tricladida: Terricola) as measured by stable isotopes. *Biology and Fertility of Soils*, 43(2), 267-270. https:/doi.org/10.1007/s00374-006-0108-3

Bogaerts, B., Van den Bossche, A., Verhaegen, B., Delbrassinne, L., Mattheus, W., Nouws, S., Godfroid, M., Hoffman, S., Roosens, N.H.C., De Keersmaecker, S.C.J., Vanneste, K. (2024). Closing the gap: Oxford Nanopore Technologies R10 sequencing allows comparable results to Illumina sequencing for SNP-based outbreak investigation of bacterial pathogens. *Journal of Clinical Microbiology*, in press. https://doi.org/10.1128/jcm.01576-23

Boll, P.K., Leal-Zanchet, A.M. (2016). Preference for different prey allows the coexistence of several land planarians in areas of the Atlantic Forest. *Zoology*, 119(3), 162-168. https://doi.org/10.1016/j.zool.2016.04.002

Bonini, A., Sauciuc, A., Maglia, G. (2024). Engineered nanopores for exopeptidase protein sequencing. *Nature Methods*, 21(1), 16-17. https://doi.org/10.1038/s41592-023-02136-y

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushki-

na, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis.

Buckley, T.R., Simon, C., Chambers, G.K. (2001). Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. *Systematic Biology*, 50(1), 67-86. https://doi.org/10.1080/10635150116786

Burbrink, F.T., Crother, B.I., Murray, C.M., Smith, B.T., Ruane, S., Myers, E.A., Pyron, R.A. (2022). Empirical and philosophical problems with the subspecies rank. *Ecology and Evolution*, 12(7), e9069. https://doi.org/10.1002/ece3.9069

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nature Biotechnology*, 31(12), 1119-1125. https://doi.org/10.1038/nbt.2727

Canovai, R. (1989). Distribuzione dei Tricladi dulcacquicoli dei Monte Pisani con riferimento ai biotipi cariologici delle diverse specie. *Atti della Società Toscana di Scienze Naturali Residente in Pisa, Memorie (Serie B)*, 96, 107-120.

Cariou, M., Duret, L., Charlat, S. (2013). Is RAD-seq suitable for phylogenetic inference? An *in silico* assessment and optimization. *Ecology and Evolution*, 3(4), 846-852. https://doi.org/10.1002/ece3.512

Carmignani, L., Conti, P., Cornamusini, G., Meccheri, M. (2004). The internal Northern Apennines, the Northern Tyrrhenian Sea and the Sardinia-Corsica block. *Special Volume of the Italian Geological Society for the IGC 32 Florence-2004*. https://www.pconti.net/doc/carmignani2004.pdf

Carranza, S., Baguñà, J., Riutort, M. (1997). Are the Platyhelminthes a monophyletic primitive group? An assessment using 18S rRNA sequences. *Molecular Biology and Evolution*, 14(5), 485-497. https://doi.org/10.1093/oxfordjournals.molbev.a025785

Carranza, S., Littlewood, T.J., Clough, K.A., Ruiz-Trillo, I., Baguñà, J., Riutort, M. (1998). A robust molecular phylogeny of the Tricladida (Platyhelminthes: Seriata) with a discussion on morphological synapomorphies. *Proceedings of the Royal Society B: Biological Sciences*, 265(1396), 631-640. https://doi.org/10.1098/rspb.1998.0341

Carsten, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D. (2013). How to fail at species delimitation. *Molecular Ecology*, 22(17), 4369-4383. https://doi.org/10.1111/mec.12413

Carvalho, P., Diniz-Filho, J.A.F., Bini, L.M. (2005). The impact of Felsenstein's "*Phylogenies and the comparative method*" on evolutionary biology. *Scientometrics*, 62(1), 53-66. https://doi.org/10.1007/s11192-005-0003-y

Ceccarelli, M., Santantonio, E., Marmottini, F., Amzallag, G.N., Cionini, P.G. (2006). Chro-

mosome endoreduplication as a factor of salt adaptation in *Sorghum bicolor*. *Protoplasma*, 227(2-4), 113-118. https://doi.org/10.1007/s00709-005-0144-0

Chambers, E.A., Hillis, D.M. (2020). The multispecies coalescent over-splits species in the case of geographically widespread taxa. *Systematic Biology*, 69(1), 184-193. https://doi.org/10.1093/sysbio/syz042

Chao, D-Y., Dilkes, B., Luo, H., Douglas, A., Yakubova, E., Lahner, B., Salt, D.E. (2013). Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science*, 341(6146), 658-659. https://doi.org/10.1126/science.1240561

Chapman, J.A., Kirknes, E.F., Simakov, O., Hampson, S.E., Mitros, T., Weinmaier, T., Rattei, T., Balasubramanian, P.G., Borman, J., Busam, D., Disbennet, K., Pfannkoch, C., Sumin, N., Sutton, G.G., Viswanathan, L.D., Walenz, B., Goodstein, D.M., Hellsten, U., Kawashima, T., Prochnik, S.E., Putnam, N.H., Shu, S., Blumberg, B., Dana, C.E., Gee, L., Kibler, D.F., Law, L., Lindgens, D., Martínez, D.E., Peng, J., Wigge, P.A., Bertulat, B., Guder, C., Nakamura, Y., Ozbek, S., Watanabe, H., Khalturin, K., Hemmrich, G., Franke, A., Augustin, R., Fraune, S., Hayakawa, E., Hayakawa, S., Hirose, M., Hwang, J.S., Ikeo, K., Nishimiya-Fujisawa, C., Ogura, A., Takahashi, T., Steinmetz, P.R.H., Zhang, X., Aufschnaiter, R., Eder, M-K., Gorny, A-K., Salvenmoser, W., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Böttger, A., Tischler, P., Wolf, A., Gojobori, T., Remington, K.A., Strausberg, R.L., Venter, J.C., Technau, U., Hobmayer, B., Bosch, T.C.G., Holstein, T.W., Fujisawa, T., Bode, H.R., David, C.N., Rokhsar, D.S., Steele, R.E. (2010). The dynamic genome of *Hydra*. *Nature*, 464(7288), 592-596. https://doi.org/10.1038/nature08830

Chen, G., Wang, L., Wu, F., Sun, X., Dong, Z., Sluys, R., Yu, F., Wu-wen, Y., Liu, D. (2022). *BMC Zoology*, 7(1), 25. https://doi.org/10.1186/s40850-022-00127-8

Chénais, B., Caruso, A., Hiard, S., Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene*, 509(1), 7-15. http://dx.doi.org/10.1016/j.gene.2012.07.042

Chinone, A., Nodono, H., Matsumoto, M. (2014). Triploid planarian reproduces truly bisexually with euploid gameted produces through a different meiotic system between sex. *Chromosoma*, 123(2), 265-272. https://doi.org/10.1007/s00412-013-0449-2

Civitelli, M., Consentino, P., Capanna, E. (1989). Inter- and intra-individual chromosome variability in *Thamnomys* (*Grammomys*) *gazellae* (Rodentia, Muridae) B-chromosomes and structural heteromorphisms. *Genetica*, 79(2), 93-105. https://doi.org/10.1007/BF00057926

Connallon, T., Clark, A.G. (2010). Gene duplication, gene conversion and the evolution of the Y chromosome. *Genetics*, 186(1), 277-286. https://doi.org/10.1534/genetics.110.116756

Cuevas-Caballé, C., Riutort, M., Álvarez-Presas, M. (2019). Diet assessment of two land planarian species using high-throughput sequencing data. *Scientific Reports*, 9(1), 8679.

https://doi.org/10.1038/s41598-019-44952-3

Dapporto, L., Wolf, H., Strumia, F. (2007). Recent geography determines the distribution of some flying Hymenoptera in the Tuscan Archipelago. *Journal of Zoology*, 272(1), 37-44. https://doi.org/10.1111/mec.12085

Darwin, C. (1839). *Journal of Researches into the Geology and Natural History of the various countries visited by HMS Beagle.* London: Henry Colburn (ed.).

Darwin, C. (1859). *On the Origin of Species by means of Natural Selection, or the Preservation of favoured Races in the Struggle for Life.* London: John Murray (ed.).

Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85, 407-415. https://doi.org/10.1111/j.1095-8312.2005.00503.x

De Coster, T., Masset, H., Tšuiko, O., Catteeuw, M., Zhao, Y., Dierckxsen, N., Larreategui Aparicio, A., Dimitriadou, E., Debrock, S., Peeraer, K., de Ruijter-Villani, M., Smits, K., Van Soom, A., Vermeesch, J.R. (2022). Parental genomes segregate into distinct blastomeres during multipolar zygotic divisions leading to mixoploid and chimeric blastocysts. *Genome Biology*, 23(1), 201-230. https://doi.org/10.1186/s13059-022-02763-2

Deamer, D., Akeson, M., Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518-524. http://dx.doi.org/10.1038/nbt.3423

Degnan, J.H., Rosenberg, N.A. (2006). Discordance of species trees with their most likely gene trees. *PLoS Genetics*, 2(5), 762-768. https://doi.org/10.1371/journal.pgen.0020068

Dejaco, T., Gassner, M., Arthofer, W., Schlick-Steiner, B.C., Steiner, F.M. (2016). Taxonomist's nightmare ... Evolutionist's delight: an integrative approach resolves species limits in jumping bristletails despite widespread hybridization and parthenogenesis. *Systematic Biology*, 65(6), 947-974. https://doi.org/10.1093/sysbio/syw003

Deri, P. (1971). Ricerche cariologiche e di biologia della riproduzione in popolazioni di *Dugesia benazzii* della Corsica. *Atti della Società Toscana di Scienze Naturali Residente in Pisa, Memorie (Serie B)*, 78, 55-74.

Deri, P. (1981). Nuove osservazioni cariologiche e di biología riproduttiva su popolazioni di *Dugesia benazzii* (triclade paludicolo) della Corsica. *Atti della Società Toscana di Scienze Naturali Residente in Pisa, Memorie (Serie B)*, 88, 61-72.

Destouni, A., Vermeesch, J.R. (2017). How can zygotes segregate entire parental genomes into distinct blastomeres? The zygote metaphase revisited. *BioEssays*, 39(4), 1600226. https://doi.org/10.1002/bies.201600226

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), 418-426. https://doi.org/10.1016/j.tig.2014.07.001

Diniz, D., Bertollo, L.A.C. (2006). Intra- and inter-individual chromosome variation in

Hoplerythrinus unitaeniatus (Pisces, Erythrinidae). A population from the Brazilian São Francisco river basin. *Genetics and Molecular Biology*, 29(3), 453-458. https://doi.org/10.1590/S1415-47572006000300009

Doolittle, W.F., Bapteste, E. (2007). Pattern pluralism and the Tree of Life hypothesis. *Proceedings of the National Academy of Sciences*, 104(7), 2043-2049. https://doi.org/10.1073/pnas.0610699104

Dorey, A., Howorka, S. (2024). Nanopore DNA sequencing technologies and their applications towards single-molecule proteomics. *Nature Chemistry*, 16(3), 314-334. https://doi.org/10.1038/s41557-023-01322-x

Dugès, A. (1830). Aperçu du quelques observations nouvelles sur les Planaires et plusienrs genres voisins. *Annales des Sciences Naturelles*, 21, 72-91. http://gallica.bnf.fr/ark:/12148/bpt6k55070078/f4.image.r . Last accessed, April 2024.

Durand, E.Y., Patterson, N., Reich, D., Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28(8), 2239-2252. https://doi.org/10.1093/molbev/msr048

Eagen, K.P. (2018). Principles of chromosome architecture revealed by Hi-C. *Trends in Biochemical Sciences*, 43(6), 469-478. https://doi.org/10.1016/j.tibs.2018.03.006

Eaton, D.A.R., Ree, R.H. (2013). Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, 62(5), 689-706. https://doi.org/10.5061/dryad.bn281

Eaton, D.A.R., Spriggs, E.L., Park, B., Donoghue, M.J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology*, 66(3), 399-412. https://doi.org/10.1093/sysbio/syw092

Edwards, S.V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution*, 63(1), 1-19. https://doi.org/10.1111/j.1558-5646.2008.00549.x

Edwards, S.V., Arctander, P., Wilson, A.C. (1991). Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proceedings of the Royal Society B*, 243(1307), 99-107. https://doi.org/10.1098/rspb.1991.0017

Evans, J.D., Brown, S.J., Hackett, K.J., Robinson, G., Richards, S., Lawson, D., Elsik, C., Coddington, J., Edwards, O., Emrich, S., Gabaldon, T., Goldsmith, M., Hanes, G., Misof, B., Muñoz-Torres, M., Niehuis, O., Papanicolaou, A., Pfrender, M., Poelchau, M., Purcell-Miramontes, M., Robertson, H.M., Ryder, O., Tagu, D., Torres, T., Zbodnov, E., Zhang, G., Zhou, X. (2013). The i5K initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, 104(5), 595-600. https://doi.org/10.1093/jhered/est050

Faria, R., Navarro, A. (2010). Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends in Ecology and Evolution*, 25(11), 660-669. http://dx.doi.

org/10.1016/j.tree.2010.07.008

Farias, I.P., Orti, G., Sampaio, I., Schneider, H., Meyer, A. (2001). The cytochrome *b* gene as a phylogenetic marker: the limits of resolution for analysing relationships among cichlid fishes. *Journal of Molecular Evolution*, 53(2), 83-109. https://doi.org/10.1007/s002390010197

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17(6), 368-376. https://doi.org/10.1007/bf01734359

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125(1), 1-15. https://doi.org/10.1086/284325

Flouri, T., Jiao, X., Rannala, B., Yang, Z. (2018). Species tree inference with BPP using genomic sequences and multispecies coalescent. *Molecular Biology and Evolution*, 35(10), 2585-2593. https://doi.org/10.1093/molbev/msy147

Frajman, B., Záveská, E., Gamisch, A., Moser, T., The STEPPE Consortium, Schönswetter, P. (2019). Integrating phylogenomics, phylogenetics, morphometrics, relative genome size and ecological niche modelling disentangles the diversification of Eurasian *Euphorbia seguieriana* s. l. (Euphorbiaceae). *Molecular Phylogenetics and Evolution*, 134, 238-252. https://doi.org/10.1016/j.ympev.2018.10.046

Frankham, R., Ballou, J.D., Dudash, M.R., Eldridge, M.D.B., Fenster, C.B., Lacy, R.C., Mendelson III, J.R., Porton, I.J., Ralls, K., Ryder, O.A. (2012). Implications of different species concepts for conserving biodiversity. *Biological Conservation*, 153, 25-31. http://dx.doi.org/10.1016/j.biocon.2012.04.034

Freedman, A.H., Gaspar, J.M., Sackton, T.B. (2020). Short paired-end reads trump long single-end reads for expression analysis. *BMC Bioinformatics*, 21, 149-160. https://doir.org/10.1186/s12859-020-3484-z

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, 27(9), 480-488. https://doi.org/10.1016/j.tree.2012.04.012

Gatesy, J., Springer, M.S. (2014). Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution*, 80, 231-266. https://doi.org/10.1016/j.ympev.2014.08.013

Giani, A.M., Gallo, G.R., Gianfranceschi, L., Formenti, G. (2020). Long walk to genomics: history and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, 18, 9-19. https://doi.org/10.1016/j.csbj.2019.11.002

Gillespie, J.H. (2013). Application of stable isotope analysis to study temporal changes in foraging ecology in a highly endangered amphibian. *PLoS ONE*, 8(1), e53041. https://doi.org/10.1371/journal.pone.0053041

Ginet, R., Puglisi, R. (1964). Écologie de *Fonticola notadena* de Beauchamp (Turbellarié, Triclade) dans la grotte de La Balme (Isère, France); survie en période de sécheresse. *International Journal of Speleology*, 1, 203-216. http://dx.doi.org/10.5038/1827-806X.1.1.16

González-Estévez, C. (2008). Autophagy in freshwater planarians. *Methods in Enzymology*, 451(8), 439-465. http://dx.doi.org/10.1016/S0076-6879(08)03227-8

González-Estévez, C., Saló, E. (2010). Autophagy and apoptosis in planarians. *Apoptosis*, 15(3), 279-292. https://doi.org/10.1007/s10495-009-0445-4

Grant, L.J. (2016). *Biodiversity, biogeography and phylogeny of Australian freshwater triclads (Platyhelminthes: Tricladida: Continenticola)*. Doctoral Thesis, James Cook University. https://researchonline.jcu.edu.au/51512/1/51512-grant-2016-thesis.pdf . Last accessed, April 2024.

Graur, D., Li, W-H. (2000). *Fundamentals of Molecular Evolution*. Sinauer Associates, Inc., Sunderland, MA.

Gregory, T.R. (2024). Animal Genome Size Database, 2024. http://www.genomesize.com . Last accessed in March, 2024

Gregory, T.R., Hebert, P.D.N., Kolasa, J. (2000). Evolutionary implications of the relationship between genome size and body size in flatworms and copepods. *Heredity*, 84(2), 201-208. https://doi.org/10.1046/j.1365-2540.2000.00661.x

Grohme, M.A., Schloissnig, S., Rozanski, A., Pippel, M., Young, G.R., Winkler, S., Brandl, H., Henry, I., Dahl, A., Powell, S., Hiller, M., Myers, E., Rink, J.C. (2018). The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature*, 554(7690), 56-61. https://doi.org/10.1038/nature25473

Grummer, J.A., Bryson Jr., R.W., Reeder, T.W. (2014). Species delimitation using bayes factors: simulations and applications to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology*, 63(2), 119-133. https://doi.org/10.1093/sysbio/syt069

Guo, L., Bloom, J.S., Dols-Serrate, D., Boocock, J., Ben-David, E., Schubert, O.T., Kozuma, K., Ho, K., Warda, E., Chui, C., Wei, Y., Leighton, D., Lemus Vergara, T., Riutort, M., Alvarado A.S., Kruglyak, L. (2022). Island-specific evolution of a sex-primed autosome in a sexual planarian. *Nature*, 606(7913), 329-334. https://doi.org/10.1038/s41586-022-04757-3

Hallez, P. (1894). *Catalogue des Rhabdocoelides, Triclades et Polyclades du Nord de la France*, 2nd edn. Lille, France: L. Daniel (ed.), 239 p + 2 pls.

Harrath, A.H., Mansour, L., Lagnika, M., Sluys, R., Boutin, C., Alwasel, S., Poch, A., Riutort, M. (2016). A molecular analysis of the phylogenetic position of the suborder Cavernicola within the Tricladida (Platyhelminthes), with the description of a new species of stygobiont flatworm from Benin. *Zoological Journal of the Linnean Society*, 178(3), 482-

491. https://doi.org/10.1111/zoj.12430

Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., Brumfield, R.T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, 65(5), 910-924. https://doi.org/10.1093/sysbio/syw036

Hay, D.H., Ball, I.R. (1979). Contributions to the biology of freshwater planarians (Turbellaria) from the Victorian Alps, Australia. *Hydrobiologia*, 62, 137-164. https://doi.org/10.1007/BF00037507

Heather, J.M., Chain, B. (2016). The sequences of sequencers: the history of sequencing DNA. *Genomics*, 107(1), 1-8. http://dx.doi.org/10.1016/j.ygeno.2015.11.003

Hedrick, P.W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18), 4606-4618. https://doi.org/10.1111/mec.12415

Hellmuth, M., Wieseke, N., Lechner, M., Lenhof, H-P., Middendorf, M., Stadler, P.F. (2015). Phylogenomics with paralogs. *Proceedings of the National Academy of Sciences of the United States of America*, 112(7), 2058-2063. https://doi.org/10.1073/pnas.1412770112

Hermansen, J.S., Sæther, S.A., Elgvin, T.O., Borge, T., Hjelle, E., Sætre, G-P. (2011). Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology*, 20(18), 3812-3822. https://doi.org/10.1111/j.1365-294X.2011.05183.x

Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, 6(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Huelsenbeck, J.P. (1998). Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Systematic Biology*, 47(3), 519-537. https://doi.org/10.2307/2585257

Huelsenbeck, J.P., Rannala, B. (1997). Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science*, 276(5310), 227-232. https://doi.org/10.1126/science.276.5310.227

Huson, D.H., Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), 254-267. https://doi.org/10.1093/molbev/msj030

Hyman, L.H. (1939). North American triclad Turbellaria. IX. The priority of *Dugesia* Girard, 1850 over *Euplanaria* Hesse, 1897 with notes on American species of *Dugesia*. *Transactions of the American Microscopical Society*, 58(3), 264-273. https://doi.org/10.2307/3222879

Hyman, L.H. (1951). *The invertebrates: Platyhelminthes and Rhynchocoela. The Acoelomate billateria*, Vol. II. New York: McGraw-Hill.

Ivanković, M., Brand, J.N., Pandolfini, L., Brown, T., Pippel, M., Rozanski, A., Schubert, T.,

Grohme, M.A., Winkler, S., Robledillo, L., Zhang, M., Codino, A., Gustincich, S., Vila-Farré, M., Zhang, S., Papantonis, A., Marques, A., Rink, J.C. (2023). A comparative analysis of planarian genomes reveals regulatory conservation in the face of rapid structural divergence. *bioRxiv*, https://doi.org/10.1101/2023.12.22.572568

Jackson, C., Christie, N., Reynolds, M., Marais, C., Tii-kuzu, Y., Caballero, M., Kampman, T., Visser, E.A., Naidoo, S., Kain, D., Whetten, R.W., Isik, F., Wegrzyn, J., Hodge, G., Acosta, J.J., Myburg, A.A. (2022). A genome-wide SNP genotyping resource for tropical pine tree species. *Molecular Ecology Resources*, 22(2), 695-710. https://doi.org/10.1111/1755-0998.13484

Janovy, J. (1963). Monsterism in *Dileptus* (Ciliata) fed on planarians (*Dugesia tigrina*). *The Journal of Protozoology*, 10(4), 428-430. https://doi.org/10.1111/j.1550-7408.1963.tb01701.x

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Hounde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M.S., Zavidovych, V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W.C., Ray, D., Green, R.E., Bruford, M., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E.P., Bertelsen, M.F., Sheldon, F., Brumfield, R.T., Mello, C., Lovell, P.V., Wirthlin, M., Samaniego, J.A., Velazquez, A.M.V., Alfaro-Núñez, A., Campos, P.F., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D., Zhou, Q., Perelman, P., Driskell, A.C., Ruby, G., Shapiro, B., Xiong, Z., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F.E., Braun, M., Fjeldsa, J., Orlando, L., Barker, K., Jønsson, K.A., Johnson, W., Koepfli, K-P., O'Brien, S., Haussler, D., Ryder, O.A., Rahbek, C., Willerslev, E., Graves, G.R., Glenn, T.C., McCormack, J., Burt, D., Ellegren, H., Alström, P., Edwards, S., Stamatakis, A., Mindell, D.P., Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T.P., Zhang, G. (2014). Whole genome analyses resolve the early branches in the Tree of Life of modern birds. *Science*, 346(6215), 1320-1331. https://doi.org/10.1126/science.1253451

Jukes, T.H., Cantor, C.R. (1969). Evolution of protein molecules, *Mammalian Protein Metabolism*, vol. III, 21-132. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., Flouri, T. (2017). Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics*, 33(11), 1630-1638. https://doi.org/10.1093/bioinformatics/btx025

Kawakatsu, M. (1968). On the origin and phylogeny of Turbellarians: Suborder Paludicola. *Japanese Society of Systematic Zoology*, circ. nos. 38-41, 11-22. [in Japanese, with English explanation of figures].

Kawakatsu, M., Oki, I., Tamura, S., Sugino, H. (1976). Studies on the morphology, kary-

ology and taxonomy of the Japanese freshwater planarian *Dugesia japonica* Ichikawa et Kawakatsu, with a description of a new subspecies, *Dugesia japonica ryukyuensis* subspec. nov. *Bulletin of Fuji Women's College,* no. 14, ser. II, 81-126.

Keeling, P.J., Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, 9(8), 605-618. https://doi.org/10.1038/nrg2386

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2), 111-120. https://doi.org/10.1007/BF01731581

Kingman, J.F.C. (1982). The coalescent. *Stochastic Processes and their Applications*, 13(3), 235-248. https://doi.org/10.1016/0304-4149(82)90011-4

Knowles, L.L., Carstens, B.C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, 56(6), 887-895. https://doi.org/10.1080/10635150701701091

Kobayashi, K., Ishizu, H., Arioka, S., Padilla Cabrera, J., Hoshi, M., Matsumoto, M. (2008). Production of diploid and triploid offspring by inbreeding of the triploid planarian *Dugesia ryukyuensis*. *Chromosoma*, 117(3), 289-296. https://doi.org/10.1007/s00412-008-0148-6

Koch, C., Flecks, M., Venegas, P.J., Bialke, P., Valverde, S., Rödder, D. (2016). Applying n-dimensional hypervolumes for species delimitation: unexpected molecular, morphological, and ecological diversity in the Leaf-Toed Gecko *Phyllodactylus reissii* Peters, 1862 (Squamata: Phyllodactylidae) from northern Peru. *Zootaxa*, 4161(1), 41-80. https://doi.org/10.11646/zootaxa.4161.1.2

Koonin, E.V., Makarova, K.S., Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology*, 55, 709-742. https://doi.org/10.1146/annurev.micro.55.1.709

Kreuzinger-Janik, B., Kruscha, S., Majdi, N., Traunspurger, W. (2018). Flatworms like it round: nematode consumption by *Planaria torva* (Müller, 1774) and *Polycelis tenuis* (Ijima, 1884). *Hydrobiologia*, 819(1), 231-242. https://doi.org/10.1007/s10750-018-3642-8

Krois, N.R., Cherukuri, A., Puttagunta, N., Neiman, M. (2013). Higher rate of tissue regeneration in polyploidy asexual versus diploid asexual freshwater snails. *Biology Letters*, 9(4). https://doi.org/10.1098/rsbl.2013.0422

Lamichhaney, S., Han, F., Webster, M.T., Andersson, L., Grant, B.R., Grant, P.R. (2018). Rapid hybrid speciation in Darwin's finches. *Science*, 359(6372), 224-228. https://doi.org/10.1126/science.359.6372.172-d

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B. (2017). PartitionFinder2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3), 772-773. https://doi.org/10.1093/molbev/msw260

Lang, A. (1884). *Die polycladen (seeplanarien) des colfes von Neapel und der angrenzenden*

meeresabschnitte. Leipzig: W. Engelmann (ed.). http://www.biodiversitylibrary.org/bibliography/10545#/summary. Last accessed, April 2024.

Laumer, C.E., Giribet, G. (2014). Inclusive taxon sampling suggests a single stepwise origin of ectolecithality in Platyhelminthes. *Biological Journal of the Linnean Society*, 111(3), 570-588. https://doi.org/10.1111/bij.12236

Laumer, C.E., Giribet, G. (2017). Phylogenetic relationships within Adiaphanida (phylum Platyhelminthes) and the status of the crustacean-parasitic genus *Genostoma*. *Invertebrate Biology*, 136(2), 184-198. https://doi.org/10.1111/ivb.12169

Lázaro, E., Harrath, A.H., Stocchino, G.A., Pala, M., Baguñà, J., Riutort, M. (2011). *Schmidtea mediterranea* phylogeography: an old species surviving on a few Mediterranean islands? *BMC Evolutionary Biology*, 11, 274. https://doi.org/10.1186/1471-2148-11-274

Lázaro, E., Riutort, M. (2013). *Dugesia sicula* (Platyhelminthes, Tricladida): the colonizing success of an asexual Planarian. *BMC Evolutionary Biology*, 13(1), 268. https://doi.org/10.1186/1471-2148-13-268

Lázaro, E., Sluys, R., Pala, M., Stocchino, G.A., Baguñà, J., Riutort, M. (2009). Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution*, 52(3), 835-845. https://doi.org/10.1016/j.ympev.2009.04.022

Lee, M.S.Y., Palci, A. (2015). Morphological phylogenetics in the genomic age. *Current Biology*, 25(19), R922-R929. http://dx.doi.org/10.1016/j.cub.2015.07.009

Lepori, N.G. (1948). Mutazione esaploide in una planaria della Sardegna appartenente al gen. *Dugesia*. *Caryologia*, 1, 92-101.

Lepori, N.G. (1951). Sulle caratteristiche morfologiche e sulla posizione sistematica della planaria di Sardegna e Corsica già ascritta a *Dugesia gonocephala* (Dugès). *Atti della Società Toscana di Scienze Naturali*, 58, 28-47.

Lepori, N.G., Pala, M. (1982). Fissioning in planarians. 1. Karyological analysis of fissiparous strains of *Dugesia gonocephala* s.l. (Turbellaria, Tricladida) collected in Sardinia, in order to determine the factors responsible for fissioning. *Monitore Zoologico Italiano*, 16, 105-131.

Leria, L. (2019). *Understanding the processes that shape the genetic diversity of freshwater planarians (Platyhelminthes, Tricladida, Dugesiidae)*. Doctoral Thesis, Universitat de Barcelona. http://hdl.handle.net/2445/147729

Leria, L., Riutort, M., Romero, R., Ferrer, X., Vila-Farré, M. (2022). Microplate tectonics and environmental factors as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography*, 49(6), 1124-1136. https://doi.org/10.1111/jbi.14373

Leria, L., Sluys, R., Riutort, M. (2018). Diversification and biogeographic history of the

Western Palearctic freshwater flatworm genus *Schmidtea* (Tricladida: Dugesiidae), with a redescription of *Schmidtea nova*. *Journal of Zoological Systematics and Evolutionary Research*, 56(3), 335-351. https://doi.org/10.1111/jzs.12214

Leria, L., Vila-Farré, M., Álvarez-Presas, M., Sánchez-Gracia, A., Rozas, J., Sluys, R., Riutort, M. (2020). Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhelminthes, Tricladida): extreme intraindividual genetic diversity, morphological stasis, and karyological variability. *Molecular Phylogenetics and Evolution*, 143, 106496. https://doi.org/10.1016/j.ympev.2019.05.010

Leria, L., Vila-Farré, M., Solà, E., Riutort, M. (2019). Outstanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. *BMC Evolutionary Biology*, 19(1), 130-149. https://doi.org/10.1186/s12862-019-1440-1

Lewis, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., Goldstein, M.M., Grigoriev, I.V., Hackett, K.J., Haussler, D., Jarvis, E.D., Johnson, W.E., Patrinos, A., Richards, S., Castilla-Rubio, J.C., van Sluys, M-A., Soltis, P.S., Xu, X., Yang, H., Zhang, G. (2018). Earth BioGenome Project: sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, 115(17), 4325-4333. https://doi.org/10.1073/pnas.1720115115

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950), 289-293. https://doi.org/10.1126/science.1181369

Liu, L., Xi, Z., Wu, S., David, C.C., Edwards, S.V. (2015). Estimating phylogenetic trees from genome-scale data. *Annals of the New York Academy of Sciences*, 1360, 36-53. https://doi.org/10.1111/nyas.12747

Liu, L., Yu, L., Edwards, S.V. (2010). A maximum pseudolikelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology*, 10(1), 302-320. https://doi.org/10.1186/1471-2148-10-302

Lock, M.A., Reynoldson, T.B. (1976). The role of interspecific competition in the distribution of two stream dwelling triclads, *Crenobia alpina* (Dana) and *Polycelis felina* (Dalyell), in North Wales. *Journal of Animal Ecology*, 45(2), 581-592. https://doi.org/10.2307/3892

Lu, L., Cox, C.J., Mathews, S., Wang, W., Wen, J., Chen, Z. (2018). Optimal data partitioning, multispecies coalescent and Bayesian concordance analyses resolve early divergences of the grape family (Vitaceae). *Cladistics*, 34, 57-77. https://doi.org/10.1111/cla.12191

Lucek, K., Giménez, M.D., Joron, M., Rafajlović, M., Searle, J.B., Walden, N., Westram, A.M., Faria, R. (2023). The impact of chromosomal rearrangements in speciation: from

micro- to macroevolution. *Cold Spring Harbor Perspective in Biology*, 15(11), a041447. https://doi.org/10.1101/cshperspect.a041447

Lucek, K., Gompert, Z., Nosil, P. (2019). The role of structural genomic variants in population differentiation and ecotype formation in *Timema cristinae* walking sticks. *Molecular Ecology*, 28(6), 1224-1237. https://doi.org/10.1111/mec.15016

Mace, G.M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1444), 711-719. https://doi.org/10.1098/rstb.2003.1454

Maciver, S.K. (2016). Asexual amoebae escape Muller's ratchet through polyploidy. *Trends in Parasitology*, 32(11), 855-862. http://dx.doi.org/10.1016/j.pt.2016.08.006

Maddison, W.P. (1997). Gene trees in species trees. *Systematic Biology*, 46(3), 523-536. https://doi.org/10.1093/sysbio/46.3.523

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, 20(5), 229-237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature*, 446(7133), 279-283. https://doi.org/10.1038/nature05706

Mallet, J., Besansky, N., Hahn, M.W. (2016). How reticulated are species? *BioEssays*, 38(2), 140-149. https://doi.org/10.1002/bies.201500149

Mammola, S., Arnedo, M.A., Pantini, P., Piano, E., Chiappetta, N., Isaia, M. (2018). Ecological speciation in darkness? Spatial niche partitioning in sibling subterranean spiders (Araneae: Linyphiidae; *Troglohyphantes*). *Invertebrate Systematics*, 32(5), 1069-1082. https://doi.org/10.1071/IS17090

Manenti, R., Barzaghi, B., Della Toffola, R., Lapadula, S. (2022). Resistance of groundwater invertebrates to droughts: two new cases in planarians and isopods. *Ecosphere*, 13(8), e4214. https://doi.org/10.1002/ecs2.4214

Manning, J.T., Dickson, D.P.E. (1986). Asexual reproduction, polyploidy and optimal mutation rates. *Journal of Theoretical Biology*, 118(4), 485-489. https://doi.org/10.1016/S0022-5193(86)80166-7

Masset, H., Tšuiko, O., Vermeesch, J.R. (2021). Genome-wide abnormalitites in embryos: origins and clinical consequences. *Parental Diagnosis*, 41(5), 554-563. https://doi.org/10.1002/pd.5895

Maxam, A.M., Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560-564. https://doi.org/10.1073/pnas.74.2.560

Mendes, J., Harris, D.J., Carranza, S., Salvi, D. (2017). Biogeographical crossroad across the Pillars of Hercules: evolutionary history of *Psammodromus* lizards in space and time. *Journal of Biogeography*, 44(12), 2877-2890. https://doi.org/10.1111/jbi.13100

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, 17(2), 240-248. https://doi.org/10.1101/gr.5681207

Mirarab, S., Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31(12), i44-i52. https://doi.org/10.1093/BIOINFORMATICS/BTV234

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szuesich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K.F., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X. (2014). Phylogenomics resolves the timing and patter of insect evolution. *Science*, 346(6210), 763-767. https:/doi.org/10.1126/science.1257570

Morita, M., Best, J.B. (1974). Electron microscopic studies of planarian regeneration II. Changes in epidermis during regeneration. *Journal of Experimental Zoology*, 187(3), 345-373. https://doi.org/https://doi.org/10.1002/jez.1401870305

Navarro, A., Barton, N.H. (2003). Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution*, 57(3), 447-459. https://doi.org/10.1111/j.0014-3820.2003.tb01537.x

Newmark, P.A., Alvarado, A.S. (2000). Not your father's planarian: a classic model enters the era of functional genomics. *Nature Reviews Genetics*, 3(3), 210-219. https://doi.org/10.1038/nrg759

Newton, L.G., Starrett, J., Hendrixson, B.E., Dekarabetian, S., Bond, J.E. (2020). Integrative species delimitation reveals cryptic diversity in southern Appalachian *Antrodiaetus unicolor* (Araneae: Antrodiaetidae) species complex. *Molecular Ecology*, 29(12), 2269-2287. https://doi.org/10.1111/mec.15483

Nichols, R. (2001). Gene trees and species trees are not the same. *Trends in Ecology and Evolution*, 16(7), 358-364. https://doi.org/10.1016/s0169-5347(01)02203-0

Nicholson, J.M., Cimini, D. (2013). Cancer karyotypes: survival of the fittest. *Frontiers in Oncology*, 3. https://doi.org/10.3389/fonc.2013.00148

Nishimura, O., Hosoda, K., Kawaguchi, W., Yazawa, S., Hayashi, T., Inoue, T., Umesono, Y., Agata, K. (2015). Unusually large number of mutations in asexually reproducing clonal planarian *Dugesia japonica*. *PLoS ONE*, 10(11), 1-23. https://doi.org/10.1371/journal.pone.0143525

Noguerales, V., Ortego, J. (2022). Genomic evidence of speciation by fusion in a recent radiation of grasshoppers. *Evolution*, 76(11), 2618-2633. https://doi.org/10.1111/evo.14508

Norén, M., Jondelius, U. (2002). The phylogenetic position of the Prolecithophora (Rhabditophora, 'Platyhelminthes'). *Zoologica Scripta*, 31(4), 403-414. https://doi.org/10.1046/j.1463-6409.2002.00082.x

Nyrén, P. (1987). Enzymatic method for continuous monitoring of DNA polymerase activity. *Analytical Biochemistry*, 167(2), 235-238. https://doi.org/10.1016/0003-2697(87)90158-8

Nyrén, P., Lundin, A. (1985). Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Analytical Biochemistry*, 151(2), 504-509. https://doi.org/10.1016/0003-2697(85)90211-8

Opatova, V., Bond, J.E., Arnedo, M.A. (2016). Uncovering the role of the Western Mediterranean tectonics in shaping the diversity and distribution of the trap-door spider genus *Ummidia* (Araneae, Ctenizidae). *Journal of Biogeography*, 43(10), 1955-1966. https://doi.org/10.1111/jbi.12838

Orr, B., Compton, D.A. (2013). A double-edged sword: how oncogenes and tumor suppressor genes can contribute to chromosomal instability. *Frontiers in Oncology*, 3. https://doi.org/10.3389/fonc.2013.00164

Pala, M., Casu, S., Lepori, N.G. (1982). Stabilized natural interspecific hybrid population of the fresh water planarians *Dugesia gonocephala* s.l. (Turbellaria, Tricladida). *Caryologia*, 35(2), 247-256. https://doi.org/10.1080/00087114.1982.10796929

Pala, M., Casu, S., Vacca, R.A. (1981). *Dugesia hepta*, nuova specie di Planaria di acqua dolce di Sardegna appartenente alla superspecie *Dugesia gonocephala* (Dugès) (Turbellaria, Tricladida). *Bolletino della Società Sarda di Scienze Naturali*, 20, 97-107.

Pala, M., Stocchino, G.A., Corso, G., Casu, S. (2000). A new species of freshwater planarian belonging to the genus *Dugesia* (Platyhelminthes, Tricladida) from Sardinia. *Italian Journal of Zoology*, 67(4), 379-383. https://doi.org/10.1080/11250000009356343

Palkopoulou, E., Lipson, M., Mallick, S., Nielsen, S., Rohland, N., Baleka, S., Karpinski, E., Ivancevic, A.M., To, T-H., Kortschak, R.D., Raison, J.M., Qu, Z., Chin, T-J., Alt, K.W., Claesson, S., Dalén, L., MacPhee, R.D.E., Meller, H., Roca, A.L, Ryder, O.A., Heiman, D., Young, S., Breen, M., Williams, C., Aken, B.L., Ruffier, M., Karlsson, E., Johnson, J., Di Palma, F., Alfoldi, J., Adelson, D.L., Mailund, T., Munch, K., Lindblad-Toh, K., Hofreiter, M., Poinar, H., Reich, D. (2018). A comprehensive genomic history of extinct and living elephants. *Proceedings of the National Academy of Sciences of the United States of America*,

115(11), E2566-E2574.https://doi.org/10.1073/pnas.1720554115

Pamilo, P., Nei. M. (1988). Relationship between gene trees and species trees. *Molecular Biology and Evolution*, 5(5), 568-583. https://doi.org/10.1093/oxfordjournals.molbev.a040517

Pandit, M.K., Pocock, M.J.O., Kunin, W.E. (2011). Ploidy influences rarity and invasiveness in plants. *Journal of Ecology*, 99(5), 1108-1115. https://doi.org/10.1111/j.1365-2745.2011.01838.x

Parisod, C., Broennimann, O. (2016). Towards unified hypotheses of the impact of polyploidy on ecological niches. *New Phytologist*, 212(3), 540-542.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhang, Y., Genschoreck, T., Webster, T., Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3), 1065-1093.

Pavlíček, T., Cohen, T., Yadav, S., Glasstetter, M., Král, P., Pearlson, O. (2016). Aneuploidy occurrence in Oligochaeta. *Ecology and Evolutionary Biology*, 1(3), 57-63. https://doi.org/10.11648/j.eeb.20160103.13

Payseur, B.A., Riesenberg, L.H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, 25(11), 2337-2360. https://doi.org/10.1111/mec.13557

Pedersen, K.J. (1963). Slime-secreting cells of planarians. *Annals of the New York Academy of Sciences*, 106(2), 424-443.

Penny, D. (1992). The comparative method in evolutionary biology. *Journal of Classification*, 9, 169-172. https://doi.org/10.1007/bf02618482

Peretti, E., Cecchin, C., Fusco, G., Gregnanin, L., Kos, I., Bonato, L. (2022). Shedding light on species boundaries in small endogeic animals through an integrative approach: species delimitation in the centipede *Clinopodes carinthiacus* (Chilopoda: Geophilidae) in the south-eastern Alps. *Zoological Journal of the Linnean Society*, 196(2), 902-923.

Pérez-Losada, M., Breinholt, J.W., Porto, P.G., Aira, M., Domínguez, J. (2011). An earthworm riddle: systematics and phylogeography of the Spanish lumbricid *Postandrilus*. *PLoS ONE*, 6(11), e28153. https://doi.org/10.1371/journal.pone.0028153

Pfenninger, M., Véla, E., Jesse, R., Arantzazu, M., Liberto, F., Magnin, F., Martínez-Ortí, A. (2010). Temporal speciation pattern in the western Mediterranean genus *Tudorella* P. Fischer, 1885 (Gastropoda, Pomatiidae) supports the Tyrrhenian vicariance hypothesis. *Molecular Phylogenetics and Evolution*, 54(2), 427-436. https://doi.org/10.1016/j.ympev.2009.09.024

Pienta, K.J., Hammarlund, E.U., Austin, R.H., Axelrod, R., Brown, J.S., Amend, S.R. (2022). Cancer cells employs an evolutionarily conserved polyploidization program to resist therapy. *Seminars in Cancer Biology*, 81, 145-159. https://doi.org/10.1016/j.semcancer.2020.11.016

Pigliucci, M. (2009). An extended synthesis for evolutionary biology. *Annals of the New York Academy of Sciences*, 1168, 218-228. https://doi.org/10.1111/j.1749-6632.2009.04578.x

Pigliucci, M., Müller, G.B. (Eds.) (2010). *Evolution: The Extended Synthesis*. MIT Press, Cambridge, MA.

Plass, M., Solana, J., Alexander Wolf, F., Ayoub, S., Misios, A., Glažar, P., Obermayer, B., Theis, F.J., Kocks, C., Rajewsky, N. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science*, 360(6391). https://doi.org/10.1126/science.aaq1723

Pongratz, N., Sharbel, T.F., Beukeboom, L.W., Michiels, N.K. (1998). Allozyme variability in sexual and parthenogenetic freshwater planarians: evidence for polyphyletic origin of parthenogenetic lineages through hybridization with coexisting sexuals. *Heredity*, 81(1), 38-47. https://doi.org/10.1046/j.1365-2540.1998.00359.x

Pongratz, N., Storhas, M., Carranza, S., Michiels, N.K. (2003). Phylogeography of competing sexual and parthenogenetic forms of a freshwater flatworm: patterns and explanations. *BMC Evolutionary Biology*, 3(1), 23. https://doi.org/10.1186/1471-2148-3-23

Pons, J., Barraclough, T., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D., Vogler, A.P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55(4), 595-609. https://doi.org/10.1080/10635150600852011

Prasniski, M.E.T., Leal-Zanchet, A.M. (2009). Predatory behaviour of the land flatworm *Notogynaphallia abundans* (Platyhelminthes: Tricladida). *Zoologia*, 26(4). https://doi.org/10.1590/S1984-46702009005000011

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, 21(8), 1864-1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W., Haussler, D., Rokhsar, D.S., Green, R.E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, 26, 342-350. https://doi.org/10.1101/gr.193474.115.Freely

de Queiroz, K. (1998). The general lineage species concept of species, species criteria, and the process of speciation. A conceptual unification and terminological recommendations. In: Howard, D.J., Berlocher, S. (eds.), *Endless Forms: Species and Speciation*. New York: Oxford University Press, 57-75.

de Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, 56(6), 879-886. https://doi.org/10.1080/10635150701701083

Rannala, B., Yang, Z. (1996). Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution*, 43(3), 304-311.

https://doi.org/10.1007/BF02338839

Rannala, B., Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, 164(4), 1645-1656. https://doi.org/10.1093/genetics/164.4.1645

Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M-R., Elmer, K.R., Vences, M., Steinfarz, S. (2019). Phylogeny and species delimitation of near Eastern *Neurergus* newts (Salamandridae) based on genome-wide RADseq data analysis. *Molecular Phylogenetics and Evolution*, 133, 189-197. https://doi.org/10.1016/j.ympev.2019.01.003

Ratnasingham, S., Hebert, P.D.N. (2007). BOLD: the Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, 7(3), 355-364. https://doi.org/10.1111/J.1471-8286.2007.01678.X

Reddien, P.W., Alvarado, A.S. (2004). Fundamentals of planarian regeneration. *Annual Review of Cell and Developmental Biology*, 20(1), 725-757. https://doi.org/10.1146/annurev.cellbio.20.010403.095114

Reynoldson, T.B. (1964). Evidence for intraspecific competition in field populations of triclads. *Journal of Animal Ecology*, 33, 187-201. https://doi.org/10.2307/2439

Reynoldons, T.B. (1983). The population biology of turbellaria with special reference to the freshwater triclads of the British Isles. *Advances in Ecology Research*, 13, 235-326. https://doi.org/10.1016/S0065-2504(08)60110-5

Reynoldson, T.B., Davies, R.W. (1970). Food niche and co-existence in lake-dwelling triclads. *Journal of Animal Ecology*, 39(3), 599-617. https://doi.org/10.2307/2857

Ribas, M. (1990). *Cariologia, sistemàtica i biogeografía de les planàries d'aigües dolces als Països Catalans*. Doctoral Thesis, Universitat de Barcelona.

Richardson, A.O., Palmer, J.D. (2007). Horizontal gene transfer in plants. *Journal of Experimental Botany*, 58(1), 1-9. https://doi.org/10.1093/jxb/erl148

Rink, J.C. (2013). Stem cell systems and regeneration in planaria. *Development Genes and Evolution*, 223(1-2), 67-84. https://doi.org/10.1007/s00427-012-0426-4

Roach, J.C., Boysen, C., Wang, K., Hood, L. (1995). Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics*, 26(2), 345-353. https://doi.org/10.1016/0888-7543(95)80219-C

Roch, S. (2010). Towards extracting all phylogenetic information from matrices of evolutionary distances. *Science*, 327(5971), 1376-1379. https://doi.org/10.1126/science.1182300

Rodríguez, F., Oliver, J.L., Marín, A., Medina, J.R. (1990). The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology*, 142(4), 485-501. https://doi.org/10.1016/S0022-5193(05)80104-3

Rokas, A., Williams, B.L., King, N., Carroll, S.B. (2003). Genome-scale approaches to re-

solving incongruence in molecular phylogenies. *Nature*, 425(6960), 798-804. https://doi.org/10.1038/nature02053

Ronaghi, M., Karamohamed, S., Petterson, B., Uhlén, M., Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry*, 242(1), 84-89. https://doi.org/10.1006/abio.1996.0432

Rosenbaum, G., Lister, G.S., Duboz, C. (2002). Reconstruction of the tectonic evolution of the western Mediterranean since the Oligocene. *Journal of the Virtual Explorer*, 8, 107-130. https://doi.org/10.3809/jvirtex.2002.00053

Rubin, B.E.R., Ree, R.H., Moreau, C.S. (2012). Inferring phylogenies from RAD sequence data. *PLoS ONE*, 7(4), e33394. https://doi.org/10.1371/journal.pone.0033394

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, 239(29), 487-491. https://doi.org/10.1126/science.2448875

Salvi, D., Pinho, C., Harris James, D. (2017). Digging up the roots of an insular hotspot of genetic diversity: decoupled mito-nuclear histories in the evolution of the Corsican-Sardinian endemic lizard *Podarcis tiliguerta*. *BMC Evolutionary Biology*, 17(1), 63. https://doi.org/10.1186/s12862-017-0899-x

Sanderson, N.D., Kapel, N., Rodger, G., Webster, H., Lipworth, S., Street, T.L., Peto, T., Crook, D., Stoesser, N. (2023). Comparison of R9.4.1/Kit10 and R10/Kit12 Oxford Nanopore flowcells and chemistries in bacterial genome reconstruction. *Microbial Genomics*, 9, 000910. https://doi.org/10.1099/mgen.0.000910

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchinson III, C.A., Slocombe, P.M., Smith, M. (1977). Nucleotide sequence of bacteriophage φX174 DNA. Nature, 265(5596), 687-695. https://doi.org/10.1038/265687a0

Sanger, F., Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), 441-448. https://doi.org/10.1016/0022-2836(75)90213-2

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, 507(7492), 354-357. https://doi.org/10.1038/nature12961

Schadt, E.E., Turner, S., Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, 19(R2), 227-240. https://doi.org/10.1093/hmg/ddq416

Schockaert, E.R., Hooge, M., Sluys, R., Schilling, S., Tyler, S., Artois, T. (2008). Global diversity of free living flatworms (Platyhelminthes, "Turbellaria") in freshwater. *Hydrobiologia*, 595, 41-48. https://doi.org/10.1007/978-1-4020-8259-7_5

Scotland, R.W., Olmstead, R.G., Bennet, J.R. Phylogeny reconstruction: the role of morphology. *Systematic Biology*, 52(4), 539-548. https://doi.org/10.1080/10635150390223613

Silva-Junior, O.B., Faria, D.A., Grattapaglia, D. (2015). A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. *New Phytologist*, 206(4), 1527-1540. https://doi.org/10.1111/nph.13322

Slatko, B.E., Gardner, A.F., Ausubel, F.M. (2018). Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology*, 122, e59. https://doi.org/10.1002/cpmb.59

Sluys, R. (1989). Phylogenetic relationships of the triclads (Platyhelminthes, Seriata, Tricladida). *Bijdragen Tot de Dierkunde*, 59(11), 3-25.

