UNIVERSITAT DE BARCELONA

Understanding diversification and dispersion patterns in planarians (Tricladida, Platyhelminthes): form one to hundreds of genes

Lisandra Benítez Álvarez

PhD. Thesis

Understanding diversification and dispersion patterns in planarians (Tricladida, Platyhelminthes): from one to hundreds of genes

Lisandra Benítez Álvarez

2022

Facultat de Biologia

Departament de Genètica, Microbiologia i Estadística

Programa de Doctorat en Genètica

Understanding diversification and dispersion patterns in planarians (Tricladida, Platyhelminthes): from one to hundreds of genes

Memòria presentada per

Lisandra Benítez Álvarez

Per optar al grau de Doctor per la Universitat de Barcelona,

Barcelona, Abril de 2022

Lisandra Benítez Álvarez
La doctoranda

Dr. Marta Riutort León
La directora i tutora

“The theory of evolution is quite rightly called the greatest unifying theory in biology”

Ernst Mayr

Acknowledgments

Agradecimientos

Muchas personas han contribuido y me han acompañado en el camino que hoy me trae a la presentación de esta tesis. Sin su apoyo, la niña que se preguntaba por qué las lagartijas del jardín tenían colores verdes y brillantes mientras las que veía de noche dentro de casa eran opacas y casi translúcidas, no estaría aquí. Y no tendría a su alcance las herramientas para responder esa pregunta y muchas más sobre la gran diversidad que nos rodea.

En primer lugar, quiero agradecer a Marta Riutort, mi tutora y cuya contribución ha sido especial en cada etapa. En lo profesional; gracias por la confianza, por llevarme de la mano en cada paso, y por darme la libertad para crecer. En lo personal; gracias por ser madre sustituta, por velar por mi bienestar, y por regalar un pedacito de tu propia vida en cada consejo.

En segundo lugar, quiero agradecer la contribución de todos los coautores de los artículos, sin los cuales no hubiese sido posible llevar a cabo todo el trabajo; y a todos los que han aportado algún punto de vista o de discusión para su desarrollo. También quiero agradecer a Julio Rozas, por permitirnos utilizar el *cluster* de su grupo de investigación para correr los análisis. A Silvia Hinojosa por guiar mis primeros pasos en el mundo de la Bioinformática y a Juan Manuel Lentijo por su valiosa ayuda con los *scripts* de python.

Gracias a los profesores de mi Alma Mater; especialmente a la profe Yoyi y Adriana por guiar mis primeros pasos y darme las primeras lecciones sobre lo que representa ser una mujer científica. Gracias también a los profesores de la Universitat de Barcelona, mi segunda casa, por su contribución a mi formación.

Personalmente, quiero agradecer a mis compañeros del lab por acogerme; Laia, Gema, Raquel, Ona, Dani, Cristian, Paula, Edu, Ricard y Oleguer. Gracias Laia por la compañía en las largas jornadas en el lab, tu solidaridad y apoyo. Gracias a todos por ser “*the lab of love”*. Joan; gracias por ser tan especial, por regalarme un pedacito de vida inolvidable en la casita del Born y la oportunidad de conocer la cultura catalana desde la confianza y el trabajo en equipo dentro de un *castell*.

Agradezco especialmente a Liannet por arrastrarme hasta este país y ayudarme a escapar de una prisión de mar y totalitarismo para perseguir sueños. Gracias a Yoe, por seguir siendo el hermano mayor resabioso, pero siempre protector que has sido desde esos años de malecón. A Faby, por estar siempre, incluso en la distancia. A esos que están desde que éramos adolescentes llenos de temores y que continúan cuidando de mis seres queridos en una pequeña isla a miles de kilómetros.

Gracias Juanma, por devolverme tanto y regalarme mucho más. Gracias por tu apoyo y por mantenerte a mi lado incluso en los más estresantes días, durante la elaboración de este documento. Gracias a Manuel y Felisa por regalarme una familia. Gracias a esa madre especial, siempre estarás con nosotros.

Finalmente quiero agradecer a mi familia. A mi padre por guiar mis primeros pasos para descubrir la naturaleza. Por llevarme al campo, a la Ciénaga. Por comprender que encontrarnos con un cocodrilo era lo mejor que nos podría pasar, más allá del temor a unas fauces feroces. A mi hermana por ser faro y guía. Y a mi madre. ¡¡Madraza ante todo!! Gracias por estar presente en cada momento, por apoyar cada decisión, cada paso. Por ser amiga y confidente. Por nunca, nunca, soltar mi mano. Y a mis abuelos, artífices de todo cuanto soy. Por darme una filosofía de vida. A ustedes siempre irá dedicado cada éxito y cada cosa buena que haga en este mundo.

Abstract

Evolutionary Biology offers a set of methods and theories to elucidate the evolutionary processes that lead to the diversification of life. With the development of sequencing technologies, the access to genetic data has increased dramatically, and with it, the development of methods and theoretical frameworks that allow studying the evolution of any branch of the Tree of Life.

Planarians (Tricladida, Platyhelminthes) are distributed across all biogeographical areas, grouping approximately 1782 species. The freshwater planarians belonging to the *Dugesiidae* family stand out for their high regenerative capability. This feature allows asexual reproduction by fission in several species, whose fissiparous individuals do not develop reproductive structures. Since the main systematic characters for this group are in the copulatory apparatus, the knowledge regarding the evolution of Tricladida is mostly based on genetic data, and many questions remain to be answered.

This thesis aims to continue unravelling the process of diversification and dispersion of Tricladida. To this end, I worked at different taxonomic levels, to answer questions about ancient and recent diversification events. Two different approaches were used. On the one hand, specific molecular markers were used to study groups for which there is almost no phylogenetic information. On the other hand, transcriptomic data was used to study a clade that has been already analysed with few markers and for which specific questions remained unanswered.

Internal relationships among the three Tricladida suborders were unveiled using fragments of the ribosomal genes 18S and 28S. The morphological hypothesis, Cavernicola is the sister group of Continenticola, was unsupported by molecular data. The resulting phylogenetic trees supported the monophyly of the Cavernicola, as well as its sister-group relationship to the Maricola. Additionally, the first molecular phylogeny of the Cavernicola suborder, including all genera described at that moment, except one was obtained. Based on these results a cautious biogeographic hypothesis was sketched, in which the Cavernicola originated, presumably before the Gondwana breakup, in a freshwater habitat and subsequently radiated and colonised both epigean and hypogean environments.

Otherwise, *Girardia* genus is native to the American continent but was reported in Europe in the 1920s and currently is present through all Palearctic regions. Despite their wide distribution and diversity, little is known regarding the evolutionary history of the genus. Here, the first *Girardia* phylogeny was obtained based on two genes, which resolved old taxonomic questions and unveiled new issues. The diversification process of *Girardia*, possibly started in South America with posterior colonisation of North American land masses, through different dispersion waves. Additionally, three introduced species were identified across the world, whose invasive potential, given by the fissiparity, can represent a risk for native populations.

Finally, the diversification of *Dugesia* in the Western Mediterranean was analysed using transcriptomic data. The use of phylogenomic methods allowed obtaining a supported phylogeny of the species from this region. The species tree supported a previous biogeographical hypothesis and added new diversification events, due to the inclusion of taxa not analysed before. Furthermore, the unexpected topology of asexual individuals and their effects on the species tree inference could be reflecting the presence of long-term fissiparity in the most ancient clade of *Dugesia* from the Western Mediterranean.