Sluys, R. (1990). A monograph of the Dimarcusidae (Platyhelminthes, Seriata, Tricladida). *Zoologica Scripta*, 19(1), 13-29. https://doi.org/10.1111/j.1463-6409.1990.tb00237.x

Sluys, R. (2001). Towards a phylogenetic classification and characterization of dugesiid genera (Platyhelminthes, Tricladida, Dugesiidae): a morphological perspective. In: Littlewood, D.T.J., Bray, R.A. (eds.), *Interrelationships of the Platyhelminthes*. London & New York: Taylor & Francis, 57-73.

Sluys, R., Kawakatsu, M., Winsor, L. (1998). The genus *Dugesia* in Australia, with its phylogenetic analysis and historical biogeography (Platyhelminthes, Tricladida, Dugesiidae). *Zoologica Scripta*, 27(4), 273-289. https://doi.org/10.1111/j.1463-6409.1998.tb00461.x

Sluys, R., Riutort, M. (2018). Planarian diversity and phylogeny. In: Rink, J.C. (ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology*, Methods in Molecular Biology, vol. 1774. Springer Science+Business Media, New York, 1-56. https://doi.org/10.1007/978-1-4939-7802-1_1

Sluys, R., Solà, E., Gritzalis, K., Vila-Farré, M., Mateos, E., Riurtort, M. (2013). Integrative delineation of species of Mediterranean freshwater planarians (Platyhelminthes: Tricladida: Dugesiidae). *Zoological Journal of the Linnean Society*, 169(3), 523-547. https://doi.org/10.1111/zoj.12077

Smocovitis, V.B. (1992). Unifying biology: the evolutionary synthesis and evolutionary biology. *Journal of the History of Biology*, 25(1), 1-65. https://doi.org/10.1007/BF01947504

Solà, E. (2014). *Systematics and historical biogeography of the genus* Dugesia *(Platyhelminthes, Tricladida)*. Doctoral Thesis, Universitat de Barcelona. http://hdl.handle.net/2445/57483

Solà. E., Álvarez-Presas, M., Frías-López, C., Littlewood, D.T.J., Rozas, J., Riutort, M. (2015). Evolutionary analysis of mitogenomes from parasitic and free-living flatworms. *PLoS ONE*, 10(3), 1-20. https://doi.org/10.1371/journal.pone.0120081

Solà, E., Leria, L., Stocchino, G.A., Bagherzadeh, R., Balke, M., Daniels, S.R., Harrath, A.H., Khang, T.F., Krailas, D., Kumar, B., Li, M-H., Maghsoudlou, A., Matsumoto, M., Naser, N., Oben, B., Segev, O., Thielicke, M., Tong, X., Zivanovic, G., Manconi, R., Baguñà, J., Riutort, M. (2022). Three dispersal routes out of Africa: a puzzling biogeographical history in freshwater planarians. *Journal of Biogeography*, 49(7), 1219-1233. https://doi.

org/10.1111/jbi.14371

Solà, E., Sluys, R., Gritzalis, K., Riurtort, M. (2013). Fluvial basin history in the northeastern Mediterranean region underlies dispersal and speciation patterns in the genus *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution*, 66(3), 877-888. https://doi.org/10.1016/j.ympev.2012.11.010

Spriggs, E.L., Eaton, D.A.R., Sweeney, P.W., Schlutius, C., Edwards, E.J., Donoghue, M.J. (2019). RAD-seq data reveal a cryptic *Viburnum* species on the North American coastal plain. *Systematic Biology*, 68(2), 187-203. https://doi.org/10.1093/sysbio/syy084

Steiner, F.M., Csösz, S., Markó, B., Gamisch, A., Rinnhofer, L., Folterbauer, C., Hammerle, S., Stauffer, C., Arthofer, W., Schlick-Steiner, B.C. (2018). Turning one into five: integrative taxonomy uncovers complex evolution of cryptic species in the harvester ant *Messor "structor"*. *Molecular Phylogenetics and Evolution*, 127, 387-404. https://doi.org/10.1016/j.ympev.2018.04.005

Stobie, C.S., Cunningham, M.J., Oosthuizen, C.J., Bloomer, P. (2019). Finding stories in noise: mitochondrial portraits from RAD data. *Molecular Ecology Resources*, 19(1), 191-205. https://doi.org/10.1111/1755-0998.12953

Stobie, C.S., Oosthuizen, C.J., Cunningham, M.J., Bloomer, P. (2018). Exploring the phylogeography of a hexaploid freshwater fish by RAD sequencing. *Ecology and Evolution*, 8(4), 2326-2342. https://doi.org/10.1002/ece3.3821

Stocchino, G.A. (2018). 80 years of research on planarians (Platyhelminthes, Tricladida) from Sardinia, Italy: an annotated checklist. *Zootaxa*, 4532(4), 539-552. https://doi.org/10.11646/zootaxa.4532.4.5

Stocchino, G.A., Corso, G., Manconi, R., Casu, S., Pala, M. (2005). Endemic freshwater planarians of Sardinia: redescription of *Dugesia hepta* (Platyhelminthes, Tricladida) with a comparison of the Mediterranean species of the genus. *Journal of Natural History*, 39(22), 1947-1960. https://doi.org/10.1080/00222930500060025

Stocchino, G.A., Dols-Serrate, D., Sluys, R., Riutort, M., Onnis, C., Manconi, R. (2021). Amphibioplanidae: a new branch and family on the phylogenetic tree of the triclad flatworms (Platyhelminthes: Tricladida), represented by a species from Sardinian caves with a remarkable lifestyle. *Zoological Journal of the Linnean Society*, 193(4), 1364-1391. https://doi.org/10.1093/zoolinnean/zlaa183

Stocchino, G.A., Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology*, 803(3), 319-328. http://dx.doi.org/10.1080/11250003.2013.822025

Sukumaran, J., Knowles, L.L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, 114(7), 1607-1612. https://doi.org/10.1073/pnas.1607921114

Tamura, S., Oki, I., Kawakatsu, M. (1991). Karyological and taxonomic studies of *Dugesia japonica* from the Southwast Islands of Japan-II. *Hydrobiologia*, 227(1), 157-162. https://doi.org/10.1007/BF00027596

Tamura, S., Yamamoto, K., Takai, M., Oki, I., Kawakatsu, M. (1998). Karyology and biogeography of *Dugesia japonica* and *Dugesia ryukyuensis* in Kyushu, Southern Japan. *Hydrobiologia*, 383(1-3), 321-327. https://doi.org/10.1023/A:1003471207447

Taylor, S.A., Larson, E.L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology and Evolution*, 3(2), 170-177. http://dx.doi.org/10.1038/s41559-018-0777-y

Thomson, S.A., Thiele, K., Conix, S., Christidis, L., Costello, M.J., Hobern, D., Nikolaeva, S., Pyle, R.L., van Dijk, P.P., Weaver, H., Zachos, F.E., Zhang, Z-Q., Garnett, S.T. (2021). Towards a global list of accepted species II. Consequences of inadequate taxonomic list governance. *Organisms Diversity and Evolution*, 21(4), 623-640. https://doi.org/10.1007/s13127-021-00518-8

Tian, Q., Guo, Q., Guo, Y., Luo, L., Kristiansen, K., Han, Z., Fang, H., Zhang, S. (2022). Whole-genome sequence of the planarian *Dugesia japonica* combining Illumina and PacBio data. *Genomics*, 114(2), 110293. https://doi.org/10.1016/j.ygeno.2022.110293

Tonzo, V., Papadopoulou, A., Ortego, J. (2019). Genomic data reveal deep structure but no support for current taxonomic designation in a grasshopper species complex. *Molecular Ecology*, 28(17), 3869-3886. https://doi.org/10.1111/mec.15189

Treangen, T.J., Salzberg, S.L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, 13(1), 36-46. http://dx.doi.org/10.1038/nrg3117

Tyler, S., Schilling, S., Hooge, M., Bush, L.F. (comp.) (2006-2024). Turbellarian taxonomic database. Version 2.07. http://turbellaria.umaine.edu/. Last accessed, April 2024.

Vacca, R.A., Casu, S., Pala, M. (1988). Popolamento planariologico dei fiumi del Nord Sardegna I: i cariotipi delle planaria d'acqua dolce del grupo *Dugesia gonocephala* (Turbellaria, Tricladida) presenti nel fiume Silis (Sassari). *Bolletino della Società Sarda di Scienze Naturali*, 26, 131-147.

Vacca, R.A., Casu, S., Pala, M. (1993). Popolamento planariologico dei fiumi del Nord Sardegna II. I cariotipi dei tricladi d'acqua dolce rinvenuti nel bacino idrografico del fiume Coghinas. *Bolletino della Società Sarda di Scienze Naturali*, 29, 59-73.

Valentini, A., Pompanon, F., Taberlet, P. (2008). DNA barcoding for ecologists. *Trends in Ecology and Evolution*, 24(2), 110-117. https://doi.org/10.1016/j.tree.2008.09.011

Van de Peer, Y., Mizrachi, E., Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics*, 18(7), 411-424. http://dx.doi.org/10.1038/nrg.2017.26

Vila-Farré, M., Rink, J.C. (2018). The ecology of freshwater planarians. In: Rink, J.C.

(ed.), *Planarian Regeneration: Methods and Protocols in Molecular Biology*, Methods in Molecular Biology, vol. 1774. Springer Science+Business Media, New York, 1-56. https://doi.org/10.1007/978-1-4939-7802-1_1

Vila-Farré, M., Rozanski, A., Ivanković, M., Cleland, J., Brand, J.N., Thalen, F., Grohme, M.A., von Kannen, S., Grosbusch, A., Vu, H.T-K., Prieto, C.E., Carbayo, F., Egger, B., Bleidorn, C., Rasko, J.E.J., Rink, J.C. (2023). Evolutionary dynamics of whole-body regeneration across planarian flatworms. *Nature Ecology and Evolution*, 7(12), 2018-2124. https://doi.org/10.1038/s41559-023-02221-7

Vila-Farré, M., Sluys, R., Almagro, I., Handberg-Thorsager, M., Romero, R. (2011). Freshwater planarians (Platyhelminthes, Tricladida) from the Iberian Peninsula and Greece: diversity and notes on ecology. *Zootaxa*, 2779, 1-38.

Vinogradov, A.E., Anatskaya, O.V. (2017). DNA helix: the importance of being AT-rich. *Mammalian Genome*, 29(9-10), 455-464. https://doi.org/10.1007/s00335-017-9713-8

de Vries, E.J. (1984). On the species of the *Dugesia gonocephala* group (Platyhelminthes, Turbellaria, Tricladida) from Greece. *Bijdragen tot de Dierkunde*, 54(1), 101-126. https://doi.org/10.1163/26660644-05401008

de Vries, E.J. (1985). The biogeography of the genus *Dugesia* (Turbellaria, Tricladida, Paludicola) in the Mediterranean region. *Journal of Biogeography*, 12(6), 509-518. https://doi.org/10.2307/2844906

de Vries, E.J. (1986). On the karyology of *Dugesia gonocephala* s.l. (Turbellaria, Tricladida) from Montpellier, France. *Hydrobiologia*, 132(1), 251-256. https://doi.org/10.1007/BF00046257

de Vries, E.J. (1988). Further contributions to the taxonomy and biogeography of the subgenus *Dugesia* (Platyhelminthes: Tricladida: Paludicola) in the Mediterranean region and the Middle East. *Israel Journal of Zoology*, 35(3), 109-136. https://doi.org/10.1080/00212210.1988.10688607

de Vries, E.J., Ball, I.R. (1980). On *Dugesia gonocephala* from Western Europe. *Bijdragen tot de Dierkunde*, 50(2), 342-350.

de Vries, E.J., Sluys, R. (1991). Phylogenetic relationships of the genus *Dugesia* (Platyhelminthes, Tricladida, Paludicola). *Journal of Zoology*, 223(1), 103-116. https://doi.org/10.1111/j.1469-7998.1991.tb04752.x

Wagner, G.E., Dabernig-Heinz, J., Lipp, M., Cabal, A., Simantzik, J., Kohl, M., Scheiber, M., Lichtenegger, S., Ehricht, R., Leitner, E., Ruppitsch, W., Steinmetz, I. (2023). Real-time nanopore Q20+ sequencing enables extremely fast and accurate core genome MLST typing and democratizes access to high-resolution bacterial pathogen surveillance. *Journal of Clinical Microbiology*, 63(4), e01631-22. https://doi.org/10.1128/jcm.01631-22

Wagner, D.E., Wang, I.E., Reddien, P.W. (2011). Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science*, 332(6031), 811-816. https://doi.org/10.1126/science.1203983

Wang, L., Dong, Z., Chen, G., Sluys, R., Liu, D. (2021). Integrative descriptions of two new species of *Dugesia* from Hainan Island, China (Platyhelminthes, Tricladida, Dugesiidae). *ZooKeys*, 1028, 1-28. https://doi.org/10.3897/zookeys.1028.60838

Watson, J., Crick, F. (1953). Molecular structure of nucleic acids. *Nature*, 171, 709-756.

Watt, F.M., Hogan, B.L.M. (2000). Out of Eden: stem cells and their niches. *Science*, 287(5457), 1427-1430. https://doi.org/10.1126/science.287.5457.1427

Weissman, J.K., Fagan, W.F., Johnson, P.L.F. (2019). Linking high GC content to the repair of double strand breaks in prokaryotic genomes. *PLoS Genetics*, 15(11), e1008493. https::/doi.org/10.1371/journal.pgen.1008493

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C-S., Phillippy, A.M., Schatz, M.C., Myers, G., DePristo, M.A., Ruan, J., Marschall, T., Sedlazeck, F.J., Zook, J.M., Li, H., Koren, S., Carroll, A., Rank, D.R., Hunkapiller, M.W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, 37(10), 1155-1162. https://doi.org/10.1038/s41587-019-0217-9

Wenne, R., Drywa, A., Kent, M., Sundsaasen, K.K., Lien, S. (2016). SNP arrays for species identification in salmonids. In: Bourlat, S.J. (ED.), *Marine Genomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1452. New York: Springer Science+Business Media New York 2016, 97-111.

Woese, C.R., Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, 74(11), 5088-5090. https://doi.org/10.1073/pnas.74.11.5088

Xia, X. (2016). PhyPA: phylogenetic method with pairwise sequence alignment outperforms likelihood methods in phylogenetics involving highly diverged sequences. *Molecular Phylogenetics and Evolution*, 102, 331-343. http://dx.doi.org/10.1016/j.ympev.2016.07.001

Yang, Z., (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39(3), 306-314. https://doi.org/10.1007/BF00160154

Yang, Z. (1996). Phylogenetic analysis using parsimony and likelihood methods. *Journal of Molecular Evolution*, 42(2), 294-307. https://doi.org/10.1007/BF02198856

Yang, Z., Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Genetics*, 164(4), 1645-1656. https://doi.org/10.1073/pnas.0913022107

Yang, Z., Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Re-*

views Genetics, 13(5), 303-314. http://dx.doi.org/10.1038/nrg3186

Zallen, D.T. (2003). Despite Franklin's work, Wilkins earned his Nobel. *Nature*, 425(6953), 15. https://doi.org/10.1038/425015b

Zardoya, R., Meyer, A. (1996). Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution*, 13(7), 933-942. https://doi.org/10.1093/oxfordjournals.molbev.a025661

Zhang, G. (2015). Bird sequencing project takes off. *Nature*, 522(7554), 34-34. https://doi.org/10.1038/522034d

Zhang, C., Rannala, B., Yang, Z. (2014). Bayesian species delimitation can be robust to guide-tree inference errors. *Systematic Biology*, 63(6), 993-1004. https://doi.org/10.1093/sysbio/syu052

Zink, R.M., McKitrick, M.C. (1995). The debate over species concepts and its implications for ornithology. *The Auk*, 112(3), 701-719. https://doi.org/10.2307/4088685

Appendix

Appendix: Other publications

Appendix I

Amphibioplanidae: a new branch and family on the phylogenetic tree of the triclad flatworms (Platyhelminthes: Tricladida), represented by a species from Sardinian caves with a remarkable lifestyle

Stocchino, G.A., Dols-Serrate, D., Sluys, R., Riutort, M., Onnis, C., Manconi, R.

2021. *Zoological Journal of the Linnean Society*, 193, 1364-1391.
doi: 10.1093/zoolinnean/zlaa183

Appendix II

Island-specific evolution of a sex-primed autosome in a sexual planarian

Guo, L., Bloom, J., Dols-Serrate, D., Ben-David, E., Schubert, O., Kozuma, K., Ho, K., Warda, E., Chui, C., Wei, Y., Leighton, D., Lemus Vergara, T., Riurtort, M., Sánchez Alvarado, A., Kruglyak, L.

2022. *Nature*, 606(7913), 329-334.
doi: 10.1038/s41586-022-04757-3

Appendix III

Building phylogenies from transcriptomic data

Benítez-Álvarez, L., Leria, L., Dols-Serrate, D., Riutort, M.

2023. In: Gentile L. (ed.), *Schmidtea mediterranea: Methods and Protocols*, Methods in Molecular Biology, vol. 2680. Springer Nature, New York.
doi: 10.1007/978-1-0716-3275-8

Zoological Journal of the Linnean Society, 2021, **193**, 1364–1391. With 15 figures.

Amphibioplanidae: a new branch and family on the phylogenetic tree of the triclad flatworms (Platyhelminthes: Tricladida), represented by a species from Sardinian caves with a remarkable lifestyle

GIACINTA ANGELA STOCCHINO[1,*], DANIEL DOLS-SERRATE[2], RONALD SLUYS[3], MARTA RIUTORT[2], CARLO ONNIS[4] and RENATA MANCONI[1]

[1]*Dipartimento di Medicina Veterinaria, Laboratorio di Zoologia, Via Vienna 2, Università di Sassari, Italy*
[2]*Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Catalonia, Spain*
[3]*Naturalis Biodiversity Center, P.O. Box 9517, 2300 RA Leiden, The Netherlands*
[4]*Via 25 Aprile 1C, I-09069 Maracalagonis CA, Italy*

Received 14 May 2020; revised 6 November 2020; accepted for publication 17 December 2020

In this study we document the phylogenetic position and the anatomical and ecological characteristics of a new taxon of cave-dwelling triclad flatworms. This species from the Mediterranean island of Sardinia (Italy) forms a separate branch on the phylogenetic tree of the planarians for which we proposed a new genus and placed it in a new family. Molecular analysis reveals a sister-group relationship between this new family, Amphibioplanidae, and the family Dimarcusidae of the suborder Cavernicola, together constituting the sister-group of the suborder Maricola. The new species ***Amphibioplana onnisi* Stocchino & Sluys, sp. nov.** is slender and furnished with an epidermis that is abundantly and uniformly provided with cilia and rhabdites. The pharynx presents an oesophageal projection pointing into the pharynx lumen. Ovaries are located far behind the brain. A relatively long common oviduct opens into the diverticulum of the bursal canal. The probursal copulatory bursa is lined with a highly vacuolated, syncytial epithelium. The habitat of this species is formed by groundwater in microcrevices of the epikarst and temporary cave pools, but the animals can survive in a basically terrestrial environment, thus exhibiting an amphibious lifestyle. Some tentative scenarios are discussed for the evolution of the new taxon and for the presence of the new species in three separate caves in Sardinia.

ADDITIONAL KEYWORDS: *Amphibioplana* – amphibious lifestyle – integrative taxonomy – insular Italy – Mediterranean – molecular systematics – new genus – stygofauna.

INTRODUCTION

Of the free-living flatworms, the triclads or planarians (Platyhelminthes: Tricladida) are perhaps the best-known representatives. They are a ubiquitous group of species that occur on land, in the sea and in freshwater environments. Species in this group are frequently used in teaching and as a model organism in studies on regeneration, stem cells and pattern formation. Based on their different habitats, triclads were originally classified in three groups, viz., the freshwater planarians or Paludicola Hallez, 1892, the land planarians or Terricola Hallez, 1892 and the marine planarians or Maricola Hallez, 1892 [see: Sluys & Riutort (2018) and references therein]. This classification remained basically unchanged for almost a century, until Sluys (1990) proposed a fourth clade, the Cavernicola Sluys, 1990, which currently

*Corresponding author. E-mail: stocchin@uniss.it; r.manconi@uniss.it
[Version of record, published online 17 February 2021; http://zoobank.org/ urn:lsid:zoobank.org:pub:35092390-CE13-4E2D-9C70-47014A4C165F]

© The Author(s) 2021. Published by Oxford University Press on behalf of The Linnean Society of London.
All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

comprises seven freshwater species and one terrestrial species, mainly inhabiting subterranean habitats and allocated in a single family, Dimarcusidae Mitchell & Kawakatsu, 1972 [see: Benítez-Álvarez *et al*. (2020) and references therein].

During the past four decades, ongoing studies on the phylogenetic relationships within Tricladida Lang, 1884 have established that Paludicola forms a paraphyletic taxon, because the freshwater family Dugesiidae Ball, 1974 appears to be more closely related to the land planarians (Terricola) than to the other freshwater planarians [see: Sluys & Riutort (2018) and references therein]. These new insights have resulted in the recognition of three suborders in the most recent classification of the higher taxa of triclad flatworms, viz., Continenticola Carranza *et al*., 1998 (including freshwater and land planarians), Maricola and Cavernicola (Sluys *et al.*, 2009).

Although subterranean habitats mainly host representatives of freshwater planarians and members of Cavernicola, recently some land planarians were reported from caves in Sardinia, Italy (Stocchino *et al*., 2008, 2015) and Brazil (Leal-Zanchet & Marques, 2018a, b; Amaral *et al.*, 2019). Furthermore, the first troglobitic maricolan planarian was recently reported also from a Brazilian terrestrial cave (Souza *et al.*, 2018).

Recent accounts on planarians from the Mediterranean island of Sardinia (Italy) highlighted that all three families of planarians found in European freshwaters are represented in the surface waters and groundwaters of the island, viz., Dendrocoelidae Hallez, 1892, Dugesiidae and Planariidae Stimpson, 1857 (Pala *et al*., 1995, 1999, 2000; Stocchino *et al*., 2005, 2008, 2013; Lázaro *et al*., 2009, 2011; Stocchino, 2018; Dols-Serrate *et al.*, 2020). It was, therefore, unexpected that recent samplings from three different Sardinian subterranean karstic areas revealed the presence of planarians that have morphological and anatomical features that not only prevented their inclusion in one of these three families known to occur on the island, but also did not allow the specimens to be comfortably included into one of the three currently recognized suborders of triclad flatworms. Apart from their unusual taxonomic features, the animals also exhibited an unusual ecology, in that the specimens were found in subterranean terrestrial as well as freshwater habitats, thus showing an amphibious lifestyle, which is unusual in planarians.

We here document the morpho-anatomical and ecological characteristics of these novel finds from Sardinia. In combination with a molecular phylogenetic analysis, this enabled us to resolve their phylogenetic affinities and taxonomic status, resulting in the description of a new species, placed in a new genus and a new family. Thus, the discovery of this new taxon changes our views on the phylogenetic relationships in the order Tricladida.

MATERIAL AND METHODS

Collection and culturing of specimens

Planarians were collected from three Sardinian caves during 2015–18 (Fig. 1). For full details on collection sites, see below (section on 'Habitat'). Some specimens were preserved in the field, just after collection, for both morphological and molecular analysis, but the majority of the animals were transferred to the zoology laboratory at the University of Sassari, Italy, where they were reared in glass bowls under semi-dark conditions at a temperature of 18 ± 2 °C, and were fed with fresh beef liver once per week. Live animals were photographed both in the field and in the laboratory (Olympus TG-3).

Morphological analysis

For morphological study, sexual specimens were fixed for 24 h in Bouin's fluid, dehydrated in an ascending ethanol series, cleared in clove oil and embedded in synthetic wax. Serial sections were made at intervals of 4, 5 and 6 µm and were stained with Mallory–Cason (cf. Winsor & Sluys, 2018). Reconstructions of the copulatory apparatus were obtained by using a compound microscope provided with a camera lucida.

Type material and other specimens examined are deposited at Naturalis Biodiversity Center, Leiden, The Netherlands (RMNH), and in the Giacinta A. Stocchino collection (CGAS), University of Sassari, Italy.

Molecular analysis

Individuals analysed, DNA extraction, quantification and sequence amplification

For total genomic DNA extractions, specimens were fixed and preserved in absolute ethanol. Although animals used for morphology and molecular analyses concerned different individuals, they were collected always from the same gathering.

We experienced problems in extracting sufficient DNA from single individuals, which may either have been due to the small size of the specimens or to suboptimal fixation. Therefore, it was not possible to obtain high-quality genomic DNA from all individuals captured and, thus, three out of the eight samples used in the present analyses are constituted by more than one individual (Table 1). Samples were digested with proteinase K (5 µL/sample) and were subsequently purified by using a Wizard Genomics DNA Purification Kit (Promega Corporation), following the manufacturer's protocol. DNA concentration was

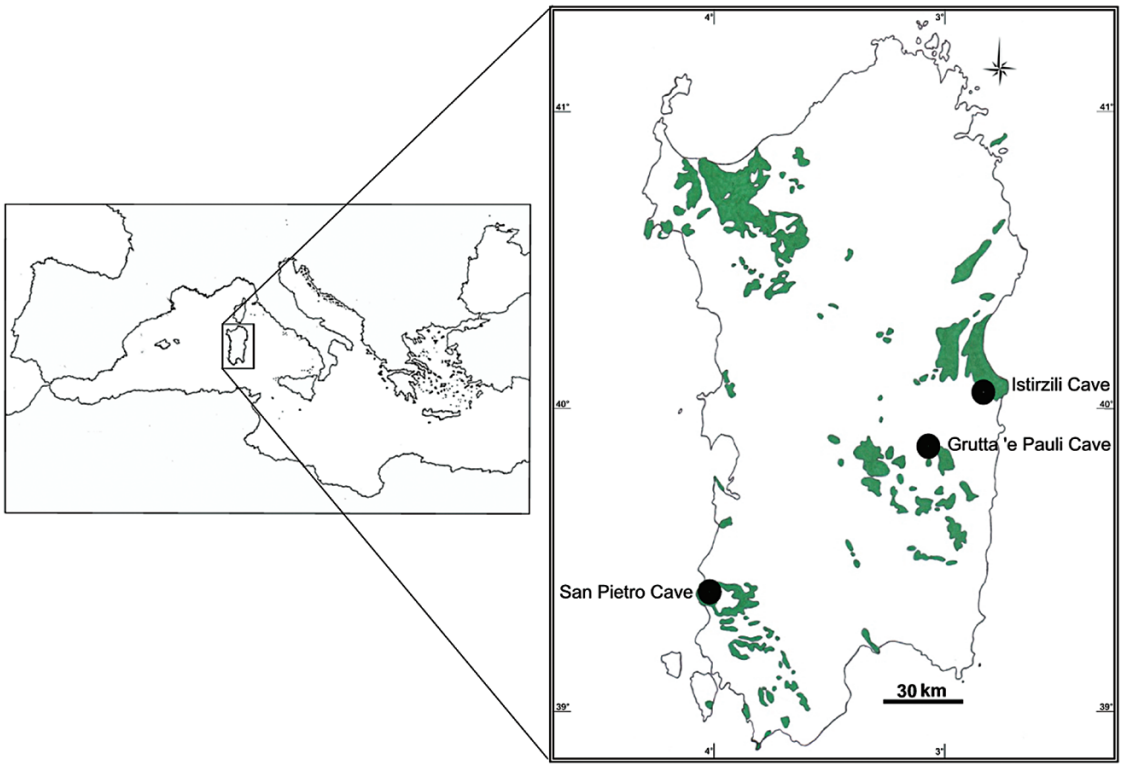


Figure 1. Geographic distribution of *Amphibioplana onnisi* in the Mediterranean region. Rectangular inset corresponds with area enlarged to right-hand side, showing the island of Sardinia. Filled black circles indicate location of caves from which populations were sampled; green areas indicate the karst regions.

quantified by analysing 2 µL per sample, using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA).

We selected 18S and 28S ribosomal RNA nuclear genes (here referred to as 18S and 28S, respectively), enabling us to place the new organisms in the phylogenetic tree of Tricladida. For analysis of the genetic diversity within and among the three populations, we tried mitochondrial markers that had been successfully amplified in previous studies, viz., cytochrome *c* oxidase subunit I (*COI*; used in most studies on Continenticola), cytochrome *b* (*Cytb*; sequenced in one continenticolan and one cavernicolan species) (cf. Lázaro *et al.*, 2011; Laumer & Giribet, 2014) and 16S ribosomal RNA (16S, used in a Maricola species) (Riesgo *et al.*, 2017). We used the primers for these markers that previously had been applied, and when these failed, universal primers were used. Amplifications of these regions were all unsuccessful, apart from those of 16S.

From each extraction we amplified and sequenced the nearly complete 18S and a fragment of about 2000 base pairs (bp) from the 28S gene, plus a fragment of about 350 bp from 16S; 18S and 28S were amplified in two overlapping amplicons. Sequencing and annealing temperatures of each pair of primers used are given in Supporting Information, Table S1. Final polymerase chain reaction (PCR) volume for all markers was 25 µL, consisting of: (1) 5 µL of Promega 5× Green GoTaq Flexi Buffer, (2) 2.5 µL of $MgCl_2$ (25 mmol/L), (3) 1.5 µL of dNTPs (0.5 mmol/L), (4) 1 µL of each primer (10 µmol/L), (5) 0.15 µL of *Taq* polymerase (5 u/µL) (GoTaq Flexi DNA Polymerase, Promega) and (6) 1 µL of DNA extraction sample. Autoclaved double-distilled deionized water (milliQ, Millipore) was added to obtain the final PCR volume. Amplification conditions were different for each gene (Supporting Information, Table S2). The resulting PCR products were run in a 1% agarose gel in order to check the correct amplification for the different molecular markers. Polymerase chain reaction products were purified before sequencing by using a vacuum system (MultiScreen HTS Vacuum Manifold of Millipore). Purified PCR products were sequenced by Macrogen Inc. (Amsterdam, The Netherlands). Both strands were sequenced by using the same primers as used for the amplification. The chromatograms thus obtained were visually checked with GENEIOUS R8 (Biomatters, http://www.geneious.com/). Reads were assembled into contigs and consensus sequences were obtained using GENEIOUS R8.

Table 1. Species names, taxonomic classification and GenBank accession numbers for the Platyhelminthes sequences used in the phylogenetic analyses. For *Amphibioplana* the number of individuals included in each sample is indicated

		GenBank accession number		
Species name	Taxon	18S	28S	16S
Acanthiella sp. [a]	Rhabditophora: Prolecithophora: Protomonotresidae	KC869786	KC869839	-
Protomonotresidae sp.[a]	Rhabditophora: Prolecithophora: Protomonotresidae	KC869820	KC869873	-
Reisingeria hexaoculata [a]	Rhabditophora: Prolecithophora: Pseudostomidae	AF065426	AY157157	-
Plagiostomum stellatum [a]	Rhabditophora: Prolecithophora: Plagiostomidae	KC869819	KC869872	-
Plagiostomum whitmani [a]	Rhabditophora: Prolecithophora: Plagiostomidae	KC869818	KC869871	-
Plicastoma cuticulata [a]	Rhabditophora: Prolecithophora: Plagiostomidae	AF065422	AY157158	-
Kronborgia isopodicola [a]	Rhabditophora: Rhabdocoela: Dalyellioida: Fecampiidae	AJ012513.1	AF022862.1	-
Notentera ivanovia [a]	Rhabditophora: Rhabdocoela: Dalyellioida: Fecampiidae	AJ287546.1	AY157167.1	-
Urastoma cyprinae [a]	Rhabditophora: Mediofusata: Urastomidae	AF065428.2	AY157165.1	-
Hausera hauseri	Tricladida: Cavernicola: Dimarcusidae	MN719501	MN719494	-
Kawakatsua pumila	Tricladida: Cavernicola: Dimarcusidae	KC869823.1	KC869876.1	-
Novomitchellia bursaelongata	Tricladida: Cavernicola: Dimarcusidae	KU096054.2	-	-
Opisthobursa mexicana	Tricladida: Cavernicola: Dimarcusidae	MN719503	MN719496	-
Rhodax sp. 2	Tricladida: Cavernicola: Dimarcusidae	MN719506	MN719499	-
Dugesia gonocephala	Tricladida: Continenticola: Geoplanoidea: Dugesiidae	DQ666002.1	DQ665965.1	-
Girardia tigrina	Tricladida: Continenticola: Geoplanoidea: Dugesiidae	AF013156.1	DQ665977.1	-
Schmidtea polychroa	Tricladida: Continenticola: Geoplanoidea: Dugesiidae	AF013154.1	DQ665993.1	-
Cephaloflexa bergi	Tricladida: Continenticola: Geoplanoidea: Geoplanidae	KJ599712.1	KC608355.1	-
Geoplana quagga	Tricladida: Continenticola: Geoplanoidea: Geoplanidae	KC608497.1	KC608380.1	-
Crenobia alpina	Tricladida: Continenticola: Planarioidea: Planariidae	M58345.1	DQ665960.1	-
Phagocata vitta	Tricladida: Continenticola: Planarioidea: Planariidae	DQ665998.1	DQ665989.1	-
Bdelloura candida	Tricladida: Maricola: Bdellouroidea: Bdellouridae	Z99947.1	AJ270167.1	-
Palombiella stephensoni	Tricladida: Maricola: Bdellouroidea: Bdellouridae	DQ666008.2	DQ665988.1	-
Pentacoelum kazukolinda	Tricladida: Maricola: Bdellouroidea: Bdellouridae	JN009784.1	JN009787.1	-
Sluysia triapertura	Tricladida: Maricola: Bdellouroidea: Uteriporidae	MF383119.1	MF383122.1	-
Uteriporus sp.	Tricladida: Maricola: Bdellouroidea: Uteriporidae	AF013148.1	-	-
Cercyra hastata	Tricladida: Maricola: Cercyroidea: Cercyridae	KM200902.1	DQ665962.1	-

Table 1. Continued

Species name	Taxon	GenBank accession number		
		18S	28S	16S
Sabussowia dioica	Tricladida: Maricola: Cercyroidea: Cercyridae	JN009785.1	JN009788.1	-
Procerodes littoralis	Tricladida: Maricola: Procerodoidea: Procerodidae	Z99950.1	DQ665985.1	-
Amphibioplana onnisi (S. Pietro-1; 4 indiv.)	Tricladida: Cavernicola: Amphibioplanidae	MT396630	MT416102	MT416094
Amphibioplana onnisi (S. Pietro-2; 7 indiv.)	Tricladida: Cavernicola: Amphibioplanidae	MT396631	MT416103	MT416096
Amphibioplana onnisi (S. Pietro-3; 1 indv.)	Tricladida: Cavernicola: Amphibioplanidae	-	-	MT416099
Amphibioplana onnisi (S. Pietro-4; 1 indv.)	Tricladida: Cavernicola: Amphibioplanidae	-	-	MT416100
Amphibioplana onnisi (S. Pietro-5; 1 indv.)	Tricladida: Cavernicola: Amphibioplanidae	MT396633	MT416105	-
Amphibioplana onnisi (Istirzili-1; 4 indv.)	Tricladida: Cavernicola: Amphibioplanidae	MT396632	MT416104	MT416097
Amphibioplana onnisi (Istirzili-2; 1 indv.)	Tricladida: Cavernicola: Amphibioplanidae	-	-	MT416098
Amphibioplana onnisi (Pauli-1; 1 indv.)	Tricladida: Cavernicola: Amphibioplanidae	MT396629	MT416101	MT416095

[a]Outgroup sequences.

Phylogenetic analysis: gene tree inference

Newly obtained 18S and 28S sequences were aligned with sequences from the three suborders of Tricladida as available from GenBank, while we also included as outgroups a wide range of taxa concerning Fecampiidae Graff, 1903, Urastomidae Marcus, 1951, and Prolecithophora Karling, 1940 (Table 1), which in recent molecular studies were identified as possibly closely related to the Tricladida (Riutort *et al.*, 2012; Laumer & Giribet, 2014; Laumer *et al.*, 2015). Alignments were constructed with the online version of MAFFT v.7 (Katoh & Standley, 2013), using the E-INS-i method and default parametrization. The subsequent alignments were revised in BioEdit v.7.2.5 (Hall, 1999). In an attempt to avoid ambiguously aligned segments and an excess of missing data, we trimmed the alignments by using ClipKIT's kcgappy filtering mode (Steenwyk *et al.*, 2020). During this procedure, the mode was selected that keeps parsimony-informative sites and conserved sites, while eliminating from the alignments those sites with more than 90% missing data. Since it has been noted that automatic filtering of alignments may result in loss of information (Tan *et al.*, 2015), we also inferred phylogenetic trees from the original alignments without any filtering, thus being able to evaluate the effect of eliminating a number of sites on the resulting trees. Subsequently, nucleotide saturation was tested for each dataset by using the Xia & Lemey (2009) test implemented in DAMBE v.7.0.28 (Xia & Xie, 2001).

Phylogenetic inferences were performed for the four datasets of individual genes (18S and 28S, with and without filtering), as well as for two concatenated datasets (filtered and non-filtered gene alignments) (Table 2) by using Bayesian inference (BI) and maximum likelihood (ML) approaches. The ML analysis was performed using the best substitution model found for each dataset and partition (in this case by gene in the concatenated datasets) with ModelFinder (Kalyaanamoorthy *et al.*, 2017) as implemented in the IQ-Tree 2.0 software (Nguyen *et al.*, 2015; Minh *et al.*, 2020) (Supporting Information, Table S3). For computing nodal supports we performed an ultrafast bootstrap (UFB) approximation (Hoang *et al.*, 2018) and the SH-aLRT test (Guindon *et al.*, 2010), with 10 000 replicates for each approach. These two support measures differ from the classic bootstrap in their cut-off levels for highly supported clades, being > 80 for SH-aLRT and > 95 for UFB (Minh *et al.*, 2013). Bayesian inference analyses were performed in PhyloBayes v.4.1 (Lartillot & Philippe, 2004; Lartillot, 2020), using a GTR+CAT+G mixture model with four gamma categories, implemented in the *pb* program and run on four independent chains, starting from

Table 2. Datasets for both single genes and the concatenated data used in the phylogenetic analyses, with options selected in the filtering step of the single gene alignments, their final length, the proportion of sites removed after filtering, and the evolutionary models selected by ModelFinder in IQ-TREE

Single gene dataset	ClipKIT option	length (bp)	% sites removed	IQtree evol model (BIC)
18S	-	1906	-	GTR+F+I+G4
18S_filtered	kcgappy	1518	20.36	GTR+F+I+G4
28S	-	1813	-	GTR+F+I+G4
28S_filtered	kcgappy	1461	19.42	GTR+F+I+G4
Concatenated dataset	partition 1 sites	IQtree subs model (BIC)	partition 2 sites	IQtree evol model (BIC)
18S+28S	1>1906	GTR+F+I+G4	1907>5539	GTR+F+I+G4
18S+28S_filtered	1>1518	GTR+F+I+G4	1519>4441	GTR+F+I+G4

random seeds until convergence criteria were met (effective size for every parameter greater than 100 and a mean discrepancy lower than 0.3). Convergence was surveyed every 100 cycles, using the *bpcomp* and *tracecomp* programs with a default burn-in equal to one-fifth of the total length of the chain to ensure exclusion of suboptimal trees when inferring the topology and the posterior probabilities (PP) of the nodes of the trees. Ultimately, the trees recovered from BI and ML analyses were visualized with FigTree v.1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Regarding 16S, sequences were blasted against the GenBank database to ensure that they were from triclads and not contaminations or artefacts. The program BLAST-nucleotide (https://blast.ncbi.nlm.nih.gov/Blast.cgi, last visited 13 October 2020) was used to search the nucleotide collection database with the discontiguous megablast option (optimization for more dissimilar sequences), leaving the rest of the parameters as default. Once validated, sequences from individuals from the three caves were aligned with MAFFT G-INS-i option (rest of parameters as default) and genetic distances were estimated by using the Kimura two parameters (K2P) correction, as implemented in MEGA X (Kumar *et al.*, 2018).

ABBREVIATIONS USED IN THE FIGURES

ag, adhesive glands; aib, anterior intestinal branch; b, brain; bc, bursal canal; bcd, bursal canal diaphragm; bm, basement membrane; c, cilia; ca, copulatory apparatus; cb, copulatory bursa; cc, caudal constriction; cg, cyanophil glands; cm, circular muscle; cs, cyanophil secretion; cod, common oviduct; csc, cup-shaped chamber; cph, convoluted pharynx; de, dorsal epidermis; di, diverticulum; e, epidermis; ed, ejaculatory duct; eg, erythrophil glands; es, erythrophil secretion; ga, genital atrium; go, gonopore; h, head; lm, longitudinal muscle; lo, left ovary; lod, left oviduct; lvd, left vas deferens; m, mouth; n, neuropil; nu, nuclei; od, oviduct; oep, oesophagal protrusion; ov, ovary; pb, penis bulb; ph, pharynx; php, pharyngeal pouch; pib, posterior intestinal branch; pp, penis papilla; r, rhabdite; rb, regenerating blastema; rfg, rhabdite-forming gland; rod, right oviduct; rph, rectilinear pharynx; rvd, right vas deferens; s, sperm; sg, shell glands; spv, spermiducal vesicle; t, testis; ta, tail; tu, tuba; va, vacuoles; vd, vas deferens; ve, ventral epidermis; vi, vitellaria; vnc, ventral nerve cord; xg, xanthophil glands.

RESULTS

MOLECULAR PHYLOGENY

Sequence characteristics

From the specimens of the new taxon from Sardinia, five sequences were obtained for 18S, five for 28S and seven for 16S, with at least one sequence from each locality (Table 1). Four datasets were built for the individual nuclear genes, which were ultimately merged into two concatenated datasets (Table 2). Because the obtained alignments showed certain regions to be extremely variable, which may generate only noise instead of information, we decided to use the ClipKIT's kcgappy filtering model in order to be able to retain both parsimony-informative and constant sites, as well as removing sites with > 90% missing data. In order to assess whether inclusion or exclusion of noisy, variable regions resulted in any difference in the generated tree topologies, and especially in the support values, we also analysed the non-filtered alignments.

Phylogenetic trees

Supporting Information, Figures S1 and S2 show the trees obtained with the four datasets of the individual genes (Table 2), with the ML topologies showing SH-aLRT, UFB (ML) and PP (BI) support values on their branches. In all trees, Maricola, Continenticola and Dimarcusidae each constitute

monophyletic groups, generally with high support, with the exception of Continenticola when it comes to the 18S gene and the BI approach, in general. In each of the four trees, the new taxon constitutes a clade (described below as family Amphibioplanidae) that is sister to the Dimarcusidae, although the support for this sister-group relationship is generally low. In the 18S trees, the clade Amphibioplanidae + Dimarcusidae forms the sister-group of the Maricola with high support (Supporting Information, Fig. S1), whereas in the 28S trees the relative phylogenetic position of the Amphibioplanidae varies according to the phylogenetic inference method used. Based on the ML approach, the clade Amphibioplanidae + Dimarcusidae forms a clade with Continenticola with low support (Supporting Information, Fig. S2), whilst the BI approach suggests a sister-group relationship with Maricola with PP supports of 0.93 and 0.94 for the 28S and 28S-filtered datasets, respectively (not shown). In this case, filtering of the alignments versus non-filtering does not have any effect on the topology of the trees obtained, with only a few support values varying between the two sets of trees, but without representing significant changes in any direction (in no instance a support value changes from low to high).

The analyses based on the two concatenated datasets resulted also in each of the three main clades (Continenticola, Dimarcusidae, Maricola) being monophyletic, all with high support values (Fig. 2; Supporting Information, Fig. S3), except for Continenticola in the Bayesian analyses. In both trees, the new family Amphibioplanidae was clearly differentiated from Dimarcusidae and Maricola, but was always the sister-group of Dimarcusidae. The clade constituted by Maricola, Dimarcusidae and Amphibioplanidae is highly supported only in the BI analyses (SH-aLRT/UFB/PP: 77.2/75/1 in Fig. 2; 69/74/1 in Supporting Information Fig. S3). In both trees Amphibioplanidae are sister to Dimarcusidae, again getting high support only from the BI (78.5/81/0.99 in Fig. 2; 74.9/78/0.99 in Supporting Information Fig. S3). Therefore, based on our present analyses, Amphibioplanidae were sister to Dimarcusidae and together share a sister-group relationship with Maricola. Because our BI analyses were executed with a mixed evolutionary model, which is expected to fit better the evolution of a complex sequence such as ribosomal genes (due to the intercalation of highly conserved regions with extremely variable ones), one may consider the BI support to be of greater significance than the absence of sufficient support in the ML analyses, which used a single model for each partition. However, it should be noted that posterior probabilities may overestimate supports, even when the models match the sequence evolution (Huelsenbeck & Rannala 2004; Morrison, 2010).

Mitochondrial gene analysis

We intended to perform a population genetic analysis of the different populations of the new species with the help of the 16S mitochondrial marker, but amplification and sequencing of this marker were challenging. Thus, in view of (1) the fairly low number of sequences obtained and (2) the fact that in three cases we had to extract DNA by pooling several individuals from the same sampling locality, the population analyses were deemed impossible. In the end, the aligned sequences obtained from animals from the different caves were used to estimate genetic distances by using K2P correction (Supporting Information, Table S3). An earlier BLAST search revealed that the best hits were the few 16S sequences of various Tricladida present in GenBank. The first hit was the complete mitochondrial genome of the continenticolan *Schmidtea mediterranea* Benazzi *et al.*, 1975 (KM821047.2) with a query cover of 92% and E value of $8e^{-23}$ and a percentage identity of 67.46%, followed by hits to the other continenticolans *S. mediterranea* and *Dugesia japonica* Ichikawa & Kawakatsu, 1964 with similar values. This demonstrated that these sequences belonged to triclads, albeit that the low identity points to a distant relationship with Continenticola. The distances found for our new species from various populations varied from 0 to 15%, the latter value concerning sequences obtained from San Pietro Cave as compared with those from the Grutta 'e Pauli Cave and Istirzili Cave. Genetic distances between sequences from the Istirzili and Grutta 'e Pauli populations are ~6%, whereas within each cave genetic differentiation is small or non-existent.

Systematic account

Order Tricladida Lang, 1884

Suborder Cavernicola Sluys, 1990

Family **Amphibioplanidae** Stocchino, Sluys, Dols-Serrate & Riutort, **fam. nov.**

*Zoobank registration: lsid:*urn:lsid:zoobank.org:act:71333CDA-DE6D-40E6-A404-53C0F61974F6

Diagnosis: This new family has the same diagnostic characters as the type genus ***Amphibioplana*** Stocchino & Sluys, **gen. nov.**, described below.