Using different genetic data sources and methodologies was possible to understand several factors that shaped the complex evolutionary process in Triclads. Processes such as new habitats adaptation, geological events, reproductive strategies, and even human intervention, have influenced the diversification and distribution of this group. Special attention deserves the fissiparity, which has played important and different roles in the evolution of freshwater planarians, leaving a genetic footprint in the asexual populations.

Table of Contents

GENERAL INTRODUCTION

GENERAL INTRODUCTION

1. Evolutionary Biology

Much has been added to the theory of evolution since a sketch of a tree with interconnected branches appeared in Darwin's B notebook in 1837. Two centuries later, the core of evolutionary biology continues being to resolve the Tree of Life, understanding the processes that drives its evolution.

Evolutionary biology as a discipline was constituted officially in 1946 by the initiative of Ernst Mayr, promoting the use of new experimental methods rather than just the descriptive or taxonomic practices used until the moment for evolutionary studies. Thus, the new discipline reinvented Darwin's theory and enriched it with empirical and experimental data, as well as new theoretic frameworks (Smocovitis, 1992).

Based on the phylogenetic hypothesis that all living organisms share a common ancestor, the phylogenetic trees have been the battle horse of this discipline, showing the evolutionary relationship between taxa (Cavalli-Sforza & Edwards, 1967; Huelsenbeck & Rannala, 1997). Therefore, the principal challenge of evolutionary biology is to reconstruct evolutionary history using only the currently available information.

From the observation in the field until the elaboration of the hypothesis that explains it, all our reasoning scheme is based on the comparative method. For a long time, the study of the evolutionary process as the adaptation was based on the comparative description of traits and their correlation with the environmental conditions (Huey, Garland, & Turelli, 2019; Martins, 2000). However, the comparative method changed radically with the development of methods based on explicit evolutionary and statistical models. These methods statistically quantify the evolutionary changes in such a way that it is possible to perform hypothesis tests about the evolution (Penny, 1992).

The introduction of the statistical and phylogenetic perspective by Felsenstein, (1985a), revolutionized evolutionary biology (Carvalho, Diniz-Filho, & Bini, 2005). Thus, the discipline has been growing with continuous technological improvement, which has allowed direct access to genomic information. Parallel to this, the development of new theoretical frameworks for the analysis of these data and the interpretation of the

results has placed evolutionary biology in a new era, characterized by high-throughput sequencing and big data analysis.

2. Molecular methods for biodiversity and evolutionary studies

During the first half of the last century, the methods to estimate relationships between species and phylogeny only included comparisons of phenotypic data (Avise, 1994). However, the development of molecular markers (sequences of DNA or protein that reflect heritable differences among individuals or populations) established a new discipline, molecular phylogeny. It is defined as the study of evolutionary relationships between organisms using molecular markers. Its 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 & Li, 2000).

Currently, the molecular phylogeny is not only used as the representation of evolutionary relationships between organisms, but also to describe relationships in gene families, population history, epidemiological dynamics of pathogens, the genealogical relationship of somatic cells during differentiation and cancer development, language evolution, and comparative genomics (Yang & Rannala, 2012).

2.1. Molecular markers

From the mid-1960s until today, a large number of molecular markers have been designed. Several of them are currently used. The first generation, the protein markers, started the revolution of phylogenetics and population genetics (Coyne, Felton, & Lewontin, 1978). However, at the end of the 70's, markers based on DNA sequences occupied the central role in phylogenetic studies; DNA-DNA hybridization (Sibley & Ahlquist, 1984), restriction fragment length polymorphism (RFLP) (Lee, Kim, Ha, & Park, 2006; Miller & Tanksley, 1990), the use of mitochondrial DNA in animals (Castresana, 2001; Marmi, López-Giráldez, & Domingo-Roura, 2004; Thomas, Schaffner, Wilson, & Pääbo, 1989) and other organelles in plants (Dong, Liu, Yu, Wang, & Zhou, 2012; Palmer, Jansen, Michaels, Chase, & Manhart, 1988; Zurawski & Clegg, 1987), single-copy nuclear DNA (Dolman & Phillips, 2004; Slade, Moritz, Heideman, & Hale, 1993; Streelman & Karl, 1997), and DNA fingerprinting (Burke & Bruford, 1987; Lynch, 1991). In the 1980s, the introduction of PCR (Polymerase Chain Reaction) revolutionized the field of Molecular Biology, allowing the amplification of specific DNA

sequences and facilitating their sequencing, a process that was previously quite complex (Saiki et al., 1988, 1985). Moreover, it allowed the development of a group of markers based on this technique, including RAPDs (Randomly Amplified Polymorphic DNAs), STR (Short Tandem Repeats) also known as microsatellites, AFLPs (Amplified Fragment-length Polymorphisms), SINEs (Short Interspersed Elements), SSCPs (Single-strand Conformational Polymorphisms), SNPs (Single Nucleotide Polymorphisms), and finally HAPSTRs and SNPSTRs referring to haplotypes and single nucleotide polymorphisms in STR regions, respectively (see Grover & Sharma, 2016 for a broader review).

There are several reasons to assert that molecular data, particularly DNA and protein sequences, are much more useful for studies of evolution than morphological and physiological data: 1) they are strictly heritable entities, 2) the description of the molecular characters and their states are not ambiguous, 3) molecular traits generally evolve more regularly than morphological and physiological ones, 4) molecular data is often more amenable to quantitative analysis, 5) it is easier to make homology assessments from them, 6) it allows us to assess evolutionary relationships between related but very distant organisms, and 7) molecular data is much more abundant than morphological data. Thus, the use of this type of data allows the estimation of genetic relationships of kinship, pedigree, genealogical affinities within geographically separated populations, genetic divergences between recently separated species, as well as to determine phylogenetic connections between ancestors, branches, and descendants on the Tree of Life (Graur & Li, 2000).

One of the main challenges of evolutionary biology is the distinction of components of biological similarity present in descendants of a common ancestor (homology) or present due to evolutionary convergence of descendants of different ancestors (analogy) (Avise, 1994). However, the phenotypic traits often involve selection mediated by responses to environmental changes leading to adaptive convergence or divergence of particular morphological attributes. On the other hand, since the molecular markers are useful for differentiating homologies and analogies, the molecular data is more useful than morphological data to study highly differentiated groups, whose shared ancestral homologies can be accessed only by analyzing genetic data. Therefore, phylogenetic inference from genetic data can be used to discover homoplasy (the same state of character is attained by convergence,

parallelism, reversals, etc. and not by descent) in morphological traits (Alvarez et al., 1999; He et al., 2018; Köhler & Criscione, 2015).