Genus ***Amphibioplana*** Stocchino & Sluys, **gen. nov.**

*Zoobank registration:*lsid:urn:lsid:zoobank.org:act:EF8453A8-B54D-4935-8F0E-7E90A97FA555

Diagnosis: Slender, unpigmented and anophthalmous triclad flatworms; epidermis uniformly ciliated and uniformly provided with numerous rhabdites;

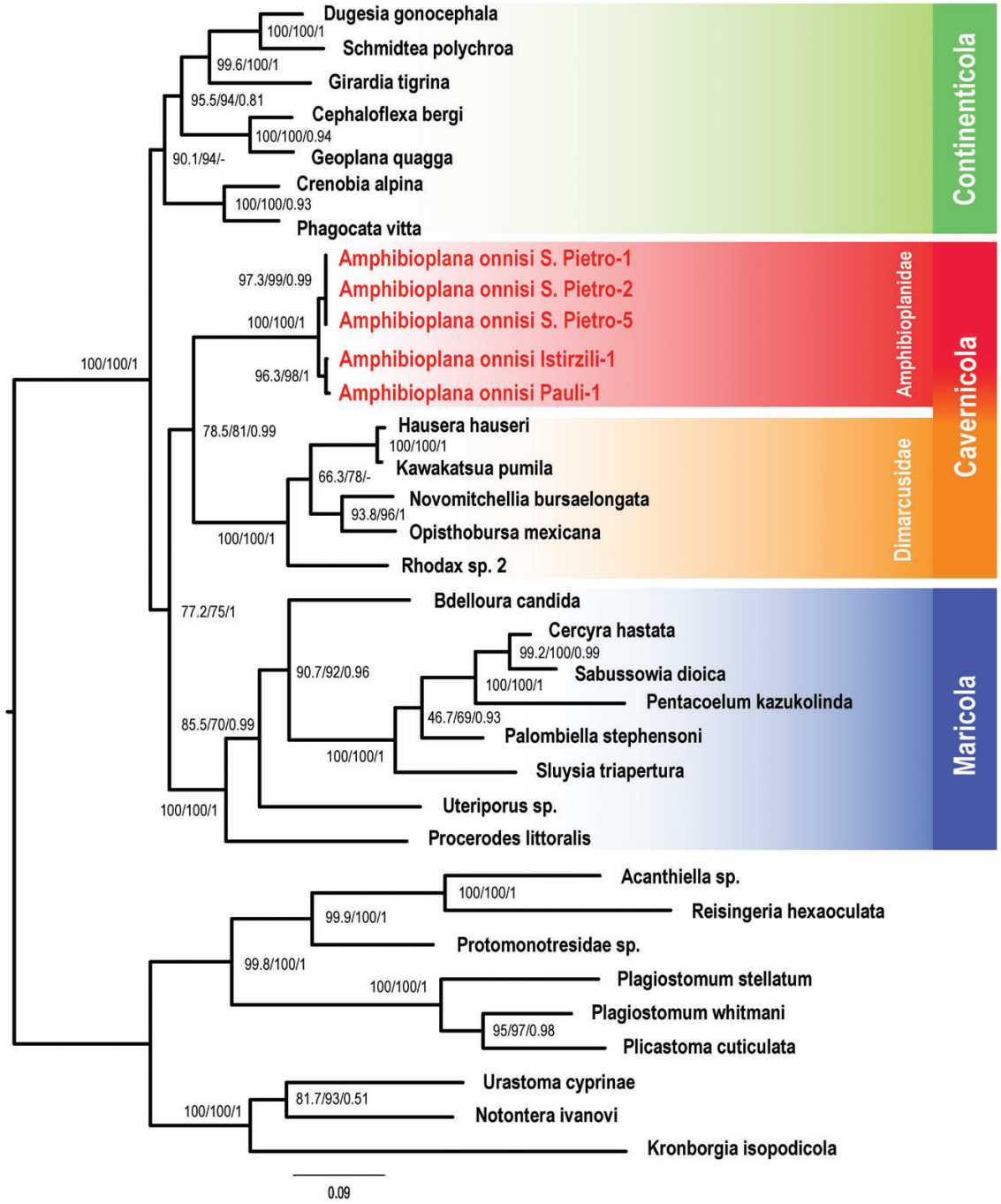


Figure 2. Maximum-likelihood tree inferred from the concatenated dataset, including the filtered 18S and 28S alignments (Table 2) of representatives of the various suborders of the Tricladida. Values at nodes correspond to UFB/SH-aLRT/PP; –: PP values below 0.5. Scale bar: number of substitutions per nucleotide position.

subepidermal musculature consisting of a thin layer of circular muscles, followed by an equally thin layer of longitudinal muscles; pharynx and copulatory apparatus located well into the posterior half of the body; convoluted pharynx provided with oesophageal projection; large and well-developed testes, occupying entire dorsoventral space of the body; testes follicles distributed in two rows, with the most anterior pair of testes situated far behind the brain and located anteriorly to ovaries, while the most posterior testes are located far anterior to the root of the pharynx; ovaries situated far behind the brain; oviducts posteriorly to the gonopore uniting to form a long common oviduct; bursal

canal giving rise to a diverticulum that communicates with the common oviduct; copulatory bursa lined with a highly vacuolated, syncytial epithelium; diaphragm between bursal canal and copulatory bursa; gonopore opening into a cup-shaped chamber that via a small dorsal pore communicates with the atrium; habitat groundwater in microcrevices of the epikarst, and temporary cave pools, but able to survive in a basically terrestrial environment (amphibious lifestyle). *Type species:* ***Amphibioplana onnisi*** Stocchino & Sluys, **sp. nov.**

Etymology: The generic name is derived from the Greek αμφιβίοσ (*amphibios*), living on both land and water, and Latin *planus*, flat, thus referring to the lifestyle and habitat of this new species, while alluding also to the fact that it belongs to the flatworms. Gender: feminine.

Amphibioplana onnisi Stocchino & Sluys, sp. nov.

(Figs 3–15)

Zoobank registration: lsid: urn:lsid:zoobank.org:act:01B99CA0-9692-477F-8913-8A476638C57B

Diagnosis: As that of the genus.

Etymology: The specific epithet honours the Sardinian biospeleologist and co-author Carlo Onnis, who discovered and sampled all populations of the new species.

Material examined: Holotype: RMNH.VER. 19956.a, Grutta 'e Pauli Cave (39°53'13.45"N, 9°20'39.47"E), terrestrial habitat, Seui, Sardinia, Italy, 18 January 2015, coll. C. Onnis, horizontal sections (hor. sect.) of anterior portion on one slide; sagittal sections (sag. sect.) of posterior portion on two slides.

Paratypes: CGAS Pla 19.1a, ibid., sag. sect. of anterior portion on two slides; CGAS Pla 19.1b, ibid., sag. sect. of posterior portion on two slides; CGAS Pla 19.2, ibid., sag. sect. on three slides; RMNH.VER. 19956.b, ibid., sag. sect. on two slides; RMNH.VER. 19956.c, ibid., sag. sect. on two slides; RMNH.VER. 19956.d, ibid., sag. sect. on two slides.

Other material: Grutta 'e Pauli Cave (39°53'13.45"N, 9°20'39.47"E), Seui, Sardinia: CGAS Pla 20.1, freshwater habitat 11 October 2015, coll. C. Onnis, sag. sect. on two slides; CGAS Pla 20.2, ibid., sag. sect. on two slides; CGAS Pla 20.3, ibid., sag. sect. on three slides; CGAS Pla 20.4, ibid., sag. sect. on three slides; CGAS Pla 20.5, ibid., sag. sect. on two slides; RMNH.VER. 19957.a, ibid., sag. sect. on two slides; CGAS Pla 20.6, ibid., sag. sect. on two slides; RMNH.VER. 19957.b, ibid., sag. sect. on two slides, RMNH.VER. 19957.c, ibid., sag. sect. on two slides, RMNH.VER. 19957.d, ibid., sag. sect. of anterior portion on two slides; RMNH.VER. 19957.e, ibid., hor. sect. of posterior portion on one slide. CGAS Pla 21, freshwater habitat, 26 March 2018, coll. C. Onnis, six entire specimens fixed in Bouin's fluid and preserved in ethanol.

Istirzili Cave (40°04'48.29"N, 9°37'12.50"E), Baunei, Sardinia: RMNH.VER. 19958.a, freshwater habitat, 22 January 2017, coll. C. Onnis, sag. sect. on two slides; CGAS Pla 22.1, ibid., sag. sect. on three slides; CGAS Pla 22.2, ibid., sag. sect. on two slides; CGAS Pla 22.3, ibid., sag. sect. on two slides; RMNH.VER. 19958.b, ibid., sag. sect. on one slide; CGAS Pla 22.4, ibid., sag. sect. on one slide; RMNH.VER. 19958.c, ibid., sag. sect. on three slides; RMNH.VER. 19958.d, ibid., sag. sect. on one slide; RMNH.VER. 19958.e, ibid., sag. sect. on two slides; RMNH.VER. 19958.f, ibid., hor. sect. on one slide; CGAS Pla 22.5, ibid., hor. sect. on two slides; CGAS Pla 22.6, ibid., transverse sections (transv. sect.) on five slides.

San Pietro Cave (39°24'27.38"N, 8°29'35.40"E), Fluminimaggiore, Sardinia: CGAS Pla 23, terrestrial habitat, 26 June 2015, coll. C. Onnis, sag. sect. on one slide. RMNH.VER. 19959.a, freshwater habitat, 5 October 2015, coll. C. Onnis, sag. sect. on one slide. CGAS Pla 24.1, freshwater habitat, May 2016, coll. C. Onnis, sag. sect. on one slide; RMNH.VER. 19960.a, ibid., sag. sect. on one slide; CGAS Pla 24.2, ibid., sag. sect. on one slide; RMNH.VER. 19960.b, ibid., sag. sect. on one slide; CGAS Pla 24.3, ibid., sag. sect. on one slide; RMNH.VER. 19960.c, ibid., sag. sect. on one slide; CGAS Pla 24.4, ibid., sag. sect. on one slide; CGAS Pla 24.5, ibid., sag. sect. on one slide; CGAS Pla 24.6, ibid., sag. sect. on one slide; CGAS Pla 24.7, ibid., sag. sect. on one slide; RMNH.VER. 19960.d, ibid., sag. sect. on one slide; CGAS Pla 24.8, ibid., sag. sect. on one slide; CGAS Pla 24.9, ibid., sag. sect. on one slide; CGAS Pla 24.10, ibid., sag. sect. on two slides; CGAS Pla 24.11, ibid., sag. sect. on two slides; CGAS Pla 24.12, ibid., sag. sect. on two slides; CGAS Pla 24.13, ibid., sag. sect. on two slides; CGAS Pla 24.14, ibid., sag. sect. on two slides; CGAS Pla 24.15, ibid., sag. sect. on two slides; RMNH.VER. 19960.e, ibid., sag. sect. on one slide; CGAS Pla 24.16, ibid., sag. sect. on one slide; CGAS Pla 24.17, ibid., sag. sect. on one slide; CGAS Pla 24.18, ibid., sag. sect. on one slide; CGAS Pla 24.19, ibid., sag. sect. on one slide; CGAS Pla 24.20, ibid., sag. sect. on one slide; RMNH.VER. 19960.f, ibid., sag. sect. on one slide; CGAS Pla 24.21, ibid., sag. sect. on two slides; CGAS Pla 24.22, ibid., sag. sect. on one slide; RMNH.VER. 19960.g, ibid., hor. sect. on one slide; CGAS Pla 24.23, ibid., hor. sect. on one slide; RMNH.VER. 19960.h, ibid., transv. sect. on four slides.

Description

External features: Body size of live specimens ranges from 4 to 6 mm long and 0.4–0.8 mm wide. Preserved

Figure 3. *Amphibioplana onnisi*. A single, curled up specimen from under a stone in the Grutta 'e Pauli Cave. Scale bar not available.

specimens measured up to 3.5–4.0 mm long and 0.4–0.8 mm wide. The animals are unpigmented and fairly transparent *in vivo*, so that the intestinal branches, brain, testes, pharynx and copulatory apparatus are clearly visible. The truncated head is devoid of eyes; posterior end rounded (Fig. 5).

Epidermis and cutaneous musculature: The entire epidermis is constituted of a cuboidal, nucleated epithelium that is completely clothed in cilia that have the same length on both dorsal and ventral body surface; dorsal and ventral epidermis have the same height. Rhabdites are abundant and uniformly distributed on the entire body surface, with erythrophil, rhabdite-forming glands being located in the parenchyma underneath the subepidermal musculature (Figs 6A, 7B, 10C). Dorsal and ventral rhabdites have the same height. The epidermis also receives a coarse granular, xanthophil secretion from glands that are mainly located in the anterior region of the body, beneath the subepidermal musculature, with several glands also lying scattered in the parenchyma (Figs 6A, 7A). The epidermis receives also an amorphous secretion from cyanophil glands located beneath the subepidermal

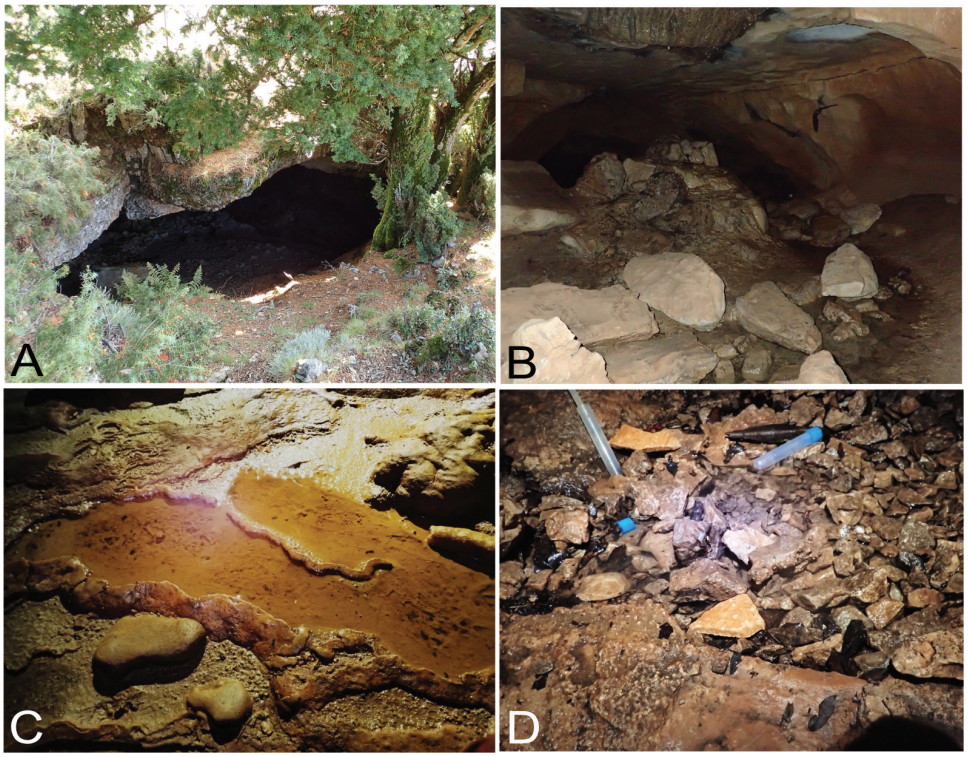


Figure 4. *Amphibioplana onnisi*. Habitat in freshwater environments. A, entrance of Grutta 'e Pauli Cave; B, collection site in Grutta 'e Pauli Cave; C, collection site in San Pietro Cave; D, collection site in Istirzili Cave. Details on precise locations of the sampling sites in the caves are omitted for reasons of species protection.

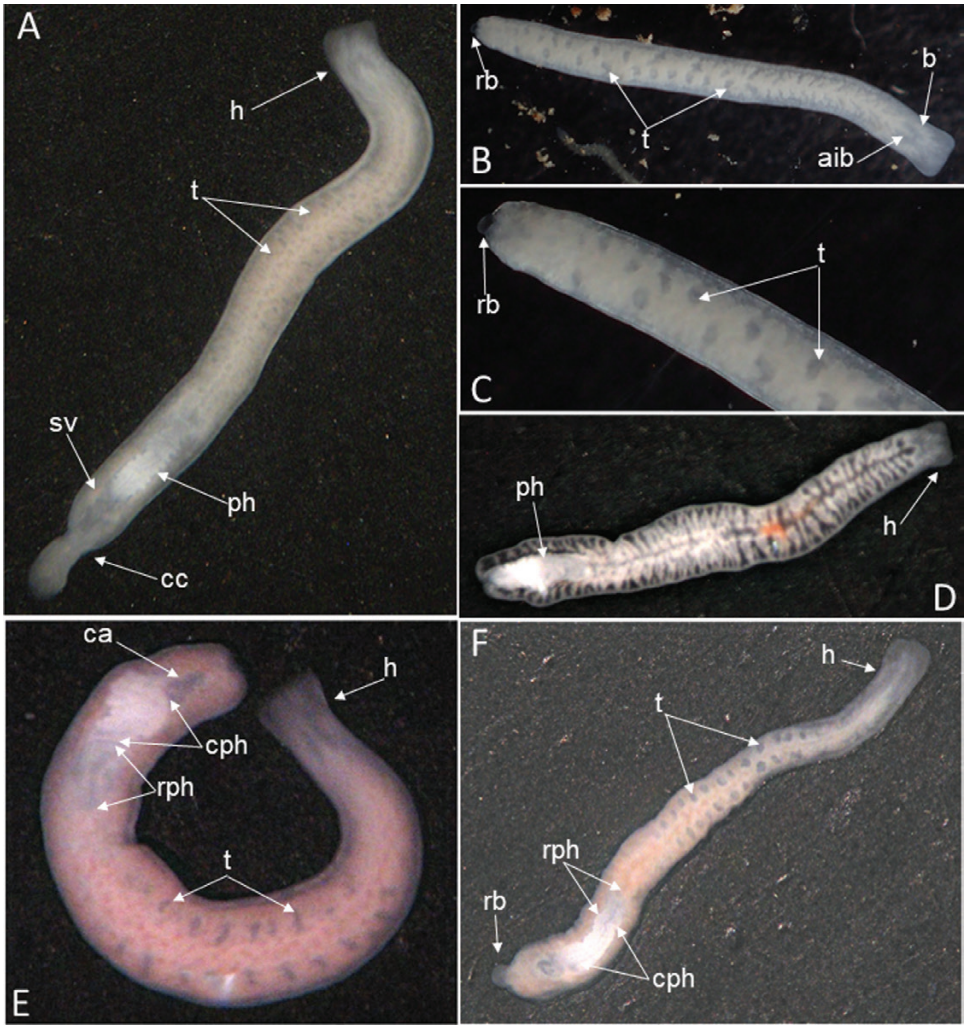


Figure 5. *Amphibioplana onnisi.* Habitus of live specimens from the Grutta 'e Pauli Cave. A, specimen with constriction just behind copulatory apparatus; B, C, anterior portion of specimen with regenerating blastema (rb) at the margin of the prepharyngeal part of the body; D, specimen with caudal tip located just behind the pharynx, suggesting a recent postpharyngeal fission; E, specimen with caudal tip located just behind the copulatory apparatus (ca), suggesting a recent caudal fission; F, specimen with a regenerating blastema behind the copulatory apparatus. Scale bars not available.

musculature (Fig. 10B, C). Underneath the epidermis, the basement membrane is evident as a pale blue, amorphous line (Fig. 10C). The openings of adhesive glands, producing a fine, granular, erythrophil secretion, are concentrated at the front end and at the posterior tip of the body (Figs 6B, 7A). Subepidermal musculature consists of only two thin layers: an outer, subepidermal layer of circular muscles and an inner layer of longitudinal muscles (Figs 10B, C, 11, 12). Dorsal and ventral musculatures show the same thickness.

Nervous system: In horizontal sections, the bilaterally symmetrical brain has the shape of an 'H' and lies on top of the two ventral nerve cords, which run in the caudal direction along the length of the animal (Figs 7A, 10C). The central neuropil of the brain is surrounded by an outer layer of neuronal cell bodies (Fig. 7A). In none of the specimens examined was it possible to discern structures related to an auricular sense organ, in view of the fact that the epidermal cells are of a uniform height and are all provided with both cilia and rhabdites.

Pharynx and intestine: The unpigmented, plicate pharynx is located in the posterior half of the body and measures ~one-tenth of the body length, at least when it is situated within the pharyngeal pocket. In point of fact, the actual size of the pharynx must be greater than this as it is considerably folded and

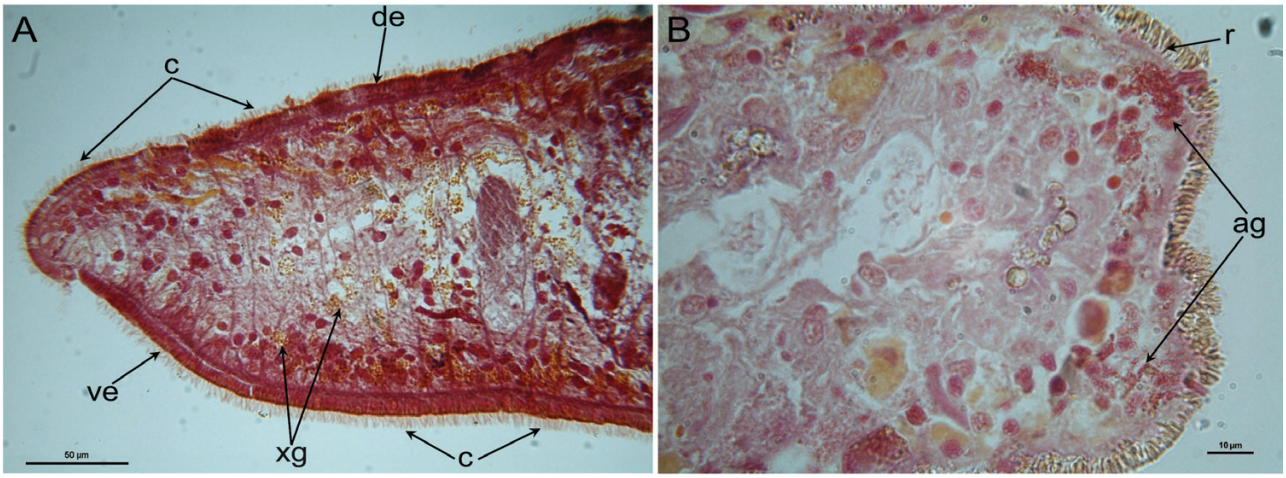


Figure 6. *Amphibioplana onnisi.* A, RMNH.VER. 19957.a, photomicrograph of sagittal section of the anterior end, showing dorsal (de) and ventral epidermis (ve) being uniformly clothed with cilia (c); B, RMNH.VER. 19960.g, photomicrograph of horizontal section of the posterior end, showing adhesive glands (ag) discharging their secretion through the epidermis.

convoluted at its median and distal tracts (Figs 5F, 8). In specimens from Grutta 'e Pauli Cave and Istirzili Cave the proximal tract of the pharynx, corresponding to one-third of its total length, has a rectilinear course and at its root it is provided with a kind of oesophagal protrusion that actually projects into the pharynx lumen, instead of the gut. In specimens from San Pietro Cave the pharynx has a less convoluted shape, while the oesophagal protrusion is smaller than in specimens of the other two populations. This oesophagal protrusion is constituted by gut tissue traversed for almost its entire length by thin, longitudinal muscle fibres (Fig. 8B). The epithelium lining the proximal, anterior part of the pharynx lumen is provided with long cilia. The muscle coat underneath the outer pharynx epithelium is bilayered, i.e. constituted by a subepithelial layer of longitudinal muscle, followed by a layer of circular muscles. The musculature underneath the epithelium lining the pharynx lumen is composed of a thick, subepithelial layer of circular muscles, followed by a layer of longitudinal muscles. Two types of pharyngeal glands are evident: one type producing a fine-grained erythrophil secretion and the other a coarse-grained xanthophil secretion. In the majority of specimens, the mouth opening is located at the posterior end of the pharyngeal pouch, but it may be shifted anteriad, so that in some cases it is located at ~one-sixth, and in one case at ~one-third, of the distance between the posterior end of the pharyngeal pouch and the root of the pharynx. The intestine is of the usual tripartite shape, with the anterior intestinal trunk terminating just behind the brain. Posterior intestinal branches unite behind the copulatory apparatus to form a single branch (Figs 5, 7, 9A).

Female reproductive system: The two ovoid ovaries are located ventrally, far behind the brain and just behind the most anterior pair of testes. The female gonads are located at ~one-third to one-quarter of the distance between the brain and the root of the pharynx, i.e. at ~one-eighth of the distance between the brain and the posterior end of the body. The ovaries occupy one-half (specimen RMNH.VER. 19956.d) or three-quarters (specimen CGAS Pla 19.2) of the dorsoventral space. The anterior portion of each oviduct is expanded to form a tuba, which communicates with the dorsolateral part of the corresponding ovary (Figs 9A, 10). The oviducts run ventrolaterally to the vitellaria and laterally to the ventral nerve cords, and extend posteriorly to beyond the level of the genital pore and up to the level of the penultimate vitellarium. Hereafter, the oviducts bend medially and subsequently unite to form a long common oviduct. The common oviduct runs anteriad and then communicates with the posterior diverticulum of the bursal canal (Figs 7A, 9A, 10–13, 15). Oviducts and common oviduct are lined with an epithelium consisting of cuboidal, nucleated and ciliated cells. The posterior part of the common oviduct is surrounded by a thin layer of longitudinal muscles that becomes even less distinct towards the point of fusion of the two oviducts. We were unable to discern muscles around the oviducts so that, apparently, these ducts are devoid of any musculature (Figs 11, 12, 15). The lumen of the common oviduct, as well as that of the oviducts, is always well discernible. Vitellaria are situated laterally to the testes and between the intestinal branches, extending from the level of the ovaries to behind the copulatory apparatus; they are well developed and occupy almost the entire dorsoventral space of the body (Figs 7, 9, 10). The copulatory bursa, located just behind the

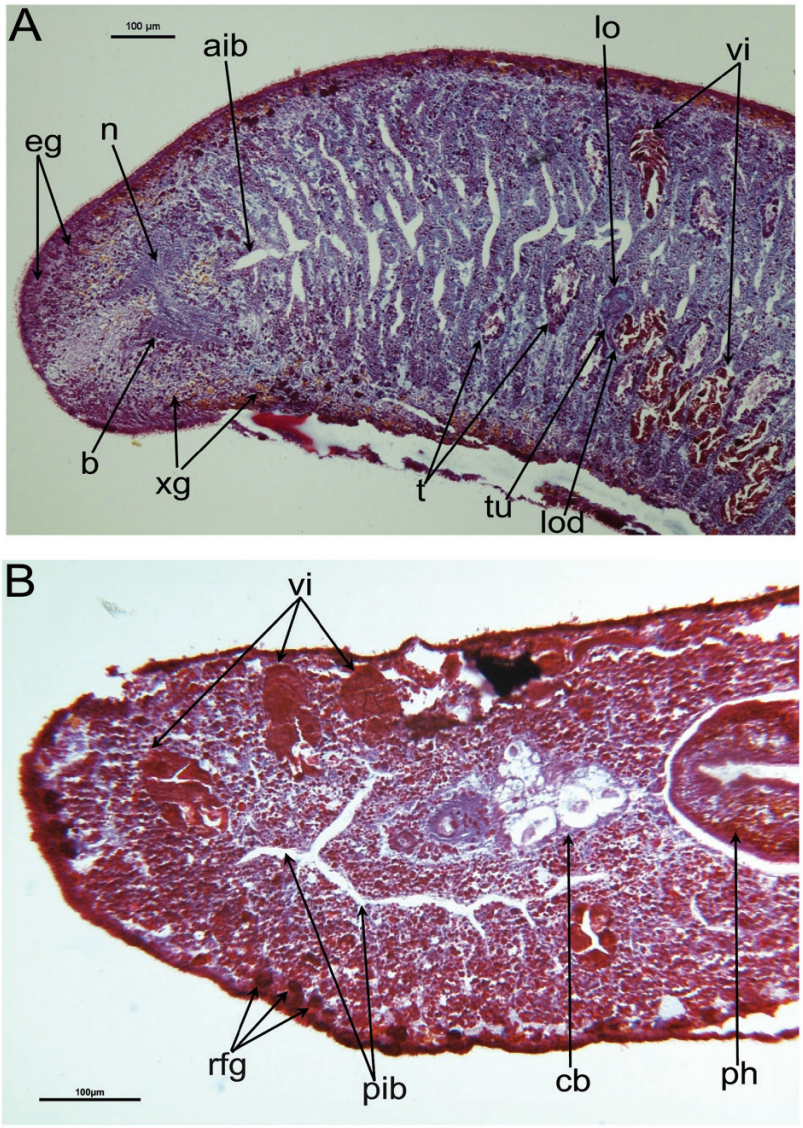


Figure 7. *Amphibioplana onnisi.* A, holotype RMNH.VER. 19956.a, photomicrograph of a horizontal section of the head, with the brain (b) and the post-cephalic region, also showing the anterior intestinal branch (aib), left ovary (lo) with its tuba (tu) and the first tract of the left oviduct (lod), the most anterior vitellaria (vi) and testes (t); B, RMNH.VER. 19958.f, photomicrograph of a horizontal section of the post-pharyngeal region with the copulatory apparatus, showing the most caudal vitellaria (vi) and the two caudal gut trunks fusing to form a single branch (pib).

pharyngeal pouch, is lined by a glandular and highly vacuolated, syncytial epithelium with scattered nuclei. Many of the vacuoles are richly provided with an amorphous, erythrophil and/or cyanophil secretion and in some cases contain sperm (Figs 7B, 11–13, 14B–E, 15). In the majority of the specimens, the bursal lumen contains a mass of free sperm and, in some cases, is so voluminous that it noticeably compresses the lining epithelium of the bursa (e.g. specimen RMNH.VER. 19958.a; see Fig. 14C). The bursa is surrounded by a layer of muscles. The size of the copulatory bursa varies between specimens, but in the majority of cases it is large and occupies the entire dorsoventral space, while occasionally it occupies only the dorsal half of the body space. The bursal canal originates from the posterior wall of the genital atrium and immediately thereafter gives rise to a posterior expansion that forms a kind of diverticulum, which receives the secretion of erythrophil shell glands and communicates with the common oviduct. From thereon, the bursal canal curves anterodorsad, while running, usually, to the right side of the rest of the copulatory apparatus and then communicates with

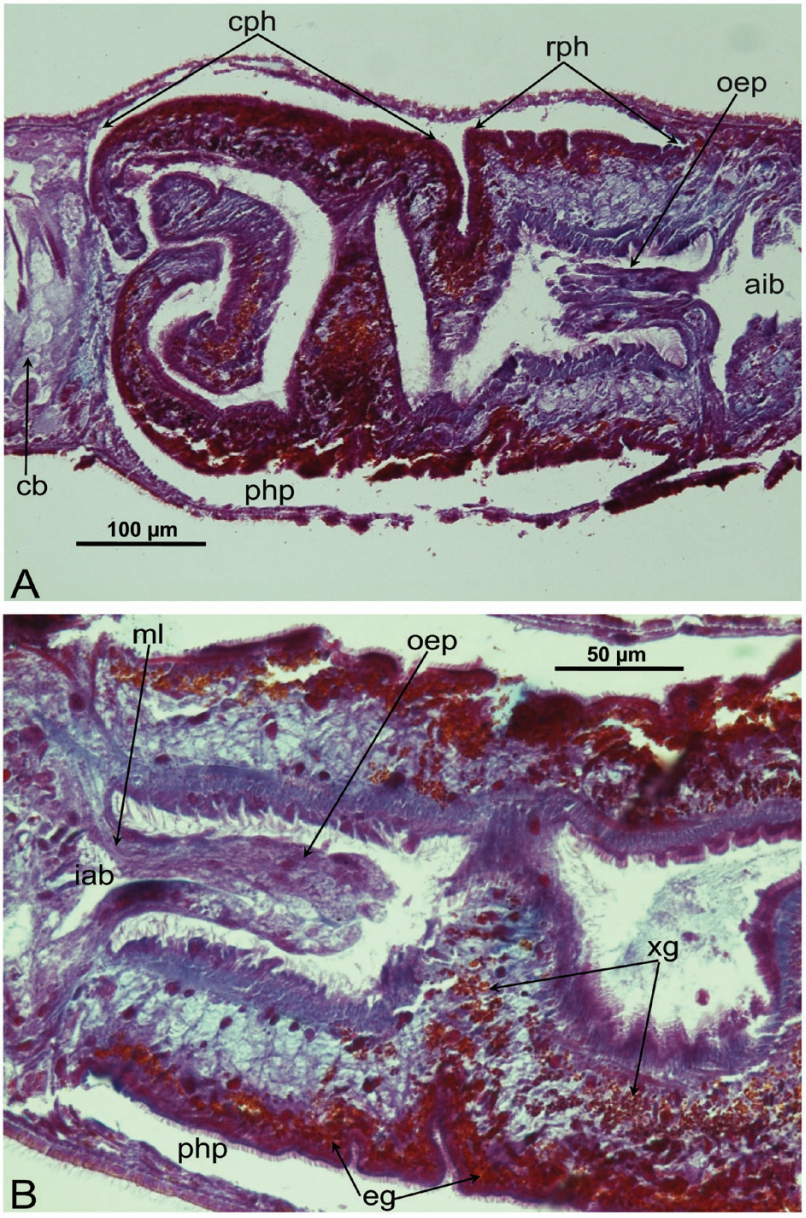


Figure 8. *Amphibioplana onnisi.* Photomicrographs of the pharynx, sagittal sections. A, CGAS Pla 19.2, oesophageal protrusion (oep) and convoluted tract of the pharynx (cph); anterior to the right; B, RMNH.VER. 19956.c, oesophageal protrusion (oep), also showing xanthophil (xg) and erythrophil (eg) secretions in the pharynx; anterior to the left.

the copulatory bursa. For a short distance, the bursal canal protrudes into the bursa, thus forming a kind of valve or diaphragm (Figs 11–13, 14E). The bursal canal is lined with an epithelium consisting of cuboidal or cylindrical, nucleated cells and is surrounded by a layer of longitudinal muscles (Figs 11, 12, 15A). The epithelial cells of the bursal canal, which are provided with long cilia, produce an erythrophil secretion that is discharged into the lumen. The point of transition between bursal canal diverticulum and common oviduct is not always clear-cut, but can be recognized by the fact that the common oviduct has a smaller diameter and is lined with a lower epithelium, provided with shorter cilia, and is surrounded by a thinner coat of muscles. In addition, oviducts and common oviduct stain less strongly than the bursal canal diverticulum.

Male reproductive system: The well-developed testes are arranged in two regular rows, one row on either side of the body, with the follicles occupying the entire

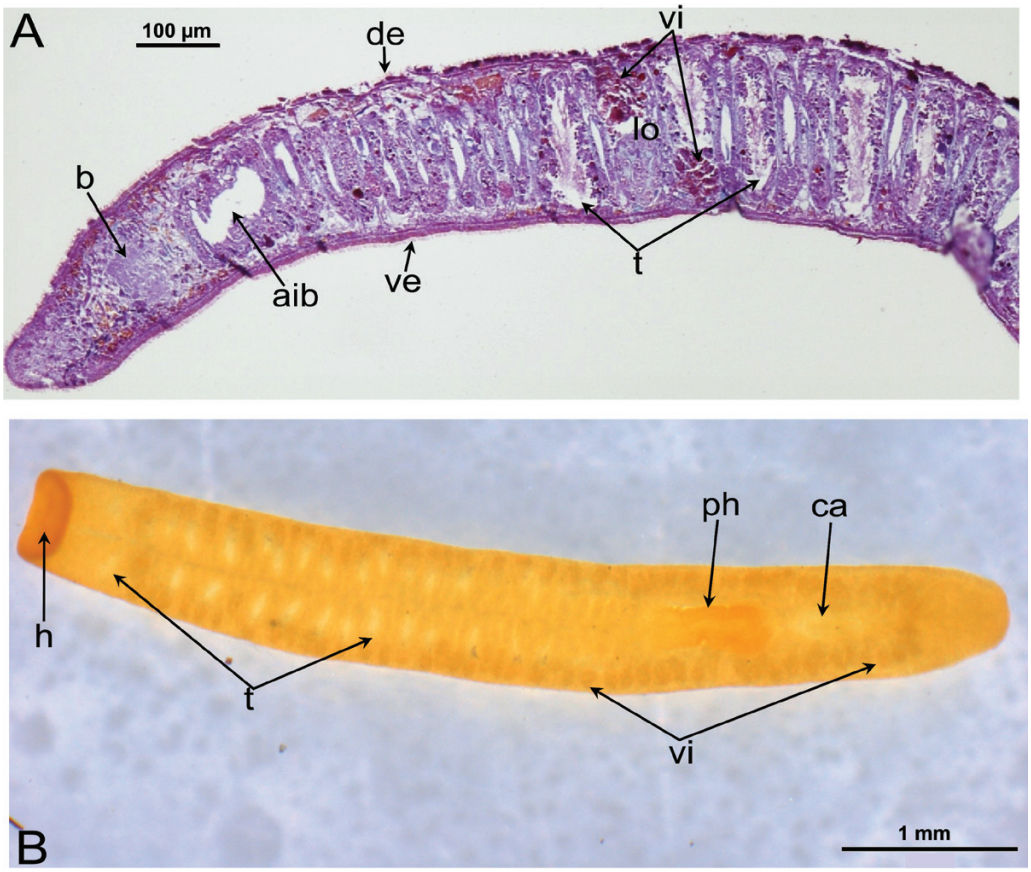


Figure 9. *Amphibioplana onnisi.* A, RMNH.VER. 19956.d, photomicrograph of a sagittal section of the prepharyngeal region, showing the brain (b), the anterior intestinal branch (aib), the most anterior testes (t) and vitellaria (vi) and the left ovary (lo); B, photograph of a preserved specimen from the Grutta 'e Pauli Cave in ventral view, showing the vitellaria (vi), testes (t), pharynx (ph), and copulatory apparatus (ca).

dorsoventral space. The most anterior testes are situated far behind the brain, albeit anteriorly to the ovaries, and from thereon extend caudally, with the most posterior follicles located far (~ 500–550 µm) from the root of the pharynx (Figs 5, 6A, 9, 10). The vasa deferentia run dorsally to the ventral nerve cords and already in the prepharyngeal part of the body they form voluminous spermiducal vesicles that are packed with sperm, at least in sexually mature specimens (Figs 5A, 11–13, 15A). The small and sometimes indistinct penis bulb is generally located just behind the copulatory bursa. However, in those specimens in which the bursa is small, the penis bulb almost touches the caudal wall of the pharyngeal pouch. At the level of the penis bulb, the spermiducal vesicles notably decrease in diameter and hereafter they curve dorsad and then abruptly penetrate the anteroventral wall of the penis bulb, after which they symmetrically ascend to reach the level of the medial part of the penis. After a while, the ducts curve again abruptly or via a knee-shaped bend and then turn posteriad, after which they run parallel and close to each other through the central portion of the penis papilla. At ~three-quarters of the length of the papilla, the vasa deferentia fuse to form a short ejaculatory duct, which runs a central course through the remainder of the penis papilla and opens at its tip (Figs 11–13, 14A–D, F, 15B). The vasa deferentia are lined with an epithelium consisting of low, cuboidal and nucleated cells and are surrounded by a layer of circular muscle. The ejaculatory duct is lined with cuboidal, nucleated cells. The symmetrical, conical penis papilla is covered by a flat infranucleated epithelium, which is underlain by a subepithelial layer of circular muscle (Figs 11–14A–D, F, 15B). Weak, finely granular, erythrophil secretions were discerned in the penis papilla. The genital atrium, which houses the penis papilla, is lined by an epithelium consisting of, cuboidal-columnar, infranucleated cells with a weak erythrophil secretion that is underlain by a subepithelial layer of circular muscle, followed by a layer of longitudinal muscle fibres (Figs 11–14A–D, F, 15). In all specimens examined, the ventral gonopore directly leads to a cup-shaped chamber, which is formed by an

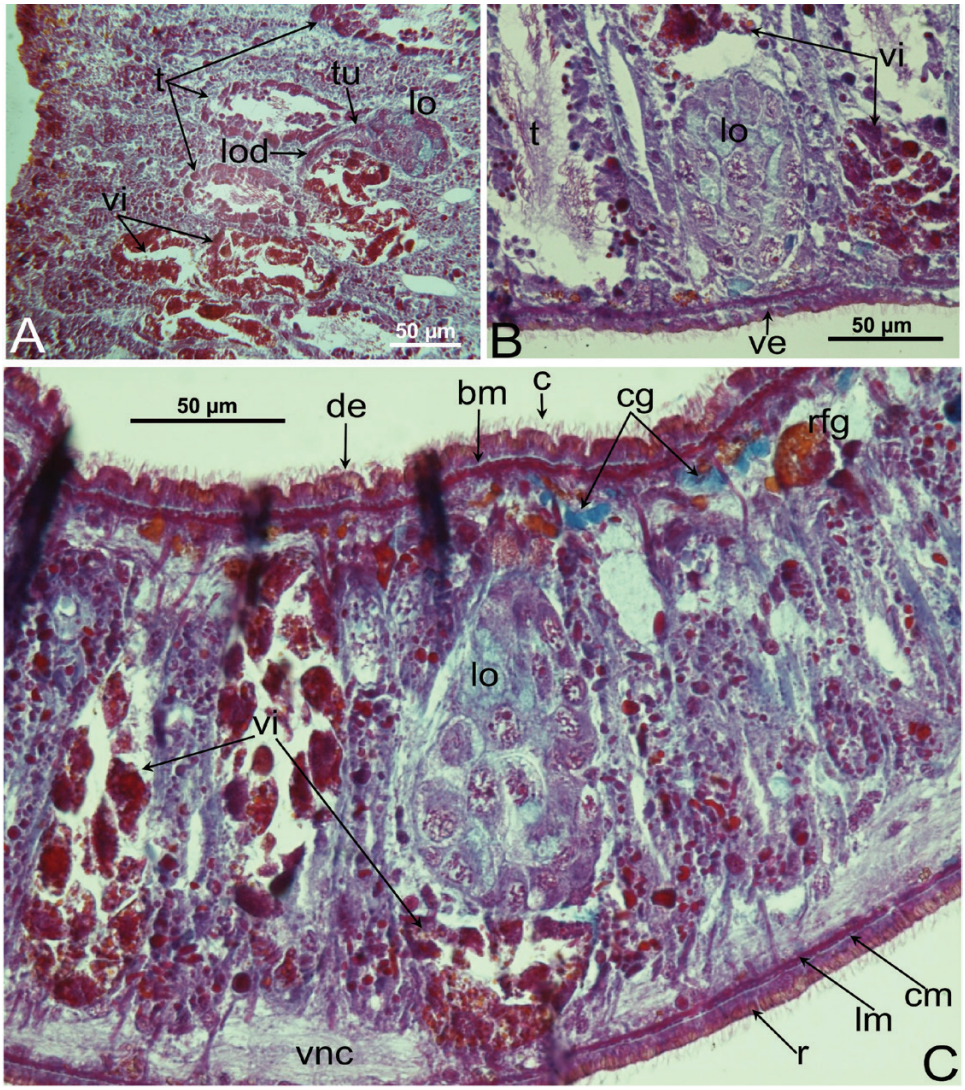


Figure 10. *Amphibioplana onnisi.* A, holotype RMNH.VER. 19956.a, photomicrograph of a horizontal section of the left ovary (lo) with tuba (tu) and origin of the left oviduct (lod), testes (t) and vitellaria (vi); B, RMNH.VER. 19956.d, photomicrograph of a sagittal section at the level of the left ovary (lo), with testes (t) and vitellaria (vi); C, CGAS Pla 19.2, photomicrograph of a sagittal section at the level of the left ovary (lo), showing in detail the dorsal (de) and ventral epidermis (ve) with the subepidermal musculature, some vitellaria (vi), and the ventral nerve cord (vnc).

invagination of the ventral epidermis, as it is lined by a ciliated epithelium that is provided with rhabdites and a subepidermal layer of circular muscles. The dorsal wall of this chamber is provided with a small dorsal pore through which it communicates with the atrium (Figs 11, 12, 14A, B, 15B). Cement glands opening into the gonopore or the chamber were not discerned.

Geographical distribution: Exclusively known from three karstic caves in Sardinia, Italy (Fig. 1).

Habitat: Planarians were found in three karstic caves, characterized by intermittent presence of dripping water. These caves are located in three karstic massifs that are greatly isolated from each other, representing distinct hydrological systems (Cassola, 1982).

The Grutta 'e Pauli Cave is located at an elevation of ~1025 m above sea level (a.s.l.) in the karstic Mesozoic (Jurassic) massif of the Tacchi d'Ogliastra in the central-eastern part of Sardinia (Figs 1, 4A). Specimens were first collected at ~40 m from the entrance from under stones that were firmly embedded in humid, clayey soil. Some individuals were found adhering to the stones, while others were found on the clayey soil, thus in a basically terrestrial habitat. Animals were motionless and tightly curled up (Fig. 3). However,

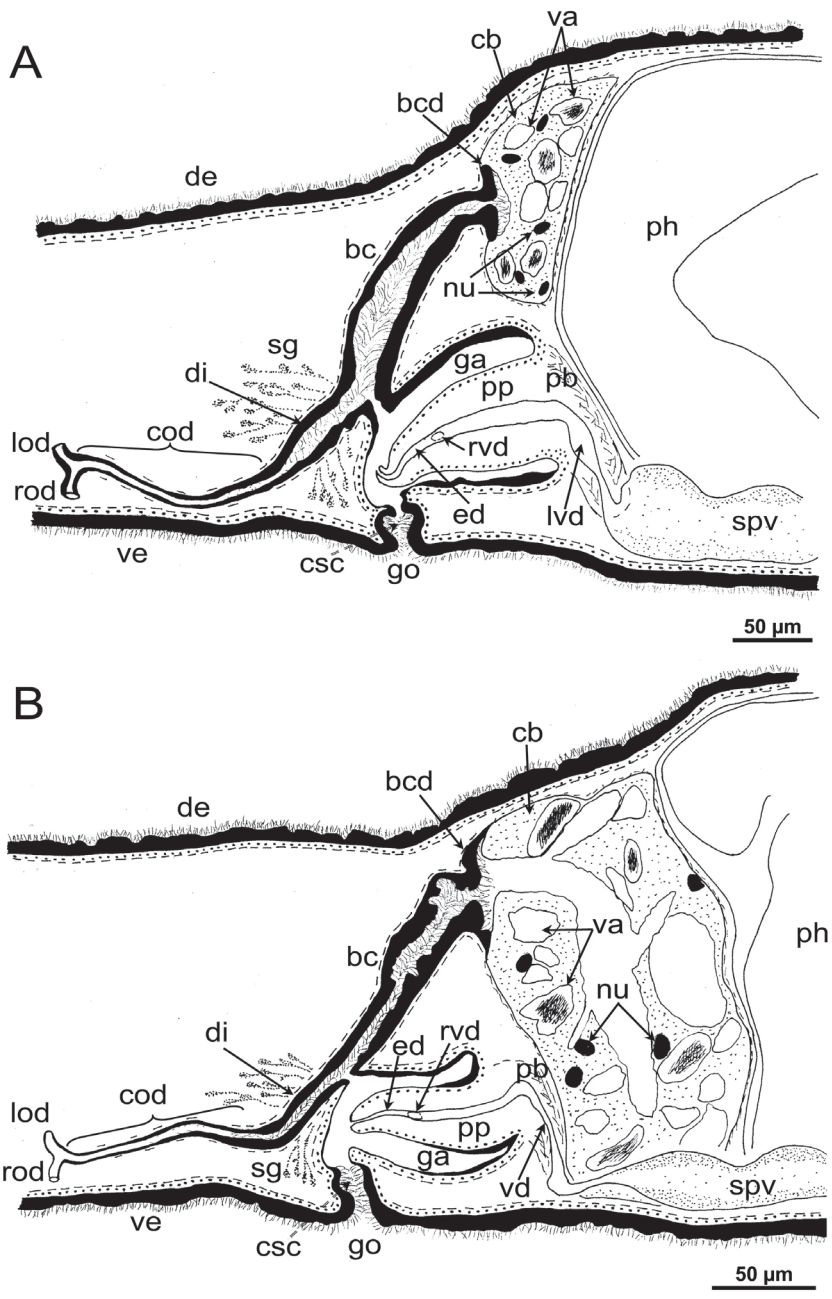


Figure 11. *Amphibioplana onnisi.* A, holotype RMNH.VER. 19956.a, sagittal reconstruction of the copulatory apparatus; anterior to the right; B, CGAS Pla 22.3, sagittal reconstruction of the copulatory apparatus; anterior to the right.

after only having been touched with a small paint brush they started to move, after which the worms were transferred to a vial with water in which, subsequently, they behaved as freshwater planarians, exhibiting the gliding manner of locomotion. Subsequent to this first collection, animals were found also in temporary pools, formed by the dripping water and provided with a clayey bottom, at ~50 m from the entrance of this cave (Fig. 4B); associated fauna was exclusively Ostracoda.

The San Pietro Cave is located at an elevation of ~225 m a.s.l. at the south-western coast of Sardinia, in the Iglesiente region, which is considered to be the most ancient karst of Italy, dating back to the Cambrian (Fig. 1). Flatworms were found under stones that were firmly embedded in a humid, clayey substrate, at ~50 m from the entrance, as well as in some rimstone basins or *gours*, with a muddy bottom, at ~260 m from the entrance of the cave (Fig. 4C).

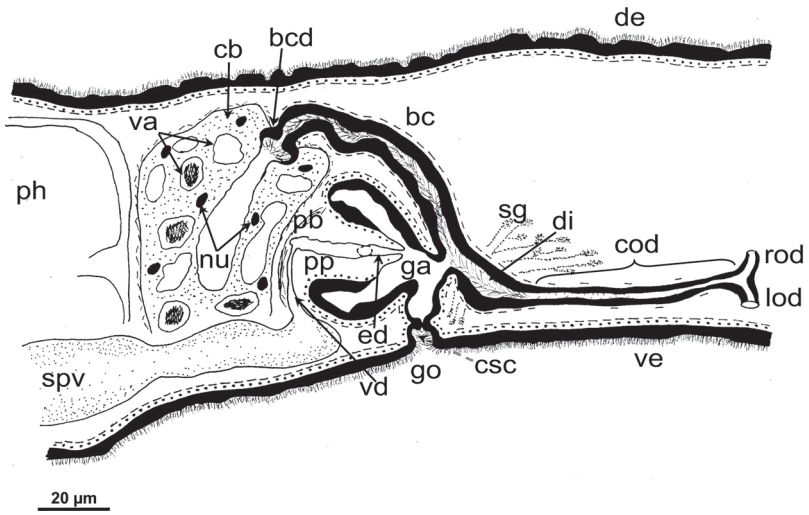


Figure 12. *Amphibioplana onnisi.* CGAS Pla 24.5, sagittal reconstruction of the copulatory apparatus; anterior to the left.

While under the stones, the worms were motionless and tightly curled up.

The Istirzili Cave is located at an elevation of ~507 m a.s.l. in the karstic Mesozoic (Jurassic) highland massif of the Supramonte of Baunei in the central-east of Sardinia (Fig. 1). Planarians were found only in a small temporary pool located near (~5 m) the cave entrance, with a clay bottom characterized by the presence of small stones and leaves (Fig. 4D).

Reproduction: The great majority of individuals were sexual at collection, but cocoons were not found in any of the three caves. Furthermore, signs of presumably asexual reproduction by fission were displayed by some sexually mature specimens from the population from Grutta 'e Pauli Cave. These animals showed five different conditions: (1) a transverse constriction (some individuals) just caudally to the copulatory apparatus, suggesting a preliminary phase of fission (Fig. 5A); (2) a small, caudal regenerating blastema (one individual), suggesting a previous process of prepharyngeal fission (Fig. 5B, C); (3) a body in which the caudal tip was located just behind the pharynx, suggesting a recent postpharyngeal fission (one individual) (Fig. 5D); (4) a body in which the caudal end was located just behind the copulatory apparatus, suggesting a presumably recent fission (one individual) just caudally to the copulatory apparatus (Fig. 5E); (5) a small, regenerating blastema at the current tail end, suggesting a previous fission (one individual) just caudally to the copulatory apparatus (Fig. 5F).

Movement and locomotion: Similar to typical freshwater planarians, the newly collected Sardinian animals characteristically exhibited a gliding motion over the substrate. However, during locomotion the worms tend to adhere temporarily to the substrate by their posterior end. The phase of adhesion was characterized by the anterior part of the animal moving rapidly with sideways swinging movements, as if exploring the surrounding environment, after which the animal releases itself from the substrate and resumes its gliding motion.

DISCUSSION

Taxonomic features and assignment

Amphibioplana onnisi shows a combination of morphological characters that does not conform with any of the subordinate taxa of the Continenticola, Maricola or Dimarcusidae. This unique combination of characters, and its mismatch with any of the known taxa among the Tricladida, will be discussed below.

With respect to freshwater planarians, presence of a planariid type of pharynx and the opening of a common oviduct into the bursal canal preclude inclusion of *A. onnisi* with either the Dendrocoelidae or the Planariidae, the first-mentioned family being characterized by a different kind of pharynx musculature and both families by a common oviduct that opens into the atrium. In the Dugesiidae, the oviducts open directly into the bursal canal, either separately or combined. An exception is formed by species of the genus *Romankenkius* Ball, 1974, in which the oviducts, separately or combined, open into a diverticulum of the bursal canal (cf. Sluys, 2001), a situation that resembles the condition in *A. onnisi*. However, *A. onnisi* differs from species of *Romankenkius* in the absence of caudally branched oviducts. *Amphibioplana onnisi* differs from representatives of the Kenkiidae in the absence of an anterior adhesive organ, as well as

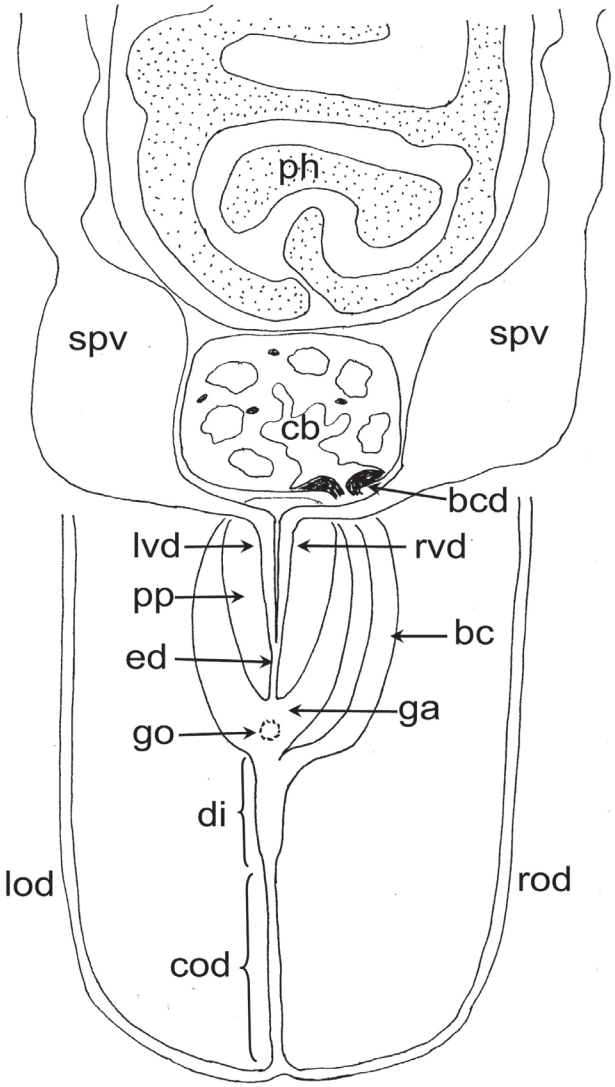


Figure 13. *Amphibioplana onnisi*. Schematic horizontal reconstruction of the copulatory apparatus.

thickened body margins, the last-mentioned feature forming the presumed apomorphy for this family (Sluys & Kawakatsu, 2006).

The genus *Amphibioplana* cannot be ranked among the terrestrial planarians of the family Geoplanidae as it lacks a creeping sole and the diploneuran nervous system. Moreover, our molecular results showed that the Amphibioplanidae clade is not closely related to the Continenticola, thus reinforcing the conclusions inferred from the morphological characters, as discussed above. The molecular phylogenetic trees revealed that the Amphibioplanidae share a sister-group relationship with the Dimarcusidae and that together they are sister to the Maricola. Therefore, it is clear that *A. onnisi* does neither belong to Maricola nor Dimarcusidae. On the basis of morphological features, assignment of *A. onnisi* to either of these groups is equally complicated, if not impossible, as will be discussed below.

Similar to the situation in *A. onnisi*, in many species of the Maricola the ovaries are situated far behind the brain and sperm is transferred free, without being packaged into a spermatophore. However, *A. onnisi* cannot be assigned to this suborder because of its secondary bursa and the condition in which the most posterior testes still remain far removed from the root of the pharynx, as, in contrast, marine planarians have a primary copulatory bursa and testes that extend backwards to the root of the pharynx or even further.

In the presence of a secondary copulatory bursa, *A. onnisi* resembles three taxa of the Dimarcusidae, as such bursae have been reported for *Kawakatsua pumila* Sluys, 2019, *Opistobursa josephinae* Benazzi, 1975 and *O. mexicana* Benazzi, 1972. In particular, *A. onnisi* shares with *K. pumila* the syncytial structure of the copulatory bursa, but the copulatory bursa of *A. onnisi* differs from that of *K. pumila* in the presence of numerous vacuoles and by having a distinct lumen. Moreover, *A. onnisi* differs from *K. pumila*, *O. josephinae* and *O. mexicana* in that its bursa exhibits the probursal condition, in contrast to the situation in these dimarcusids in which the bursae are positioned posterior to the male copulatory apparatus. The presence of numerous vacuoles in the bursa of *A. onnisi*, which are often richly provided with an amorphous secretion and, in some cases, also contain sperm, suggests a sperm resorptive function of the bursa.

One of the major differences between *A. onnisi* and the Dimarcusidae concerns the situation that in the latter the bursal canal runs horizontally and exhibits a T-junction at which it receives the openings of the oviducts or the common oviduct; this feature forms one of the defining characters of the Dimarcusidae (see: Sluys, 1990; Benítez-Álvarez *et al.*, 2020). Clearly, this condition does not apply to *A. onnisi* and thus prevents it from being included in the Dimarcusidae.

Apart from the features discussed above, *A. onnisi* exhibits a number of other characters that are worthy of comparison with conditions found in other triclad flatworms. For example, as compared with other species of planarians, the body dimensions of *A. onnisi* are unusual, as the worms are slender. Generally, the smallest planarians are found among the Maricola. Although some marine planarians, such as *Paucumara falcata* Wang & Li, 2019, are slender, they are also short (*P. falcata*, length: 3.8–4.1 mm; width: 0.48–0.58 mm; Chen *et al.*, 2019). This is usually the case in marine triclads, i.e. when their body width is small, they are also short, and vice versa, thus contrasting with the unusually slender habitus of *A. onnisi*.

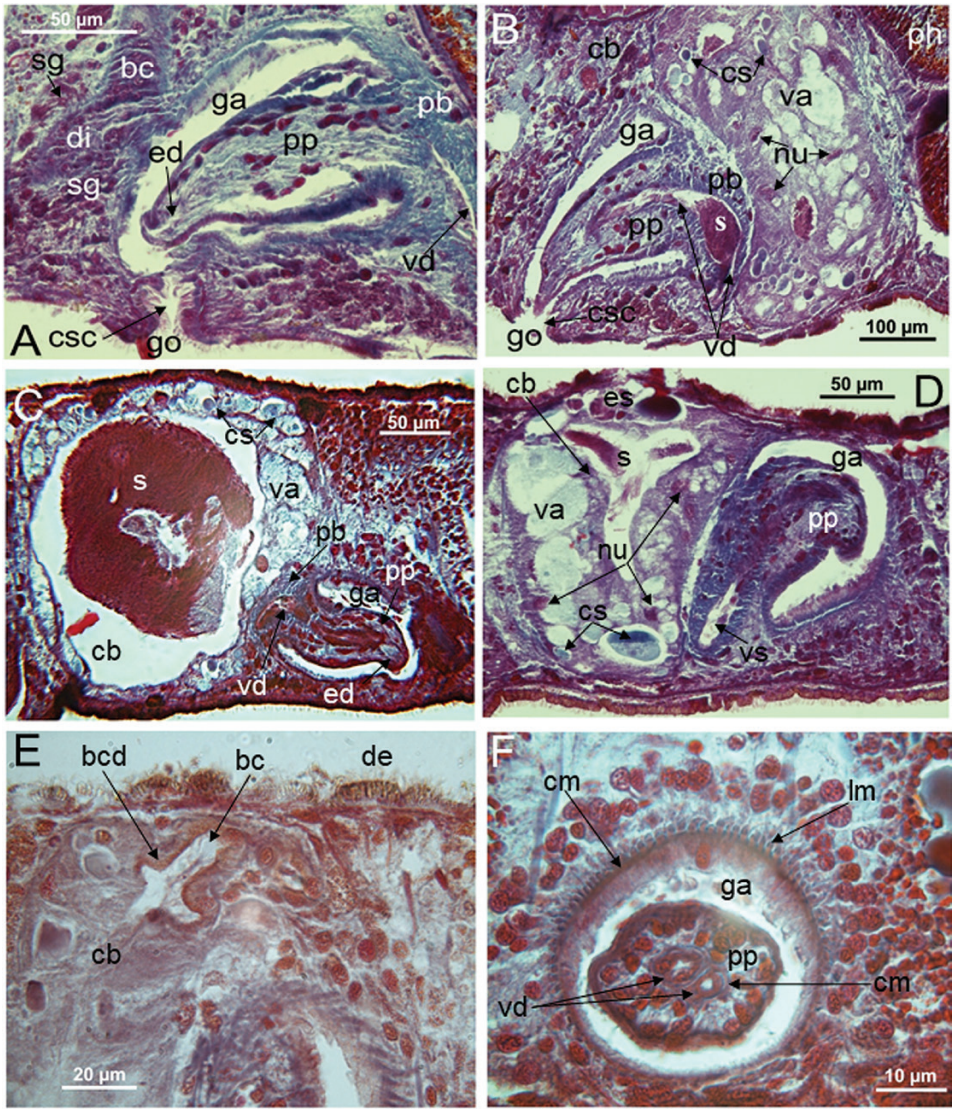


Figure 14. *Amphibioplana onnisi.* Photomicrographs of the copulatory apparatus; sagittal sections. A, holotype RMNH. VER. 19956.a, showing penis bulb (pb), penis papilla (pp), genital atrium (ga), and the gonopore (go) opening into the cup-shaped chamber (csc); anterior to the right; B, CGAS Pla 19. 2, showing copulatory bursa (cb), penis bulb, penis papilla and genital atrium; anterior to the right; C, RMNH.VER. 19958.d, showing copulatory bursa with a voluminous mass of sperm (s) in its lumen, penis bulb, penis papilla and genital atrium; anterior to the left; D, CGAS Pla 19.1b, showing copulatory bursa with sperm in its lumen, penis bulb, penis papilla and genital atrium; anterior to the left; E, CGAS Pla 24.5, sagittal section showing the proximal tract of the bursal canal (bc) with its diaphragm-like protrusion (bcd) into the copulatory bursa; anterior to the left; F, CGAS Pla 22. 6, transverse section showing the penis papilla with the two vasa deferentia (vd) running parallel before their fusion to form the ejaculatory duct.

Another unusual character of *A. onnisi* is the uniformity of its epidermis, both dorsally and ventrally. Ventrally and dorsally it has the same height and is uniformly clothed with cilia and provided with abundant rhabdites. Usually, in triclads the epidermis is thicker on the dorsal side of the body and often also at the body margins, including the anterior and posterior ends (Hyman, 1951; Sluys, 1989). Furthermore, cilia are generally well developed only on the lateral and ventral surfaces, but are usually absent or short on the dorsal body surface, except the dorsal sensory areas (Hyman, 1951; Ball & Reynoldson, 1981; Sluys, 1989).

Among triclads, rhabdites usually are not uniformly distributed on the body surface, in that they may be absent or reduced in number on the ventral surface

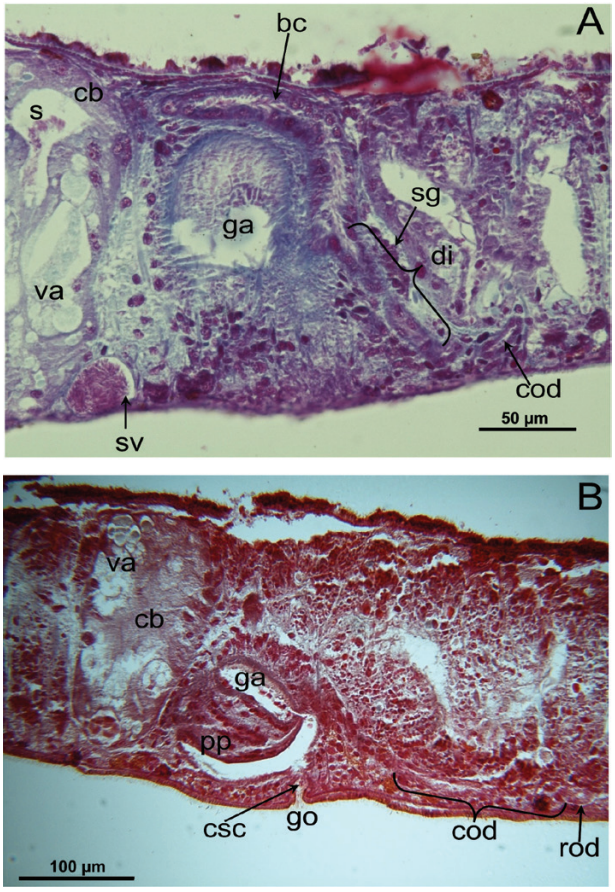


Figure 15. *Amphibioplana onnisi.* Photomicrographs of the copulatory apparatus. A, CGAS Pla 19.1b, sagittal section showing copulatory bursa (cb), bursal canal (bc) with its diverticulum (di), and common oviduct (cod); anterior to the left; B, RMNH.VER. 19957.a, sagittal section showing copulatory bursa (cb), penis papilla (pp), genital atrium (ga), cup-shaped chamber (csc), common oviduct (cod), and right oviduct (rod); anterior to the left.

(Ball & Reynoldson, 1981). Furthermore, rhabdites on the dorsal and lateral body surfaces are generally larger than those in the ventral epidermis (Hyman, 1951; Sluys, 1989). As epidermal cells are clothed with more cilia, the number of rhabdites is usually reduced, thus contrasting with the situation in *A. onnisi*.

Interestingly, in none of the specimens of *A. onnisi* examined was it possible to discern any structure resembling an auricular sense organ, although according to Hyman (1951) freshwater planarians are constantly provided with this kind of sensory structure. These sense organs are usually located in infolded regions of the dorsal epidermis that are devoid of rhabdites and gland cells, but richly provided with nerves and short cilia.

As a rule, among triclads the pharynx leads directly into the intestine; sometimes, such as in the marine planarian *Procerodes dahli* Marcus & Marcus, 1959, the pharynx has a short cone at its root that protrudes into the intestinal lumen and that may be designated as the oesophagus (Sluys, 1989: fig. 5). In land planarians, the term oesophagus is used for a particular portion of the gut near its point of communication with the pharynx, as is also the case in certain other turbellarians (cf. Von Graff, 1912–17: fig. 47; Meixner, 1938: fig. 36; Hyman, 1951: fig. 18B). In view of this broad application of the term oesophagus, we designated the protrusion in the pharynx of *A. onnisi* with the same terminology, albeit that in this case it projects into the pharynx lumen instead of the gut. Variability in the position of the mouth opening within the same population, as found to be the case in *A. onnisi*, is a condition that recently was reported also for the freshwater planarian *Girardia tigrina* (Girard, 1850) (Stocchino *et al.*, 2019).

In view of (1) the sister-group relationship between Amphibioplanidae and Dimarcusidae, (2) the topology of the tree of the entire suborder of the Tricladida (Fig. 2) and (3) the fact that *A. onnisi* shares some features with the Dimarcusidae (in particular, an unusual secondary bursa, and ovaries located at a distance far behind the brain), taxonomic classification is best served, in our opinion, by ranking the Amphibioplanidae as a second family under the suborder Cavernicola. With respect to the copulatory bursa, an obvious difference between *A. onnisi* and representatives of the Dimarcusidae is that in the latter the bursae are situated posteriorly to the penis (retrobursal condition), excepting *Hausera hauseri* Leal-Zanchet & Souza, 2014 and *Rhodax evelinae* Marcus, 1946, since these lack a bursa. In contrast, in *A. onnisi* the bursa exhibits the probursal condition, as it is located anteriorly to the male copulatory complex. In general, secondary bursae are characterized by an aberrant lining epithelium (e.g. syncytial) and/or a direct opening to the exterior, in contrast to the usual primary bursae (for further discussion, see: Sluys, 1990). The latter opening may be the only exit of the bursa to the exterior or it may represent one that is additional to the usual, primary connection between bursa and genital atrium facilitated by the bursal canal. Therefore, the copulatory bursae of the dimarcusids are secondary bursae because they have a second opening to the exterior and/or have a lining that differs from the usual high, vacuolated and nucleated epithelium of primary bursae. Exceptions may be formed by *Balliania thetisae* Gourbault, 1978, *Novomitchellia bursaelongata* Harrath *et al.*, 2016 and *N. sarawakana* (Kawakatsu & Chapman, 1983), because in these dimarcusids the bursae do not have a secondary opening to the exterior, while their lining epithelium seems to resemble that of primary bursae.

However, for *Balliania* it was mentioned that the cells of the bursa contain a granular substance, which differs from the usual condition in primary bursae. A granular secretion was present also in the cells lining the bursa of *O. mexicana*. The bursa of the dimarcusid *K. pumila* has neither a secondary opening, but its lining is syncytial, as in *A. onnisi*, which suggests that it is indeed a secondary bursa.

Possible diagnostic or apomorphic features of the Dimarcusidae were recently reviewed by Benítez-Álvarez *et al*. (2020), who expressed their doubt about Sluys' (1990) hypothesis that the character state 'ovaries situated at some distance posterior to the brain' forms a synapomorphy for the dimarcusids. Their opinion might well be true, as in *A. onnisi* the ovaries are located far behind the brain, thus implying a different level of universality for this character state. In other words, it may be the case that ovaries being located at some considerable distance behind the brain forms a synapomorphy for Dimarcusidae and Amphibioplanidae. Similarly, presence of a secondary bursa, in one form or the other, may represent another apomorphy for the Cavernicola in its present extension, i.e. Amphibioplanidae + Dimarcusidae.

Apomorphic, as well as diagnostic, features of the Dimarcusidae are formed by (1) the horizontally oriented bursal canal, provided with a T-junction, the latter formed by a diverticulum and/or communication of the canal with the common oviduct, located perpendicularly to the canal (Sluys, 1990) and (2) oviducts running medially to the ventral nerve cords (Benítez-Álvarez *et al.*, 2020). In *A. onnisi*, the bursal canal has a different course, lacks the T-junction, while the course of the oviducts expresses the symplesiomorphic character state, i.e. running laterally to the ventral nerve cords.

It is difficult to distinguish apomorphic characters for the Amphibioplanidae, which is partly due to the fact that, presently, the family is represented by only one species. Possible apomorphies for the Amphibioplanidae are the following: (1) uniformly ciliated epidermis, both ventrally and dorsally provided with numerous rhabdites, (2) oesophageal projection into the lumen of the pharynx, (3) syncytial lining epithelium of the copulatory bursa, albeit that such is present also in the dimarcusid *K. pumila* and (4) gonopore opening into a cup-shaped chamber.

Reproduction

The great majority of the specimens of *A. onnisi* that we examined were sexually mature, having testes with abundant sperm and showing voluminous spermiducal vesicles filled with sperm. Furthermore, presence of a mass of sperm in the copulatory bursa of all sexually mature individuals indicates that they all had previously mated. As the animals were collected during winter, spring and autumn (excepting the dry season July to September), this suggests that there is an almost continuous mating season. Such a reproductive cycle, which is independent from the seasons, may well form part of a stygobiont adaptive strategy of these worms.

Interestingly, animals from at least one population of *A. onnisi* (viz., Grutta 'e Pauli Cave) could be able to undergo fission during their sexual state. Traces of presumable fission in sexual individuals have been detected (1) at the prepharyngeal level, (2) just caudally to the pharynx and (3) caudally to the copulatory apparatus. Unfortunately, we did not find posterior portions of animals with signs of fission, so that with this data on hand, the exact modalities of asexual reproduction in *A. onnisi* cannot be described. Fission in sexually mature planarians is a rare phenomenon and among freshwater flatworms has been reported for ex-fissiparous individuals of *Dugesia aethiopica* Stocchino *et al.*, 2002 and *D. etrusca* Benazzi, 1946 (Stocchino & Manconi, 2013). However, in these *Dugesia* species, fission always occurs at the same level (by architomy slightly behind the pharynx in *D. aethiopica* and by distal architomy at the caudal tip in *D. etrusca*), in contrast to the various levels at which fission apparently may take place in *A. onnisi*. In contrast to fission in sexually mature animals, there are several species with populations in which the sexual and asexual reproductive modes alternate, usually seasonally, e.g. *Atrioplanaria aquabellae* Bromley, 1983, *A. morisii* Benazzi & Gourbault, 1977, *Atrioplanaria* sp. De Beauchamp, 1932 from Sardinia, *Dugesia bengalensis* Kawakatsu, 1983, *D. japonica*, *D. subtentaculata* (Draparnaud, 1801), *Girardia tigrina, Phagocata fawcetti* Ball & Gourbault, 1975, *P. morgani* (Stevens & Boring, 1906), *P. velata* (Stringer, 1909) and *Rhodax evelinae* Marcus, 1946 (Kenk, 1935, 1937; Benazzi, 1942; Marcus, 1946; Ogukawa & Kawakatsu, 1956; Ball & Gourbault, 1975; Benazzi & Gourbault, 1977; Ball *et al.*, 1981; Bromley, 1983; Aditya & Mahapatra, 1991; Leria *et al.*, 2020).

Freshwater and terrestrial eco-ethology

Amphibioplana onnisi shows great ecological plasticity, as it exhibits a peculiar lifestyle in that, even though it is a freshwater animal, under adverse environmental conditions, i.e. in the absence of water, it behaves like a terrestrial flatworm. In the freshwater habitats of the caves, worms were always found in pools fed by intermittent dripping water. This strongly suggests that their primary habitat is formed by microcrevices and small voids in the epikarst, i.e. the uppermost layer of the karst zone overlying the caves.

Presence of the animals in only temporary waters and their occurrence under stones that are firmly set in the humid clayey substrate, suggest that with the disappearance of temporary pools, these flatworms likely reach the relatively humid refuges under stones by moving through the small spaces at the interface of the stones and the substrate, as flatworms generally are unable to burrow into the substrate, which in the present case is highly compact. In this way, the planarians can survive in a temporary terrestrial habitat, despite the fact that they are basically freshwater animals. In particular, the slender body of *A. onnisi* suggests adaptation to such small spaces. The same may be true of the abundant rhabdites uniformly distributed on the body surface and the abundant ciliation of both body surfaces. Expelled rhabdites, which can be rich in glycosaminoglycans [Smith *et al.* (1982) and references therein; Hayes (2017)], and secretion of the subepidermal cyanophil glands form a layer of mucus that lowers the friction to which the body is subjected when moving through the small spaces of the epikarst, while it may form also a moisture-retaining film around the animal when it lies curled up in its refuge and that thus prevents it from desiccation.

It is well known that different distribution of cilia is directly related to locomotion, since the basic locomotory mechanism in planarians is formed by the combined action of the ventral longitudinal muscles and ventral cilia (Ball & Reynoldson, 1981). From that perspective, the unusual uniform ciliation, and notably the presence of abundant cilia on the dorsal surface of *A. onnisi*, suggests that this facilitates not only the usual gliding locomotion in aquatic environments, but also enables the animals to move through the microcrevices of the epikarst and the small spaces at the interface between the clay and the embedded stones under which they were found. Moreover, the uniform ciliation could be involved in the maintenance of the moisture-retaining film around the animal (see above) and might, perhaps, facilitate also the process of curling up. However, although uniform ciliation seems to represent the rule among interstitial microturbellarians, it is noteworthy that in many interstitial, sand-dwellers the dorsal ciliation frequently is reduced, with ciliation being restricted to the ventral creeping sole and small dorsolateral portions of the front end (Ax, 1956; Rieger *et al.*, 1991).

Recently, it was suggested that in land planarians a 'watery' layer of mucus may facilitate also the uptake of oxygen, as water is required to dissolve oxygen and carbon dioxide in order to cross cell membranes (Sluys, 2019). This adaptation may possibly occur also in *A. onnisi* and thus enable it to survive adverse, comparatively dry terrestrial conditions. Furthermore, the presence of abundant rhabdites in *A. onnisi* may be related also to respiration under terrestrial conditions, as it has been suggested that particular substances present in the rhabdoids of land planarians may have a respiratory role [Sluys (2019) and references therein].

Evolution

On the basis of its phylogenetic position, the morphological characteristics and adaptive strategies of *A. onnisi*, we offer the following tentative hypotheses for the evolution of this newly discovered branch on the phylogenetic tree of the triclad flatworms and for its distribution in three isolated caves on the Sardinian island at the middle of the western Mediterranean Sea.

The few calibrated timetrees available for some groups of Continenticola (Lázaro *et al.*, 2011; Solà *et al.*, 2013) suggest that continenticolans, and, therefore, also the other suborders of the Tricladida, are ancient clades. From that perspective, Sluys (1990) has already suggested that the Dimarcusidae had differentiated on Gondwana and, perhaps, even on Pangaea. The notion that Cavernicola in its present extension (Amphibioplanidae + Dimarcusidae) may indeed have diversified already on Pangaea, rather than Gondwana, is supported by the finding of *A. onnisi* near the southern margin of Laurasia. When Pangaea broke up, the Dimarcusidae diversified on Gondwana, while their sister-group, the Amphibioplanidae, evolved on Laurasia and are currently only represented by a single species, *A. onnisi*.

However, some geological accounts have suggested that Sardinia, or at least the portion of the island south of the Asinara–Posada Line, actually formed part of the northern margin of Gondwanaland during the Palaeozoic [Carmignani *et al.* (2016) and references therein; Cruciani *et al.* (2019)]. It is precisely in this southern part of Sardinia where the three caves are located from where we obtained *A. onnisi*. This geological configuration may imply that Cavernicola in their current extension did indeed diversify on Gondwana, as suggested by Sluys (1990) for Dimarcusidae, and that Amphibioplanidae evolved on its northern margin. Subsequently, representatives of Amphibioplanidae became isolated on Sardinia when it broke off from the northern margin of Gondwana and eventually collided with the Laurasian plate.

In either case, *A. onnisi* may be a phylogenetically relict species, i.e. representing the last survivor of an ancient radiation, in view of the fact that stygofauna originates from ancestors either living in surface freshwater or in a marine habitat (cf. Humphreys, 2000; Danielopol & Rouch, 2012; Trajano & Cobolli, 2012), both being environments that harbour no known close relatives of Amphibioplanidae.

The possible ancestral habitat of Dimarcusidae was discussed and analysed by Harrath *et al.* (2016) and

Benítez-Álvarez *et al*. (2020). The comprehensive last-mentioned study found that the ancestral habitat of Dimarcusidae probably was epigean and freshwater, implying that dimarcusids evolved from worms originally adapted to continental epigean freshwater habitats. In view of the topology of the phylogenetic tree and the sister-group relationship between Dimarcusidae and Amphibioplanidae, the same must hold true for the common ancestor of these two families, i.e. for the ancestor of Cavernicola in their current circumscription.

Although the few molecular data that we were able to generate for the 16S sequence did not allow for a full population genetic analysis, it does provide some clues on the possible history of the various cave populations of *A. onnisi*. As shown above, populations from the two caves Istirzili and Grutta 'e Pauli, which are geographically close, show a lower genetic distance between them (~6% difference), despite being in different karstic systems, than any of these populations exhibits when compared to sequences from San Pietro Cave (~15% genetic distance). This coincides with the fact that the last-mentioned population exhibits some minor morphological differences (less convoluted pharynx, smaller oesophagal protrusion and smaller cup-shaped chamber) as compared to animals from the other two caves. This situation may result from an isolation-by-distance process, according to a scenario in which a population became stranded in one groundwater system and then spread through interconnections between the three karstic systems. Although the three caves are greatly isolated from each other, dispersal cannot be excluded for these animals, due to the fact that the primary habitat of *A. onnisi* is not the cave itself but consists of microcrevices in the epikarst. Therefore, the animals may be able to move from the epikarst into microfissures of the substrates separating the karstic regions and thus may be able to migrate from one epikarst system to the other. Nevertheless, these displacements may interfere with a constant gene flow among caves, thus resulting in the observed genetic differences between the three populations.

The high differentiation exhibited by the 16S gene, as well as the clear reciprocal monophyly in the nuclear genes-based trees between the populations of *A. onnisi* in the eastern and western caves (Fig. 1), may induce one to consider the possibility that the San Pietro Cave population represents a different species from the animals in the other two caves. Nevertheless, such data are weak when it comes to species delimitation. Future research involving new samplings, also from other Sardinian caves, and next-generation sequencing data (either transcriptomes or some type of reduced genome data) will allow a more robust molecular delimitation methodology. In this way, it will be possible to take into account possible flaws, such as incomplete lineage sorting or introgression events that may affect phylogenetic inference and delimitation procedures at such a shallow level. At this stage, we prefer to follow a conservative approach to taxonomy until new data will indisputably reveal the presence of two separate species. Our present conservative approach agrees with the fact that only slight morphological differences were observed in the San Pietro Cave population, which are not sufficient for describing this population as a separate species, different from the other populations examined.

ACKNOWLEDGEMENTS

This research was supported by Regione Autonoma della Sardegna project "Conservazione e valorizzazione delle grotte sarde: biodiversità e ruolo socio-economico-culturale" (RAS2012-LR7/2007-CRP-60215) and Fondazione di Sardegna (FdS-2016 CUP J86C1800082005). G. A. Stocchino was additionally supported by a Temminck Fellowship from Naturalis Biodiversity Center, Leiden, The Netherlands, and by grants from SYNTHESYS (European Union-funded Integrated Activities grants: NL-TAF 3687 and NL-TAF 8228). Molecular work at University of Barcelona was supported by Ministerio de Ciencia, Innovación y Universidades, project PGC2018-093924-B-100. Professor M. Pala is thanked for her kind support. We gratefully thank Dr F. Stoch for his suggestions on ecological aspects, Dr J. de Waele and Dr L. Sanna for geological information and Dr F. Melis, E. Onnis and N. Ibba for their kind support during the fieldwork. We are grateful to C. Laumer, three anonymous reviewers and the editorial team for their useful suggestions that helped to improve the manuscript.

REFERENCES

Aditya AK, **Mahapatra S. 1991.** Notes on the biology of the freshwater planarian *Dugesia bengalensis* (Platyhelminthes Turbellaria Tricladida). *Hydrobiologia* **227:** 145.

Amaral SV, **Leal-Zanchet AM**, **Ferreira R. 2019.** A new species of land planarian (Platyhelminthes: Continenticola) from a limestone cave in a Neotropical semi-arid environment. *Studies on Neotropical Fauna and Environment* **54:** 87–96.

Ax P. 1956. Monographie der Otoplanidae (Turbellaria). *Akademie der Wissenschaften und Literatur, Abhandlungen der Mathematisch-Naturwissenschaftlichen Klasse* **1956:** 1–298.

Ball IR, **Gourbault N. 1975.** The morphology, karyology and taxonomy of a new freshwater planarian of the genus *Phagocata* from California (Platyhelminthes: Turbellaria). *Royal Ontario Museum, Life Science Contributions* **105:** 1–19.

Ball IR, **Reynoldson TB. 1981.** *British planarians*. Cambridge: Cambridge University Press.

Ball IR, **Gourbault N**, **Kenk R. 1981.** The planarians (Turbellaria) of temporary waters in eastern North America. *Royal Ontario Museum, Life Science Contributions* **127:** 1–27.

Benazzi M. 1942. Ulteriori osservazioni su una *Fonticola* (*Atrioplanaria* sec. De Beauchamp) della Sardegna. *Bollettino di Zoologia* **13:** 28–34.

Benazzi M. 1946. Sopra una nuova planaria d'acqua dolce. *Archivio Zoologico Italiano* **31:** 93–102.

Benazzi M. 1972. Notizie preliminari sulle planarie raccolte nella seconda spedizione Lincea in Messico. *Accademia Nazionale dei Lincei* **52:** 403–405.

Benazzi M. 1975. *Opistobursa josephinae*, a new troglobitic planarian from Chiapas, Mexico. *Accademia Nazionale dei Lincei* **59:** 533–536.

Benazzi M, **Gourbault N. 1977.** *Atrioplanaria morisii* n. sp., a new cave planarian from Italy. *Italian Journal of Zoology* **44:** 327–335.

Benazzi M, **Baguña J**, **Ballester R**, **Puccinelli**, **Del Papa R. 1975.** Further contribution to the taxonomy of the "*Dugesia lugubris-polychroa* group" with description of *Dugesia mediterranea* n. sp. (Tricladida, Paludicola). *Bollettino di Zoologia* **42:** 81–89.

Benítez-Álvarez L, **Leal-Zanchet AM**, **Oceguera-Figueroa A**, **Ferreira RL**, **de Medeiros Bento D**, **Braccini J**, **Sluys R**, **Riutort M. 2020.** Phylogeny and biogeography of the Cavernicola (Platyhelminthes: Tricladida): relicts of an epigean group sheltering in caves? *Molecular Phylogenetics and Evolution* **145:** 106709.

Bromley HJ. 1983. The morphology and reproduction of a new species of *Atrioplanaria* De Beauchamp (Turbellaria: Tricladida) from Israel. *Israel Journal of Zoology* **32:** 171–184.

Carmignani L, **Oggiano G**, **Funedda A**, **Conti P**, **Pasci S. 2016.** The geological map of Sardinia (Italy) at 1:250,000 scale. *Journal of Maps* **12:** 826–835.

Cassola F. 1982. Il popolamento cavernicolo della Sardegna. *Biogeographia* **7:** 615–755.

Chen J-J, **Li W-X**, **Sluys R**, **Wu M-Q**, **Wang L**, **Li S-F**, **Wang A-T. 2019.** Two new species of marine flatworm from southern China facilitate determination of the phylogenetic position of the genus *Nerpa* Marcus, 1948 and the histochemical structure of the nervous system in the genus *Paucumara* Sluys, 1989 (Platyhelminthes, Tricladida, Maricola). *Zootaxa* **4568:** 149–167.

Cruciani G, **Fancello D**, **Franceschelli M**, **Musumeci G. 2019.** Geochemistry of the Monte Filau orthogneiss (SW Sardinia, Italy): insight into the geodynamic setting of Ordovician felsic magmatism in the N/NE Gondwana margin. *Italian Journal of Geosciences* **138:** 136–152.

Danielopol DL, **Rouch R. 2012.** Invasion: active versus passive. In: Culver DC, White WB, eds. *Encyclopedia of caves*, *2nd edn*. Oxford: Academic Press, 404–409.

Dols-Serrate D, **Leria L**, **Aguilar JP**, **Stocchino GA**, **Riutort M. 2020.** *Dugesia hepta* and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex? *Organisms Diversity and Evolution* **20:** 369–386.

Draparnaud JPR. 1801. *Tableau des Mollusques terrestres et fluviatiles de la France*. Montpellier & Paris: Renaud. Doi.org/10.5962/bhl.title.40947.

Girard C. 1850. A brief account of the fresh water species of planariae inhabiting the United States. *Proceedings of the Boston Society of Natural History* **3:** 264–265.

Gourbault N. 1978. Une nouvelle planaire primitive d'origine marine, Turbellarié Triclade de Polynésie. *Cahiers de Biologie Marine* **19:** 23–36.

Guindon S, **Dufayard J**, **Lefort V**, **Anisimova M**, **Hordijk W**, **Gascuel O. 2010.** New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59:** 307–321.

Hall T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95–98.

Harrath AH, **Mansour L**, **Lagnika M**, **Sluys R**, **Boutin C**, **Alwasel S**, **Poch A**, **Riutort M. 2016.** A molecular analysis of the phylogenetic position of the suborder Cavernicola within the Tricladida (Platyhelminthes), with the description of a new species of stygobiont flatworm from Benin. *Zoological Journal of the Linnean Society* **178:** 482–491.

Hayes MJ. 2017. Sulphated glycosaminoglycans support an assortment of planarian rhabdite structures. *Biology Open* **6:** 571–581.

Hoang DT, **Chernomor O**, **von Haeseler A**, **Minh BQ**, **Vinh LS. 2018.** UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35:** 518–522.

Huelsenbeck JP, **Rannala B. 2004.** Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology* **53:** 904–913.

Humphreys WF. 2000. Relict faunas and their derivation. In: Wilkins H, Culver DC, Humphreys WF, eds. *Ecosystems of the world 30. Subterranean ecosystems*. Amsterdam: Elsevier Academic Press, 319–343.

Hyman LH. 1951. *The invertebrates: Platyhelminthes and Rhynchocoela. The acoelomate bilateria*, *Vol. II*. New York: McGraw-Hill Book Company, Inc.

Ichikawa A, **Kawakatsu M. 1964.** A new freshwater planarian, *Dugesia japonica*, commonly but erroneously known as *Dugesia gonocephala* (Dugès). *Annotationes Zoologicae Japonenses* **37:** 185–194.

Kalyaanamoorthy S, **Minh BQ**, **Wong T**, von **Haeseler A**, **Jermiin L. 2017.** ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14:** 587–589.

Katoh K, **Standley DM. 2013.** MAFFT multiple sequence alignment software v.7: improvements in performance and usability. *Molecular Biology and Evolution* **30:** 772–780.

Kawakatsu M, **Chapman P. 1983.** *Mitchellia sarawakana* gen. et sp. nov. (Turbellaria, Tricladida, Maricola), a new freshwater planarian from Water Polo Cave, Sarawak, East Malaysia. *Journal of the Speleological Society of Japan* **8:** 21–34.

Kawakatsu M, **Oki I**, **Tamura S**, **Yamayoshi T**, **Aditya AK. 1983.** A new freshwater planarian from West Bengal. *Bulletin of the Biogeographical Society of Japan* **38:** 1–10.

Kenk R. 1935. Studies on Virginian triclads. *Journal of the Elisha Mitchell Scientific Society* **51:** 79–125.

Kenk R. 1937. Sexual and asexual reproduction in *Euplanaria tigrina* (Girard). *The Biological Bulletin* **73:** 280–294.

Kumar S, **Stecher G**, **Li M**, **Knyaz C**, **Tamura K. 2018.** MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35:** 1547–1549.

Lartillot N. 2020. PhyloBayes: Bayesian phylogenetics using site-heterogenous models. In: Scornavacca C, Delsuc F, Galtier N, eds. *Phylogenetic in the genomic era*. Published online: HAL archives ouvertes, open access book, pp. 1.5:1–1.5:16. hal-02535342. Available at: https://hal.inria.fr/PGE/hal-02535342. Accessed October 2020.

Lartillot N, **Philippe H. 2004.** A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* **21:** 1095–1109.

Laumer CE, **Giribet G. 2014.** Inclusive taxon sampling suggests a single, stepwise origin of ectolecithality in Platyhelminthes. *Biological Journal of the Linnean Society* **111:** 570–588.

Laumer CE, **Hejnol A**, **Giribet G. 2015.** Nuclear genomic signals of the 'microturbellarian' roots of platyhelminth evolutionary innovation. *ELife* **2015:** e05503.

Lázaro EM, **Sluys R**, **Pala M**, **Stocchino GA**, **Baguñà J**, **Riutort M. 2009.** Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution* **52:** 835–845.

Lázaro EM, **Harrath AH**, **Stocchino GA**, **Pala M**, **Baguñà J**, **Riutort M. 2011.** *Schmidtea mediterranea* phylogeography: an old species surviving on a few Mediterranean Islands? *BMC Evolutionary Biology* **11:** 274.

Leal-Zanchet AM, **Marques AD. 2018a.** Searching for shelter in a ferruginous cave? A new species of *Pasipha* from a plateau in the Brazilian savanna (Platyhelminthes, Tricladida). *ZooKeys* **776:** 13–25.

Leal-Zanchet AM, **Marques AD. 2018b.** Coming out in a harsh environment: a new genus and species for a land flatworm (Platyhelminthes: Tricladida) occurring in a ferruginous cave from the Brazilian savanna. *PeerJ* **6:** e6007.

Leal-Zanchet AM, **Teles de Souza S**, **Ferreira RL. 2014.** A new genus and species for the first recorded cave-dwelling Cavernicola (Platyhelminthes) from South America. *ZooKeys* **442:** 1–14.

Leria L, **Vila-Farré M**, **Álvarez-Presas M**, **Sánchez-Gracia A**, **Rozas J**, **Sluys R**, **Riutort M. 2020.** Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhelminthes, Tricladida): Extreme intraindividual genetic diversity, morphological stasis, and karyological variability. *Molecular Phylogenetics and Evolution* **143:** 106496.

Marcus E. 1946. Sôbre Turbellaria Brasileiros. *Boletim da Faculdade de Filosofia, Ciências e Letras da Universidade de São Paulo, Zoologia* **11:** 5–254.

Marcus E, **Marcus E. 1959.** Turbellaria from Madeira and the Azores. *Boletim do Museu Municipal do Funchal* **12:** 15–42.

Meixner J. 1938. Turbellaria (Strudelwürmer). *Tierwelt Nord- und Ostsee* **33:** 1–146.

Minh BQ, **Nguyen MAT**, von **Haeseler A. 2013.** Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30:** 1188–1195.

Minh BQ, **Schmidt H**, **Chernomor O**, **Schrempf D**, **Woodhams MD**, **von Haeseler A**, **Lanfear R**, **Teeling E. 2020.** IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37:** 1530–1534.

Morrison DA. 2010. *Bayesian posterior probabilities: revisited*. Available at: http://arxiv.org/abs/1001.3638. Accessed October 2020.

Nguyen L, **Schmidt H**, von **Haeseler A**, **Minh BQ. 2015.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32:** 268–274.

Okugawa KI, **Kawakatsu M. 1956.** Studies on the fission of Japanese fresh-water planaria, *Dugesia gonocephala* (Dugès). IV. Comparative Studies on breeding and fission frequencies of sexual and assumed asexual races which were observed in laboratory cultures and natural habitats. *Bulletin of the Kyoto Gakugei University* **8:** 23–42.

Pala M, **Vacca RA**, **Casu S**, **Stocchino GA. 1995.** The freshwater planarian *Dugesia sicula* Lepori from Sardinia (Platyhelminthes, Tricladida). *Hydrobiologia* **310:** 151–156.

Pala M, **Casu S**, **Stocchino GA. 1999.** Karyology and karyotype analysis of diploid freshwater planarian populations of the *Dugesia gonocephala* group (Platyhelminthes, Tricladida) found in Sardinia. *Hydrobiologia* **392:** 113–119.

Pala M, **Stocchino GA**, **Corso G**, **Casu S. 2000.** A new species of freshwater planarian belonging to the genus *Dugesia* (Platyhelminthes, Tricladida) from Sardinia. *Italian Journal of Zoology* **67:** 379–383.

Rieger RM, **Tyler S**, **Smith JPS III**, **Rieger G. 1991.** Platyhelminthes: Turbellaria. In: Harrison FW, Bogitsh BJ, eds. *Microscopic anatomy of invertebrates*, *Vol. 3*. New York: Wiley-Liss, 7–140.

Riesgo A, **Burke EA**, **Laumer CE**, **Giribet G. 2017.** Genetic variation and geographic differentiation in the marine triclad *Bdelloura candida* (Platyhelminthes, Tricladida, Maricola), ectocommensal on the American horseshoe crab *Limulus polyphemus*. *Marine Biology* **164:** 111.

Riutort M, **Álvarez-Presas M**, **Lazaro EM**, **Solà E**, **Paps J. 2012.** Evolutionary history of the Tricladida and the Platyhelminthes: an up-to-date phylogenetic and systematic account. *International Journal of Developmental Biology* **56:** 5–17.

Sluys R. 1989. *A monograph of the marine triclads*. Rotterdam & Brookfield: A. A. Balkema.

Sluys R. 1990. A monograph of the Dimarcusidae (Platyhelminthes, Seriata, Tricladida). *Zoologica Scripta* **19:** 13–29.

Sluys R. 2001. Towards a phylogenetic classification and characterization of dugesiid genera (Platyhelminthes, Tricladida, Dugesiidae): a morphological perspective. In: Littlewood DTJ, Bray RA, eds. *Interrelationships of the Platyhelminthes, Systematics Association Special volume Ser. 60*. London and New York: Taylor & Francis, 57–73.

Sluys R. 2019. The evolutionary terrestrialization of planarian flatworms (Platyhelminthes, Tricladida, Geoplanidae): a review and research programme. *Zoosystematics and Evolution* **95:** 543–556.

Sluys R, **Kawakatsu M. 2006.** Towards a phylogenetic classification of dendrocoelid freshwater planarians (Platyhelminthes): a morphological and eclectic approach. *Journal of Zoological Systematics and Evolutionary Research* **44:** 274–284.

Sluys R, **Laumer C. 2019.** A new genus and species of a terrestrial cavernicolan planarian from Barro Colorado Island, Panama (Platyhelminthes, Tricladida, Cavernicola). *Zootaxa* **4586:** 187–193.

Sluys R, **Riutort M. 2018.** Planarian diversity and phylogeny. In: Rink JC, ed. *Planarian regeneration – methods and protocols [Methods in Molecular Biology, Vol. 1774]*. New York: Humana Press, Springer Science+Business Media, 1–56.

Sluys R, **Kawakatsu M**, **Riutort M**, **Baguñà J. 2009.** A new higher classification of planarian flatworms (Platyhelminthes, Tricladida). *Journal of Natural History* **43:** 1763–1777.

Smith J III, **Tyler S**, **Thomas MB**, **Rieger RM. 1982.** The morphology of Turbellarian rhabdites: phylogenetic implications. *Transactions of the American Microscopical Society* **101:** 208–228.

Solà E, **Sluys R**, **Gritzalis K**, **Riutort M. 2013.** Fluvial basin history in the northeastern Mediterranean region underlies dispersal and speciation patterns in the genus *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution* **66:** 877–888.

Souza S, **Riutort M**, **Ferreira RL**, **Leal-Zanchet AM. 2018.** An integrative approach reveals the first marine triclad (Platyhelminthes) trapped in a cave from a semiarid Neotropical environment. *Invertebrate Systematics* **32:** 627–638.

Steenwyk JL, Buida TJ III, Li Y, Shen X-X, **Rokas A. 2020.** ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biology* 18: e3001007.

Stevens NM, Boring AM. 1906. *Planaria morgani* n. sp. *Proceedings of the Academy of Natural Sciences of Philadelphia* **58:** 7–9.

Stocchino GA, **2018.** 80 years of research on planarians (Platyhelminthes, Tricladida) from Sardinia, Italy: an annotated checklist. *Zootaxa* **4532:** 539–552.