2.1.1. Species identification based on DNA sequences

Several DNA markers are variable and informative enough to be used for taxonomic identification at different levels. The sequence of these markers retains phylogenetic information that can act as a “barcode” at an individual level, allowing the assignment of individuals to species. The mitochondrial gene *Cytochrome Oxidase I* (COI) is an excellent barcode marker for animals. Due to its high evolution rate, this gene allows the assignment to species level or even phylogeographic groups within a single species. In addition, COI retains a strong phylogenetic signal that enables the assignation of an unidentified organism to higher levels of taxonomic classification (Hebert, Cywinska, Ball, & R, 2003). International scale DNA Barcoding projects such as BOLD (Barcode of Life Data System, (Ratnasingham & Hebert, 2007), FISH-BOL (Fish Barcode of Life, (Ward, Hanner, & Hebert, 2009) stand out due to their large databases based on a region of COI and a standardized species identification process. Additionally, the DNA metabarcoding method allows the high-throughput DNA-based identification of multiple species from a complex sample (Cristescu, 2014).

Although COI is the most used barcoding marker (Valentini, Pompanon, & Taberlet, 2008), other genes with similar characteristics can be used for species identification, depending on the taxonomic group or the level of taxonomic assignation required (Epp et al., 2012; Guardiola et al., 2015; Wangensteen, Palacín, Guardiola, & Turon, 2018). Moreover, DNA barcoding methods based on genomic sequencing techniques have been developed (Bohmann, Mirarab, Bafna, & Gilbert, 2020; Palumbo, Scariolo, Vannozzi, & Barcaccia, 2020; Yang et al., 2020).

DNA barcoding does not replace the classic taxonomy, on the contrary, it emerges as a valuable complementary tool for the taxonomic assignment of unknown specimens and the discovery of new species. Also, it constitutes a good alternative for the classification of the organisms with inaccessible morphological diagnosis, being a rapid and effective method to place any unknown taxa in a phylogenetic context. In addition, DNA barcoding is very useful in other fields such as medicine, the food industry, and forensic sciences (Frézal & Leblois, 2008).

The application of DNA barcoding in ecological studies for conservation has been broadly extended. This technique can be used for the identification of commercial or endangered species (Ardura, Linde, Moreira, & Garcia-Vazquez, 2010; Li et al., 2017), invasive alien species (Armstrong & Ball, 2005; Bezeng & Van Der Bank, 2019; Briski, Cristescu, Bailey, & MacIsaac, 2011; Dejean et al., 2012) and the assessing of biodiversity in endangered ecosystems (Neigel, Domingo, & Stake, 2007; Trivedi, Aloufi, Ansari, & Ghosh, 2016; Robert D Ward, Holmes, White, & Last, 2008; Witt, Threloff, & Hebert, 2006).

2.2. Sequencing technologies

The development of molecular markers has been possible because of the parallel development of DNA sequencing methods from the late 1970s. The first generation of sequencing methods started with the Maxam and Gilbert technique; based on the cleavage of a terminally labeled DNA molecule with chemical agents, producing a set of radioactive fragments that were separated in an electrophoresis Polyacrylamide gel (Maxam & Gilbert, 1977).

However, has been the development of the Sanger method, based on the dideoxy technique (Sanger, Nicklen, & Coulson, 1977; Zimmermann, Voss, Schwager, Stegemann, & Ansorge, 1988) (see https://www.youtube.com/watch?v=e2G5zx-OJIw) the driving force behind the first generation of sequencing methods. This method has been improved with time until the current stand-alone 96-capillary sequencers with a capacity to produce about half a million nucleotides of DNA sequence per day. The basic principle is simple (Fig. 1A) and consists of the use of DNA dideoxynucleotides (ddNTPs) labelled with different fluorescent dyes that allow identifying the extension products separated by capillary electrophoresis (Men, Wilson, Siemering, & Forrest, 2008). This method remains useful for the sequencing of particular regions using specific primers and currently is broadly used for the sequencing of molecular markers for evolutionary studies.

The second generation of DNA sequencing arrives with the use of another method for nucleotide identification. The pyrosequencing method measures the amount of pyrophosphate produced during the polymerization reaction of a dNTP (Hyman, 1988). Several improvements of this pyrosequencing system led to the Roche

454 sequencing systems, the first commercial platform for the “next-generation sequencing” technology (Rothberg & Leamon, 2008).

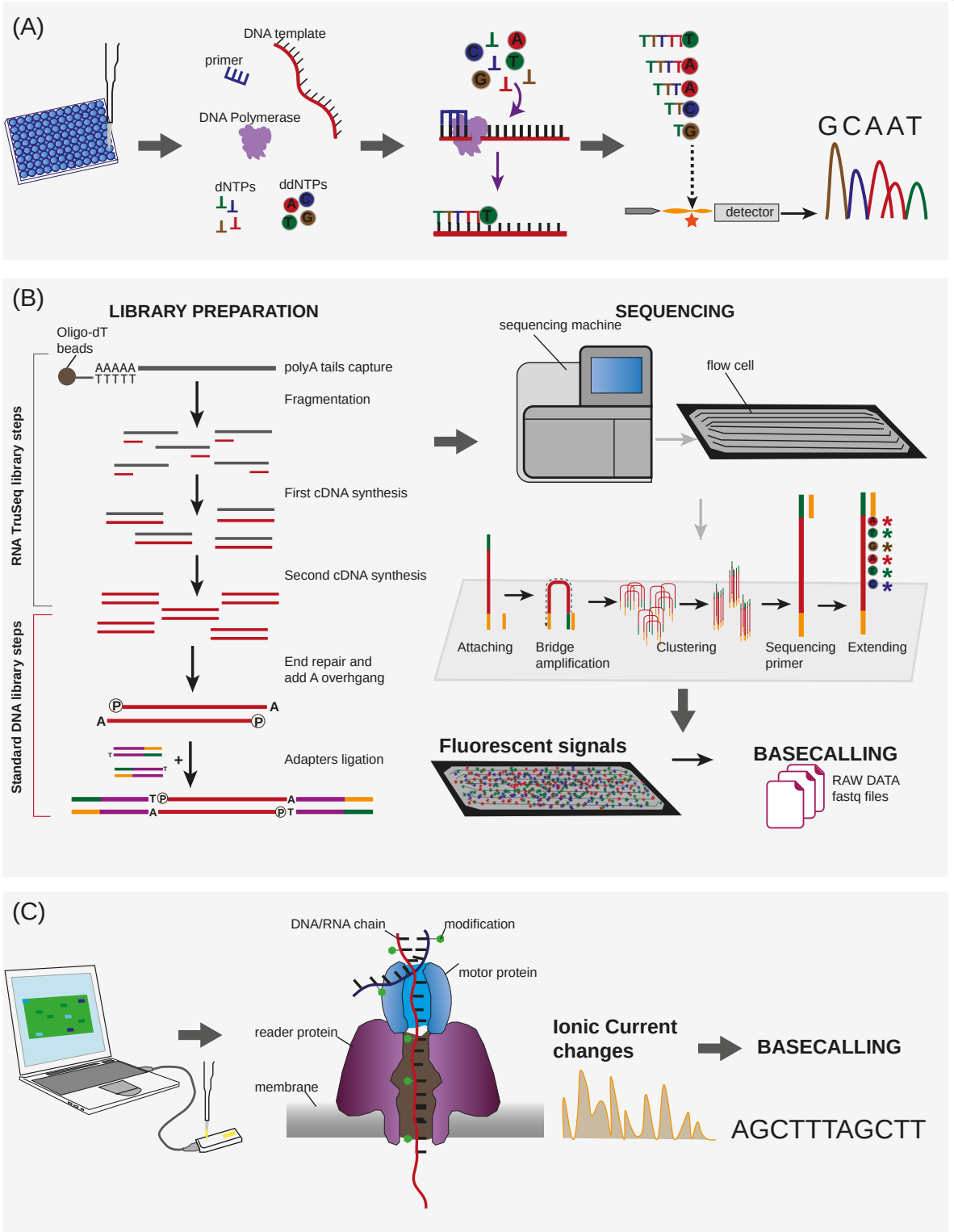


Figure 1. Sequencing systems. (A): Sanger method based on dideoxy technique. DNA template, primers, DNA polymerase, deoxyribonucleotides of DNA (dNTPs), and dideoxynucleotides (ddNTPs) are the components of the reaction. The ddNTPs lack the 3′ hydroxyl group that is required for the extension of DNA chains. *The figure caption continues in the next page*