Stocchino GA, **Manconi R. 2013.** Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology* **80:** 319–328.

Stocchino GA, **Corso G**, **Manconi R**, **Pala M. 2002.** African planarians: *Dugesia aethiopica* sp. n. (Platyhelminthes, Tricladida) from Lake Tana (NW Ethiopia). *Italian Journal of Zoology* **69:** 45–51.

Stocchino GA, **Corso G**, **Manconi R**, **Casu S**, **Pala M. 2005.** Endemic freshwater planarians of Sardinia: Redescription of *Dugesia hepta* (Platyhelminthes, Tricladida) with a comparison of the Mediterranean species of the genus. *Journal of Natural History* **39:** 1947–1960.

Stocchino GA, **Marcia P**, **Grafitti G**, **Manconi R**, **Corso G**, **Casu S**, **Pala M. 2008.** Tricladi cavernicoli di Sardegna. *Memorie dell'Istituto Italiano di Speleologia Ser. 2* **21:** 121–124.

Stocchino GA, **Sluys R**, **Marcia P**, **Manconi R. 2013.** Subterranean aquatic planarians of Sardinia, with a discussion on the penial flagellum and the bursal canal sphincter in the genus *Dendrocoelum* (Platyhelminthes, Tricladida, Dendrocoelidae). *Journal of Cave and Karst Studies* **75:** 93–112.

Stocchino GA, **Sluys R**, **Onnis C**, **Cadeddu B**, **Demurtas D**, **Marcia P**, **Manconi R. 2015.** Reproductive, behavioural, and ecological peculiarities of cave-dwelling planarians from Sardinia (Platyhelminthes, Tricladida). In: *Book of Abstracts XIII International Symposium on Flatworm Biology*, *Oxford*, 66.

Stocchino GA, **Sluys R**, **Harrath AH**, **Mansour L**, **Manconi R. 2019.** The invasive alien freshwater flatworm *Girardia tigrina* (Girard, 1850) (Platyhelminthes, Tricladida) in Western Europe: new insights into its morphology, karyology and reproductive biology. *Contributions to Zoology* **88:** 236–256.

Stringer CE. 1909. Notes on Nebraska Turbellaria with descriptions of two new species. *Zoologischer Anzeiger* **34:** 257–262.

Tan G, **Muffato M**, **Ledergerber C**, **Herrero J**, **Goldman N**, **Gil M**, **Dessimoz C. 2015.** Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Systematic Biology* **64:** 778–791.

Trajano E, **Cobolli M. 2012.** Evolution of lineages. In: Culver DC, White WB, eds. *Encyclopedia of caves, 2nd edn*. Oxford: Academic Press, 295–297.

Von Graff L. 1912–17. Tricladida. In: Bronn HG, ed. *Klassen und Ordnungen des Tier-Reichs, Bd. IV Vermes Abt. IC: Turbellaria, II Abt.: Tricladida*. Leipzig: C. F. Winter, 2601–3369.

Winsor L, **Sluys R. 2018.** Basic histological techniques for planarians. In: Rink JC, ed. *Planarian regeneration – methods and protocols [Methods in Molecular Biology vol. 1774]*. New York: Humana Press, Springer Science+Business Media, 285–351.

Xia X, **Lemey P. 2009.** Assessing substitution saturation with DAMBE. In: Lemey P, Salemi M, Vandamme AM, eds. *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*, *2nd edn*. Cambridge: Cambridge University Press, 615–630.

Xia X, **Xie Z. 2001.** DAMBE: software package for data analysis in molecular biology and evolution. *Journal of Heredity* **92:** 371–373.

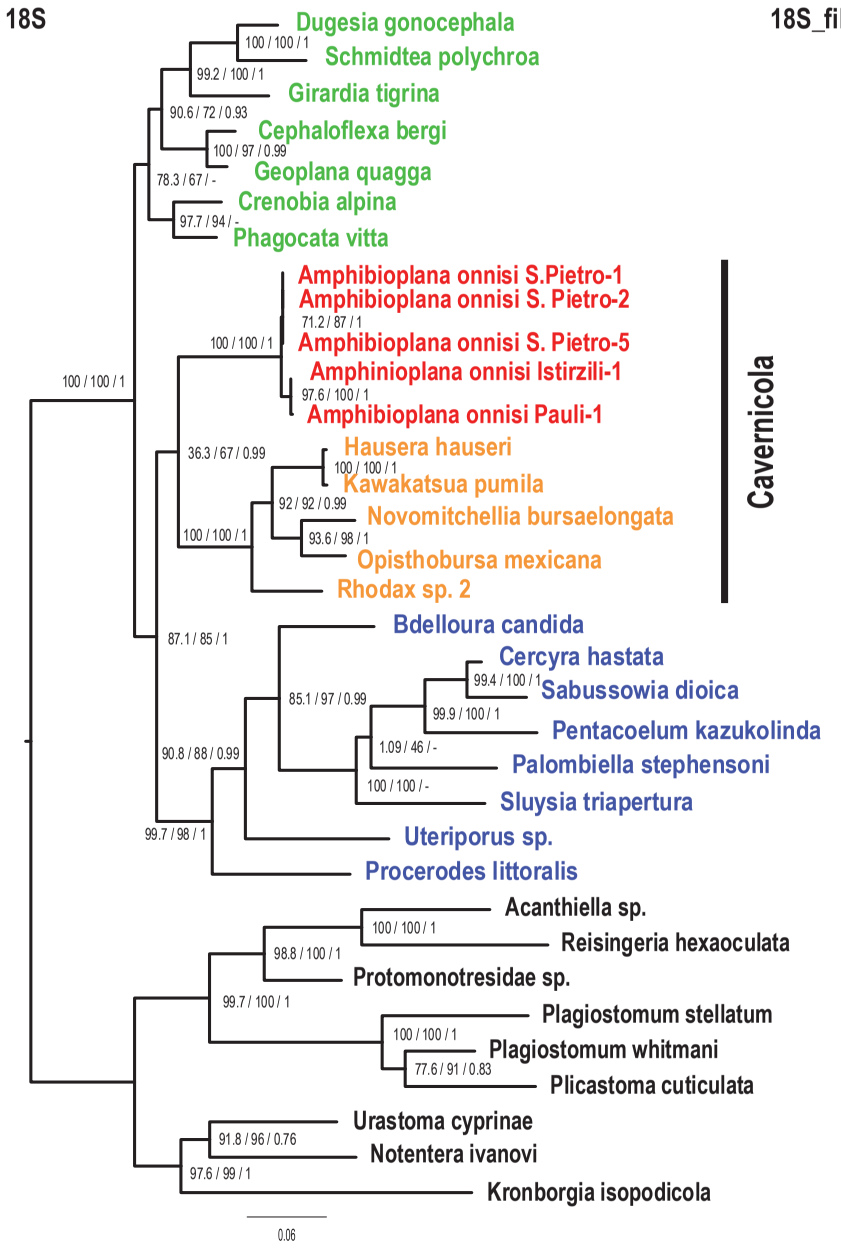
SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Figure S1. Maximum-likelihood trees inferred from the 18S and 18S-filtered datasets (Table 2). Values at nodes correspond to UFB/SH-aLRT/PP. –: PP values below 0.5. Scale bar: number of substitutions per nucleotide position.
Figure S2. Maximum likelihood trees inferred from the 28S and 28S-filtered datasets (Table 2). In the BI tree an alternative topology is obtained in which the Cavernicola form a clade with the Maricola with PP support values of 0.93 for 28S and 0.94 for 28S_filtered. Values at nodes correspond to UFB/SH-aLRT/PP. –: PP values below 0.5 or clade not-pressent on the BI inferred tree. Scale bar: number of substitutions per nucleotide position.
Figure S3. Maximum likelihood tree inferred from the concatenated dataset, including the complete 18S and 28S alignments (Table 2). Values at nodes correspond to UFB/SH-aLRT/PP –: PP values below 0.5. Scale bar: number of substitutions per nucleotide position.
Table S1. Primer sequences, annealing temperatures and references.
Table S2. PCR conditions for each gene.
Table S3. Kimura-2 parameter distances between *Amphibioplana* 16S sequences obtained for populations from the three caves.

Supplementary Materials: figures S1 to S3

Figure S1. Maximum likelihood trees inferred from the 18S and 18S-filtered datasets (Table 2). Values at nodes correspond to UFB/ SH-aLRT/ PP. –: PP values below 0.5. Scale bar: number of substitutions per nucleotide position.


18S
Dugesia gonocephala
Schmidtea polychroa
Girardia tigrina
Cephaloflexa bergi
Geoplana quagga
Crenobia alpina
Phagocata vitta
Amphibioplana onnisi S.Pietro-1
Amphibioplana onnisi S. Pietro-2
Amphibioplana onnisi S. Pietro-5
Amphinioplana onnisi Istirzili-1
Amphibioplana onnisi Pauli-1
Hausera hauseri
Kawakatsua pumila
Novomitchellia bursaelongata
Opisthobursa mexicana
Rhodax sp. 2
Cavernicola
Bdelloura candida
Cercyra hastata
Sabussowia dioica
Pentacoelum kazukolinda
Palombiella stephensoni
Sluysia triapertura
Uteriporus sp.
Procerodes littoralis
Acanthiella sp.
Reisingeria hexaoculata
Protomonotresidae sp.
Plagiostomum stellatum
Plagiostomum whitmani
Plicastoma cuticulata
Urastoma cyprinae
Notentera ivanovi
Kronborgia isopodicola
0.06

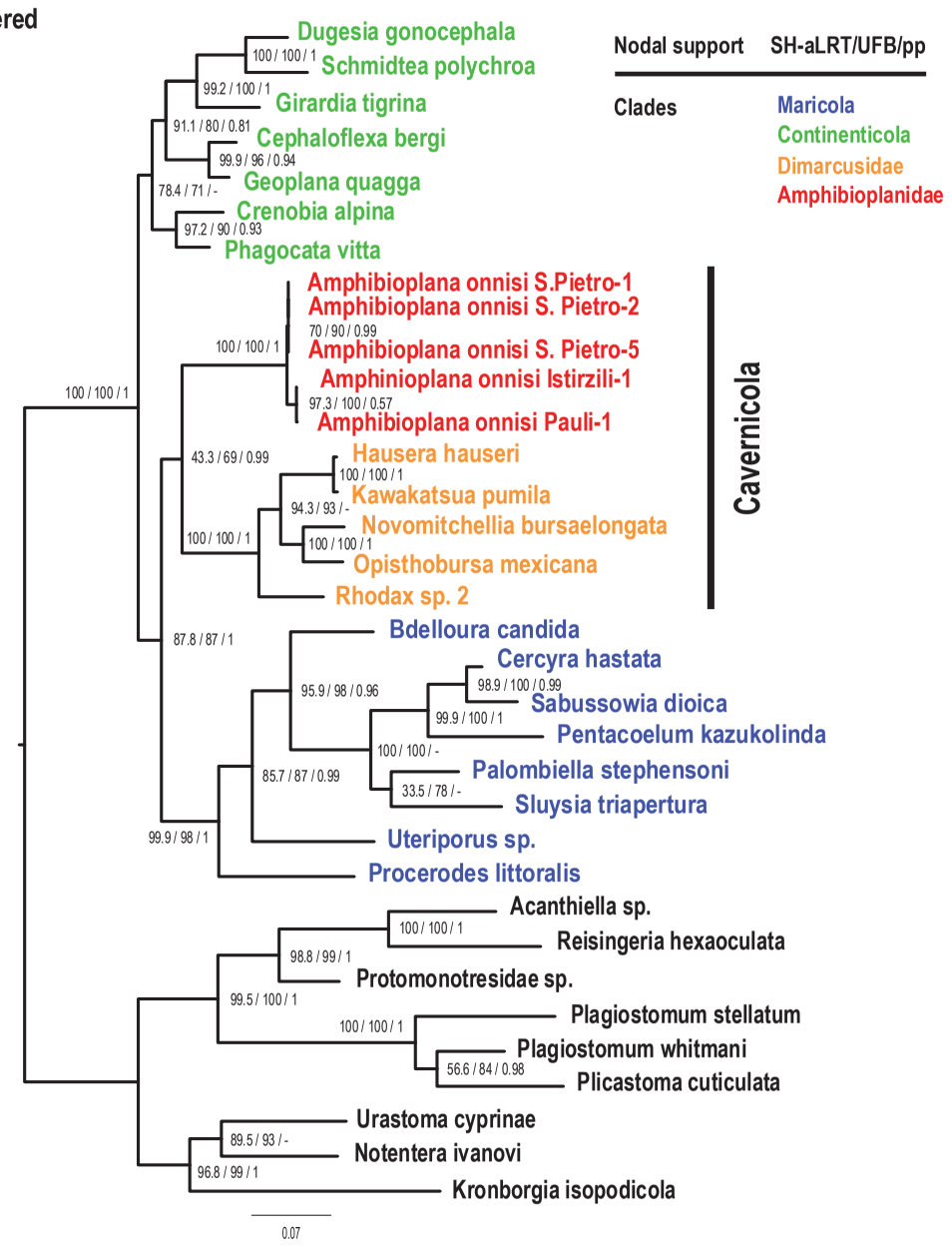
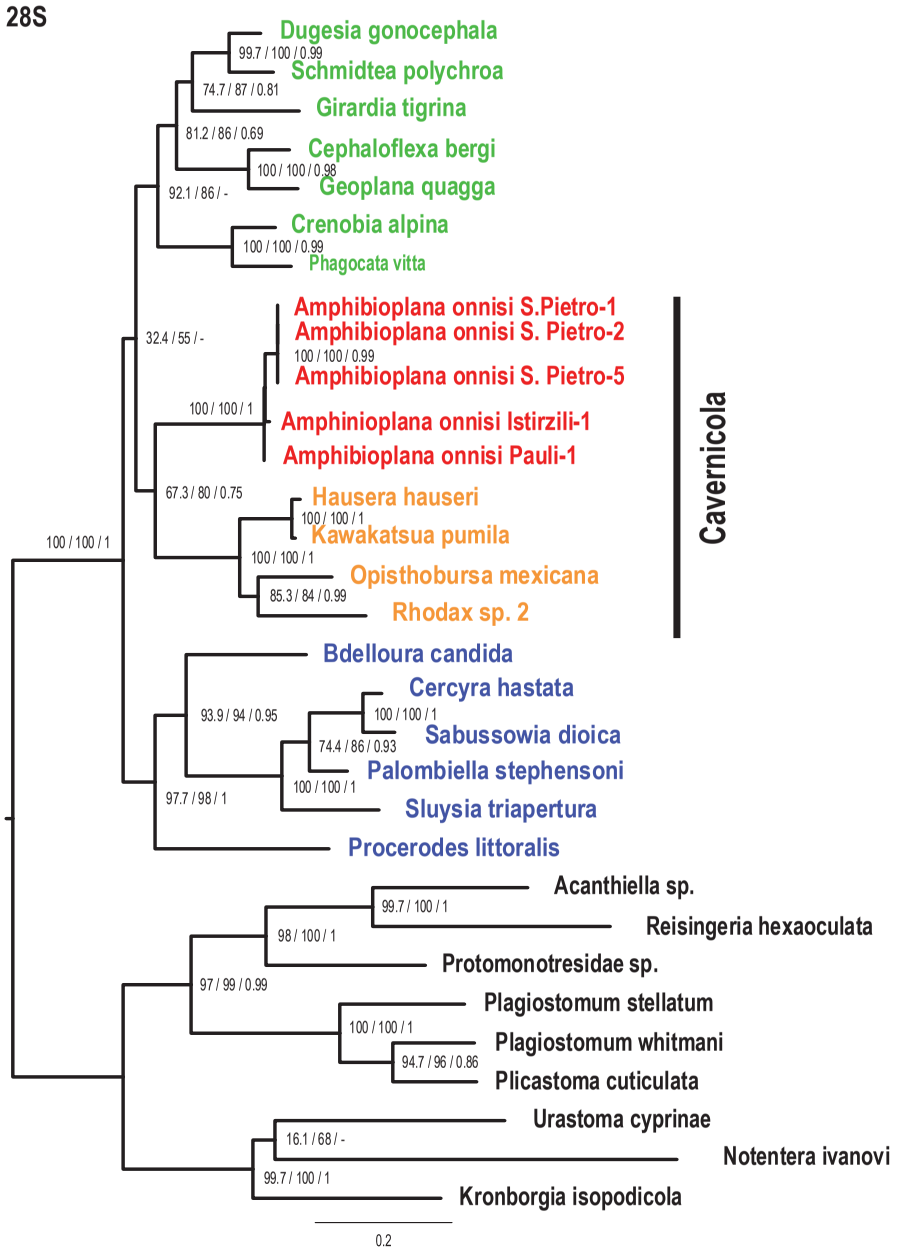

18S_filtered
Nodal support SH-aLRT/UFB/pp
Clades
Maricola
Continenticola
Dimarcusidae
Amphibioplanidae
Dugesia gonocephala
Schmidtea polychroa
Girardia tigrina
Cephaloflexa bergi
Geoplana quagga
Crenobia alpina
Phagocata vitta
Amphibioplana onnisi S.Pietro-1
Amphibioplana onnisi S. Pietro-2
Amphibioplana onnisi S. Pietro-5
Amphinioplana onnisi Istirzili-1
Amphibioplana onnisi Pauli-1
Hausera hauseri
Kawakatsua pumila
Novomitchellia bursaelongata
Opisthobursa mexicana
Rhodax sp. 2
Cavernicola
Bdelloura candida
Cercyra hastata
Sabussowia dioica
Pentacoelum kazukolinda
Palombiella stephensoni
Sluysia triapertura
Uteriporus sp.
Procerodes littoralis
Acanthiella sp.
Reisingeria hexaoculata
Protomonotresidae sp.
Plagiostomum stellatum
Plagiostomum whitmani
Plicastoma cuticulata
Urastoma cyprinae
Notentera ivanovi
Kronborgia isopodicola
0.07

Figure S2. Maximum likelihood trees inferred from the 28S and 28S-filtered datasets (Table 2). In the BI tree an alternative topology is obtained in which the Cavernicola form a clade with the Maricola with pp support values of 0.93 for 28S and 0.94 for 28S_filtered. Values at nodes correspond to UFB/ SH-aLRT/ PP. –: PP values below 0.5 or clade not-pressent on the BI inferred tree. Scale bar: number of substitutions per nucleotide position.


28S
Dugesia gonocephala
99.7 / 100 / 0.99
Schmidtea polychroa
74.7 / 87 / 0.81
Girardia tigrina
81.2 / 86 / 0.69
Cephaloflexa bergi
100 / 100 / 0.98
Geoplana quagga
92.1 / 86 / -
Crenobia alpina
100 / 100 / 0.99
Phagocata vitta
Amphibioplana onnisi S.Pietro-1
Amphibioplana onnisi S. Pietro-2
100 / 100 / 0.99
Amphibioplana onnisi S. Pietro-5
32.4 / 55 / -
100 / 100 / 1
Amphinioplana onnisi Istirzili-1
Amphibioplana onnisi Pauli-1
67.3 / 80 / 0.75
Hausera hauseri
100 / 100 / 1
Kawakatsua pumila
100 / 100 / 1
Opisthobursa mexicana
85.3 / 84 / 0.99
Rhodax sp. 2
Cavernicola
100 / 100 / 1
Bdelloura candida
Cercyra hastata
93.9 / 94 / 0.95
100 / 100 / 1
Sabussowia dioica
74.4 / 86 / 0.93
Palombiella stephensoni
100 / 100 / 1
97.7 / 98 / 1
Sluysia triapertura
Procerodes littoralis
Acanthiella sp.
99.7 / 100 / 1
Reisingeria hexaoculata
98 / 100 / 1
Protomonotresidae sp.
97 / 99 / 0.99
Plagiostomum stellatum
100 / 100 / 1
Plagiostomum whitmani
94.7 / 96 / 0.86
Plicastoma cuticulata
Urastoma cyprinae
16.1 / 68 / -
Notentera ivanovi
99.7 / 100 / 1
Kronborgia isopodicola
0.2

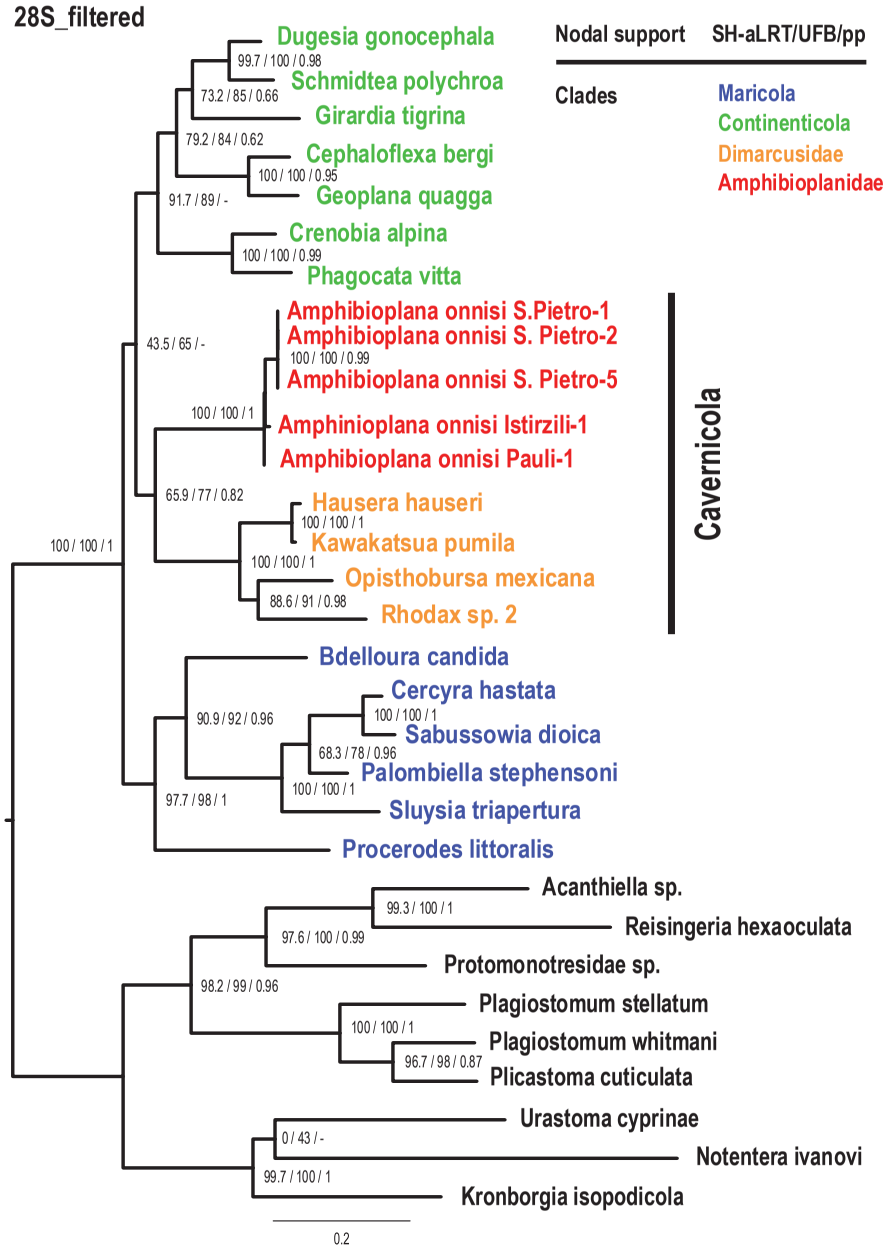
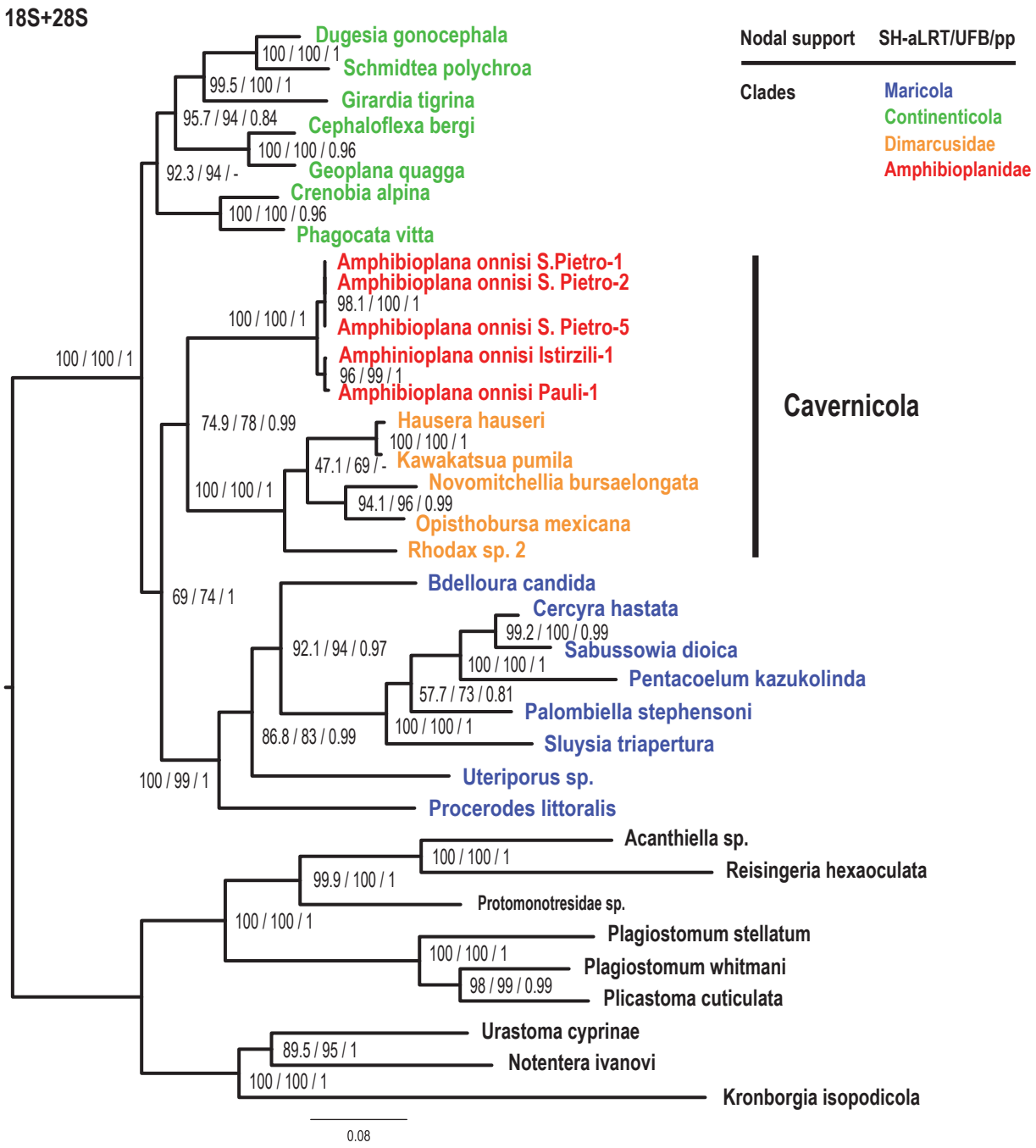

28S_filtered
Nodal support
SH-aLRT/UFB/pp
Clades
Maricola
Continenticola
Dimarcusidae
Amphibioplanidae
Dugesia gonocephala
99.7 / 100 / 0.98
Schmidtea polychroa
73.2 / 85 / 0.66
Girardia tigrina
79.2 / 84 / 0.62
Cephaloflexa bergi
100 / 100 / 0.95
Geoplana quagga
91.7 / 89 / -
Crenobia alpina
100 / 100 / 0.99
Phagocata vitta
Amphibioplana onnisi S.Pietro-1
Amphibioplana onnisi S. Pietro-2
100 / 100 / 0.99
Amphibioplana onnisi S. Pietro-5
43.5 / 65 / -
100 / 100 / 1
Amphinioplana onnisi Istirzili-1
Amphibioplana onnisi Pauli-1
65.9 / 77 / 0.82
Hausera hauseri
100 / 100 / 1
Kawakatsua pumila
100 / 100 / 1
Opisthobursa mexicana
88.6 / 91 / 0.98
Rhodax sp. 2
Cavernicola
100 / 100 / 1
Bdelloura candida
Cercyra hastata
90.9 / 92 / 0.96
100 / 100 / 1
Sabussowia dioica
68.3 / 78 / 0.96
Palombiella stephensoni
100 / 100 / 1
97.7 / 98 / 1
Sluysia triapertura
Procerodes littoralis
Acanthiella sp.
99.3 / 100 / 1
Reisingeria hexaoculata
97.6 / 100 / 0.99
Protomonotresidae sp.
98.2 / 99 / 0.96
Plagiostomum stellatum
100 / 100 / 1
Plagiostomum whitmani
96.7 / 98 / 0.87
Plicastoma cuticulata
Urastoma cyprinae
0 / 43 / -
Notentera ivanovi
99.7 / 100 / 1
Kronborgia isopodicola
0.2

Figure S3. Maximum likelihood tree inferred from the concatenated dataset, including the complete 18S and 28S alignments (Table 2). Values at nodes correspond to UFB/ SH-aLRT/ PP –: PP values below 0.5. Scale bar: number of substitutions per nucleotide position.


18S+28S
Nodal support SH-aLRT/UFB/pp
Clades
Maricola
Continenticola
Dimarcusidae
Amphibioplanidae
Dugesia gonocephala
Schmidtea polychroa
Girardia tigrina
Cephaloflexa bergi
Geoplana quagga
Crenobia alpina
Phagocata vitta
Amphibioplana onnisi S.Pietro-1
Amphibioplana onnisi S. Pietro-2
Amphibioplana onnisi S. Pietro-5
Amphinioplana onnisi Istirzili-1
Amphibioplana onnisi Pauli-1
Hausera hauseri
Kawakatsua pumila
Novomitchellia bursaelongata
Opisthobursa mexicana
Rhodax sp. 2
Cavernicola
Bdelloura candida
Cercyra hastata
Sabussowia dioica
Pentacoelum kazukolinda
Palombiella stephensoni
Sluysia triapertura
Uteriporus sp.
Procerodes littoralis
Acanthiella sp.
Reisingeria hexaoculata
Protomonotresidae sp.
Plagiostomum stellatum
Plagiostomum whitmani
Plicastoma cuticulata
Urastoma cyprinae
Notentera ivanovi
Kronborgia isopodicola
100 / 100 / 1
99.5 / 100 / 1
95.7 / 94 / 0.84
100 / 100 / 0.96
92.3 / 94 / -
100 / 100 / 0.96
98.1 / 100 / 1
100 / 100 / 1
96 / 99 / 1
74.9 / 78 / 0.99
100 / 100 / 1
47.1 / 69 / -
100 / 100 / 1
94.1 / 96 / 0.99
69 / 74 / 1
92.1 / 94 / 0.97
99.2 / 100 / 0.99
100 / 100 / 1
57.7 / 73 / 0.81
86.8 / 83 / 0.99
100 / 100 / 1
100 / 99 / 1
100 / 100 / 1
99.9 / 100 / 1
100 / 100 / 1
100 / 100 / 1
98 / 99 / 0.99
89.5 / 95 / 1
100 / 100 / 1
0.08

Supplementary Table S1. Primer sequences, annealing temperatues and references

Gene	Primers	Sequence	Annealing temperature (ºC)	References
18S	18S1F	TACCTGGTTGATCCTGCCAGTAG	50	Giribet et al. 1996
	18S5R	GCGAAAGCATTTGCCAAGAA	50	Giribet et al. 1996
	18S4F	CCAGCAGCCGCGCTAATTC	50	Giribet et al. 1996
	18S9R	GATCCTTCCGCAGGTTCACCTAC	50	Giribet et al. 1996
28S	28S1F	TATCAGTAAGCGGAGGAAAAG	55	Álvarez-Presas et al. 2008
	28S4R	CCAGCTATCCTGAGGG	55	Álvarez-Presas et al. 2008
	28S2F	CTGAGTCCGATAGCAAACAAG	55	Álvarez-Presas et al. 2008
	28S6R	GGAACCCCTTCTCCACTTCAGT	55	Álvarez-Presas et al. 2008
16S	Tor_16S_F	TACTCTGACTGTGCTAAGGT	51	This study
	Lara_16S_R	GATCCAACATCGAGGTA	51	This study
	16Sa	CGC CTG TTT ATC AAA AAC AT	47	Xiong and Kocher 1991
	16Sb	CTC CGG TTT GAA CTC AGA TCA	47	Xiong and Kocher 1991

Supplementary Table S2. PCR conditions for each gene

18S		94º	2'
		94º	45''
	x30	50º	50"
		72º	35''
		72º	3'
28S		94º	1'
		94º	45''
	x35	55º	45''
		72º	45''
		72º	4'
16S		94º	1'
		94º	45''
	x35	51º	35''
		72º	45''
		72º	2'

Supplementary Table S3. Kimura-2 parameter distances between *Amphibioplana* 16S sequences obtained for populations from the three caves

	Name in tables & trees							
	A. onnisi-Spietro-1							
MR0567-01 to MR0567-07	A. onnisi-Spietro-2	0,000						
MR0562-01	A. onnisi-Spietro-3	0,000	0,000					
MR0562-02	A. onnisi-Spietro-4	0,000	0,000	0,000				
MR1296-01 to MR1296-04	A. onnisi-Istirzili-1	0,147	0,147	0,147	0,147			
MR1296-05	A. onnisi-Istirzili-2	0,148	0,148	0,148	0,148	0,004		
MR0568-01	A. onnisi-Pauli-1	0,150	0,150	0,150	0,150	0,059	0,058	
		S. Pietro-1	S. Pietro-2	S. Pietro-3	S. Pietro-4	Istirzili-1	Istirzili-2	Pauli-1

Island-specific evolution of a sex-primed autosome in a sexual planarian

https://doi.org/10.1038/s41586-022-04757-3

Received: 15 September 2021

Accepted: 12 April 2022

Published online: 1 June 2022

Open access

Check for updates

Longhua Guo[1,2]✉, Joshua S. Bloom[1,2], Daniel Dols-Serrate[3], James Boocock[1,2], Eyal Ben-David[4], Olga T. Schubert[1,2], Kaiya Kozuma[1], Katarina Ho[1], Emily Warda[1], Clarice Chui[1], Yubao Wei[5], Daniel Leighton[1,2], Tzitziki Lemus Vergara[1,2], Marta Riutort[3], Alejandro Sánchez Alvarado[2,6] & Leonid Kruglyak[1,2]✉

The sexual strain of the planarian *Schmidtea mediterranea*, indigenous to Tunisia and several Mediterranean islands, is a hermaphrodite[1,2]. Here we isolate individual chromosomes and use sequencing, Hi-C[3,4] and linkage mapping to assemble a chromosome-scale genome reference. The linkage map reveals an extremely low rate of recombination on chromosome 1. We confirm suppression of recombination on chromosome 1 by genotyping individual sperm cells and oocytes. We show that previously identified genomic regions that maintain heterozygosity even after prolonged inbreeding make up essentially all of chromosome 1. Genome sequencing of individuals isolated in the wild indicates that this phenomenon has evolved specifically in populations from Sardinia and Corsica. We find that most known master regulators[5–13] of the reproductive system are located on chromosome 1. We used RNA interference[14,15] to knock down a gene with haplotype-biased expression, which led to the formation of a more pronounced female mating organ. On the basis of these observations, we propose that chromosome 1 is a sex-primed autosome primed for evolution into a sex chromosome.

Sex chromosomes evolve from homologous autosomes that acquire sex-determining genes and lose their ability to recombine[16–21]. As such, sex chromosome evolution and recombination suppression are closely associated[16–21]. However, because direct evidence of such homologous autosomes primed for evolution into sex chromosomes is difficult to capture, little is known about the molecular signatures associated with the evolution of recombination suppression.

The freshwater planarian *Schmidtea mediterranea*, an important model organism for studies of regeneration[22,23], exists as asexual and sexual reproductive strains. The sexual strain is distributed mostly in Tunisia and on the islands of Sardinia, Corsica and Sicily[1]. The sexual strain is a simultaneous hermaphrodite that develops both male and female reproductive systems in the same adult individual and obligately outcrosses to fertilize other individuals[24,25]. Individuals in the asexual strain do not develop sexual reproductive systems. We considered that studying chromosome evolution in a simultaneous hermaphrodite might provide insight into the early evolution of a primitive sex chromosome.

S. mediterranea has four pairs of chromosomes, which are stably diploid. The genome is reported to comprise 774 Mb assembled as 481 non-contiguous series of genomic sequences, or scaffolds[26–28]. A previous study found that approximately 300 Mb of the genome remained heterozygous even after extensive inbreeding of laboratory strains, and that this phenomenon also occurs naturally in wild populations in Sardinia[24]. The two sets of heterozygous alleles were collectively named J and V haplotypes. To define the chromosomal locations of these alleles and investigate the reasons underpinning the persistence of heterozygosity in *S. mediterranea*, a detailed assembly of all four chromosomes is needed.

Chromosome-scale genome assembly

To transform the 481 scaffolds[26] into a chromosome-scale genome reference, we carried out chromosome sequencing (ChrSeq)[29,30] of a laboratory strain, as well as chromatin proximity sequencing by Hi-C[3] (Fig. 1a). To do so, we dissected individual chromosomes from mitotic cells using laser capture and amplified and sequenced each chromosome individually (Fig. 1b). We examined the sequencing depth of each scaffold across multiple samples of the same and different chromosomes to ensure reproducibility and specificity, respectively (Extended Data Fig. 1). Overall, we successfully amplified and confidently assigned 740 Mb of the 774-Mb genome to one of four chromosomes (Supplementary Table 1).

We used ChrSeq information and chromatin interaction data generated by Hi-C to correct and connect the individual scaffolds within a chromosome into a chromosome-scale genome, hereafter referred to as Smed_chr_ref_v1. The Hi-C data analysed using the SALSA scaffolding algorithm[3] resolved the 481 scaffolds into 57 super-scaffolds and 104 singletons. The ChrSeq data indicated that 3 of the 57 super-scaffolds were disjointed inter-chromosomal fragments, consistent with the Hi-C contact heat map (Fig. 1c). We split the disjointed super-scaffolds, merged the scaffolds into chromosomes, and ordered and oriented

[1]Department of Human Genetics and Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA. [2]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [3]Departament de Genètica, Microbiologia i Estadística, Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain. [4]Department of Biochemistry and Molecular Biology, Institute for Medical Research Israel-Canada, Hebrew University of Jerusalem–Hadassah Medical School, Jerusalem, Israel. [5]Institute of Reproductive Medicine, Henan Provincial People's Hospital, Zhengzhou University, Zhengzhou, China. [6]Stowers Institute for Medical Research, Kansas City, MO, USA. ✉e-mail: longhuaguo@mednet.ucla.edu; lkruglyak@mednet.ucla.edu

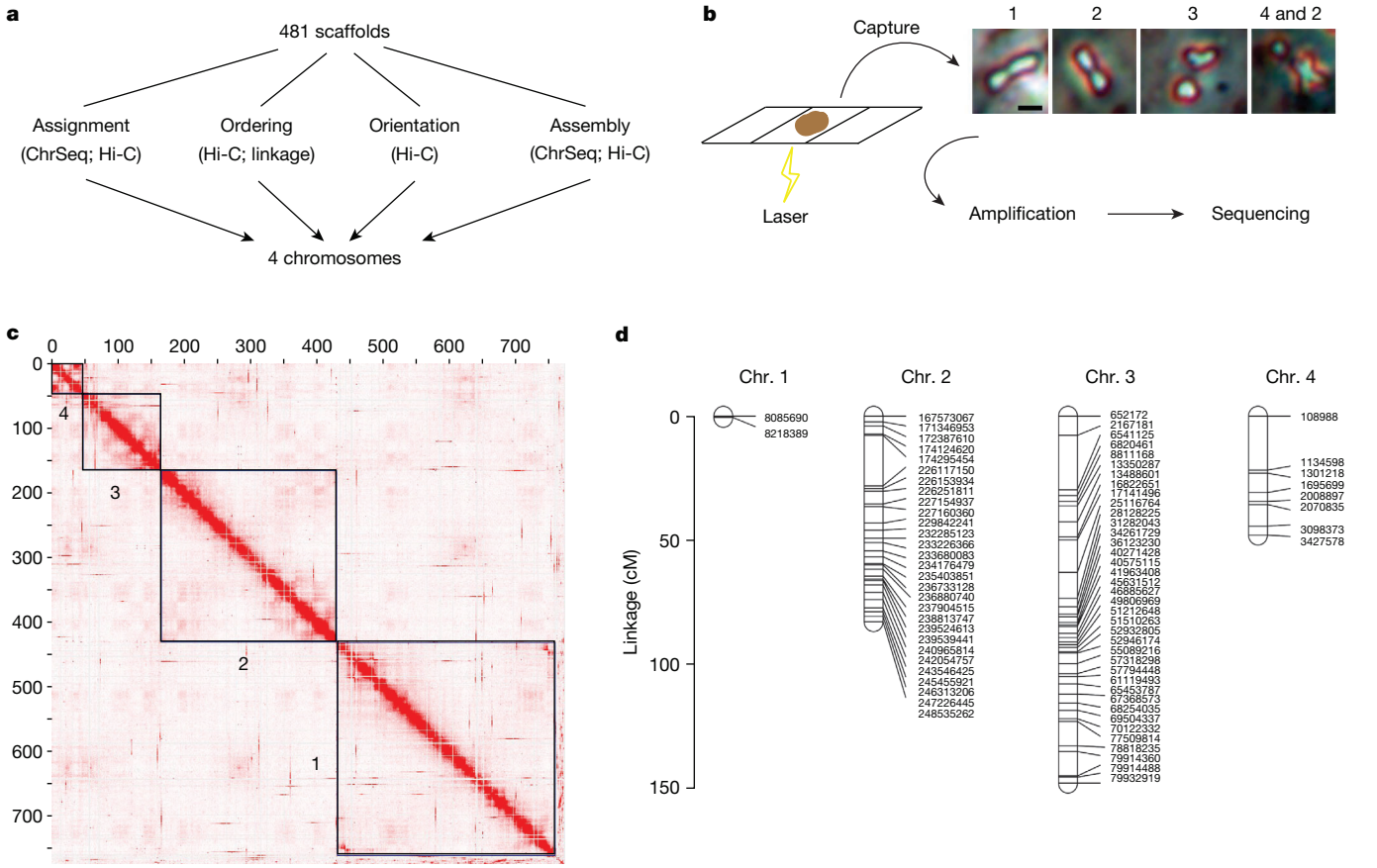


Fig. 1 | Chromosome-scale genome assembly. a, Schematic of the use of Hi-C, ChrSeq and a linkage map to transform 481 scaffolds[26] into a chromosome-scale genome assembly. **b**, Chromosomes from mitotic cells spread onto membrane slides for laser capture and sequencing. Numbers denote the identity of chromosomes present in the representative samples. **c**, Contact heat map of chromatin interactions indicated by Juicebox[4]; black boxes denote the four chromosomes in the final assembly. **d**, Linkage map generated through LinkageMapView[68], showing a lack of recombination on chromosome 1. Tick marks and labels indicate genetic markers.

all scaffolds within the chromosomes using Juicebox visualization software[4] (Methods). Chromosome assignments by ChrSeq alone (Supplementary Table 1) and by Hi-C alone were inconsistent for only 3 of 384 (0.8%) scaffolds (Supplementary Table 2). We manually assigned these three scaffolds to chromosomes according to the Hi-C data. Hi-C also detected 26 inter-chromosomal or intra-chromosomal assembly errors (Supplementary Table 2) in the previous assembly[26], 5 of which were confirmed by ChrSeq to be inter-chromosomal disjointed scaffolds (Supplementary Table 2). The final genome assembly, Smed_chr_ref_v1 (Fig. 1c), had four chromosomes with a total size of 764 Mb, which is 98.4% of that reported previously[26]. Of the 1.6% of the previous assembly not contained in Smed_chr_ref_v1 by Hi-C, about half (52.6%) of the scaffolds could not be assigned to a chromosome or were assigned to two chromosomes by ChrSeq, indicating lower assembly quality (Supplementary Table 3). Moreover, 33.0% and 12.4% of the scaffolds had chromatin interaction signals with structurally complex regions of chromosomes 1 and 2, respectively (Extended Data Fig. 2), suggesting that they might be alternative assemblies or repetitive sequences. In our new assembly, Smed_chr_ref_v1, the two ends of chromosome 4 were capped by >1,000 copies of the telomere repeat TTAGGG, indicating high assembly quality.

To validate the linearity of the chromosomes in Smed_chr_ref_v1, we generated a linkage map. We crossed two divergent laboratory strains of *S. mediterranea*, S2F10b and D5, to produce an F_2 population, and we genotyped individual worms with RADseq[31]. Eighty markers that were evenly distributed and genotyped in at least 98% of the F_2 segregants (91 of 93) were used to establish four linkage groups (Fig. 1d and Supplementary Table 4) representing the four chromosomes. The ordering of the 80 markers in the linkage map is consistent with Smed_chr_ref_v1, independently supporting the quality of our chromosome-scale genome assembly. This highly contiguous and complete genome assembly and linkage map together facilitate further genetic and epigenetic investigations of the functions of the genome in this model planarian.

Chromosome 1 recombination suppression

We next re-examined the heterozygous alleles that define the J and V haplotypes in the newly assembled genome. We found that 87.7% of the genetic markers that remained heterozygous were located on chromosome 1, spanning 333 Mb at a density of 30,148 variants per 10 Mb (Fig. 2a). The remaining 12.3% of the heterozygous markers were located on the other three chromosomes at a density of 3,274 variants per 10 Mb and probably correspond to differences between highly similar copies of repetitive elements rather than true polymorphisms (Supplementary Table 5). All F_2 worms ($n = 93$) in this study were heterozygous for chromosome 1 but not for chromosomes 2–4 (Supplementary Table 6), which is consistent with the previous study[24]. Hence, we concluded that the J/V haplotypes were on chromosome 1.

Our linkage map revealed an extremely low rate of recombination on chromosome 1 of only 0.5 cM for the entire chromosome (Fig. 1d). This is particularly notable because, at 333 Mb, chromosome 1 was the largest of the four chromosomes, containing more than 40% of the genome.

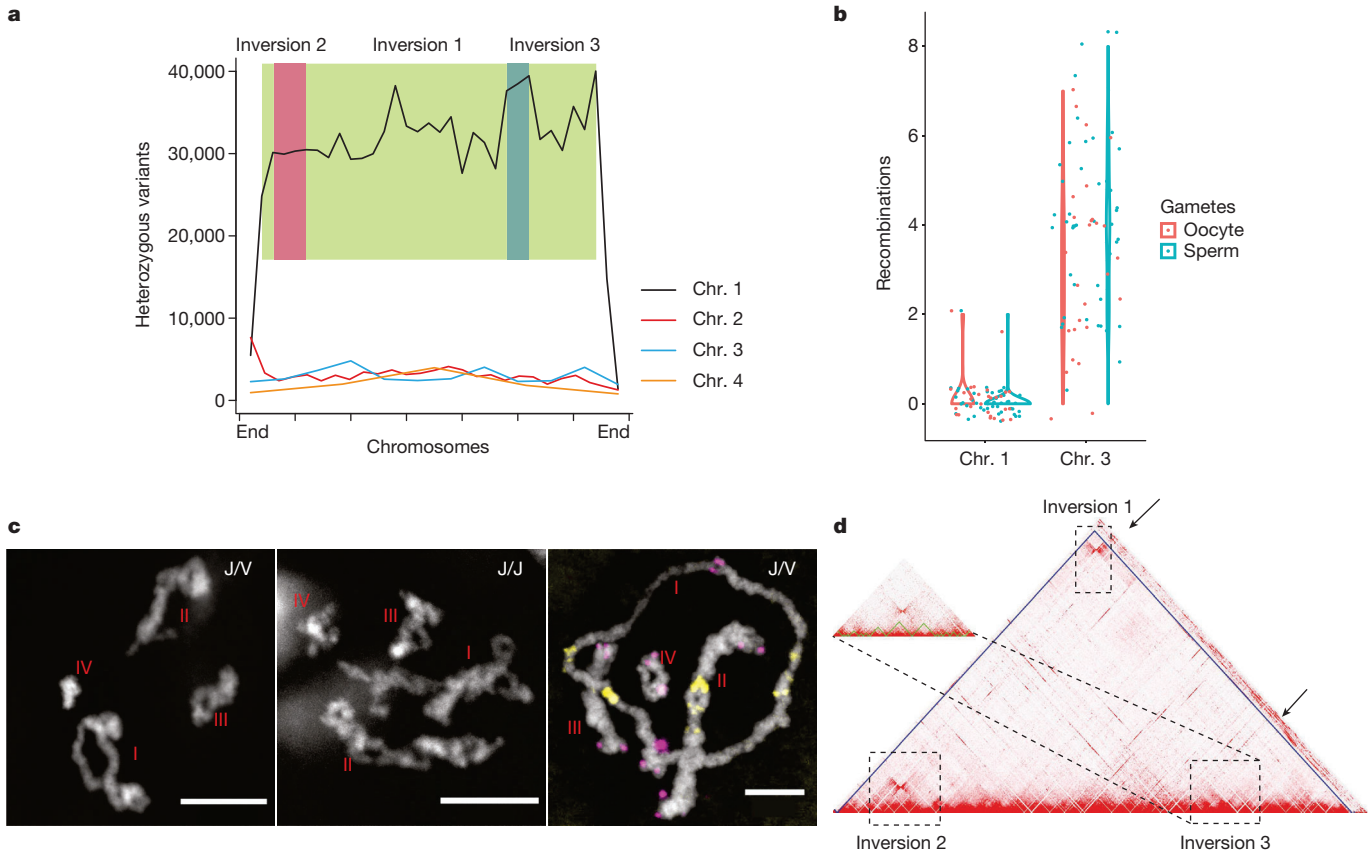


Fig. 2 | Recombination suppression and inversions on chromosome 1. a, Distribution of heterozygous variants maintained in the S2 inbreeding pedigree along the four chromosomes. The *y* axis shows variant counts per 10 Mb. Green, pink and blue boxes represent inversions 1, 2 and 3, respectively. **b**, Number of meiotic recombination events on chromosomes 1 and 3 in oocytes (red) and sperm cells (cyan). Dots represent individual gametes, summarized with violin plots. **c**, Chromosome 1 in late prophase I in oocytes from J/V worms (left and right panels) shows fewer crossovers than in oocytes from J/J worms (middle panel). This experiment was repeated more than 10 times independently with similar results. The right panel shows FISH[42] results with probes for the telomeric repeat TTAGGG (magenta) and a repeat located near the centromere (yellow). This experiment was repeated more than three times independently with similar results. **d**, Chromatin contact heat map for chromosome 1. Dashed-line rectangles indicate the locations of potential inversions. Arrows indicate chromatin contact signals with inversion regions from unassigned scaffolds.

To directly examine whether chromosome 1 can recombine, we sequenced 45 single sperm cells and 28 single oocytes from J/V line S2 (Fig. 2d). Gamete sequencing is preferred because recombination events in hatchlings can be selected for by differential fertilization or embryonic lethality. We identified 3,197 single-nucleotide variants (SNVs) on chromosome 1 and 3,312 SNVs on chromosome 3, covering 99% of the length of each chromosome (Supplementary Table 7). The SNVs were distributed at a similar density across 20-Mb windows, with coefficients of variation of 0.38 and 0.31 for chromosomes 1 and 3, respectively (Supplementary Table 7). We observed that 98% of the sperm cells (44 of 45) and 93% of the oocytes (26 of 28) had no crossovers on chromosome 1. By contrast, most sperm cells and oocytes had crossovers on chromosomes 2, 3 and 4 (Fig. 2b, Extended Data Fig. 3a, b and Supplementary Table 7). We thus concluded that recombination on chromosome 1 was strongly suppressed.

Consistent with this conclusion, we found that during prophase I, when other chromosomes had numerous crossovers, chromosome 1 formed a ring structure in the oocyte in J/V worms but not in J/J worms (Fig. 2c and Extended Data Fig. 4). Fluorescence in situ hybridization (FISH) with telomere probes suggested that crossovers between the two homologous pairs of chromosome 1 occurred only in regions close to the telomeres, leading to the observed ring conformation rather than the side-by-side pairing observed for chromosomes 2, 3 and 4. Furthermore, Hi-C analysis showed that chromosome 1 had three putative inversions, each >20 Mb in size (Fig. 2d); such inversions can cause crossover suppression[32–34]. The rest of the genome had only one large inversion of approximately 10 Mb on chromosome 2 (Extended Data Fig. 2b).

Island-specific evolution of chromosome 1

To investigate the genetic diversity of *S. mediterranea* in its natural environment and determine whether chromosome 1 J/V heterozygosity occurs throughout the entire species, we used RADseq[31] to sample the genomes of 70 sexual individuals from Sardinia, Corsica, Sicily and Tunisia and 2 asexual individuals from Menorca[1] (Fig. 3a). To look for genetic relationships between the individuals, we determined clustering of relatedness measured by identity-by-state pairwise distances[35] and identified two superclusters. Animals from Sicily were closely related to those from Tunisia, and animals from Sardinia were closely related to those from Corsica (Fig. 3b). Phylogenetic clustering (Extended Data Fig. 5), fixation index (F_{ST}) values (Extended Data Fig. 6) and analyses with Structure (Extended Data Fig. 7) further supported this observation.

The relatedness of the populations from Sardinia and Corsica suggests that they may share genome characteristics that differ from those of the populations from Sicily and Tunisia. Indeed, animals in the populations from Sardinia and Corsica had greater heterozygosity on chromosome 1 than expected under Hardy–Weinberg equilibrium (Fig. 3c), whereas the heterozygosity on the other three chromosomes

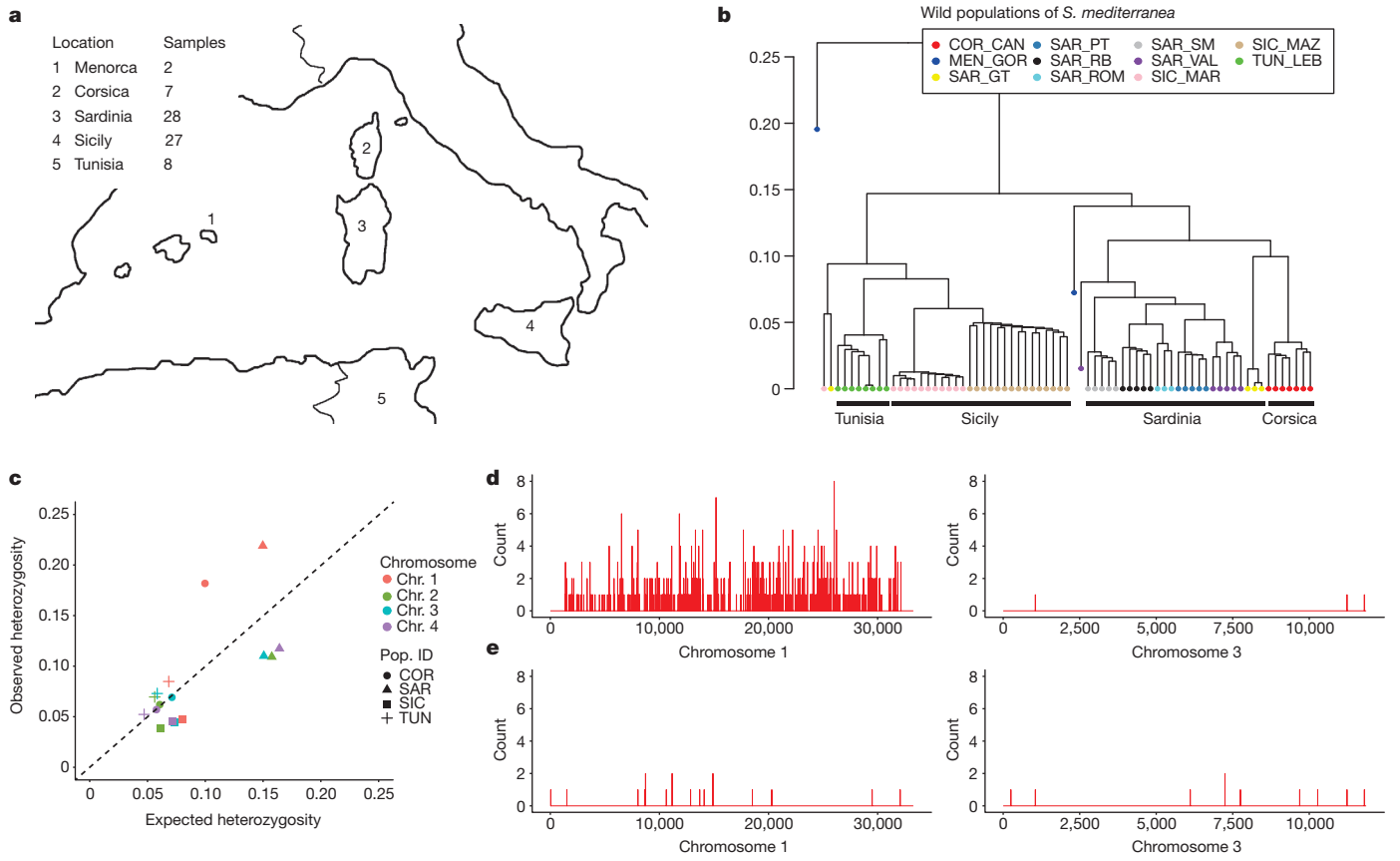


Fig. 3 | Island-specific evolution of chromosome 1 heterozygosity. **a**, Collection sites of 72 wild isolates from the Mediterranean islands and Tunisia. **b**, Maximum-likelihood tree of 2 asexual isolates from Menorca (MEN_GOR) and 70 sexual isolates from Sardinia (SAR), Corsica (COR), Sicily (SIC) and Tunisia (TUN). Dots denote individual worms, with colours corresponding to their collection sites. **c**, Observed versus expected heterozygosity for each chromosome (denoted by colour) in each population (denoted by shape). The dashed line corresponds to equality between observed values and those expected under Hardy–Weinberg equilibrium; the deviation of the points corresponding to chromosome 1 in the COR and SAR samples from the expectation is notable. **d**, **e**, Number of sites with heterozygosity in >80% of the population per 10-kb window along chromosome 1 (left) and chromosome 3 (right) in Sardinia (**d**) and Sicily (**e**).

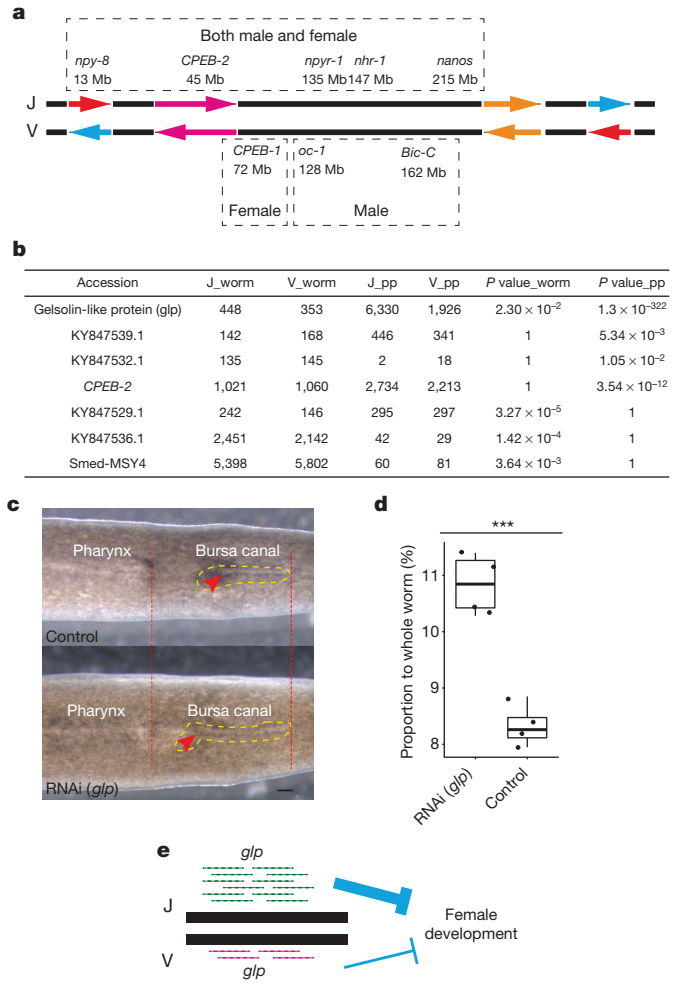
in these populations and on all four chromosomes in the other populations closely followed expectations. By analysing the J/V haplotype markers in wild populations, we found that animals from Sardinia and Corsica were heterozygous J/V, whereas those from Sicily and Tunisia were homozygous J/J. Moreover, the animals from Sardinia ($n = 28$) had many sites that were heterozygous in more than 80% of the individuals and were distributed along the length of chromosome 1 except near the ends; conversely, the animals from Sicily ($n = 27$) had very few such heterozygous sites (Fig. 3d, e and Supplementary Table 8). Few such heterozygous sites were observed on chromosomes 2, 3 and 4 in either population (Fig. 3d, e and Extended Data Fig. 3c, d). These analyses suggest that chromosome 1 specifically evolved and diverged on the islands of Sardinia and Corsica.

A sex-primed chromosome

To gain insight into the island-specific suppression of recombination on chromosome 1, we examined the genes located on this chromosome, which contained 39% of all annotated genes. Five of the seven known master regulators of the reproductive system were found on chromosome 1, including *nanos*[5], *nhr-1* (refs. [8,9]), *npy-8* (ref. [6]), *npyr-1* (ref. [11]) and *CPEB-2* (ref. [12]) (*ophis*[11] and *boule-2* (ref. [10]) were the two exceptions) (Fig. 4a). Knockdown of any one of these master regulators leads to depletion of both male and female reproductive tissues[5,6,8–12]. The presence of these genes with crucial roles in sexual development on chromosome 1 suggests that chromosome 1 integrity is important for the maintenance of sexual reproduction. Indeed, the asexual lineage of *S. mediterranea*, which is devoid of any reproductive organs[23], had a translocation from chromosome 1 to chromosome 3 and probably evolved through loss of function of one or more of these genes.

Chromosome 1 also contained three of the four known master regulators of male or female reproductive tissues (*CPEB-1* (ref. [12]), *onecut*[13] and *Bic-C*[5]; *dmd-1* (ref. [36]) was the exception) (Fig. 4a). The *CPEB-1* gene is specifically required for the development of oocytes and yolk glands[12]. Loss of *onecut* and *Bic-C* expression leads to a 'no-testes' phenotype without affecting the ovary[5,13]. The presence of a female-determining gene on a chromosome that does not recombine provides an ideal foundation for the evolution of a sex chromosome. A loss-of-function mutation in the gene on one of the two homologous chromosomes would turn the chromosome with the mutation into a male-determining chromosome. Similar logic applies to a male-determining gene.

Haplotype-specific expression of sex-specific genes is a predicted signature of a sex-primed autosome. To test this prediction, we examined the expression of 20 genes in sexually mature adult worms and in the male copulatory organ, the penis papilla[7], in a J/V line. The 20 genes were chosen because they were characterized as having sex-specific function and/or expression in the literature and include the 8 master regulators mentioned above[5,12] (Supplementary Table 9). Of these 20 genes, 13 contained at least one heterozygous variant in the coding sequence, which allowed us to determine whether these genes showed biased expression from the J or V haplotype (Extended Data Fig. 8). We found that a gelsolin-like protein[12], *Smed-glp*, was expressed predominantly



Accession	J_worm	V_worm	J_pp	V_pp	*P* value_worm	*P* value_pp
Gelsolin-like protein (glp)	448	353	6,330	1,926	2.30×10^{-2}	1.3×10^{-322}
KY847539.1	142	168	446	341	1	5.34×10^{-3}
KY847532.1	135	145	2	18	1	1.05×10^{-2}
CPEB-2	1,021	1,060	2,734	2,213	1	3.54×10^{-12}
KY847529.1	242	146	295	297	3.27×10^{-5}	1
KY847536.1	2,451	2,142	42	29	1.42×10^{-4}	1
Smed-MSY4	5,398	5,802	60	81	3.64×10^{-3}	1

Fig. 4 | Acquisition and haplotype-specific expression of sex-related genes. **a**, Schematic diagram of the inferred structure of the J and V haplotypes on chromosome 1. Arrows denote putative inversions. Names, positions and sex specificity are shown for the eight genes on chromosome 1 with known key roles in the development of the reproductive system. **b**, Comparison of read counts for the J and V alleles of key genes in the transcriptomes of whole sexually mature worms and penis papillae (PP). Bonferroni-corrected *P* values from a two-sided binomial test of equal expression are shown. **c**, Dorsal view of control (top) and *Smed-glp* knockdown (bottom) sexually mature planarians. Red dashed lines and arrows indicate the posterior and anterior ends of the pharynx and the bursa canal, respectively. Scale bar, 400 um. Yellow dashed lines outline the bursa canal. **d**, Quantification of the length of the bursa canal relative to that of the whole worm. The *y* axis shows the percentage of whole-worm length spanned by the bursa canal as a box plot, with individual data points shown ($n = 4$ each for control and knockdown worms). Statistical significance was assessed with a two-sided, two-sample Student's *t*-test (***$P = 0.00029$). RNAi (*glp*): minimum, 10.28; maximum, 11.4; median, 10.845; first quartile, 10.375; third quartile, 11.31. Control: minimum, 7.95; maximum, 8.85; median, 8.26; first quartile, 8.06; third quartile, 8.6. **e**, Model of negative regulation of bursa canal development by *Smed-glp* alleles.

from the J haplotype in both whole worms and the penis papilla, with more haplotype-specific expression in the male copulatory organ. Four other genes showed biased expression from the J haplotype in whole worms (KY847529.1 and KY847536.1) or in the penis papilla (KY847539.1 and *CPEB-2*), and two additional genes showed biased expression from the V haplotype in the male copulatory organ (KY847532.1) or in whole worms (Smed-MSY4) (Fig. 4b and Supplementary Table 9). All seven of these genes were localized within the inversions or close to inversion breakpoints (Supplementary Table 5).

We used FISH with hybridization chain reaction to confirm that the expression of *Smed-glp* is highly enriched in both the penis papilla and the bursa canal[7,37] (Extended Data Fig. 9), which is a female organ used to receive sperm from mating partners. To examine the function of *Smed-glp*, we used RNA interference (RNAi)[14,15] to knock down its expression in young hatchlings. After 8 weeks of feeding, both the knockdown and control hatchlings reached sexual maturity. Notably, the bursa canal was much more pronounced in the knockdown animals (Fig. 4c, d). No obvious morphological differences were observed in the male copulatory organ. These results suggest that the J allele of *Smed-glp* may have a greater role in preventing over-development of the female bursa canal (Fig. 4e), probably by controlling the number of organ-specific muscle fibres[38]. This is consistent with a previous observation that J/J individuals show higher egg production, suggesting a larger contribution to female reproduction (Fig. 4b in ref. [24]; Supplementary Information).

We have thus identified a chromosome in the hermaphrodite planarian *S. mediterranea* that does not recombine, is enriched in master regulators of reproductive systems, shows allele-specific expression of sex-related genes, and contains genes orthologous to those on the sex chromosome of *Schistosoma mansoni* (Supplementary Tables 9 and 10, and Supplementary Information), which is the only known Platyhelminthes species with differentiated sex chromosomes[39]. These observations collectively led us to propose that chromosome 1 of *S. mediterranea* is primed for evolution into a sex chromosome.

Chromosome 1 of *S. mediterranea* marks an intriguing system for studying sex chromosome evolution[40]. Planarians from the islands of Sardinia and Corsica are heterozygous for the J and V haplotypes, whereas those in Sicily and Tunisia exist as J/J homozygotes. In laboratory crosses involving J/V lines, all hatchlings were J/V heterozygotes[24]. No V/V planarians have been identified either in nature or in laboratory crosses. We genotyped single zygotes from crosses between J/V individuals and determined that homozygous zygotes (J/J or V/V) exist (Extended Data Fig. 10a and Supplementary Information). We propose that early embryonic lethality leads to the loss of J/J and V/V adults in both nature and laboratory crosses. Such lethality may arise from degeneration of coding sequences on the J and V haplotypes as a consequence of crossover suppression. Indeed, chromosome 1 has an elevated rate of transposable elements and mutations introducing stop codons relative to the rest of the genome (Supplementary Table 11) as well as higher nonsynonymous substitution rates (Supplementary Table 13 and Extended Data Fig. 10b). As a consequence, the J and V haplotypes each carry unique sets of functional genes that are silenced or truncated on the other haplotype (Supplementary Table 11).

The presence of three nested inversions on chromosome 1 suggests the possibility that recombination suppression on this chromosome may have evolved in a stepwise manner similar to that reported for the human X and Y chromosomes[17]. In support of this possibility, we observed evidence for three evolutionary strata corresponding to the inverted regions. We used PacBio genome sequencing data[26] to identify long reads that bridged the inversion breakpoints, and we validated the three inversions identified by Hi-C (Supplementary Table 12 and Supplementary Information). We observed that the rates of heterozygous sites and synonymous substitutions were elevated within the three inversions in a pattern consistent with the evolutionary strata (Fig. 2a, Supplementary Tables 5 and 13, and Extended Data Fig. 10b, c). We used deep whole-genome sequencing data of individuals from laboratory crosses[24] to estimate the de novo mutation rate in sexual planarians to be approximately 1.0×10^{-8} mutations per nucleotide per generation. On the basis of this estimate, it is likely that inversions 3, 1 and 2 evolved approximately 450,000, 320,000 and 260,000 generations ago, respectively.

Although it is not possible to know how chromosome 1 will evolve, our findings provide a snapshot of an incipient stage that supports the

hypothesis that sex chromosomes evolve from homologous autosomes that acquire sex-specific roles and cease to recombine[16–19]. The locked J/V heterozygous system may facilitate this process by maintaining sex-specific alleles in the planarian population before the evolution of dioecy[41]. We propose that the planarian chromosome 1 haplotypes provide a unique opportunity to directly examine the molecular characteristics of a sex-primed autosome.

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/s41586-022-04757-3.

1. Lazaro, E. M. et al. *Schmidtea mediterranea* phylogeography: an old species surviving on a few Mediterranean islands? *BMC Evol. Biol.* **11**, 274 (2011).
2. Benazzi, M., Baguná, J., Ballester, R., Puccinelli, I. & Papa, R. D. Further contribution to the taxonomy of the «*Dugesia Lugubris-Polychroa Group*» with description of *Dugesia Mediterranea* N.S.P. (Tricladida, Paludicola). *Boll. Zool.* **42**, 81–89 (1975).
3. Burton, J. N. et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
4. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
5. Wang, Y., Zayas, R. M., Guo, T. & Newmark, P. A. *Nanos* function is essential for development and regeneration of planarian germ cells. *Proc. Natl Acad. Sci. USA* **104**, 5901–5906 (2007).
6. Collins, J. J. III et al. Genome-wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS Biol.* **8**, e1000509 (2010).
7. Chong, T., Stary, J. M., Wang, Y. & Newmark, P. A. Molecular markers to characterize the hermaphroditic reproductive system of the planarian *Schmidtea mediterranea*. *BMC Dev. Biol.* **11**, 69 (2011).
8. Tharp, M. E., Collins, J. J. III & Newmark, P. A. A lophotrochozoan-specific nuclear hormone receptor is required for reproductive system development in the planarian. *Dev. Biol.* **396**, 150–157 (2014).
9. Zhang, S. et al. A nuclear hormone receptor and lipid metabolism axis are required for the maintenance and regeneration of reproductive organs. Preprint at *bioRxiv* https://doi.org/10.1101/279364 (2018).
10. Iyer, H., Issigonis, M., Sharma, P. P., Extavour, C. G. & Newmark, P. A. A premeiotic function for boule in the planarian *Schmidtea mediterranea*. *Proc. Natl Acad. Sci. USA* **113**, E3509–E3518 (2016).
11. Saberi, A., Jamal, A., Beets, I., Schoofs, L. & Newmark, P. A. GPCRs direct germline development and somatic gonad function in planarians. *PLoS Biol.* **14**, e1002457 (2016).
12. Rouhana, L., Tasaki, J., Saberi, A. & Newmark, P. A. Genetic dissection of the planarian reproductive system through characterization of *Schmidtea mediterranea* CPEB homologs. *Dev. Biol.* **426**, 43–55 (2017).
13. Li, P. et al. Single-cell analysis of *Schistosoma mansoni* identifies a conserved genetic program controlling germline stem cell fate. *Nat. Commun.* **12**, 485 (2021).
14. Newmark, P. A., Reddien, P. W., Cebrià, F. & Alvarado, A. S. Ingestion of bacterially expressed double-stranded RNA inhibits gene expression in planarians. *Proc. Natl Acad. Sci. USA* **100**, 11861–11865 (2003).
15. Rouhana, L. et al. RNA interference by feeding in vitro-synthesized double-stranded RNA to planarians: methodology and dynamics. *Dev. Dynam.* **242**, 718–730 (2013).
16. Bachtrog, D. A dynamic view of sex chromosome evolution. *Curr. Opin. Genet. Dev.* **16**, 578–585 (2006).
17. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
18. Rice, W. R. Evolution of the Y sex chromosome in animals. *BioScience* **46**, 331–343 (1996).
19. Charlesworth, B. The evolution of sex chromosomes. *Science* **251**, 1030–1033 (1991).
20. Muller, H. J. Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics* **3**, 422–499 (1918).
21. Charlesworth, D. Evolution of recombination rates between sex chromosomes. *Philos. Trans. R. Soc. Lond. B* **372**, 20160456 (2017).
22. Reddien, P. W. & Sanchez Alvarado, A. Fundamentals of planarian regeneration. *Annu. Rev. Cell Dev. Biol.* **20**, 725–757 (2004).
23. Newmark, P. A. & Sanchez Alvarado, A. Not your father's planarian: a classic model enters the era of functional genomics. *Nat. Rev. Genet.* **3**, 210–219 (2002).
24. Guo, L., Zhang, S., Rubinstein, B., Ross, E. & Alvarado, A. S. Widespread maintenance of genome heterozygosity in *Schmidtea mediterranea*. *Nat. Ecol. Evol.* **1**, 19 (2016).
25. Zayas, R. M. et al. The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: analysis of ESTs from the hermaphroditic strain. *Proc. Natl Acad. Sci. USA* **102**, 18491–18496 (2005).
26. Grohme, M. A. et al. The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature* **554**, 56–61 (2018).
27. Brandl, H. et al. PlanMine—a mineable resource of planarian biology and biodiversity. *Nucleic Acids Res.* **44**, D764–D773 (2016).
28. Robb, S. M., Ross, E. & Sanchez Alvarado, A. SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res.* **36**, D599–D606 (2008).
29. Weise, A. et al. High-throughput sequencing of microdissected chromosomal regions. *Eur. J. Hum. Genet.* **18**, 457–462 (2010).
30. Ma, L. et al. Direct determination of molecular haplotypes by chromosome microdissection. *Nat. Methods* **7**, 299–301 (2010).
31. Bayona-Vasquez, N. J. et al. Adapterama III: quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ* **7**, e7724 (2019).
32. Dobzhansky, T. & Epling, C. The suppression of crossing over in inversion heterozygotes of *Drosophila pseudoobscura*. *Proc. Natl Acad. Sci. USA* **34**, 137–141 (1948).
33. Miller, D. E. et al. The molecular and genetic characterization of second chromosome balancers in *Drosophila melanogaster*. *G3* **8**, 1161–1171 (2018).
34. Sun, Y., Svedberg, J., Hiltunen, M., Corcoran, P. & Johannesson, H. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat. Commun.* **8**, 1140 (2017).
35. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
36. Chong, T., Collins, J. J. III, Brubacher, J. L., Zarkower, D. & Newmark, P. A. A sex-specific transcription factor controls male identity in a simultaneous hermaphrodite. *Nat. Commun.* **4**, 1814 (2013).
37. Hyman, L. H. in *The Invertebrates*, vol. II (ed. Boell, E. J.) 52–458 (McGraw-Hill Book Company, 1951).
38. Bertin, B. et al. *Gelsolin* and *dCryAB* act downstream of muscle identity genes and contribute to preventing muscle splitting and branching in *Drosophila*. *Sci. Rep.* **11**, 13197 (2021).
39. Buddenborg, S. et al. Assembled chromosomes of the blood fluke *Schistosoma mansoni* provide insight into the evolution of its ZW sex-determination system. Preprint at *bioRxiv* https://doi.org/10.1101/2021.08.13.456314 (2021).
40. Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997 (1978).
41. Charlesworth, D. & David, S. in *Sex Determination in Plants*, 1st ed. (ed. Ainsworth, C. C.) 25–50 (Garland Science, 1999).
42. Guo, L. et al. An adaptable chromosome preparation methodology for use in invertebrate research organisms. *BMC Biol.* **16**, 25 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

Methods

Planarian husbandry, RNAi and phenotyping

Sexual planarians were fed organic beef liver once a week. All animals used for experiments were selected randomly. For RNAi, hatchlings around 2 weeks old from strain S2F10 were used. RNAi food was prepared by mixing 1 µg of double-stranded RNA with 10 µl of liver paste[15]. To examine the RNAi phenotypes, photographs were taken with a stereomicroscope (Zeiss) when the animals were freely swimming. The lengths of the worms or bursa canal were measured with ImageJ software by multiple researchers in a double-blind experimental design. The ratio of the bursa canal to the whole worm was used to mitigate potential variations in worm size and their degree of relaxation caused by their soft bodies.

Chromosome sequencing

Chromosomes were collected from multiple animals of one clonal line, S2, maintained in the laboratory by amputation and regeneration. Chromosome spreads were prepared on nuclease-free membrane slides (Zeiss) according to a previously developed protocol except that, at the last step, the tissues were dissociated into single nuclei and placed onto the slides without squashing with a coverslip[42]. Single chromosomes were identified under a ×40 lens and were collected into the caps of single PCR tubes by PALM MicroBeam laser microdissection (Zeiss). The collected chromosomes were spun down with a tabletop centrifuge in 4 µl PBS, and the DNA in the pellets was amplified with a REPLI-g Single Cell Kit (Qiagen) for sequencing on a MiSeq or HiSeq 3000 sequencing system (Illumina).

Chromosome-scale genome assembly

A Hi-C sequencing library was prepared from multiple animals of the S2 strain using the enzyme DpnII. Sequenced reads were aligned to dd_Smes_g4.fasta[26] with bwa mem (version 0.7.17)[43]. An assembly file was prepared from the SALSA[3] output FINAL.fasta with juicebox_scripts (Phase Genomics). The .hic file was prepared by run-assembly-visualizer.sh from three-dimensional de novo assembly[44]. The two files were loaded into Juicebox[4] for scaffold manipulation using split, merge, order and orient commands and for chromosome assembly. The modified assembly file (chromosome-scale) was converted to fasta (Smed_chr_ref_v1) with juicebox_assembly_converter.py.

Hi-C library construction

A Hi-C library was generated using Phase Genomics Proximo Animal Kit version 3.0. Approximately four worms were finely chopped and were then cross-linked for 20 min at room temperature with end-over-end mixing in 1 ml of Proximo cross-linking solution. The cross-linking reaction was terminated with quenching solution for 15 min at room temperature, again with end-over-end mixing. The quenched tissue was rinsed once with 1× chromatin rinse buffer (CRB), transferred to a liquid nitrogen-cooled mortar and ground to a fine powder. The powder was resuspended in 700 µl Proximo lysis buffer 1 and lysed with glass beads for 20 min at room temperature on a vortex mixer. A low-speed spin was used to clear the large debris, and the chromatin-containing supernatant was transferred to a new tube. After a second spin at higher speed, the supernatant was removed, and the pellet containing the nuclear fraction of the lysate was washed with 1× CRB. After removal of the 1× CRB wash, the pellet was resuspended in 100 µl Proximo lysis buffer 2 and incubated at 65 °C for 15 min. Chromatin was bound to the recovery beads for 10 min at room temperature. The beads were placed on a magnetic stand and washed with 200 µl of 1× CRB.

The chromatin bound on the beads was resuspended in 150 µl Proximo fragmentation buffer, and 2.5 µl of Proximo fragmentation enzyme was added. The reaction was incubated for 1 h at 37 °C, cooled to 12 °C and then incubated with 2.5 µl of finishing enzyme for 30 min. After the addition of 6 µl of Stop Solution, the beads were washed with 1× CRB and were resuspended in 100 µl of Proximo ligation buffer supplemented with 5 µl of proximity ligation enzyme. The proximity ligation reaction was incubated at room temperature for 4 h with end-over-end mixing. To this, 5 µl of reverse cross-linking enzyme was added, and the reaction was incubated at 65 °C for 1 h.

After reversing the cross-links, the free DNA was purified with recovery beads, and the Hi-C junctions were bound to streptavidin beads and washed to remove unbound DNA. The washed beads were used to prepare paired-end deep sequencing libraries using Proximo library preparation reagents.

Oocyte and sperm cell sequencing

Sperm cells were released from sexually mature S2 strain animals into calcium- and magnesium-free buffer (1% BSA). The cell dissociation solution was placed onto a slide and examined under a phase-contrast microscope to identify single sperm cells. Oocytes were released from egg capsules. Single sperm cells or oocytes were transferred into single PCR tubes for amplification with a REPLI-g Single Cell Kit (Qiagen). RADseq libraries and whole-genome libraries were prepared and sequenced on a HiSeq 3000 or NovaSeq S2 sequencing system (Illumina). RADseq sequencing data were analysed as described in the linkage map section. Whole-genome sequencing data were analysed as described in the recombination section.

Synteny analysis

For candidate genes, the protein sequences of genes of interest were obtained from *Schmidtea* specimens. Such genes were aligned to *S. mansoni* genomes[39] using the Protein to Nucleotide BLAST (tblastn) tool. To systematically examine the synteny of the whole genome, we used SonicParanoid (version 1.3.8)[45] to identify one-to-one protein orthologues between *S. mediterranea* and *S. mansoni*. The *S. mansoni* genome assembly (V9) and its protein annotations are available at https://zenodo.org/record/5149023#.Ybk9jn3MK3I.

Linkage map

A J/J line, D5, was crossed to a J/V line, S2F8b, to build an F_2 population of 93 animals. Genomic DNA was extracted from a fragment of each animal using an Easy-DNA gDNA Purification Kit (K180001, ThermoFisher). Sequencing libraries for RADseq were prepared according to the procedures of Adapterama III[31] with a few modifications. Genetic variants were identified using Stacks (version 2.41)[46,47]. All variants were filtered with VCFtools (version 0.1.14)[48] to remove insertions and deletions and to select biallelic SNVs. Clusters of markers located within 200 bp were removed because they were likely to correspond to repetitive elements. Markers disobeying Mendelian segregation were also removed. Only markers that were homozygous in both parents were used. The linkage map was built with R/QTL[49].

Quantifying recombination

Sequencing reads from S2 and its oocytes and sperm cells were aligned to Smed_chr_ref_v1 with bwa mem (version 0.7.17). Genetic variants were jointly called by using the Genome Analysis Toolkit (GATK, version 4.1.4.1) with GenomicsDB and GenotypeGVCFs[50]. Biallelic heterozygous markers in S2 were further filtered by removing abnormal markers, including those with no segregation of two alleles in the gametes, clusters of markers in close proximity (<200 bp apart) and markers of heterozygosity in sperm cells. The J and V haplotypes were manually phased for oocytes without recombination and were phased by MPR.genotyping[51] for all gametes. The MPR.genotyping package was also used to impute or correct missing or erroneous genotypes. The final genotype bins were used to identify and visualize recombination with customized R code. Quantification of recombination was focused on crossovers between long tracks of haplotypes along a chromosome. Putative gene conversion events such as short tracks of haplotype switches encompassing <1% of the chromosome length were not included.

Gene expression analysis

To examine the gene content on chromosome 1, transcriptome data were downloaded from the NCBI Sequence Read Archive (SRA). To examine genes related to sexual development, transcriptomes from sexual adults[9,12,52], sexual juveniles[9] and sexual adults with *nhr-1* RNAi[9] were used. To examine stem cell-enriched genes, transcriptomes from sorted X1 cells and CIW4 were used[53,54].

All sequencing data were aligned to dd_Smed_v6 (ref. [27]) with bwa mem (version 0.7.17). Differential gene expression was analysed with DESeq2 (version 1.26.0)[55]. Expression was quantified at the transcript level with kallisto (version 0.44.0)[56] and was imported and summarized to gene-level count matrices by tximport[57].

Haplotype-specific expression

To examine the haplotype-specific expression of critical regulators of the reproductive system, mRNA was extracted from six sexually mature animals of the S2J/V line and was analysed as three biological replicates, with two animals pooled into each replicate. Nine penis papillae were dissected from nine sexual adult animals of the same line and were analysed as three biological replicates, with penis papillae from three animals pooled into each replicate. All mRNA samples were extracted on the same day and were processed at the same time for library preparation and sequencing to minimize technical variation. Libraries for RNA-seq were prepared with a Clontech SMARTer Stranded Total RNA-seq (Pico) Kit. The workflow consisted of converting total RNA to cDNA and then adding adaptors for Illumina sequencing through PCR. The PCR products were purified, and ribosomal DNA was depleted. The cDNA fragments were further amplified with primers universal to all libraries. Lastly, the PCR products were purified again to yield the final cDNA library. Different adaptors were used to multiplex samples in one lane. Sequencing was performed on an Illumina NovaSeq 6000 with a 150-bp paired-end-read run. Data quality checks were conducted using the Illumina Sequencing Analysis Viewer. Demultiplexing was performed using Illumina Bcl2fastq2 version 2.17.

All sequencing data were aligned to dd_Smed_v6 (ref. [27]) with bwa mem (version 0.7.17). To ensure accuracy, haplotype-specific expression of the 14 genes of interest was manually examined using the Integrative Genomics Viewer. J or V allele counts were identified for each biallelic variant in the exons. For a particular gene, the allele counts were aggregated from all variants on all exons for the J and V haplotypes. The allele counts were then subjected to binomial testing and Bonferroni correction[58] to determine whether the observed allele bias was statistically significant. $P > 1$ after Bonferroni correction was set to 1.

De novo map, parameter optimization and phylogenetic inference

All datasets were run through the de novo pipeline as implemented in Stacks version 2.52 (refs. [46,47]). First, paired-end reads were demultiplexed and filtered for quality using PROCESS_RADTAGS and were truncated to a length of 135 bp. Individual reads with Phred scores below 30 or uncalled bases were discarded (96.9% of reads passed the quality filters). Optimal parameters were identified following the guidelines of ref. [59] by running multiple iterations of the de novo pipeline and varying just one parameter with each new iteration on a subset of 13 samples from the same population, [Sic_mar], following recommendations[60]. We varied the minimum stack depth (-m) between 1 and 6 (m1–m6), the number of mismatches allowed between stacks (-M) between 1 and 10 (M1–M10), and the number of mismatches allowed to merge catalogue loci (-n) while keeping all other parameters constant (m3, M2 and n0). We then compared the number of polymorphic assembled loci across samples using a sample representation cut-off of 80% (r80) and gain or loss of polymorphic loci with each new iteration. Once -m and -M were optimized, we assessed -n by evaluating the change in the number of polymorphic loci for $n = M - 1$, $n = M$ and $n = M + 1$. RAD loci were then assembled using the denovo_map.pl wrapper in Stacks, and the following parameters were set: $m = 3$, $M = 2$ and $n = 3$.

Assembled loci that were present in 75% of all individuals were kept from POPULATIONS, a minor allele frequency (MAF) filter of 0.04 (--min-maf) was used to filter out singleton SNPs that could mask population structure and a maximum observed heterozygosity (--max-obs-het) filter of 0.99 was used to remove potentially paralogous loci[61]. Additionally, to build the phylogenetic tree, we concatenated all RADseq loci after filtering (--phylip-var-all). Phylogenetic trees were built by maximum likelihood using RAxML-NG version 0.9.0 (ref. [62]), starting from a random seed and applying a GTR+G substitution model and 1,000 bootstrap replicates. The sample from Menorca (Sme7-5_men) was used as an outgroup to root the tree.

Population structure analysis

For this analysis, we excluded the outgroup sequence and ran POPULATIONS to retain loci present in all populations (-p 10) and 75% of the individuals present in each population (-r 0.75). On the basis of the loci that passed our filtering criteria, a random whitelist of 1,000 loci was generated and again run through POPULATIONS with the same criteria but retaining the first SNP at each locus (--write-single-snp). The output was exported in Structure format, and Structure version 2.3.4 (ref. [63]) was used to infer population structure with 10,000 chains as burn-in and 100,000 MCMC chains with 20 iterations for $K = [1–11]$. The resulting files were run through Structure Harvester[64], and the optimal K was determined[65].

DNA FISH

For telomere FISH on oocyte chromosomes, ovaries were dissected as reported previously[66,67]. Hybridization was carried out as reported with chromosome spreads[42] except that the ovaries were kept in suspension in washing buffer or hybridization buffer. Before hybridization, the dissected ovaries were treated with a digestion buffer (0.1% SDS and 10 μg ml^{-1} proteinase K (Qiagen) in 0.3% Triton X-100 in PBS) for 10 min at room temperature. The repeat located near centromeres in Fig. 2c had the following sequence: TCTGGACGGAAATTTTTTAATCTTTATAGGCTTGTATCTCTGTCAATTTTTATTTGTTTTCATAATCTTTGATATATTTCTCGATAACTTTTGATTCTCTACATGATAGCATTTTAAAAATTGCAAAAATCATAACGGGCTCGTCAAACACAAGTCAT.

Hybridization chain reaction and RNA FISH

To examine tissue expression of the genes *glp* and *smedwi-1*, we used probes and buffers for third-generation ISH chain reaction purchased from Molecular Instruments. Sexually mature planarians were treated with 7.5% *N*-acetyl-L-cysteine (Sigma-Aldrich) for 10 min and were then fixed in 4% paraformaldehyde (Electron Microscopy Sciences; 16% solution diluted 1:4 in PBS) for 20 min. The copulatory apparatus was dissected into a 1.5-ml RNase-free tube. The rest of the procedures followed the hybridization chain reaction RNA FISH protocol of Molecular Instruments for whole-mount mouse embryos.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The authors confirm that all data underlying the findings are fully available without restriction. All sequencing data are available from the NCBI SRA database (accession number PRJNA731187). The chromosome-scale genome assemblies for sexual *S. mediterranea* (including phased genomes and genomes with repetitive elements) are openly available on the Planosphere (https://planosphere.stowers.org/), PlanMine (https://planmine.mpibpc.mpg.de/planmine/begin.do), GenBank (GCA_022537955.1) and Zenodo (https://doi.org/10.5281/

zenodo.5807415) databases. We used publicly available NCBI PacBio sequencing data (accession numbers SRX2700681–SRX2700684) and planarian transcriptome data (accession numbers SRR2658118–SRR2658125, SRR2658134–SRR2658141, SRR3473955–SRR3473957, SRR3629945–SRR3629952, SRR6351185–SRR6351188, SRR6351201–SRR6351204, SRR6351213–SRR6351216, SRR6363910–SRR6363927 and SRR6364586–SRR6364588).

Code availability

Custom code for data analysis can be openly accessed on the Zenodo database at https://doi.org/10.5281/zenodo.5807415.

43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
44. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
45. Cosentino, S. & Iwasaki, W. SonicParanoid: fast, accurate and easy orthology inference. *Bioinformatics* **35**, 149–151 (2019).
46. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
47. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3* **1**, 171–182 (2011).
48. Danecek, P. et al. The variant call format and VCF tools. *Bioinformatics* **27**, 2156–2158 (2011).
49. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
50. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
51. Xie, W. et al. Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proc. Natl Acad. Sci. USA* **107**, 10578–10583 (2010).
52. Davies, E. L. et al. Embryonic origin of adult stem cells required for tissue homeostasis and regeneration. *eLife* **6**, e21052 (2017).
53. Zeng, A. et al. Prospectively isolated tetraspanin$^+$ neoblasts are adult pluripotent stem cells underlying planaria regeneration. *Cell* **173**, 1593–1608 (2018).
54. Duncan, E. M., Chitsazan, A. D., Seidel, C. W. & Sanchez Alvarado, A. Set1 and MLL1/2 target distinct sets of functionally different genomic loci in vivo. *Cell Rep.* **13**, 2741–2755 (2015).
55. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
56. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
57. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
58. Etymologia: Bonferroni correction. *Emerg. Infect. Dis.* **21**, 289 (2015).
59. Paris, J. R., Stevens, J. R. & Catchen, J. M. Lost in parameter space: a road map for STACKS. *Methods Ecol. Evol.* **8**, 1360–1373 (2017).
60. Rochette, N. C. & Catchen, J. M. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* **12**, 2640–2659 (2017).
61. Stobie, C. S., Oosthuizen, C. J., Cunningham, M. J. & Bloomer, P. Exploring the phylogeography of a hexaploid freshwater fish by RAD sequencing. *Ecol. Evol.* **8**, 2326–2342 (2018).
62. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
63. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
64. Earl, D. A. & vonHoldt, B. M. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).
65. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
66. Guo, L. et al. Subcellular analyses of planarian meiosis implicates a novel, double-membraned vesiculation process in nuclear envelope breakdown. Preprint at *bioRxiv* https://doi.org/10.1101/620609 (2019).
67. Guo, F. et al. Planarian ovary dissection for ultrastructural analysis and antibody staining. *J. Vis. Exp.* **175**, e62713 (2021).
68. Ouellette, L. A., Reid, R. W., Blanchard, S. G. & Brouwer, C. R. LinkageMapView—rendering high-resolution linkage and QTL maps. *Bioinformatics* **34**, 306–307 (2018).

Acknowledgements We thank L. Rouhana, G. Bruni and S. Zdraljevic for discussions; Z. Kashif, E. Belfer and B. Pakfar for planarian husbandry; and X. Li and the UCLA Technology Center for Genomics & Bioinformatics for next-generation sequencing. We thank Y. Zeng, E. Noskova and D. Pasha for discussions on demographic history. We acknowledge Life Science Editors for professional editing services. Confocal laser-scanning microscopy was performed at the Advanced Light Microscopy/Spectroscopy Laboratory and the Leica Microsystems Center of Excellence at the California Nano Systems Institute at UCLA with funding from NIH Shared Instrumentation grant S10OD025017 and NSF Major Research Instrumentation grant CHE-0722519. This work was supported by grants from the Howard Hughes Medical Institute (L.K.) and the Helen Hay Whitney Foundation (L.G.).

Author contributions L.K. and L.G. conceptualized the study. L.G. coordinated the project, conducted experiments and wrote the original draft. L.G., E.B.-D. and T.V.L. developed methods for single-sperm sequencing. L.G., K.K., K.H., E.W., C.C. and Y.W. collected the data. L.G., J.S.B., D.D.-S., M.R., J.B. and D.L. analysed the data. K.K., K.H., E.W., C.C. and Y.W. maintained planarian colonies. L.K., L.G., O.T.S. and A.S.A. edited the manuscript. A.S.A. provided the D5 and S2 strains, and sequencing data of a dozen oocytes. M.R. provided wild isolate DNA. L.K. acquired funding and supervised the project.

Competing interests The authors declare no competing interests.

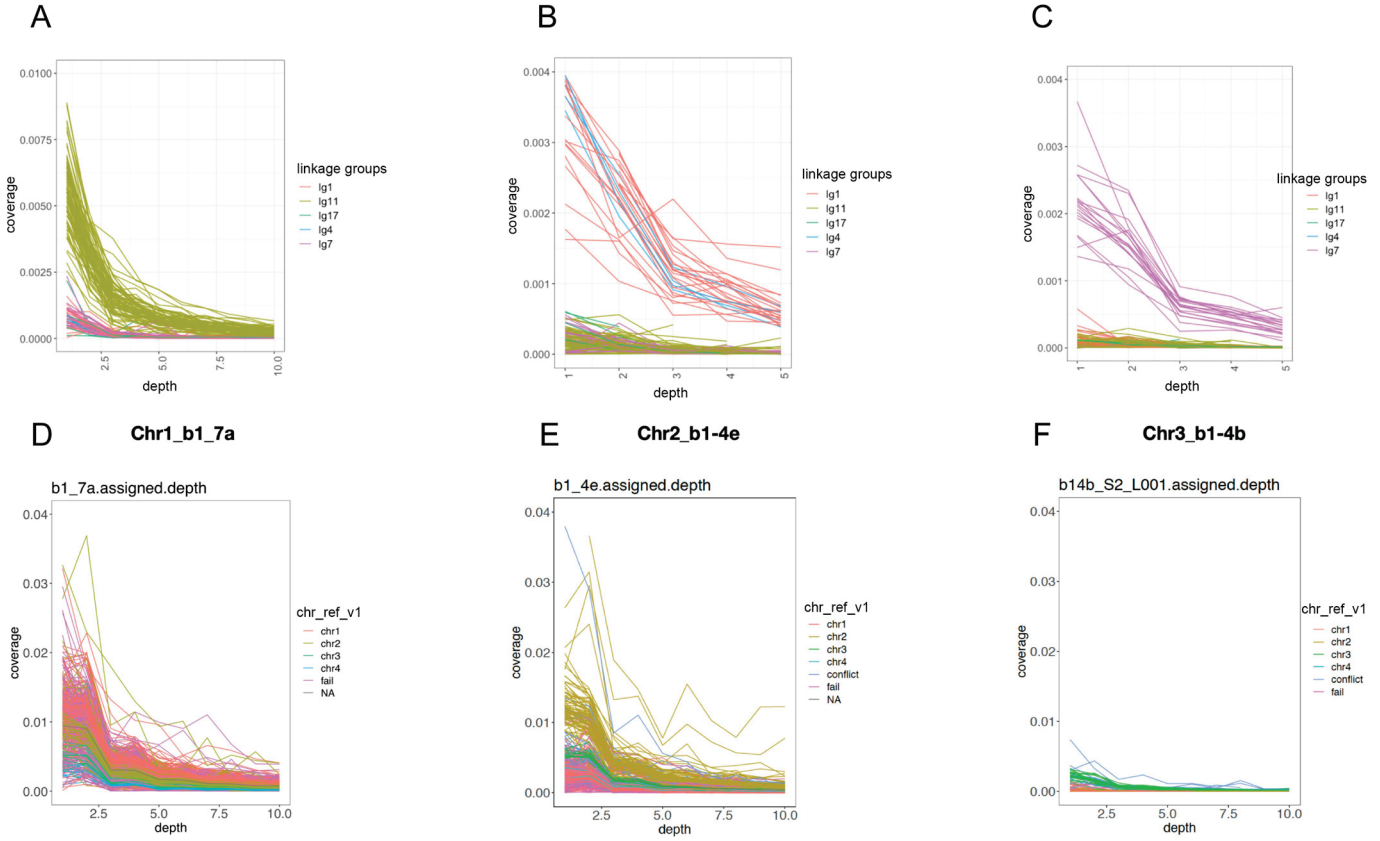
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04757-3.