Figure 1 ***cont***. Therefore, the bond with the 5′ phosphate region of the next dNTP cannot be formed. When the ddNTPs are incorporated to the chain, the reaction is stopped generating fragments of different length that are separated by capillary electrophoresis. Every ddNTP is labelled with different fluorescent dyes and generate a specific signal when the fluorophores are excited by the laser at the end of the capillary. The signals are interpreted by a detector, and the DNA sequence can be interpreted from the colour that corresponds to a particular nucleotide. (B): Second generation sequencing system based on sequencing by synthesis. The first step is the generation of the library. In the case of RNAseq, the first steps convert the RNA to cDNA using two amplification steps, the first one using a reverse transcriptase. After these steps, the general workflow follows in the same way for RNA and DNA sequencing. DNA fragments are repaired and prepared for adapters ligation. The adapter-ligated fragments are amplified by PCR and purified. The library is loaded into the sequencing platform, and all process occurs in the flow cell. The fragments are attached to the surface by complementary sequences to the adapters. Each fragment is repeatedly amplified creating clonal clusters of thousands of copies. The sequencing runs through the incorporation of proprietary modified nucleotides with different fluorescent labels during the synthesis, which generate specific fluorescent signals in the flow cell. These signals are interpreted by the sequencing machine that at the end generates the raw data output in fastq format, indicating a quality score for every called base. (C): Long read sequencing based on nanopore technology. A simple library without any tagmentation step is first prepared and loaded into a small device connected to the computer. The pass of DNA or RNA molecules through the pore in the membrane generate changes of ionic current in the system. Each base, and modification, generate a specific electrical signal that is interpreted by the software and traduced to base sequence.

The sequencing technology has been under a continuous upgrade aimed at improving the performance, data throughput, and quality (see Heather & Chain, 2016 for broader review). Diverse methodologies such as the sequencing by oligonucleotide ligation and detection (SOLiD) and the Ion Torrent, a descendant of 454 sequencing, have arisen in the first decade of 2000. However, the Illumina platform is the most successful sequencing system in the last decades (Slatko, Gardner, & Ausubel, 2018).

The Illumina technology is based on the sequencing by synthesis (SBS) method, which supports massively parallel sequencing. The basic process (Fig. 1B) include firstly the library preparation, to obtain the pre-processed DNA fragments that will be loaded into a flow cell (see https://www.youtube.com/watch?v=fCd6B5HRaZ8 and Illumina documentation for more information). Thereby, this technology also allows the amplification of both ends of DNA fragments. Paired-end (PE) sequencing produces twice the number of reads investing the same time and effort in library preparation. It enables more accurate read alignment and the ability to detect insertion-deletion (indel) variants, which is not possible with single-read data. In addition, PE sequencing facilitates the detection of genomic rearrangements and repetitive sequence elements, as well as gene fusions and novel transcripts (Illumina, 2022a).

Illumina sequencing supports several methods developed for the analysis of different sample sources and data analysis options. The library preparation steps are the principal differentiation among methods, the rest of the sequencing stages are the same. For genomic sequencing, several methods have been developed; Whole-Genome Sequencing (WGS), Exome Sequencing, *De novo* Sequencing, and targeted sequencing. In the case of transcriptomic sequencing, the ribosomal RNA is removed and the total RNA is converted to cDNA by a reverse transcriptase before the standard library preparation. The principal methodologies are focused on total RNA and mRNA sequencing, target RNA sequencing, small RNA sequencing, non-coding RNA sequencing, and single cell RNA sequencing. For epigenomics, the methods are aimed at methylation sequencing, ChIp sequencing, and ribosome profile (Illumina, 2022b). In addition, Illumina offers a wide suite of instruments. From the MiniSeq series of benchtop sequencers to the production-scale sequencer NovaSeq6000 with the capacity to output up to 6 Tb and 20 billion reads in less than two days.

Despite its high development, the short-read sequencing technology is insufficient to resolve long genomic regions with high structural complexity. This took to the development of the third generation, that enables the sequencing of single molecules of DNA or RNA without previous amplification steps (Slatko et al., 2018; Walker, Gurven, Burger, & Hamilton, 2008). The current top technology for long reads sequence is based on single molecule real time (SMRT) sequencing (Eid et al., 2009) and is housed by Pacific Biosciences (PacBio) (see https://www.pacb.com/videos/video-introduction-to-smrt-sequencing/). This technology enables the acquisition of approximately 55,000 to 365,000 reads, depending on the platform, with 10-16Kb of average length (see Ardui, Ameur, Vermeesch, & Hestand, 2018 for a broader revies).

Recently, new long-reads technology has been developed by Oxford Nanopore Technologies (ONT) (Jain et al., 2016; Olasagasti et al., 2010). The method is based on the identification of individual nucleotide by a specific change in the electrical conductivity when the DNA molecule passes through a pore (Fig. 1C) (see https://www.youtube.com/watch?v=RcP85JHLmnI). The most important attribute of this technology is the portable sequencing devices, which have a relatively low cost and can produce data in real-time (Laver et al., 2015). Although the length of the reads depends on the DNA fragments in the library, Nanopore sequencing has reported a

maximum read length of 2273 Mb, with a throughput of 10-15 Gb per flow cell for MinION devices, and 153 Gb for a single PromethION flow cell with an average sequencing speed of approximately 430 bases per second (Wang, Zhao, Bollas, Wang, & Au, 2021).

Despite their advantages (access to structural arrangements in long genomic and transcriptomic regions, the identification of methylation sites, real time sequence), both long-read methods have high error rates. Several tools are available for long read error correction, but the combination of long and short reads data is the best strategy regarding the correction quality and computing resources (Zhang, Jain, & Aluru, 2020). Hence, recent studies use a combination of both sequencing technologies to improve the data analysis process (Mahmoud, Zywicki, Twardowski, & Karlowski, 2019; Zimin & Salzberg, 2022).

The development of all these sequencing systems, together with the reduction of their costs, has allowed access to a large amount of genomic information. Now, the processing and analysis of this data represent a great challenge in terms of software and hardware resources.

2.3. Phylogenetic inference

2.3.1. Traditional inference methods

The development and sophistication of phylogenetic inference methods has been driven by the need to give statistical support to phylogenetic hypotheses and the capability to analyze large genetic data sets under diverse sequence and population evolution models.

The first developed methods to reconstruct phylogenies were based on distance or parsimony. With the distance methods, the trees are obtained based on the genetic distance between pairs of taxa. These methods are very fast, but they are not statistically consistent on highly divergent data (Bruno, Socci, & Halpern, 2000). However, methods based on distance are useful for haplotype network building (Paradis, 2018) in population genetics, and phylogenetic network inference. The last used to discover reticulate evolutionary processes such as horizontal gene transfer, hybrid speciation, hybrid introgression, and recombination (Huson & Bryant, 2006;

Janssen & Liu, 2021), or to follow the evolution of pathogens causing major pandemics worldwide (Forster, Forster, Renfrew, & Forster, 2020).

On the other hand, the maximum parsimony principle always favours the simplest hypothesis. Thus, under this method, it is obtained the topology that requires the minor number of changes in the character states (Fitch, 1971). Although the parsimony methods have their followers, principally for the analysis of morphological data (Goloboff, Torres Galvis, & Arias, 2018) their disadvantages against probabilistic methods have been demonstrated (O'Reilly, Puttick, Pisani, & Donoghue, 2018; Puttick, O'Reilly, Pisani, & Donoghue, 2019). In the case of genetic data, the parsimony lacks a robust statistical base, especially when the evolutionary rates are not small and differ sufficiently in different lineages (Felsenstein, 1983; Felsenstein, 1978, 1985b). Thus, the basic assumptions of this method fail against realistic and complex evolutionary models (Yang, 1996).

Contrary to previous methodologies, the probabilistic methods infer the best tree based on probability distributions, assigning statistical support to the phylogenetic hypothesis. In addition, these methods deal efficiently with complex models of sequence evolution. The most used probabilistic approaches in Evolutionary Biology are Maximum Likelihood (ML) and Bayesian Inference (BI).

ML method estimates the phylogenetic tree with the maximum likelihood value. The likelihood of a tree is the probability of the data given the tree. It is a function of the topology and the parameters of the sequence evolution model. Thus, the tree with the maximum likelihood is not the tree with a higher probability to be the correct one, but it is the tree that yields the highest probability of representing the evolution of the observed data (Felsenstein, 1981). The ML approach allows testing different phylogenetic hypotheses under different models of sequence substitution (Huelsenbeck & Crandall, 1997), and even today, constitutes one of the most robust frameworks for the development of new methodologies (Morel et al., 2022). Despite its high computational cost, many programs have implemented this method for phylogenetic inference. Currently, the most widely used are RaxML (Stamatakis, 2014) and RaxML-NG (Kozlov, Darriba, Flouri, Morel, & Stamatakis, 2019), with high performance for large datasets, diverse substitution models and supported data types. In addition, the implementation of ML approach in IQ-TREE (Minh et al., 2020) with high computational efficiency on genomic data analysis and notable features for

evolutionary hypothesis testing, also constitute one of the favourites. Both RAxML suit and IQTREE implementations surpass other ML software in terms of efficiency and accuracy on phylogenomic dataset analyses (Zhou, Shen, Hittinger, & Rokas, 2017).

The BI method allows obtaining the phylogenetic tree with the maximum posterior probability for a certain data set and a substitution model (Rannala & Yang, 1996). Whereas in ML the parameters in the model are considered unknown fixed constants, in BI the parameters are considered random variables with statistical distributions. Parameters are first assigned a prior distribution and combined with the data to generate the posterior distribution. Its integration with the Markov Chain Monte Carlo (MCMC) algorithms allows the inference of independent branch lengths on unrooted trees, which makes it popular (Yang & Rannala, 2012). This method is broadly used for inferring trees from large datasets, detecting natural selection, and choosing among models of DNA substitution (Huelsenbeck, Ronquist, Nielsen, & Bollback, 2001). In addition, new methodologies have been developed based on the Bayesian framework (Barido-Sottani, Vaughan, & Stadler, 2020; Liu, Edrisi, Ogilvie, & Nakhleh, 2022; Meyer, Dib, Silvestro, & Salamin, 2019). MrBayes (Ronquist et al., 2012) has been the most used software for a long time. Currently, the BI implementation in BEAST (Bouckaert et al., 2019) enable the inference of time-calibrated phylogenies by integrating multiple types of data such as genetic sequences, phenotypic character states, fossil records, and biogeographic range information under different evolutionary models. Also, software such as ExaBayes (Aberer, Kobert, & Stamatakis, 2014) and PhyloBayes (Lartillot & Philippe, 2004, 2006; Lartillot, Brinkmann, & Philippe, 2007) for the analysis of large datasets of genomic data have been developed recently, with the handicap that both have a high demand of computational requirements and time.

Both ML and BI approaches are designed to work with concatenated alignments, where all genes are concatenated in a super matrix and a single species tree is inferred. In these cases, partition schemes are applied for the independent estimation of model parameters for genes or codons. It has been demonstrated that combining data from different molecular markers produce a more resolved phylogeny than when a single gene is used because the increase of the sample site's number improves the phylogenetic accuracy (Gadagkar, Rosenberg, & Kumar, 2005). However, large data sets of genomic data demand more efficient methods of data matrix analysis.

2.3.2. Sequence evolution models

Probabilistic based phylogenetic methods demand the use of a sequence evolutionary model (Holder & Lewis, 2003). Thus, several models have been developed to describe the sequence evolution and improve the accuracy of tree inference.

Sequence evolution models describe the rates at which a nucleotide or amino acid (in the case of protein sequences) is substituted by another during the evolution. The complexity of each model depends on the number and type of parameters (frequency, exchangeability or heterogeneity) used to estimate the substitution rates.

The simplest nucleotide substitution model is the JC (Jukes & Cantor, 1969), which considers equal nucleotide frequency and only one substitution rate among all nucleotides. However, this model is not realistic since the probability of transitions (changes between bases with equal chemical structures) and transversions (changes between bases with different chemical structures) are different. K80, allows different rates of change between transitions and transversions (Kimura, 1980), while F81 only allows unequal nucleotide frequency (Felsenstein, 1981). On the other hand, HKY model admits unequal transition/transversion rates and unequal base frequency (Hasegawa, Kishino, & Yano, 1985). Other models, based on these four, have been developed integrating only parameters regarding nucleotide frequency and substitution rates. The most complex, the General Time Reversible (GTR) model integrate unequal rates for every change and unequal nucleotide frequency (Tavaré, 1986). These substitution models can be combined with models of rate heterogeneity across sites (+G) and the proportion of invariable sites (+I).

Aminoacid substitution models can be classified into two principal groups; empirical and parametric models. The empirical models are based on a matrix of exchangeability rates of 20 × 20, and 20 amino acid frequencies. Based on this principle, have been developed specific models for specific taxonomic groups or organelle, but the general matrix Dayhoff (Dayhoff, Schwartz, & Orcutt, 1978) , LG (Le & Gascuel, 2008), JTT (Jones, Taylor, & Thornton, 1992), WAG (Whelan & Goldman, 2001) and VT (Müller & Vingron, 2000) are the most used empirical methods. On the other hand, the parametric models are based on parameters that describe structural protein evolution. These methods integrate the structural constraints in the substitution

models. However, their computational implementation is complicated and they are difficult to integrate into the standard programs used for phylogenetic inference (Arenas, 2015).