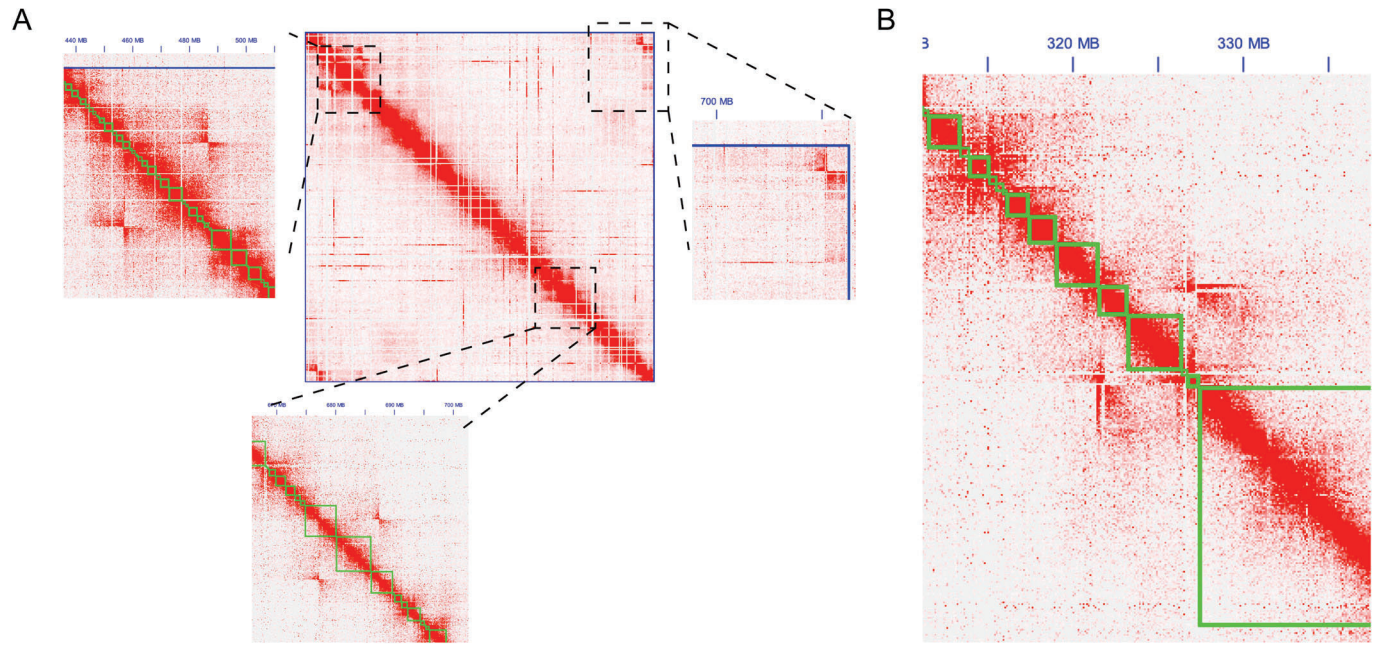
Correspondence and requests for materials should be addressed to Longhua Guo or Leonid Kruglyak.

Peer review information *Nature* thanks Bo Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

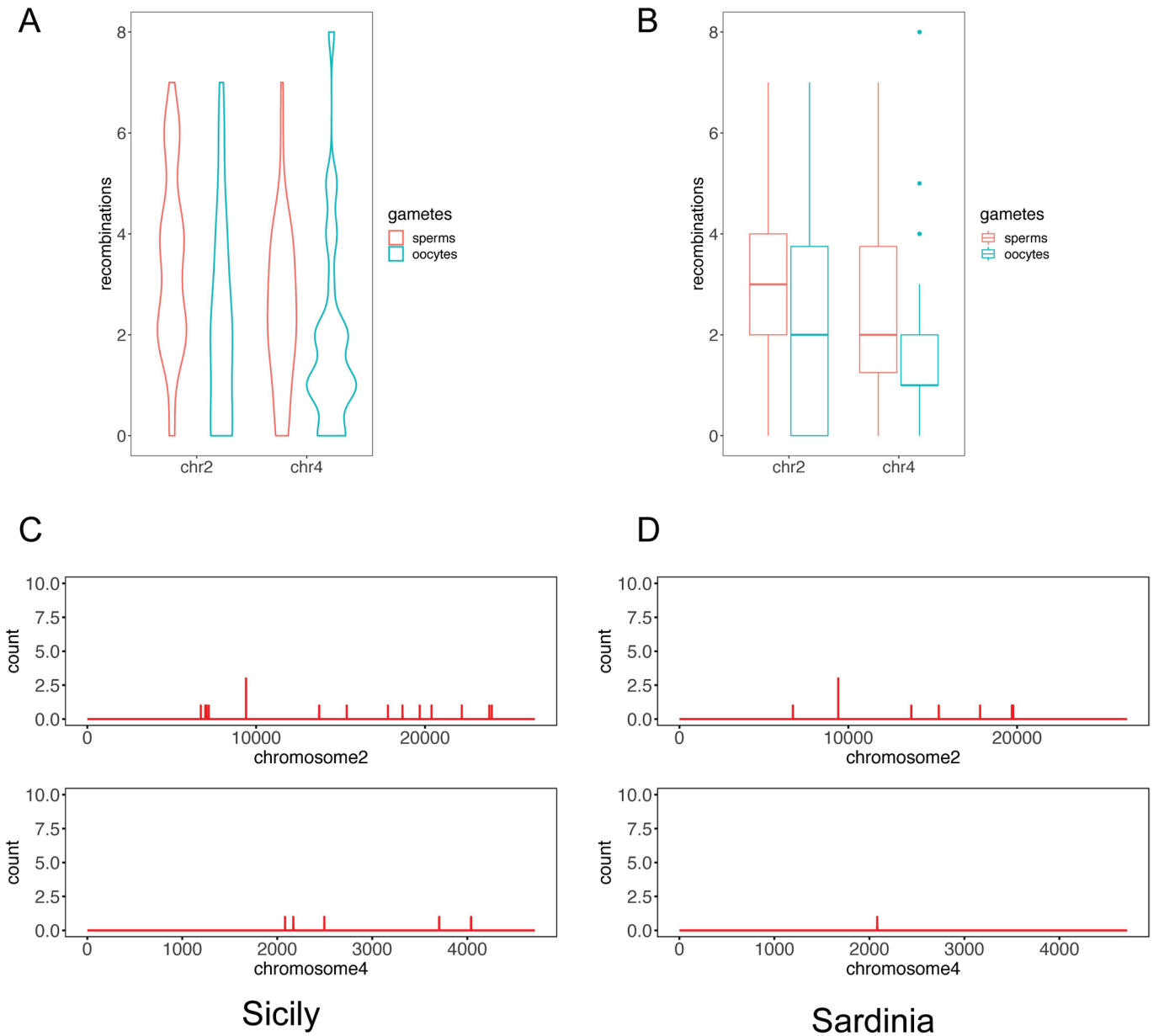
Reprints and permissions information is available at http://www.nature.com/reprints.



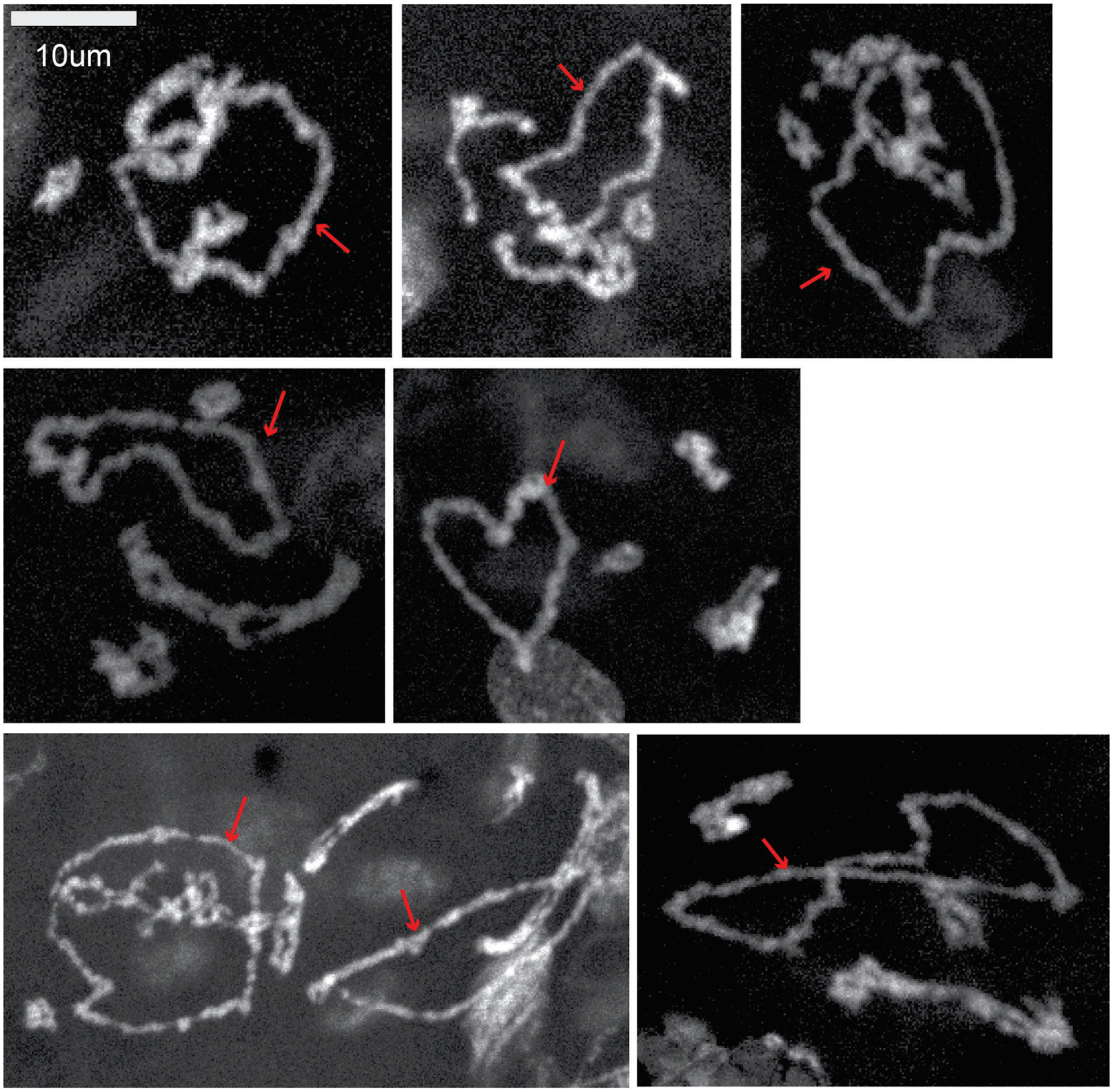
Extended Data Fig. 1 | Chromosome assignment by ChrSeq. (A-C) Chromosome sequencing depth of different scaffolds is consistent with linkage group assignment. Linkage groups: lg11 = chr1, lg1+lg4 = chr2, lg7 = chr3, lg17 = chr4. (D-F) Chromosome sequencing depth of different scaffolds is consistent with chromosome assignment in the genome assembly smed_chr_ref_v1. Chromosome sequencing samples: chromosome 1 (A, D); chromosome 2 (B, E); chromosome 3 (C, F).



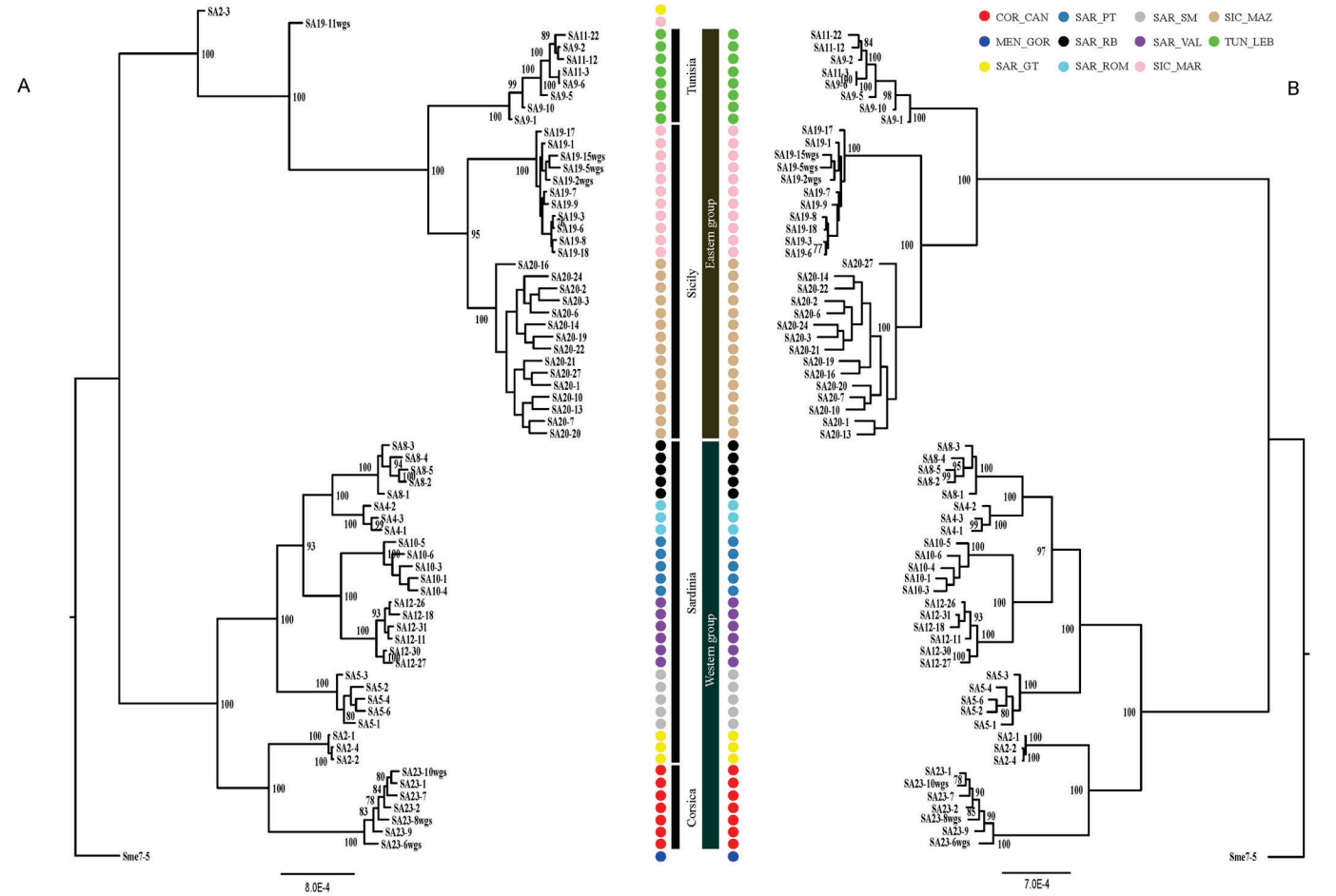
Extended Data Fig. 2 | Putative inversions in the planarian genome detected by chromatin contact heatmap. (A) chromosome 1 chromatin contact heatmap showing 3 potential inversions (dashed-line rectangles) visualized in Juicebox. (B) A complex structural variant on chromosome 2. Green outline: dd_Smes_g4 scaffolds. Blue outline: chromosome boundaries.



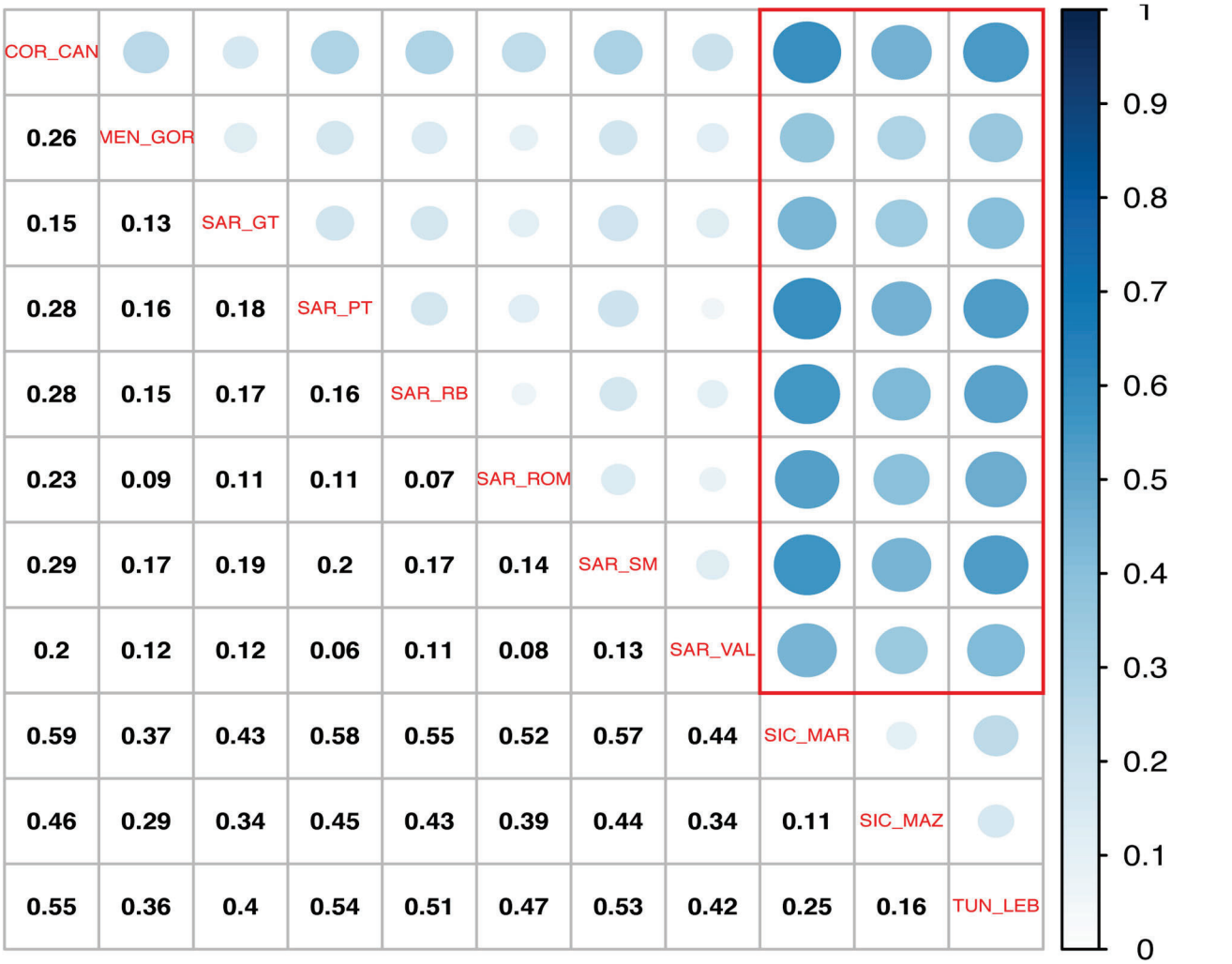
Extended Data Fig. 3 | Recombination in gametes and heterozygosity distribution in the wild on chromosomes 2 and 4. (A-B) Number of recombinations in oocytes and sperm on chromosomes 2 and 4 summarized as a violin plot (A) and a box plot (B). The median number of recombinations on chromosome 2 is 3 for sperm and 2 for oocytes, derived from a total of 49 sperm and 38 oocytes. The median number of recombinations on chromosome 4 is 2 for sperm and 1 for oocytes, derived from a total of 50 sperm and 26 oocytes. Chr2 sperm: min = 0, max = 7, median = 3, first quartile = 2, third quartile = 4. Chr2 oocytes: min = 0, max = 7, median = 2, first quartile = 0, third quartile = 4; Chr4 sperm: min = 0, max = 7, median = 2, first quartile = 1, third quartile = 4; Chr4 oocytes: min = 0, max = 3, median = 1, first quartile = 1, third quartile = 2. (C-D) Very few sites are consistently heterozygous in more than 80% of the wild isolates on chromosomes 2 and 4 in Sicily (C) and in Sardinia (D).



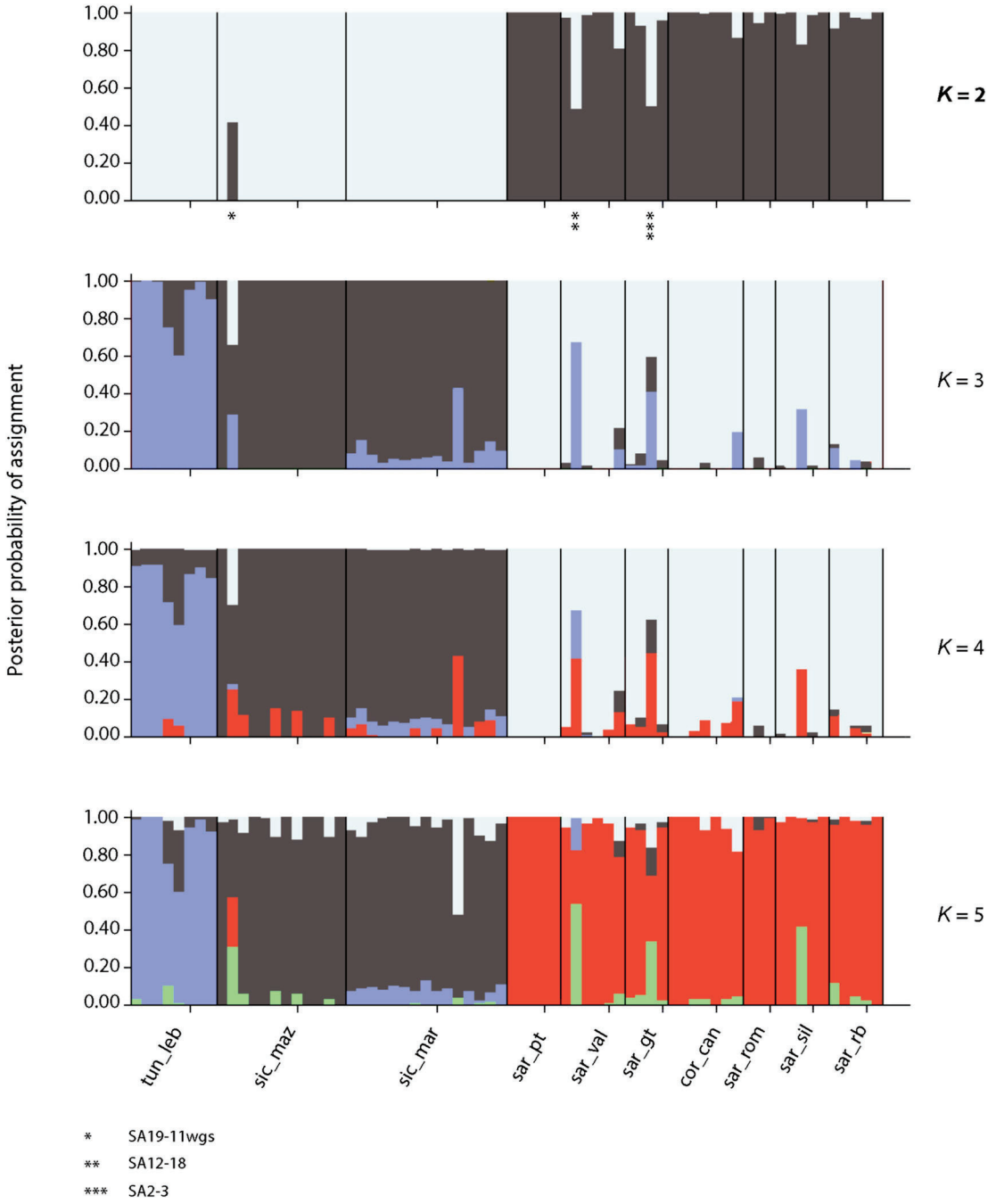
Extended Data Fig. 4 | Representative images of late prophase I oocytes with ring-shaped chromosome 1. More than 100 oocytes from > 40 ovaries and > 20 sexual strains were examined for the presence of a ring-shaped chromosome 1. The experiment was repeated >10 times independently with similar results. Chromosome 1 always forms a ring in J/V strains in late prophaseI oocytes. Shown in the figure are chromosomes from 8 different oocytes atlate prophase I. Arrows point to the ring-shaped chromosome 1.



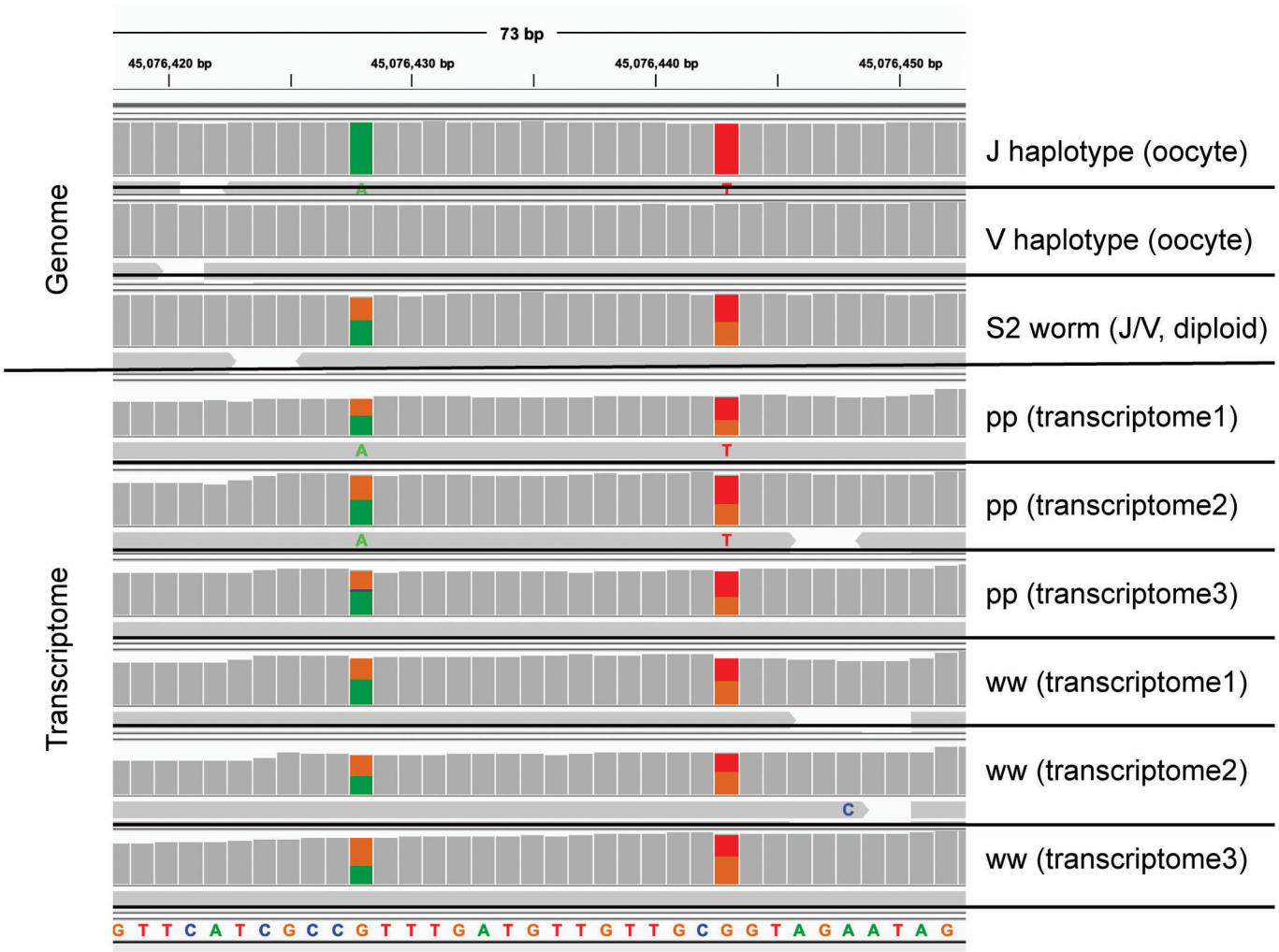
Extended Data Fig. 5 | Phylogenetic tree of wild isolates of *Schmidtea mediterranea*. The tree is rooted with one asexual animal from Menorca (Sme7-5). The tree on the left (A) includes two animals from SAR and SIC with higher genome heterozygosity than the cohorts from the same collection sites. Phylogenetic relations agree with the previous designation of eastern and western populations[1]. SAR: Sardinia; COR: Corsica; SIC: Sicily; TUN: Tunisia.



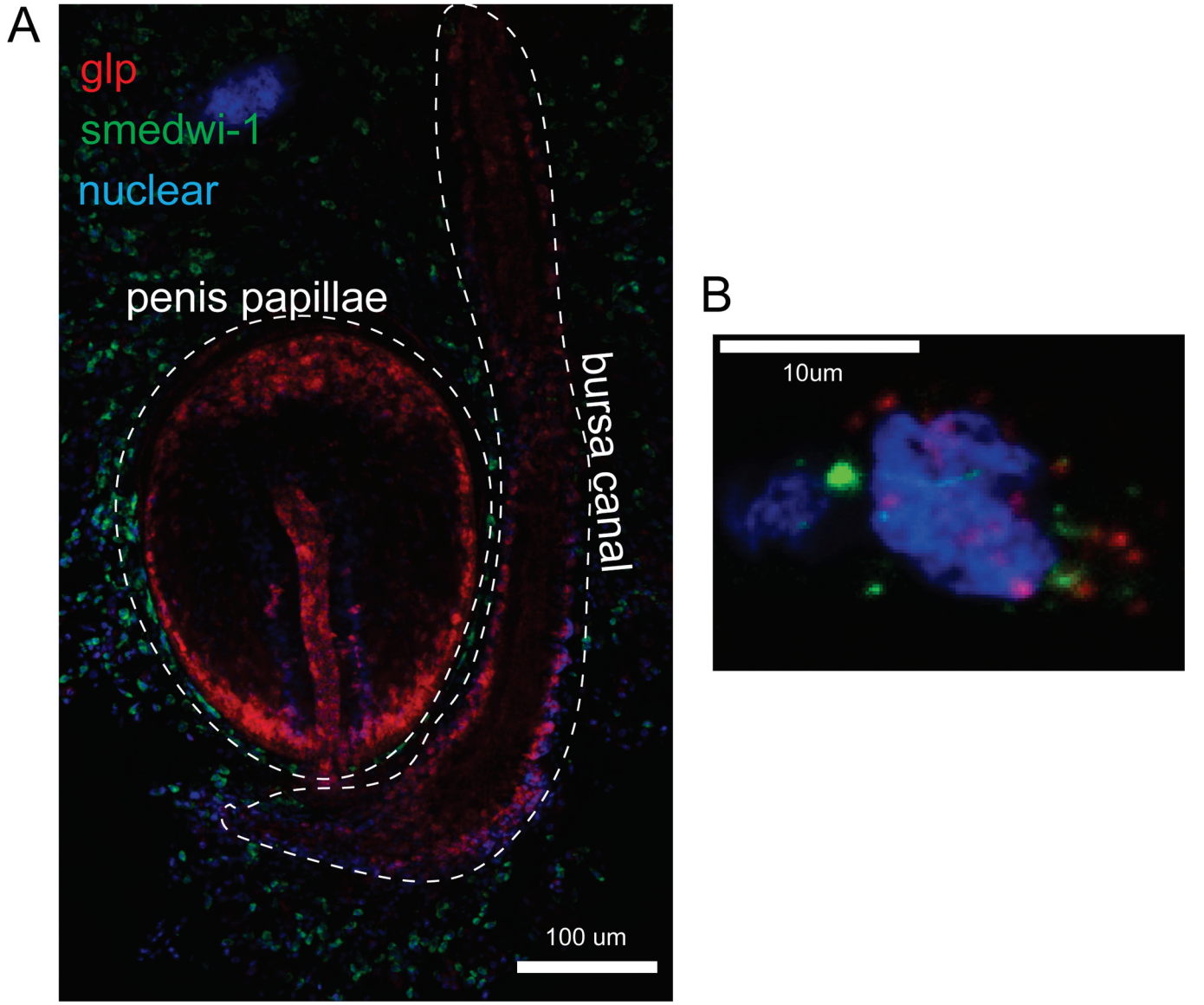
Extended Data Fig. 6 | Fixation index (F_{ST}) analysis of wild isolates. Pairwise population F_{ST} for all collection sites from Sardinia (SAR), Corsica (COR), Menorca (MEN), Sicily (SIC), and Tunisia (TUN). The size and darkness of the blue circles (top right) correspond to the values of F_{ST} (bottom left).



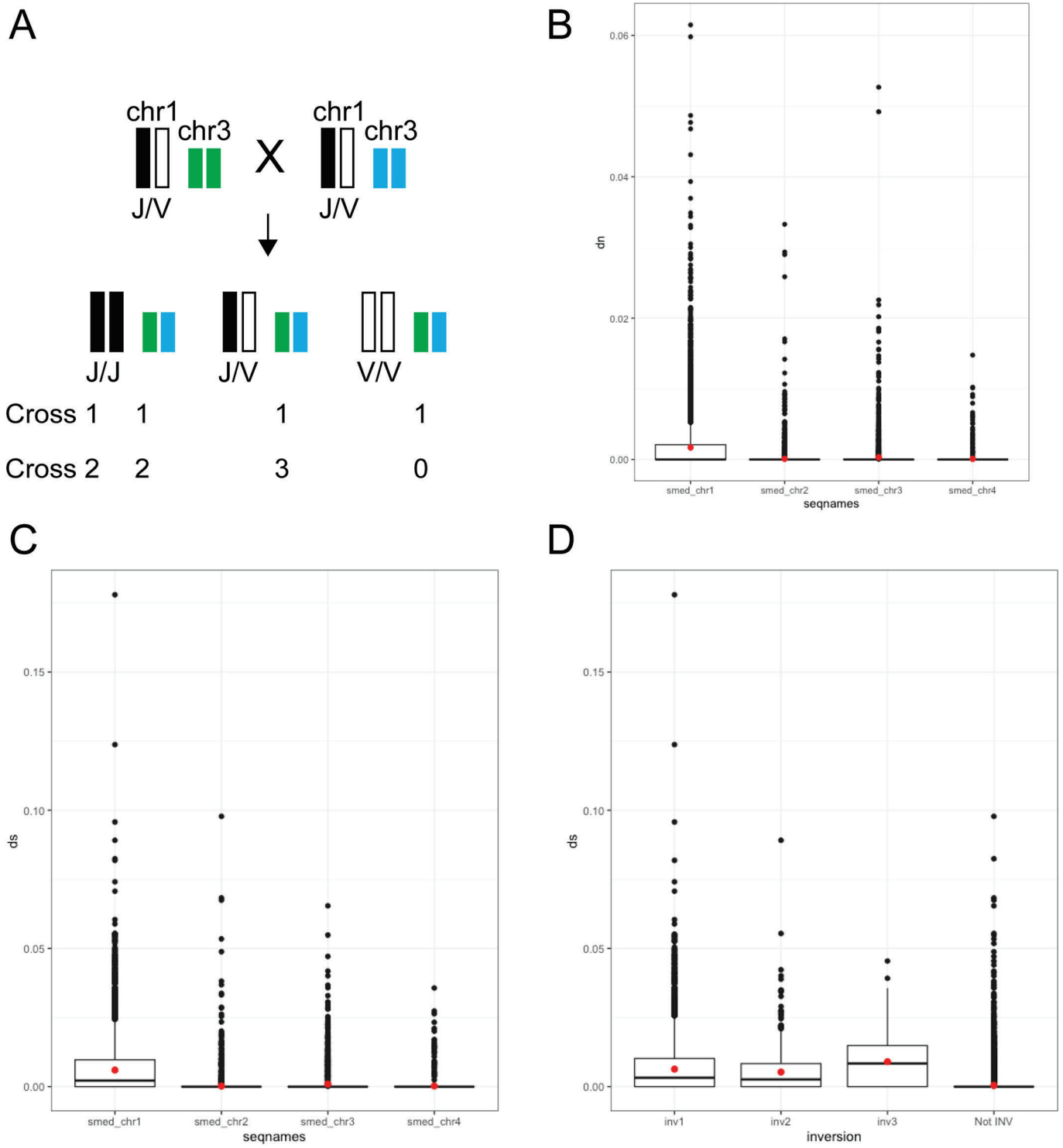
Extended Data Fig. 7 | STRUCTURE analysis of wild isolates. Evanno analysis supports a K = 2 split, dividing the samples geographically into Eastern and Western groups. With K = 3, a Tunisian signal was found to be present in all other 3 groups (light blue).



Extended Data Fig. 8 | Datasets for the examination of haplotype-specific expression. The sequencing reads from genomes (J-haplotype oocytes, V-haplotype oocytes, diploid S2 worms) and transcriptomes (pp, ww from the diploid S2 worms) were aligned to Smed_chr_ref_v1. Shown in the figure are reads for the gene CPEB-2. Two variants with J and V alleles are marked by colored bars. In the genome of the adult worm S2 (i.e., DNA), both J and V alleles are present at equal ratios. In the genome of oocytes (J or V haplotype, i.e., DNA), the variants have only one allele. These observations validate the heterozygous variants. In the transcriptomes (i.e., RNA) of penis papillae (pp) and whole worms (ww) from S2 adults, both J and V alleles are present. In the three transcriptomes of pp, J alleles were more highly expressed. In the three transcriptomes of ww, the J:V allele ratios vary and do not significantly depart from equal expression.



Extended Data Fig. 9 | Specific expression of *smed-glp* in the male reproductive organ (penis papillae) and female reproductive organ (bursa canal). Expression of the *glp* gene (in red color) and the stem cell marker *smedwi-1* (in green color) in the dissected copulatory apparatus was visualized by Fluorescence *in situ* hybridization chain reaction. (A) White dashed lines outline penis papillae and bursa canal. Nuclei were stained by DAPI (in blue color). No *glp* expression in tissues surrounding penis papillae or bursa canal. Less than 1% of *glp*+ cells express the stem cell marker *smedwi-1*. (B) Cells expressing the two markers do exist in both bursa canal and penis papillae. This experiment was repeated 3 times independently with 6 sexually mature worms, producing similar results.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Embryonic lethality and synonymous divergence in *Schmidtea mediterranea*. (A) Two J/V lines with differential SNP markers on chromosome 3 were crossed. Zygotes were collected for genotyping. Heterozygous chromosome 3 markers validate the occurrence of fertilization. All three genotypes of chromosome 1 (J/J, J/V and V/V) were observed in zygotes. (B-C) Non-synonymous divergence (B) and Synonymous divergence (C) for all heterozygous sites in the J/V strain S2 and their distribution in different chromosomes summarized as boxplots. Chromosome 1 has the highest mean (red dots) and median dN or dS values. The min, max, median, mean, first quartile, and third quartile of dN (B) for each chromosome are as following: chr1 = (0, 0.0052, 0, 0.0017, 0, 0.0021), chr2 = (0, 0, 0, 0.000092, 0, 0), chr3 = (0, 0, 0, 0.00031, 0, 0), and chr4 = (0, 0, 0, 0.000088, 0, 0). The min, max, median, mean, first quartile, and third quartile of dS (C) for each chromosome are as following: chr1 = (0, 0.024, 0.0022, 0.006, 0, 0.0098), chr2 = (0, 0, 0, 0.00021, 0, 0), chr3 = (0, 0, 0, 0.00093, 0, 0), and chr4 = (0, 0, 0, 0.00026, 0, 0). (D) Synonymous divergence for all heterozygous sites in the J/V strain S2, and their distribution in the three chromosome 1 inversions and the rest of the genome, summarized as boxplots. Mean (red dots) and median dS decreases in the following order: inversion 3 (inv3), inversion 1 (inv1), inversion 2 (inv2), and the rest of the genome (Not INV). One way ANOVA and Turkey Honest Significant Differences tests showed that dS were significantly different between all 3 inversions (inv2-inv1: p.adj = 0.005; inv3-inv1: p.adj < 1e-06; inv3-inv2: p.adj < 1e-06). The min, max, median, mean, first quartile, and third quartile of dS for each inversion are as following: inversion 1 = (0, 0.026, 0.0034, 0.0064, 0, 0.01), inversion 2 = (0, 0.021, 0.0026, 0.0053, 0, 0.0083), inversion 3 = (0, 0.036, 0.0084, 0.0091, 0, 0.015), and non-inversion = (0, 0, 0, 0.0005, 0, 0). A total of 8317 genes from chromosome 1, 6568 genes from chromosome 2, 3394 genes from chromosome 3, and 1834 genes from chromosome 4 were plotted for non-synonymous divergence (B) and synonymous divergence (C). A total of 6633 genes from inversion 1, 754 genes from inversion 2, 298 genes from inversion 3, and 12428 genes from the rest of the genome were plotted for synonymous divergence (D).

Chapter 1

Building Phylogenies from Transcriptomic Data

Lisandra Benítez-Álvarez, Laia Leria, Daniel Dols-Serrate, and Marta Riutort

Abstract

Transcriptomic data (obtained from RNA sequencing) has become a very powerful source of information to reconstruct the evolutionary relationships among organisms. Although phylogenetic inference using transcriptomes retains the same core steps as when working with few molecular markers (viz., nucleic acid extraction and sequencing, sequence treatment, and tree inference), all of them show significant differences. First, the needed quantity and quality of the extracted RNA has to be very high. Although this may not represent a challenge when working with certain organisms, it may well be a headache with others, especially for those with small body sizes. Second, the tremendous increase in the quantity of sequences obtained requires a high computational power for both treating the sequences and inferring the subsequent phylogenies. This means that transcriptomic data can no longer be analyzed using personal computers nor local programs with a graphical interface. This, in turn, implies the requirement of an increased set of bioinformatic skills from the researchers. Finally, the genomic peculiarities of each group of organisms, such as the level of heterozygosity or the percentage of base composition, also need to be considered when inferring phylogenies using transcriptomic data.

Key words Freshwater planarians, Sampling, Transcriptome, Phylotranscriptomics, Molecular phylogeny

1 Introduction

The present procedure is designed to infer phylogenies from transcriptomic data in planarians (Fig. 1). We present the basic methodologies and some advice regarding sampling, storage, RNA extraction, and sequencing. Additionally, we describe with full detail the bioinformatic workflow, including in-house developed scripts. Although we explain how to collect field samples and fix them to keep the RNA in a good state until its posterior extraction, the transcriptome sequencing and all the bioinformatic procedures can equally be applied to any transcriptome obtained from laboratory strains or simply downloaded. We give some details on the best libraries to be used to sequence the transcriptomes, recommending

Luca Gentile (ed.), *Schmidtea Mediterranea: Methods and Protocols*, Methods in Molecular Biology, vol. 2680,
https://doi.org/10.1007/978-1-0716-3275-8_1,
© The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

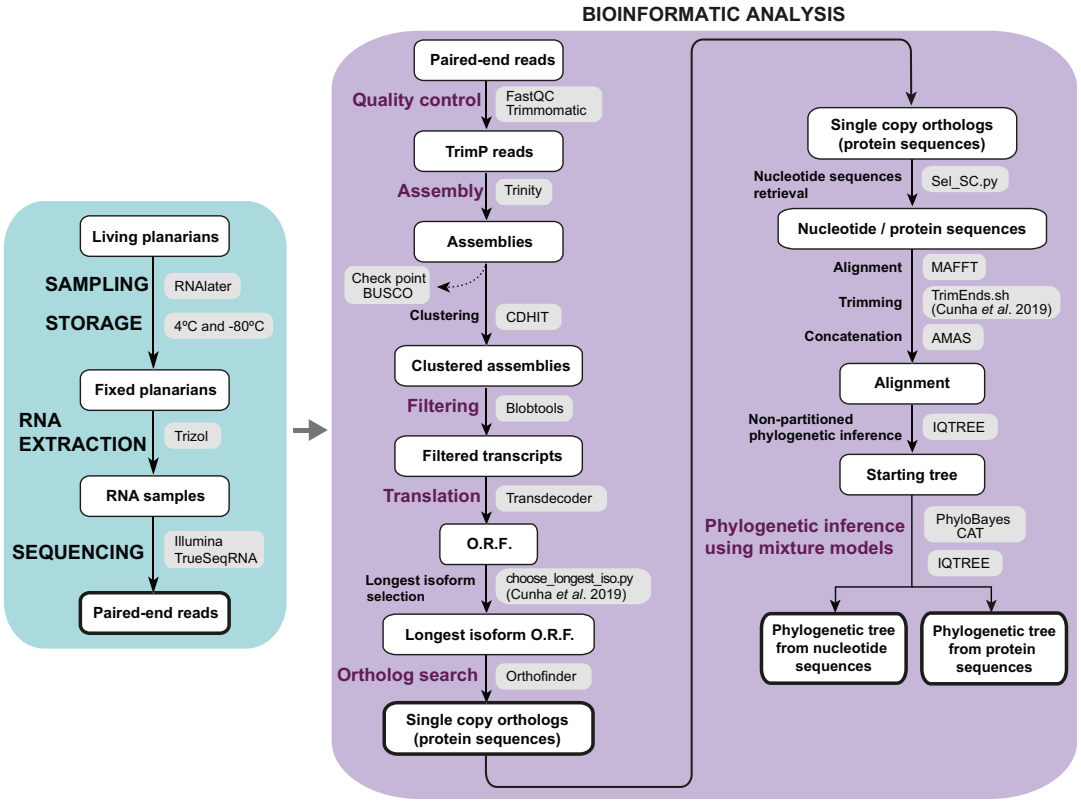


Fig. 1 Workflow of the main procedures: sampling, storage, RNA extraction, sequencing (green panel), and bioinformatic analysis (lilac panel). In gray bubbles are highlighted the main reagent, condition, technology, or software necessary to perform each step indicated by the arrows. In the lilac panel are highlighted the main steps of the bioinformatic workflow: (1) quality control, (2) assembly, (3) filtering, (4) translation, (5) orthologue search, and (6) phylogenetic inference

paired-end reads with a library that guarantees the elimination of ribosomal RNAs. In our case the libraries were obtained and sequenced in a specialized company, but they can be built and sequenced in-house if the adequate laboratory and skills are available. To run the bioinformatic protocol it is necessary to have access to powerful computational resources as well as basic skills to work with the terminal in Unix systems and shell script programming. This protocol has six main steps (Fig. 1): (1) screening the quality of the reads and filtering using FastQC [1] and Trimmomatic [2], (2) de novo assembly of transcriptomes with Trinity [3, 4], (3) filtering of contaminants based on BlobTools results [5, 6], (4) translation to protein using TransDecoder [7], (5) orthologue search with OrthoFinder [8], and (6) phylogenetic inference with maximum likelihood and Bayesian inference methodologies using the mixture models implemented in IQ-TREE [9] and PhyloBayes [10], respectively.

This step-by-step procedure enables the obtention of phylogenetic trees from transcriptomic data, without requiring expertise in bioinformatics, only basic skills are necessary.

Phylogenetic tree inference methodologies constitute a growing and complex field, in the present chapter we only provide a "black box" strategy for obtaining correctly inferred trees. Nonetheless, we refer the readers to specific literature [11–16] to understand the procedures and correctly interpret the results. We only use concatenated dataset analyses. Other types of analyses, such as coalescence based on each gene, are possible and may be more adequate in certain circumstances that we do not discuss here.

2 Materials

2.1 Sampling

1. Plastic tray.
2. Watercolor brushes.
3. Ice blocks and portable freezer.
4. Plastic screw cap tubes (2 mL, 15 mL, and 50 mL).
5. Permanent markers.
6. Petri dishes.
7. Tweezers.
8. Lamp.
9. Tissues.
10. Gloves.
11. Pasteur pipettes.
12. Automatic pipette.
13. Pipette tips (1 mL).
14. Sterile razor or scalpel blades.
15. Absolute ethanol (molecular biology grade).
16. RNase away™ surface decontaminant.
17. RNAlater stabilization solution.

2.2 Storage

Refrigerator (4 °C), freezers (−20 °C and −80 °C).

2.3 RNA Extraction

1. RNase away™ surface decontaminant.
2. Trizol (or any guanidine thiocyanate-based extraction method).
3. Chloroform (molecular biology grade).
4. Isopropanol (molecular biology grade).
5. Ethanol 70% (molecular biology grade).
6. Diethyl pyrocarbonate (DEPC)-treated water.

7. Tissue homogenizer.
8. Refrigerated centrifuge.
9. Pipette tips with and without filter.
10. 1.5 mL tubes with safe lock.
11. Rack (1.5 mL tubes).
12. Box with ice.