Further, the coding regions are under differential selective pressure that determines heterogeneity substitution rates among sites. Mutations at coding regions can be classified as synonymous (mutations that do not lead to amino acid change, *dS*) and non-synonymous (mutations that implicate amino acid replacing, *dN*). The rate of *dN/dS* is commonly used as an estimator of molecular selection. *dN/dS* >1 indicates positive selection, *dN/dS* < 1 is signal of purifying selection, and *dN/dS* = 1 can be interpreted as neutral evolution (Kimura, 1983). Several codon models have been formulated integrating different sources of information just as *dN/dS* variation across sites and branches, physicochemical properties of the encoded amino acids, codon bias, GC contents, and mutational biases (Arenas, 2015). For example, GY94 is a full codon model that uses a Markov process to describe substitutions between codons, allowing transition/transversion rate bias and codon usage bias. The selective constraint at the protein level is integrated into the model using physicochemical distances between amino acids (Goldman & Yang, 1994). However, such models are computationally expensive and are not integrated into the general software used for model selection or phylogenetic inference. As an alternative, the data can be partitioned in genes and codon positions to assign the more appropriate nucleotide substitution model, in such a way that the estimation of the parameters is independent among genes and codon positions.

The use of an inappropriate substitution model can impact the accuracy of phylogenetic analyses, affecting the topology inference, the branch length estimation, and the calculation of erroneous support values. Generally, the programs identify the most appropriate model to be used for a specific data set by iterating through a hierarchical set of models and testing the fit of the models to the data. Two methods have been used to develop the programs for model selection. The hierarchical likelihood ratio tests (hLRTs) method consists of comparing the likelihood (the probability of the data given the model) of a null model and the alternative. With a P-value below the significance level, the null model is rejected, assuming that the alternative model fits the data significantly better than the null model. Then, the accepted model becomes the null model of the next likelihood ratio tests (LRT) in a

hierarchical scheme. On the other hand, Bayesian theory provides a notable framework for the development of selection model tools. The Bayes factor is the analogue of LRT but calculates the probability of the model given the data and chooses the model with the highest posterior probability. The Bayesian Criterion Information (BIC) constitutes an approximation to the Bayes factor but it is computationally less complex. Another approach for model selection is the Akaike Information Criterion (AIC). It calculates the fit of the model based on the amount of information lost when the model is used to describe the real process of nucleotide substitution. Both AIC and BIC allow for ranking the candidate models based on their values (models with lower values of BIC or AIC are preferred) and the model selection uncertainty estimates. In addition, both introduce a penalization for the number of parameters estimated by the model (Posada & Buckley, 2004; Posada & Crandall, 2001).

Several software have been developed for model selection. Many of them have evolved to efficiently handle large data sets. One of the most used is jModelTest (Darriba, Taboada, Doallo, & Posada, 2012) which implements LRT based methods, Bayesian based methods, and AIC. PartitionFinder is a broadly used tool that allows selecting the model along with partition schemes in molecular and morphological data based on AIC and BIC criteria (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017). ModelFinder additionally incorporates a model of rate heterogeneity across sites. The latter is implemented in IQ-TREE and allows for comparing models of sequence evolution inferred on the same or different trees (Kalyaanamoorthy, Minh, Wong, Von Haeseler, & Jermiin, 2017). SMS, a tool integrated into the PhyML program, is based on heuristic strategies that avoid testing all models and options, simplifying some calculations to save computing time (Lefort, Longueville, & Gascuel, 2017).

Although it is the most extended practice, only the selection of the most fitted model cannot guarantee that this model captures efficiently the dynamics of sequence evolution to provide an unbiased phylogenetic tree. One alternative is also testing if the chosen model is plausible given the data. An implementation of posterior predictive tests of model fit is available in the RevBayes software, which can be applied to nucleotide evolution, continuous trait evolution, and lineage diversification models (Höhna, Coghill, Mount, Thomson, & Brown, 2018). In addition, tools for selecting branch-specific models have been developed (Dutheil et al., 2012).

Currently, the analysis of large genomic datasets dominates the field of phylogenetic inference. Alignments of thousands of genes and millions of positions have to be analysed, and partitions by gene and codon are not so efficient. One solution is to work with mixture model profiles, which also solves the problem of the rates of heterogeneity across sites. Contrary to partition schemes, that assign a specific model to the partition, mixture models allow several substitutions models across sequences and compute the probability that a site fits a category of the mixture model. For example, the discrete-gamma model, integrated with other substitution models to allow variable rates of substitution across nucleotide sites (Yang, 1994), is a type of mixture model constrained to take a specific distribution allowing several categories with equal probability.

The CAT model is one of the most used mixture models. It assumes the existence of a number K of classes differentiated by their equilibrium frequencies and each site is described by the class that fit its substitutional history. This model allows inferring simultaneously the mixture parameters, the rates at each site, the branch lengths, and the topology of the underlying phylogenetic tree (Lartillot & Philippe, 2004). Currently, CAT model is implemented in both; the Bayesian and ML frameworks in PhyloBayes (Lartillot & Philippe, 2004) and PhyML (Guindon et al., 2010) respectively, with several components for protein data ranging from 10 to 60 (C10 to C60) (Quang, Gascuel, & Lartillot, 2008). On the other hand, IQ-TREE (Minh et al., 2020) has an excellent implementation of C10-C60 profiles for protein data, and also a posterior mean site frequency (PMSF) model, as a rapid approximation, to save time and memory consumption (Wang, Minh, Susko, & Roger, 2018). In addition, IQTREE allows defining mixture models using specific components (nucleotide substitution models) for nucleotide data.

Although the mixture model methods are computationally expensive, new tools continue to be presented (Dang & Kishino, 2019; Jayaswal, Wong, Robinson, Poladian, & Jermiin, 2014; Schrempf, Lartillot, & Szöllősi, 2020), since their good performance in phylogenomics has been demonstrated (Redmond & McLysaght, 2021; Venditti, Meade, & Pagel, 2008).

2.3.3. Systematic error sources

Although the use of genomic data substantially minimizes the stochastic errors, systematic errors tend to be more evident. Inconsistency in tree building methods arises from models that do not describe properly the variable evolutionary rates across lineages, heterotachy and heterogeneous nucleotide composition. Thus, these systematic biases introduce a non phylogenetic signal, that the inference method misinterpreted as supporting an alternative topology (Philippe, Delsuc, Brinkmann, & Lartillot, 2005).

When the evolutionary rates across lineages vary significantly, the taxa with higher rates are grouped together because numerous convergent changes accumulated along these long branches are interpreted as false synapomorphies. This artefact is called long branch attraction (LBA) (Felsenstein, 1978) and has a noticeable effect on tree topology. Moreover, heterotachy is the variation of evolutionary rates of a given position through time (Fitch & Markowitz, 1970; Lopez, Casane, & Philippe, 2002). Simulated and empirical data have shown that probabilistic methods such as ML and BI outperform Parsimony approach under realistic combinations of heterotachy and variation of evolutionary rates across lineages. In addition, the effects of heterotachy on tree inference methods can be avoided by the development of mixture models in a probabilistic framework (Philippe et al., 2005).

Also, heterogeneous nucleotide composition across genomic regions can result in the artificial grouping of taxa with similar nucleotide frequencies. Thus, variation in the GC content through genomic regions with important implications for relationships inference have been reported in several taxa (Huttener et al., 2019; Nekrutenko & Li, 2000; Weber, Boussau, Romiguier, Jarvis, & Ellegren, 2014). GC-biased gene conversion (gBGC) can lead to high GC content regions. It is caused by a biochemical bias in the meiotic recombination that favours the incorporation of G and C nucleotides during the repair of mismatches of DNA heteroduplex. This gene conversion bias increases the rate of strong substitutions (A/T to G/C) and leads to the fixation of GC alleles in areas of high recombination frequency (Duret & Galtier, 2009; Webster & Hurst, 2012). Fortunately, although nucleotide composition heterogeneity is an important issue for substitution rates estimation (Kostka, Hubisz, Siepel, & Pollard, 2012), it does not seem to be so problematic for phylogenetic inference if a large amount of data is analysed (Rosenberg & Kumar, 2003).

Two strategies can help to minimise the systematic errors; the use of genomic regions with low substitution rates, which should be homoplasy-free or the use of many genes combined with inference methods that efficiently address multiple substitutions models to describe the heterogeneity of the evolutionary process across positions (Lartillot & Philippe, 2008). Therefore, in phylogenomics it is imperative to use efficient phylogenetic inference methods and substitution models that describe the evolution of sequences across sites.

2.3.4. Sources of discordance between gene trees and species trees

In the light of the great amount of genomic data available, the conflicts among individual gene trees and the species tree are more evident (Degnan & Rosenberg, 2006; Morales-Briones et al., 2021; Nichols, 2001; Rosenberg, 2013). Biological processes such as gene duplication/loss, horizontal gene transfer, introgression, and deep coalescence lead to gene tree heterogeneity (Maddison, 1997).

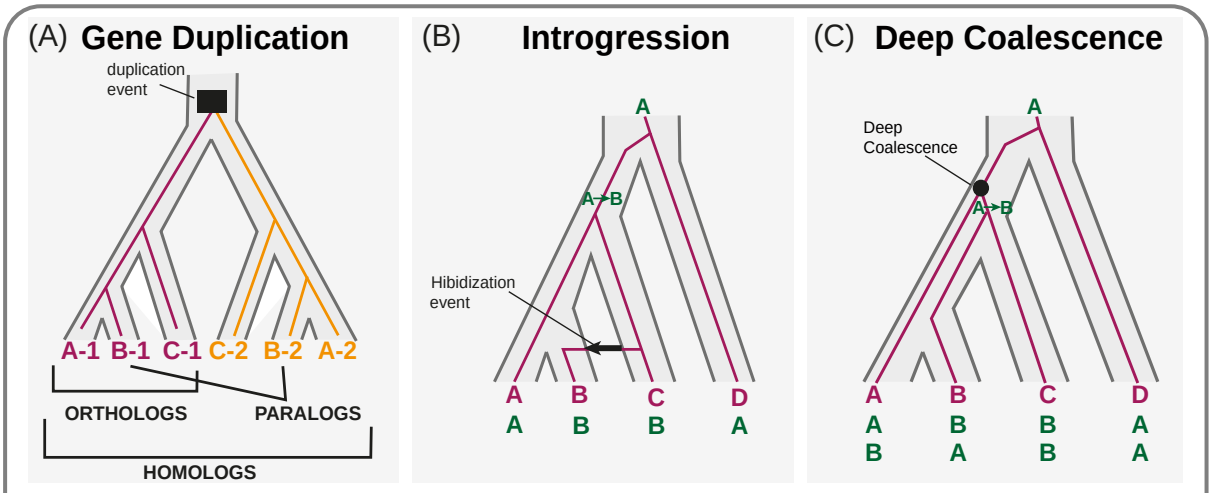


Figure 2. Schematic representation of principal sources of gene tree discordance. The species tree is highlighted in grey and the gene tree in yellow and/or garnet (A): Gene duplication. A duplication event (black square) lead to two gene copies (yellow and garnet gene trees). The concept of orthologs, paralogs, and homologs are also represented. (B): Introgression. A hibridization event leads to gene flow and introgression between differentiated species changing the allele frequencies in the population. (C): Deep coalescence. The alleles coalesce before the speciation event. The premises of ABBA-BABA test are highlighted in green in (B) and (C). For an ancestral allele A and a derived allele B (in green), the frequences of both allele configurations (ABBA and BABA) are equal under deep coalescence, but under gene flow one pattern occurs with more frequency.

Gene duplication events are an important source of gene tree discrepancy because they lead to paralogy. Two characters are homologs if they descend from the same common ancestor (Fig. 2A). The common ancestor can be shared by duplication

(paralogs) or by speciation (orthologs) (Fitch, 2000). Because paralogs do not reflect genealogical relationships, distinguishing between paralogs and orthologs is critical to describing evolutionary processes with accuracy (Altenhoff, Glover, & Dessimoz, 2019). Thus, a correct ortholog inference is an essential step for species tree inference.

Horizontal gene transfer (HGT) implies the transfer of genetic material between distinct evolutionary lineages and constitutes one of the most important processes in microbial evolution (Lawrence, 1999). Most studies have been centred on HGT in prokaryotic genome evolution because of the role that it plays in the origins of new functions, emergence and spread of virulence, and resistance to antibiotics (Eisen, 2000). However, it should not be underestimated in eukaryotic evolution due to the important massive transfer throughout endosymbiosis processes, and their implications in adaptation to specialised niches (Keeling & Palmer, 2008).

Hybridization is defined as the outcrossing of individuals from two different species (Fig. 2B) (Harrison, 1990, 1993). It leads to introgression if alleles from one species are incorporated into the genetic pool of another by backcrossing the hybrids with their parents (Anderson, 1949). Although the traditional biological species concept implicates the cancellation of genetic exchange between different species through reproductive isolation, more recent definitions establish the species as separately evolving metapopulation lineages (De Queiroz, 2007) supporting the permeability of species boundaries (Harrison & Larson, 2014). Hybridization constitutes an important source of genetic variation. At least 25% of plants and 10% of known animal species hybridize (Mallet, 2005). Genetic variation is enhanced by introgression, which occurs regularly and especially in rapidly adaptive radiation groups (Mallet, 2007). Hence several species of hybrid origin are currently recognized (Goss et al., 2011; Meier et al., 2021; Ottenburghs, 2018; Runemark, Vallejo-Marin, & Meier, 2019; Tripp, Fatimah, Darbyshire, & Mcdade, 2013).

Deep coalescence (also called Incomplete Lineage Sorting, ILS) is one of the most common sources of gene discordance. It takes place when ortholog gene copies from two species coalesce into a common ancestral copy before the speciation event (Fig. 2C) (Maddison, 1997). Then, because species can share ancestral genetic diversity even long after their divergence, this gene does not reflect the species' phylogeny. The amount of ILS is correlated with effective population sizes (N_e) and the generation time (Pamilo & Nei, 1988). Therefore, in ancestral populations with high N_e

and recent divergence times, the genetic drift is unlikely to have had time to fix the loci before later divergences, leading to high levels of ILS (Maddison & Knowles, 2006). ILS have been detected in the evolution of large lineages such as mammals (Scornavacca & Galtier, 2017), primates (Rogers & Gibbs, 2014), birds (Suh, Smeds, & Ellegren, 2015), bryophites (Meleshko et al., 2021), neotropical fishes (Alda et al., 2019).