2.4 Library Construction and Sequencing

Library preparation can be performed in an external company or in the lab using commercial kits. Illumina sequencing services are offered in multiple companies or may be available at research center services.

2.5 Bioinformatic Analysis

Basic skills to work in a terminal in Unix systems and shell script programming are required to run the bioinformatic protocol.

Hardware

Access to a powerful computational cluster running under Linux systems.

≥24 cores

>126 GB RAM memory

>1 TB available disk space

It is recommended to use a queuing system for submitting jobs.

Software

1. Conda [17] available at https://anaconda.org/, https://docs.conda.io/projects/conda/en/latest/user-guide/install/, https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#activating-an-environment.
2. FastQC [1] available at https://github.com/s-andrews/FastQC, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Help/.
3. MultiQC [18] available at https://multiqc.info/docs/#running-multiqc.
4. Trimmomatic [2] available at https://github.com/timflutre/trimmomatic.
5. Trinity [3, 4] available at https://github.com/trinityrnaseq/trinityrnaseq/releases, https://github.com/trinityrnaseq/trinityrnaseq/wiki.
6. assembly-stats [19] available at https://github.com/rjchallis/assembly-stats.
7. CDHIT [20, 21] available at https://github.com/weizhongli/cdhit, http://weizhong-lab.ucsd.edu/cd-hit/.

8. bwa [22] available at https://github.com/lh3/bwa.
9. Samtools [23, 24] available at https://www.htslib.org, https://github.com/samtools/samtools.
10. BLAST+ [25] available at https://blast.ncbi.nlm.nih.gov/Blast.cgi.
11. BlobTools [5, 6] available at https://github.com/blobtoolkit/blobtools2/tree/2.6.1, https://github.com/DRL/blobtools.
12. Seqtk [26] available at https://github.com/lh3/seqtk.
13. TransDecoder [7] available at https://github.com/TransDecoder.
14. choose_longest_iso.py [27, 28] available at https://github.com/tauanajc/Cunha_Giribet_2019_ProcRSocB/blob/master/choose_longest_iso.py.
15. OrthoFinder [8] available at https://github.com/davidemms/OrthoFinder.
16. capt_SCseq_by_sp.py [29] available at https://github.com/lisy87/dugesia-transcriptome/tree/main/scripts.
17. interleave_to_oneline.pl [29] available at https://github.com/lisy87/dugesia-transcriptome/tree/main/scripts.
18. sel_sc.py [29] available at https://github.com/lisy87/dugesia-transcriptome/tree/main/scripts.
19. MAFFT [30] available at https://mafft.cbrc.jp/alignment/software/linux.html.
20. trimEnds.sh [28] available at https://github.com/tauanajc/Cunha_Giribet_2019_ProcRSocB/blob/master/trimEnds.sh.
21. trimAL [31] available at http://trimal.cgenomics.org/, https://github.com/scapella/trimal, https://vicfero.github.io/trimal/.
22. AMAS.py [32] available at https://github.com/marekborowiec/AMAS.
23. IQ-TREE [9] available at http://www.iqtree.org/.
24. PhyloBayes [10] available at http://www.atgc-montpellier.fr/phylobayes/.
25. newick_utils [33] available at https://github.com/tjunier/newick_utils/.
26. FigTree available at https://github.com/rambaut/figtree/releases.
27. Optional: Transcriptome phylogeny workflow available at https://github.com/lisy87/dugesia-transcriptome.

3 Methods

3.1 Sampling (See Note 1)

To manipulate animals in the field:

1. Use RNase away to clean the space and set a portable laboratory in the plastic tray.
2. Prepare a small Petri dish to do the absolute ethanol fixation and another to do the RNAlater fixation.
3. Have a dedicated brush for each step (manipulating the animals in their water, ethanol fixation, and fixing in RNAlater, have them clearly identified) to avoid killing animals accidentally.
4. Select the animals to be fixed in the Petri dishes with river water.
5. Cut the selected animal's tail.
6. Transfer the small tail piece to the Petri with ethanol and the rest to the Petri containing RNAlater.
7. Leave the two fragments there for a few seconds.
8. Transfer each piece to the tubes containing the corresponding fixation liquid. If the portion of the animal to be fixed in RNAlater is very big, cut it into small pieces. Make sure you correctly identify the tubes with the name of the individual and the fixation liquid.
9. Place the tube in the portable refrigerator with ice or ice packs and try to keep them cold during all the transport (either in car, plane, etc.).

In case you decide to take the living animals to your laboratory:

1. Fill 50 mL conical bottom tubes with around 45 mL river water.
2. Do not overload the tube (around 10 worms/tube).
3. Place the tubes in the portable refrigerator with ice or ice packs and try to keep them cold during all the transport.
4. Open the tubes to aerate the animals at least twice a day.

3.2 Storage

The samples in RNAlater can be stored at room temperature (15–25 °C) for 1 week, 4 weeks at 2–8 °C, and indefinitely at −20 °C or −80 °C. For long-term storage it is recommendable at −80 °C.

Caution: do not freeze samples in RNAlater solution immediately. Store the samples at 4 °C or room temperature at least one night; the solution must penetrate the tissue before freezing.

Extract with pipette all the solution from the tube before freezing at −80 °C, at this temperature the solution crystallizes, and the crystals may break the tissue.

3.3 RNA Extraction (See Note 2)

Protocol of RNA extraction with Trizol following the manufacturer's instructions:

1. Place the fixed planaria in a 1.5 mL tube.
2. Add 500 μL of Trizol.
3. Use a tissue homogenizer to break the tissues.
4. Centrifuge 10 min at 12,000 × *g*. Use a refrigerated centrifuge at 4 °C.
5. Transfer the upper phase to a new 1.5 mL tube.
6. Add 100 μL of chloroform.
7. Invert the tube 10–15 times.
8. Incubate 3 min on ice.
9. Centrifuge 15 min at 12,000 × *g*. Use a refrigerated centrifuge at 4 °C.
10. Transfer the upper phase to a new tube.
11. Add 250 μL of isopropanol.
12. Invert the tube 10–15 times.
13. Incubate 10 min on ice.
14. Centrifuge 10 min at maximum speed. Use a refrigerated centrifuge at 4 °C.
15. Remove supernatant.
16. Add 500 μL of ethanol at 75%.
17. Centrifuge 5 min at 7,500 × *g*. Use a refrigerated centrifuge at 4 °C.
18. Remove supernatant.
19. Dry pellet at room temperature (not above 25 °C).
20. Add 20 μL of DEPC water preheated at 65 °C.
21. Dilute the pellet without pipetting; just tap gently with your fingers on the tube.
22. For long-term storage keep at −80 °C. But try to avoid freezing and thawing multiple times for that will easily degrade the RNA.

3.4 Library Construction and Sequencing

The bioinformatic workflow presented here is designed to analyze Illumina paired-end reads obtained from a Truseq, stranded, and ribo-zero library. We recommend using this sequencing strategy for the transcriptome obtention (*see* **Note 3**). Nonetheless, library preparation methodologies and kits evolve rapidly and researchers may consider at each experiment what are their best options when preparing their RNA to obtain a transcriptome.

Follow the indications of the sequencing service to process the samples for library preparation and sequencing.

3.5 Bioinformatic Analysis

The following steps are sufficient to execute the workflow. However, if you wish to expand your understanding of the workflow you can find a set of complementary scripts to run available at https://github.com/lisy87/dugesia-transcriptome/tree/main/scripts.

In addition, we provide the transcriptomic raw data of six *Dugesia* species that have been used to test and ensure the proper functioning of the bioinformatic workflow. They are available at NCBI-SRA's repository with the following accession codes: *D. subtentaculata* (SRR17642690), *D. vilafarrei* (SRR17642676), *D. aurea* (SRR17642743), *D. corbata* (SRR17642750), *D. benazzii* (SRR17642753), and *D. hepta* (SRR17642736). At different steps we show the results obtained for this dataset. You can download it and follow the bioinformatic workflow to check everything is working fine for you.

3.5.1 Install Conda

Install conda in the cluster. Consult conda's documentation (*see* **Note 4**).

3.5.2 Download the Data

Download the transcriptomic data.

The raw data format is *fastq.gz. If you have paired-end data, you will have two files per sample: e.g., samplename_1.fastq.gz and samplename_2.fastq.gz (*see* **Note 5**).

Save all the *fastq.gz files in the same folder.

Use the *wget* command to download it directly from a server:

```
wget http://path_to_raw_data
```

Or, if you have the raw data in a local machine, you can use the *scp* command to copy it to the cluster:

```
scp -rp PATH_from_your_machine/ user@ip.cluster:~/PATH_to_the_cluster
```

3.5.3 Quality Assessment of the Raw Data

1. Install FastQC and MultiQC with conda.
2. Check the quality of the data with FastQC using the default parameters (*see* **Note 6**):

   ```
   fastqc samplename_1.fastq.gz
   ```

3. Use MultiQC to generate a single report for all samples (Fig. 2) (*see* **Note 7**). In a folder with all outputs of FastQC run:

   ```
   multiqc .
   ```

4. Check the results for all samples. It is important that all samples show homogeneous and acceptable values for the following parameters:

Fig. 2 Sequence quality histograms (quality score by position) from FastQC report of 6 *Dugesia* species summarized with MultiQC

- Sequence quality >15.
- GC content homogeneous among samples and similar to the reported for the studied taxon.
- N content per base <5.
- Correct sequence length (it must be mostly the length defined by the library).
- Sequence duplication levels <20%.
- Overrepresented sequences <0.1%.
- Adapter content (it is not a problem; you will remove them in the next step).

Consult the FastQC help page for information about the different modules.

3.5.4 Adapter Removing and Filtering (See ***Note 8***)

1. Install Trimmomatic with conda.
2. Remove adapters and filter the reads using Trimmomatic. For paired-end Illumina reads, use the following options:

```
trimmomatic PE -threads 8 -phred33 -trimlog samplename_trimmo.log samplename_1.fastq.gz
samplename_2.fastq.gz  samplename_1.trimP.fq.gz samplename_1.trimUP.fq.gz
samplename_2.trimP.fq.gz samplename_2.trimUP.fq.gz ILLUMINACLIP:TruSeq3-PE-
2.fa:2:30:10:1:keepBothReads LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36
```

If new adapter sequences or Unique Dual indexes are provided by the sequencing service, add them to the TruSeq3-PE-2.fa file.

3. Retain the *trimP.fq.gz files for downstream analysis.

Optional: Run FastQC again to corroborate the improvement in quality of your data.

*3.5.5 Assembly with Trinity (*See ***Note 9****)*

Follow the steps carefully to obtain the filtered transcripts:

1. Install Trinity with conda in a new environment with python ≥3.6.
2. Use Trinity to assemble the transcriptome of each sample independently. Use default options.

```
Trinity --seqType fq --max_memory ">=100GB" --left samplename_1.trimP.fq.gz --right
samplename_2.trimP.fq.gz --CPU ">=24" --output samplename.trinity_out
```

Only change these options if you are sure about its functionality and how it can affect your analysis. If that is the case, consult Trinity's documentation for a detailed explanation on the parameters.

3. Retain the Trinity.fasta and Trinity.gene_trans_map files for all samples from the output folders (e.g. samplename.trinity_out).
4. Change the names in a way that you can easily identify all samples, for example, samplename.trinity.fasta and samplename.gene_trans_map. You have to correctly identify every assembly and its corresponding gene_trans_map.
5. Install assembly-stats with conda.
6. Look at the stats of every assembly with assembly-stats or use the script TrinityStats.pl from the Trinity package to obtain the stats by gene and isoforms.

```
assembly-stats -t -u samplename.trinity.fasta > assemblies_stats_samplename.tsv
PATH_to/trinityrnaseq-v2.9.1/util/TrinityStats.pl samplename.trinity.fasta >>
assemblies_stats_samplename.tsv
```

7. Check the completeness of the assemblies with BUSCO. Use the Metazoa database (ODB10) to predict the orthologous genes in planarian transcriptomes. Install the latest version of BUSCO with conda. Be sure to install all dependencies, and run the following command:

```
busco -i samplename.trinity.fasta -l metazoa_odb10 -o busco.out -m transcriptome -c ">=24"
```

3.5.6 Transcript Clustering

Use CDHIT-EST to cluster each transcriptome using a sequence identity threshold of 0.99.

1. Install CDHIT-EST with conda and run the following command:

```
cd-hit-est -i samplename.trinity.fasta -o samplename.trinity.cdhit.fasta -c 0.99 -G 1 -M
“>=120000” -T “>=24” -d 0 -g 1
```

3.5.7 Map Reads Against Assembly

Mapping the reads against the assembly is a way to explore the quality of the assembly and one of the steps to perform the BlobTools analysis.

1. Install bwa and samtools with conda in a new environment with python ≥3.6.
2. Map the forward and reverse reads of every sample against its corresponding assembly. Use the following commands:

```
bwa index samplename.trinity.cdhit.fasta
bwa mem -t “>=24” -o samplename_Trinity.cdhit.sam samplename.trinity.cdhit.fasta
samplename_1.trimP.fq.gz samplename_2.trimP.fq.gz
samtools view -S samplename_Trinity.cdhit.sam -bo samplename_Trinity.cdhit.bam
samtools sort -o samplename_Trinity.cdhit.sort.bam samplename_Trinity.cdhit.bam
samtools index samplename_Trinity.cdhit.sort.bam
```

The intermediate files samplename_Trinity.cdhit.sam and samplename_Trinity.cdhit.bam can be deleted if necessary, to save space.

3.5.8 Map Transcripts Against Nucleotide Database

1. Download the update_blastdb.pl script from ncbi page.
2. Use it to download the nucleotide database:

```
wget
https://www.ncbi.nlm.nih.gov/IEB/ToolBox/CPP_DOC/lxr/source/src/app/blast/update_blastd
b.pl
chmod u+x update_blastdb.pl
update_blastdb.pl --showall
update_blastdb.pl --decompress nt
```

3. Install BLAST+ with conda and run the following command:

```
blastn -db PATH_to_nucleotide_database/nt -query samplename.trinity.cdhit.fasta -
num_threads ">=24" -max_target_seqs 10 -max_hsps 1 -outfmt '6 qseqid staxids bitscore
std' -evalue 1e-25 > samplename.blastn.outfmt6
```

3.5.9 Filter Transcripts with BlobTools

1. Install BlobTools. Follow the steps described in the GitHub page of the program.
2. Put the following files in a folder:

```
The assembly: samplename.trinity.cdhit.fasta
The bam file of mapping (reads VS assembly): samplename_Tri-
nity.cdhit.sort.bam
The blastn file (assembly VS nt): samplename.blastn.outfmt6
```

3. Run BlobTools using the following commands:

```
#Creating blobDB
PATH_to_BlobTools_directory/blobtools create -i samplename.trinity.cdhit.fasta -b
samplename_Trinity.cdhit.sort.bam -x bestsumorder -t samplename.blastn.outfmt6 -o
./samplename .blobt.out
#Creating view
PATH_to_BlobTools_directory/blobtools view -i *.blobDB.json  -x bestsumorder
#Creating plot
PATH_to_BlobTools_directory/blobtools plot -i *.blobDB.json -x bestsumorder
```

4. Explore the graphics in the output of BlobTools and check the taxonomic affiliation of the sequences and their coverage (Fig. 3a) as well as the total percentage of mapped reads and the percentage assigned to different taxonomic groups (Fig. 3b). You can summarize these values in a table for all samples. The percentage of no-hit can be high because there is not much information for Platyhelminthes in the NCBI.

3.5.10 Capture Filtered Transcripts

1. Select the names of the transcripts with hit against Platyhelminthes and with no-hit from the samplename.blobt.out.blobDB.table.txt output file.

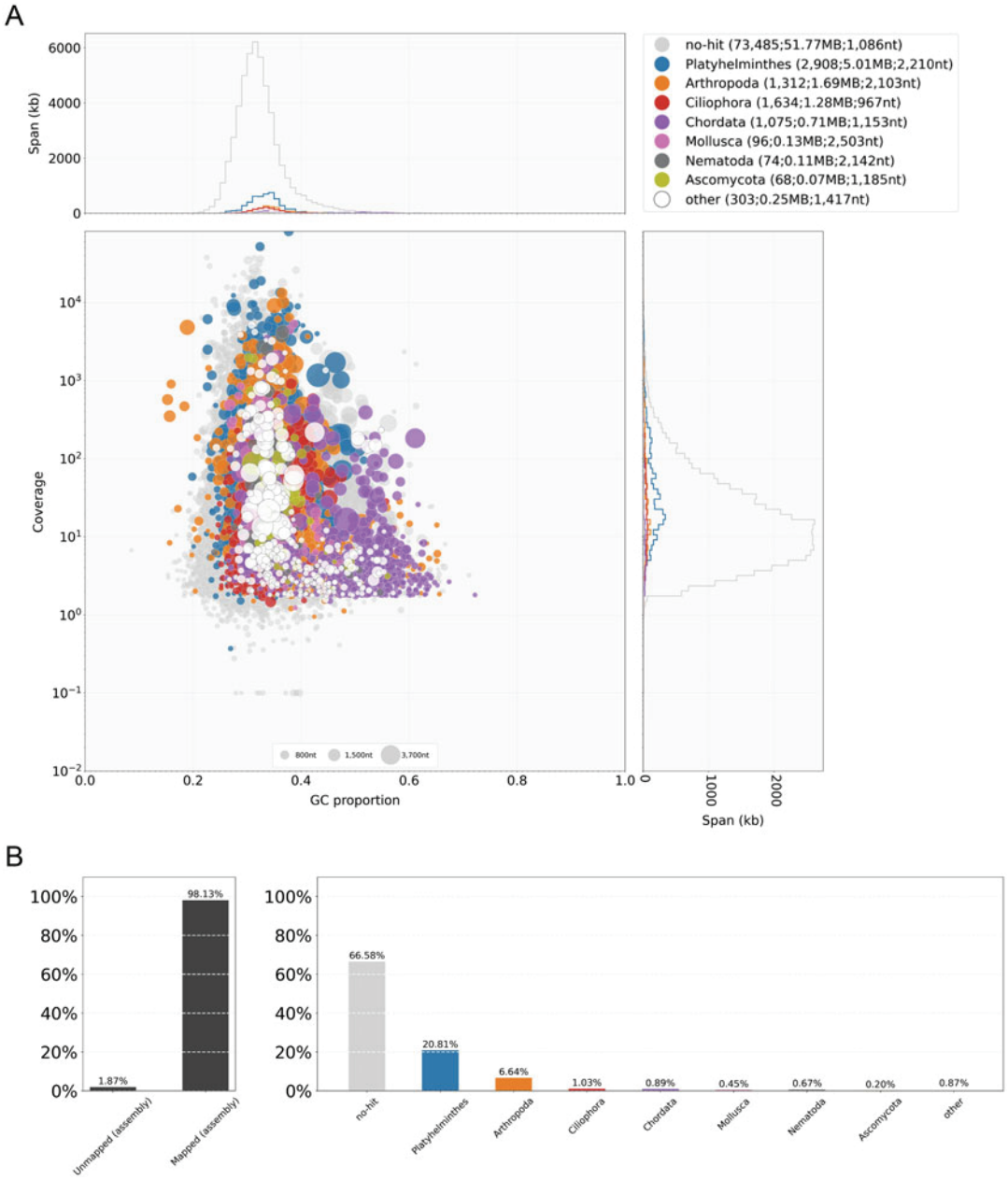


Fig. 3 BlobTools plots for *Dugesia aurea* sample. (**a**) BlobPlot, dimensional scatterplots showing the GC content and coverage. The sequences are represented by circles with diameters proportional to sequence length. The assignment to taxonomic groups is identified with colors. (**b**) ReadCovPlot histogram showing the proportion of unmapped/mapped reads against the assembly and percentage of mapped reads by taxonomic group

```
cat samplename .blobt.out.blobDB.table.txt | grep "Platyhelminthes" | cut -f 1 >
samplename.filtered.txt
cat samplename .blobt.out.blobDB.table.txt | grep "no-hit" | cut -f 1 >>
samplename.filtered.txt
```

2. Install seqtk with conda.
3. Capture the sequences of selected transcripts from the assembly. Check that the names in the samplename.filtered.txt have the same format as in the samplename.trinity.cdhit.fasta and they can match.

```
seqtk subseq samplename.trinity.cdhit.fasta samplename.filtered.txt >
samplename.filt.transc.fasta
```

4. Extract the gene_transc_map for the filtered transcripts.

```
grep -wFf samplename.filtered.txt samplename. gene_transc_map >
samplename.gene_trans_map.filtered
```

3.5.11 Translate Transcripts to Protein

1. Install TransDecoder with conda in a new environment with python ≥3.6.
2. Run:

```
TransDecoder.LongOrfs -t samplename.filt.transc.fasta --gene_trans_map
samplename.gene_trans_map.filtered
TransDecoder.Predict -t samplename.filt.transc.fasta
```

3. Retain the *.pep and *.cds files for downstream analysis.

3.5.12 Select the Longest Isoform

1. Download the choose_longest_iso.py script:

```
wget
https://github.com/tauanajc/Cunha_Giribet_2019_ProcRSocB/blob/master/choose_longest_i
so.py
```

2. Use python2 to run the script as follows:

```
python PATH_to_script/choose_longest_iso.py -l -i= samplename.filt.transc.fasta.pep -o=
sample_A.filt.transc.pep.longest.fasta
```

3.5.13 Orthologue Search

1. Install OrthoFinder following the indications in the web page (*see* **Note 10**).
2. Put all the fasta files with the longest isoforms selected (one by sample) into a folder.
3. Run OrthoFinder just as follows:

```
PATH_to/OrthoFinder/orthofinder -f ./ -t ">=24" -M msa -S diamond_ultra_sens
```

For more information about OrthoFinder's parameters, please consult its web page.

3.5.14 Extract the Protein Sequences of Single-Copy Genes (SC)

Follow the path to Orthofinder/Results_"date"/Single_Copy_Orthologue_Sequences/ and copy the protein sequences of the SC genes directly from this folder.

3.5.15 Extract the Nucleotide Sequences of Single-Copy Genes (SC)

Pay attention to all the steps in this section. Especially to those regarding the format in the names of the files.

1. Capture the name of the sequences that form each SC:

```
grep -Fwf Orthogroups/Orthogroups_SingleCopyOrthologues.txt
Orthogroups/Orthogroups.tsv > SingCopy_ID.txt
```

2. Generate the species/sample list file.

```
ls -1 PATH_to_sequences/ | grep ".fasta" > species_list.txt
sed -i 's/all_that_is_not_the_name_of_the_sample/_seq.txt/' species_list.txt
```

3. Capture the names of single-copy sequences by sample.
   Download the script capt_SCseq_by_sp.py available at https://github.com/lisy87/dugesia-transcriptome/tree/main/scripts
   Run the following script:

```
python3 capt_SCseq_by_sp.py-f=species_list.txt -i= SingCopy_ID.txt
```

4. Copy the *.cds files obtained from TransDecoder's run (**step 2** in Subheading 3.5.11) into a folder.

5. Convert the *.cds files from interleave format to sequential. You can use any tool of your liking for that, or you can use the script interleave_to_oneline.pl available at https://github.com/lisy87/dugesia-transcriptome/tree/main/scripts.
6. Put the *_seq.txt files obtained in **step 3** of this section in the same folder with the *.cds files and the Orthogroups_SingleCopyOrthologues.txt file, obtained from OrthoFinder run (**step 7** in Subheading 3.5.13).
7. Download the sel_sc.py script available at https://github.com/lisy87/dugesia-transcriptome/tree/main/scripts.
8. Run the sel_sc.py script in the folder that contains all these files: *.cds, *_seq.txt, and Orthogroups_SingleCopyOrthologues.txt.

```
python3 sel_sc.py
```

Refer to the help of the script for more information about it:

```
python3 sel_sc.py --help
```

9. Retain the OG*.fasta files for downstream analyses. Every file contains the nucleotide sequences of all the samples for the single-copy gene in fasta format (*see* **Note 11**).

*3.5.16 Alignment (See **Note 12**)*

1. Install MAFFT with conda.
2. Align the single-copy genes independently using MAFFT:

```
mafft --auto --maxiterate 1000 --thread ">=24" OG*.fasta > OG*_alig.fasta
```

3. Convert the OG*_alig.fasta files from interleaved to sequential format using the interleave_to_oneline.pl script.

3.5.17 Trim the Ends

1. Download the script trimEnds.sh.

```
wget https://github.com/tauanajc/Cunha_Giribet_2019_ProcRSocB/blob/master/trimEnds.sh
```

2. Use the script directly to trim the protein sequences:

```
trimEnds.sh OG*_alig.oneline.fasta
```

3. Edit the script to trim the nucleotide sequences; for that change the aa letters by a, c, g, t (lines 31 and 44) in the original script.
4. Run:

```
trimEnds_edit.sh OG*_alig.oneline.fasta
```

*3.5.18 Delete Poorly Aligned Regions (See **Note 13**)*

1. Install trimAL with conda.
2. Run trimAL using the automated1 option:

```
trimal -in OG*_alig.oneline.trimmed.fasta -out OG*_alig.oneline.trimmed.trimal.fasta -automated1
```

3.5.19 Concatenation

1. Install AMAS with conda.
2. Put all the OG*_alig.oneline.trimmed.trimal.fasta files in a folder.
3. Be sure that the names of the sequences are the same in all files. They have to match perfectly. If that is not the case, use this bash command:

```
cat OG*_alig.oneline.trimmed.trimal.fasta | sed 's/_DN/@/' | cut -d "@" -f 1 > OG*_alig.oneline.trimmed.trimal.samename.fasta
```

4. Run AMAS in the folder with all the fasta files:

```
#For protein sequences

AMAS.py concat -i *.fasta -f fasta -d aa -t concat_prot -p concat_prot.part -u fasta -y raxml -c 8

#For nucleotide sequences

AMAS.py concat -i *.fasta -f fasta -d dna -t concat_nuc -p concat_nuc.part -u fasta -y raxml -c 8

# Obtaining the phylip format

AMAS.py convert -i concat_prot.fasta -f fasta -d aa -u phylip -c 8

AMAS.py convert -i concat_nuc.fasta -f fasta -d dna -u phylip -c 8
```

From those command lines a partition file and two concatenate files (fasta and phylip format) are obtained for nucleotide and protein data.

3.5.20 Phylogenetic Inference

1. Exploratory analysis without any partition scheme.

   Install IQ-TREE with conda in a new environment and run:

   ```
   iqtree -s concat_nuc.fasta -m MFP -bb 10000 -nt ">=24"

   iqtree -s concat_prot.fasta -m MFP -bb 10000 -nt ">=24"
   ```

   Use newick_utils to visualize and handle the trees from the terminal (Fig. 4):

   ```
   nw_display *.treefile
   ```

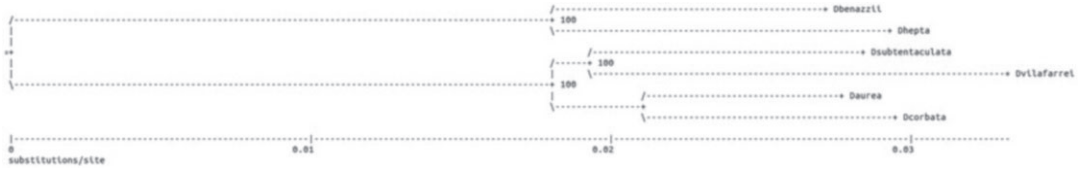


Fig. 4 Screenshot of a tree obtained with the maximum likelihood method, visualized with newick_utils in the terminal. The bootstrap values are shown on the nodes. Scale bar represents the substitutions per site

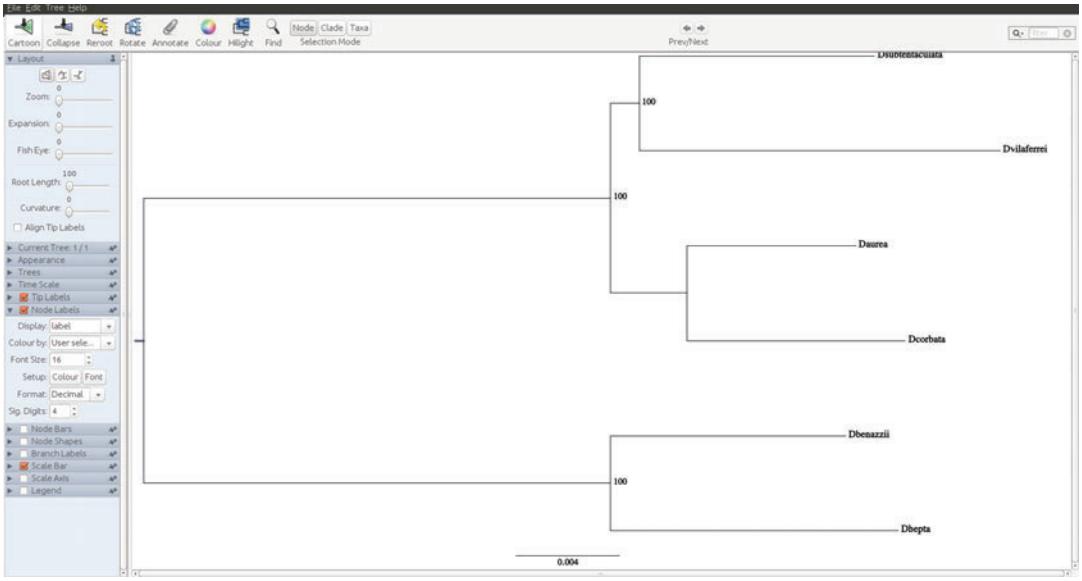


Fig. 5 Screenshot of a tree obtained with the maximum likelihood method, visualized with FigTree. The bootstrap values are shown on the nodes. Scale bar represents the substitutions per site. The options in the upper and left panels enable the tree's customization, with several exporting formats fit for publication

Reroot the tree with the outgroup using newick_utils:

```
nw_reroot *.treefile "outgroup" > *.treefile.rooted
```

If the outgroup is formed by more than one terminal, use their names separated by spaces.

```
nw_display *.treefile.rooted
```

Or FigTree in a local machine to work with a graphic interface (Fig. 5).

Retain the *.treefile files as starting trees for downstream analyses.

2. Phylogenetic Inference using mixture models (*see* **Note 14**).
   Using mixture models with nucleotide data in IQ-TREE:

```
iqtree -s concat_nuc.phylip -m "MIX{JC,HKY,GTR}+G4" -bb 1000000 -nt “>=24”
```

   Using site-specific frequency models with protein data in IQ-TREE:

```
iqtree -s concat_prot.phylip -m LG+C20+F+G -ft concat_prot.fasta.treefile -bb 1000000 -nt
“>=24”
```

   Using Bayesian CAT models implemented in PhyloBayes for protein data.
   Install phylobayes-mpi with conda and run:

```
mpirun -n “>=24” pb_mpi -d concat_prot.phylip -catfix C20 -lg -t concat_prot.fasta.treefile
chain_0
mpirun -n “>=24” pb_mpi -d concat_prot.phylip -catfix C20 -lg -t concat_prot.fasta.treefile
chain_1
```

   Using Bayesian CAT models implemented in PhyloBayes for nucleotide data:

```
mpirun -n “>=24” pb_mpi -d concat_nuc.phylip -cat gtr -t concat_nuc.fasta.treefile chain_0
mpirun -n “>=24” pb_mpi -d concat_nuc.phylip -cat gtr -t concat_nuc.fasta.treefile chain_1
```

   These last two analyses can take a long time. Check the chain periodically:

```
tail chain_0.trace
tail chain_1.trace
```

   Stop the chains softly if you estimate that the chains have reached convergence:

```
echo 0 > chain_0.run
echo 0 > chain_1.run
```

   Check the convergence of parameters:

```
tracecomp -x 1000 chain_0.trace chain_1.trace
```

   Use the following criteria to evaluate them:

maxdiff < 0.1 and minimum effective size>300: good run

maxdiff < 0.3 and minimum effective size > 50: acceptable run

```
bpcomp -x 1000 chain_0 chain_1
```

maxdiff < 0.1: good run

maxdiff < 0.3: acceptable: gives a good qualitative picture of the posterior consensus

0.3 < maxdiff < 1: the sample is not yet sufficiently large, and the chains have not converged yet, but this is on the right track.

maxdiff = 1 even after 10,000 points; this indicates that at least one of the runs is stuck in a local maximum.

If the values of maxdiff parameter are not good, restart the runs using the same commands as in the previous steps with the same name for the chains and the analysis will continue retaining the previous results.

Once the parameters indicate the runs have converged, visualize the trees (bpcomp.con.tre file) in newick_utils or FigTree.

At the end of the process, four trees based on mixture models are obtained. An ML tree and a Bayesian tree for each dataset, nucleotide and protein. Just as with Sanger data-based trees, it is possible to take one of these trees and represent the bootstrap and probability values obtained in the four analyses. (Fig. 6).

4 Notes

1. Sampling

The best option to have a good-quality nucleic acid extraction is always to work with living planarians. However, when these have to be collected and transported for a long time over long distances to the laboratory, it is better to fix them in situ lest we want to risk losing them in the process.

To obtain a good yield of high-quality RNA, the best fixation is accomplished using RNAlater. Another important point to take into account when sampling in the field is the fact that some localities may be occupied by more than one species, some being externally extremely similar to our target species. If we are interested in one specific species, then it is highly recommended to cut a small piece of the animal (commonly the tail) in situ and fix it in a microtube with absolute ethanol, while fixing the rest in another microtube containing RNAlater. Both tubes must be clearly identified to know which are the two pieces of each individual. It is a good practice to use tubes with screw caps of different colors for each

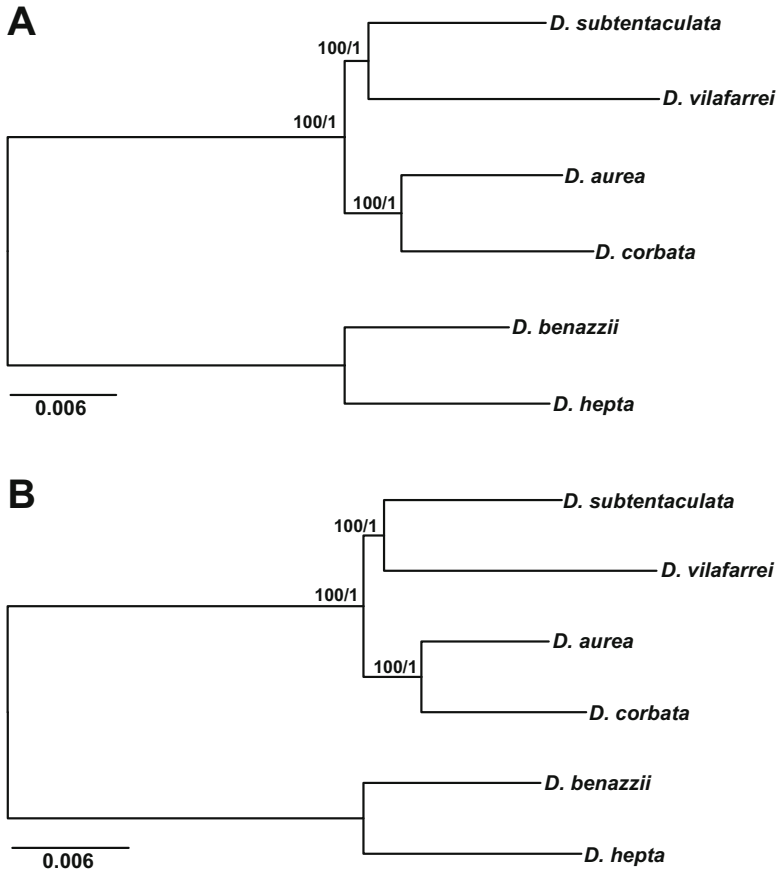


Fig. 6 Maximum likelihood trees based on the example dataset of the *Dugesia* genus. Inferred using mixture models and 1,000,000 replicates of ultrafast bootstrap approximation implemented in IQ-TREE. (**a**) Tree obtained from protein data (LG + C20 + F + G); (**b**) tree obtained from nucleotides (MIX{JC,HKY, GTR} + G4). Numbers over nodes indicate bootstrap/posterior probability support values. The probability values were obtained from analyses in PhyloBayes running two chains with (**a**) protein data (-catfix C20 -lg) and (**b**) nucleotide data (-cat -gtr)

fixation procedure (we regularly use red caps for alcohol and yellow ones for RNAlater).

Once in the laboratory we can check the identity of the species by obtaining DNA from the ethanol preserved fragment and sequencing its cytochrome oxidase I gene (COI) to give an example.

2. Extraction of nucleic acids in planarians

The experience of many years in our laboratory demonstrates that the use of commercial kits based on column technology is not a good strategy for nucleic acid extraction from planarian tissue. It is

our hypothesis that the pigments form a molecular structure with the nucleic acids that is very difficult to separate. The impossibility to separate pigments from RNA has been reported previously [34]. Actually, when we extract DNA or RNA from pigmented planarians, it is very frequent to obtain a dark solution from the final step of the protocol, probably due to the formation of this complex between pigments and nucleic acids. For *Dugesia ryukyuensis,* an absorption spectrum of body pigment characterized by three absorption peaks of 256, 367, and 463 nm has been reported [35], the first peak being in the same range of nucleic acid absorption (260 nm) which may interfere the quantification of DNA and RNA in spectrophotometer-based apparatuses. We have witnessed this phenomenon when quantifying dark samples using a Nanodrop machine, showing high absorbance and overestimating concentration values when compared to quantification by Qubit. We think that the macromolecular complex formed by the nucleic acids and pigments cannot be separated by the current column methods and, in the final elution step, this structure does not properly elute. As a consequence, the yield of the extraction based on column methods is very low in planarians. We recommend using traditional methods of nucleic acid extraction not based on columns, such as phenol/chloroform for DNA [12] and Trizol for RNA.

3. Library construction and sequencing

Transcriptome sequencing can be achieved in different companies nowadays. Most RNAseq is still performed by Illumina sequencing (short reads) although some long read technology (as Nanopore, Oxford) options are beginning to be offered. The latter ensures full-length transcript sequences without the need of assembling the short reads, allowing an accurate identification of isoforms and alleles. Nonetheless, they are way more expensive options.

Library construction may preferably avoid including the highly repetitive and abundant ribosomal RNAs. This can be achieved using a polyA-based capture kit to include only protein coding RNAs or, presently more common, a Ribo-Zero library kit preparation. The latter uses probes that selectively bind rRNAs to eliminate them before sequencing, but keeps noncoding RNAs (ncRNAs) that may also be of interest. Also using stranded information identifies from which of the two DNA strands a given RNA transcript was derived. This information provides increased confidence in transcript annotation.

4. Software installation

A high number of informatic programs are necessary to execute the bioinformatic workflow. We describe briefly how to install and

use the software implicated in every step. Please, consult the documentation of all of them and study the options carefully.

Use conda to access and install packages. We recommend creating new environments to install the software. In some cases, it is mandatory because the software has specific requirements. Please, consult the conda's manual as the first step.

Not all the software or the latest versions are available in conda. In those cases, carefully consult the installation guide of the software documentation. Take into account the characteristics of the cluster you are using.

5. Raw data from Illumina technology

For paired-end Illumina raw data, you obtain two files by sample: samplename_1.fastq.gz (forward/right reads) and samplename_2.fastq.gz (reverse/left reads). Please see [36] for information about fastqc format and [37] for the gzip data compression utility.

From this point we use "samplename" to denote a hypothetical sample.

6. Working with loops to automate steps

Here, we describe the steps to build a phylogeny with transcriptomic data. If you are working with a high number of samples, you can use loops to automate steps to avoid working with the samples one by one. If you have skills with bash, python, or perl, you can do it easily. However, if you are a beginner with these languages and the command line, you can use the scripts available at https://github.com/lisy87/dugesia-transcriptome/tree/main/scripts.

7. Quality control of the raw data

The MultiQC software allows us to summarize the results of FastQC of all samples in a nice html interface. Use this tool to compare the quality scores of your samples.

Take into account that the GC percentage in planarians is lower when compared to most organisms [38], for instance, GC content in the transcriptome of *Schmidtea mediterranea* is around 35% [39].

8. Remove adapters

Use Trimmomatic program to remove universal Illumina adapters and low-quality bases from the ends (LEADING:3 TRAILING:3). The sliding window method is used to trim bases with a mean quality under 15 (SLIDINGWINDOW:4:15). In addition, reads with lengths shorter than 36 bp are dropped (MINLEN:36). From paired-end reads, Trimmomatic outputs the paired (*trimP.fq.gz) and the unpaired (*trimUP.fq.gz) files for both forward and reverse reads. Retain the paired forward and reverse reads (*trimP.fq.gz) for downstream analyses.

After trimming, it is advisable to run FastQC again on the paired reads to corroborate the improvement in quality after Trimmomatic filtering.

9. Assembly with Trinity

Trinity is one of the most popular softwares to assemble a de novo transcriptome without a reference genome. Trinity has shown better performance against other transcriptome assemblers such as Bridger or SOAPdenovo-Trans [40]. Here, we use Trinity with default options, but it is recommended to consult the broad extension of information available about Trinity and the detailed pipelines described in its GitHub page.

10. Installing OrthoFinder

We recommend creating an environment with python ≥3.6 and install NumPy and SciPy. Then, follow the directions on the web page of OrthoFinder to download and install it.

Please pay attention to all specifications of OrthoFinder and the characteristics of your system. You may have to use the installer designed for glibc-2.15 libraries.

In later versions of OrthoFinder, the software first checks if the number of files that the user can open at the same time is enough to open all the required files to run. If not, OrthoFinder fails immediately and the user has to adjust the limits (see https://github.com/davidemms/OrthoFinder/issues/384). However, if the cluster works under a queuing system, it is possible that this failure happens even for a user with high limits of opened files. In this case, the job has to be submitted directly to the cluster without using the queuing system.

11. Select nucleotide sequences of single-copy orthogroups

After running the script sel_sc.py, you get a fasta file for each of the species with all the OGs (samplename_sc.fasta) and a fasta file for each of the OGs with all the species (OG*.fasta), ready to be aligned.

You get as many "samplename_sc.fasta" files, as samples you have, and as many "OG*.fasta" files as single-copy orthogroups you have. We recommend to move the OG*.fasta files to another folder to work in.

12. Alignment

All through this section be sure to keep the original input files in all intermediate steps and save the files with an informative name related to the step, e.g., OG000625.fasta –> OG000625_alig.fasta –> OG000625_alig_oneline.fasta –> OG000625_alig_oneline_trim.fasta. Follow the same steps for the protein and nucleotide datasets.

13. Delete poorly aligned regions

Please, consult the manual of trimAl and the usage options for the command line. We recommend using the -automated1 option to choose the best options.

14. Phylogenetic inference using mixture models

When using probabilistic-based phylogenetic inference methods, it is compulsory to implement an evolutionary model [13]. Initially, the same model was applied to all the datasets; later on some strategies to partition the datasets and apply different models depending on the fragments of analyzed sequences were developed. Nowadays, and especially ever since the upraise of genomic datasets, the use of mixture models has picked up in the field of phylogenomic analyses, replacing the traditional partition scheme by gene or codon. In this new strategy the phylogenetic inference algorithm evaluates for each site the more adequate model to be applied and groups similar sites in categories [14] making it unnecessary to give beforehand any partition scheme. Overall, the good performance of this method when applied to phylogenetic inference has been demonstrated [15, 16] and its implementation in a variety of software makes it easy to use.

To apply these models to the phylogenetic inference with protein and nucleotide data, we will use IQ-TREE and PhyloBayes. The IQ-TREE software has an excellent implementation of mixture models as well as a posterior mean site frequency (PMSF) model that could be likened to PhyloBayes' CAT model. However, PMSF is less consuming than the latter both timewise and memory-wise [12]. Furthermore, IQ-TREE benefits from the ultrafast bootstrap approximation [41] which in turn helps make this analysis relatively easy without an excessive consumption of either time or computer resources. The implementation of the CAT model in PhyloBayes [14] allows it to work with both protein and nucleotide data implementing a Bayesian approach. Alas, PhyloBayes is much more needy than IQ-TREE both computationally and timewise and the analyses may take much longer.

References

1. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data
2. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 2114–2120: 2114. https://doi.org/10.1093/bioinformatics/btu170
3. Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. https://doi.org/10.1038/nprot.2013.084
4. Grabherr MG, Haas BJ, Yassour M et al (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29:644. https://doi.org/10.1038/NBT.1883
5. Laetsch DR, Blaxter ML (2017) BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Research 6. https://doi.org/10.12688/f1000research.12232.1

6. Challis R, Paulini M (2021) blobtoolkit/blobtools2: v2.6.1
7. Haas B, Papanicolaou A (2019) TransDecoder 5.5.0
8. Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238. https://doi.org/10.1186/s13059-019-1832-y
9. Minh BQ, Schmidt HA, Chernomor O et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
10. Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 21:1095–1109. https://doi.org/10.1093/molbev/msh112
11. Lemey P, Salemi M, Vandamme A-M (2018) The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing, 2nd edn. Cambridge University Press
12. Wang H-C, Minh BQ, Susko E, Roger AJ (2018) Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst Biol 67:216–235. https://doi.org/10.1093/sysbio/syx068
13. Holder M, Lewis PO (2003) Phylogeny estimation: traditional and Bayesian approaches. Nat Rev Genet 4:275–284. https://doi.org/10.1038/nrg1044
14. Quang LS, Gascuel O, Lartillot N (2008) Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24:2317–2323. https://doi.org/10.1093/bioinformatics/btn445
15. Venditti C, Meade A, Pagel M (2008) Phylogenetic mixture models can reduce node-density artifacts. Syst Biol 57:286–293. https://doi.org/10.1080/10635150802044045
16. Schrempf D, Lartillot N, Szöllősi G (2020) Scalable empirical mixture models that account for across-site compositional heterogeneity. Mol Biol Evol 37:3616–3631. https://doi.org/10.1093/molbev/msaa145
17. Anaconda Software Distribution (2020) Anaconda
18. Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics:3047–3048. https://doi.org/10.1093/bioinformatics/btw354
19. Challis R (2017) rjchallis/assembly-stats. Zenodo
20. Fu L, Niu B, Zhu Z et al (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565
21. 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. https://doi.org/10.1093/bioinformatics/btl158
22. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324
23. Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAMtools and BCFtools. Gigascience 10. https://doi.org/10.1093/gigascience/giab008
24. SAMtools (2020) SAMtools, Version 1.11
25. Camacho C, Coulouris G, Avagyan V et al (2009) BLAST+: architecture and applications. BMC Bioinform 10. https://doi.org/10.1186/1471-2105-10-421
26. Li H (2012) Seqtk. https://github.com/lh3/seqtk
27. Fernández R, Laumer CE, Vahtera V et al (2014) Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. Mol Biol Evol 31:1500–1513. https://doi.org/10.1093/MOLBEV/MSU108
28. Cunha TJ, Giribet G (2019) A congruent topology for deep gastropod relationships. Proc R Soc B Biol Sci 286:20182776. https://doi.org/10.1098/rspb.2018.2776
29. Benítez-Álvarez L, Leria L, Fernández R et al (2023) Phylotranscriptomics interrogation uncovers a complex evolutionary history for the planarian genus Dugesia (Platyhelminthes, Tricladida) in the Western Mediterranean. Mol Phylogenet Evol 178:107649. https://doi.org/10.1016/j.ympev.2022.107649
30. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010
31. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scalephylogenetic analyses. Bioinformatics 25:1972–1973. https://doi.org/10.1093/bioinformatics/btp348
32. Borowiec ML (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4:e1660. https://doi.org/10.7717/peerj.1660

33. Junier T, Zdobnov EM (2010) The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. Bioinformatics 26:1669–1670. https://doi.org/10.1093/bioinformatics/btq243
34. Stubenhaus BM, Dustin JP, Neverett ER et al (2016) Light-induced depigmentation in planarians models the pathophysiology of acute porphyrias. eLife e14175:10.7554/eLife.14175
35. Hase S, Wakamatsu K, Fujimoto K et al (2006) Characterization of the pigment produced by the planarian. Pigment Cell Res, Dugesia ryukyuensis. https://doi.org/10.1111/j.1600-0749.2006.00306.x
36. Support Illumina. https://emea.support.illumina.com/bulletins/2016/04/fastq-files-explained.html. Accessed 26 Jul 2021
37. Gzip. https://www.gzip.org/. Accessed 26 Jul 2021
38. Martín-Durán JM, Ryan JF, Vellutini BC et al (2017) Increased taxon sampling reveals thousands of hidden orthologs in flatworms. Genome Res 27:1263–1272. https://doi.org/10.1101/gr.216226.116
39. Abril JF, Cebrià F, Rodríguez-Esteban G et al (2010) Smed454 dataset: unravelling the transcriptome of *Schmidtea mediterranea*. BMC Genomics 11. https://doi.org/10.1186/1471-2164-11-731
40. Hölzer M, Marz M (2019) De novo transcriptome assembly: a comprehensive cross-species comparison of short-read RNA-Seq assemblers. Gigascience 8. https://doi.org/10.1093/gigascience/giz039
41. Hoang DT, Chernomor O, von Haeseler A et al (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281

Daniel Dols Serrate
Barcelona, 2024