Since both ILS and introgression can lead to the same gene topology, identifying ILS from introgression processes constitutes one of the challenges to decipher the evolutionary history of certain lineages. The ABBA-BABA test (or D statistic) is one of the most used methods to detect introgression, using genome-scale SNP data. It is based on the four-taxon test, with an ancestral allele A and a derived allele B. In this situation, three hypothetical allele configurations can take place (BBAA, ABBA, and BABA) (Fig. 2B and C). Under ILS, are spectated equal frequencies of ABBA and BABA configuration. On the contrary, under gene flow one of the patterns is expected to occur with more frequency (Martin, Davey, & Jiggins, 2015). Thus, ILS acts as a null hypothesis for introgression tests (see Hibbins & Hahn, 2022 for a broader review). Additionally, other parametric approaches, based on the idea that the expected minimum genetic distance between sequences from two species is smaller for some hybridization events than for incomplete lineage sorting scenarios have been proposed (McLenachan, & Lockhart, 2009).

2.3.5. Coalescence method to resolve gene tree discordance

Since the population is the basic unit of evolution, modelling the microevolutionary process inside populations constitutes an important issue in evolutionary biology. The Coalescent theory (Kingman, 1982a, 1982b) has become the core of population genetics. It allows the estimation of population genealogies from empirical genetic data using efficient simulation algorithms and testing various population genetics models with high statistical support (Fu & Li, 1999). The Coalescent theory constitutes the framework of the Multispecies Coalescent (MSC) method (Edwards, Liu, & Pearl, 2007; Rannala & Yang, 2003) that comes to reconcile the discordance between gene trees and species trees.

The coalescence theory at a population level proposes that for a neutral non-recombining locus, all alleles in the population must coalesce in a common ancestor,

providing the probability distribution of the genealogical history of the alleles. In a multispecies species context, the MSC theory describes the genealogical relationships of sequences coming from different species affected by divergence times and population size parameters. Thus, the gene trees are under independent coalescent processes (Rannala et al., 2020).

Because coalescence is an intrinsic process of the sequence evolution and acts independently of the divergence times, ILS patterns can be predicted by coalescent events with accuracy. Thereby, under ILS, the MSC became one of the most robust frameworks, providing parametric multi-locus statistical methods for phylogenomic analysis. In addition, using MSC as a null model, other biological processes that lead to gene trees discordance as introgression can be incorporated into the model (Degnan, 2018; Flouri, Jiao, Rannala, & Yang, 2020; Mirarab, Nakhleh, & Warnow, 2021).

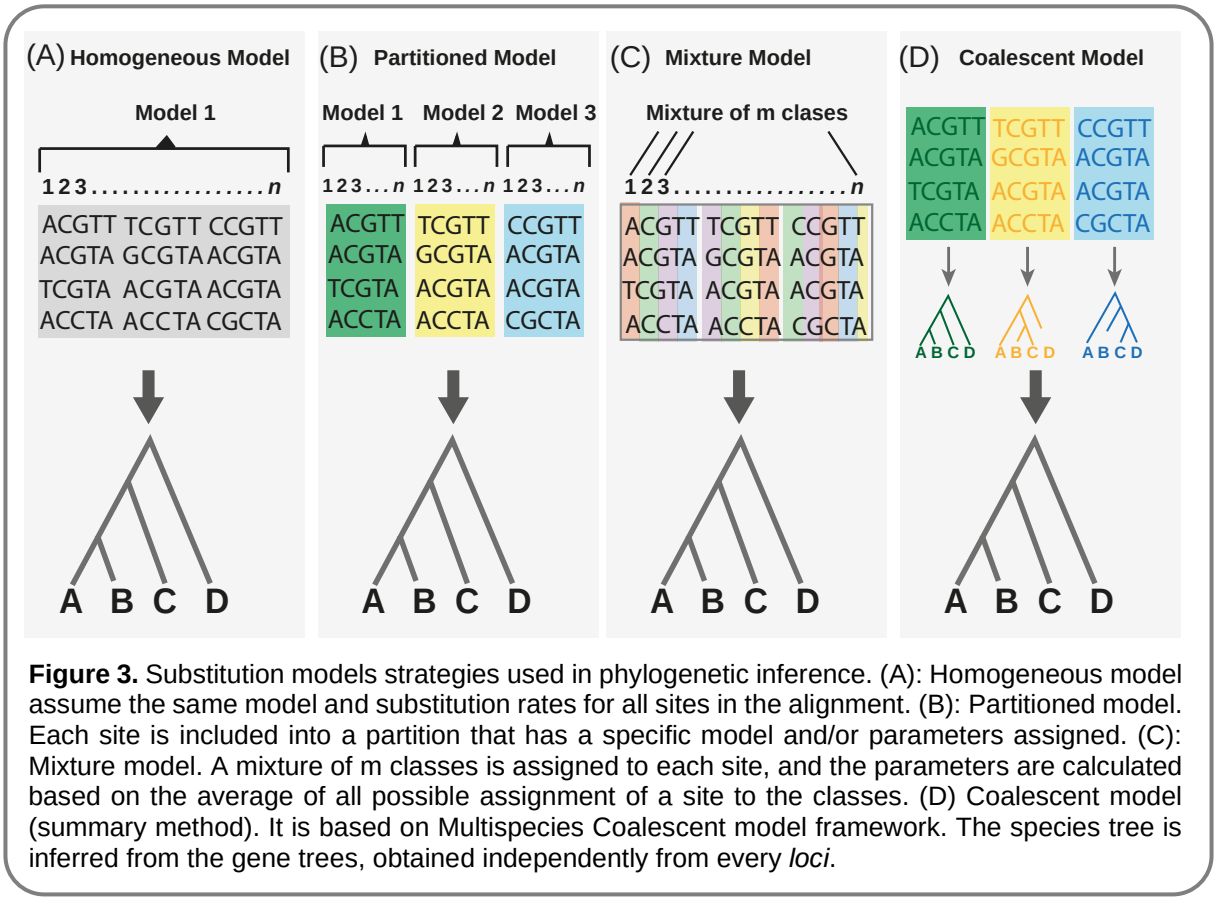


Figure 3. Substitution models strategies used in phylogenetic inference. (A): Homogeneous model assume the same model and substitution rates for all sites in the alignment. (B): Partitioned model. Each site is included into a partition that has a specific model and/or parameters assigned. (C): Mixture model. A mixture of m classes is assigned to each site, and the parameters are calculated based on the average of all possible assignment of a site to the classes. (D) Coalescent model (summary method). It is based on Multispecies Coalescent model framework. The species tree is inferred from the gene trees, obtained independently from every *loci*.

Even though the use of concatenated multigene matrices (Fig. 3B,C) remains useful, under high gene discordance this approach can fail to recover the accurate species tree (Kubatko & Degnan, 2007; Liu, Wu, & Yu, 2015; Liu, Yu, Kubatko, Pearl,

& Edwards, 2009). So, instead of the concatenation scheme used with ML or BI methods, the coalescent methods assume that every gene tree is an independent variable (Fig. 3D). Thus, the coalescence framework allows the calculation of the probabilities of individual gene tree topologies for a given phylogeny that represents the historical relationships among species (Degnan, Rosenberg, & Stadler, 2012).

MSC based methods can combine the information from gene trees in two ways. The summary or two-step methods first build gene trees by traditional methods (ML or BI) and then, the species tree is inferred based on a likelihood function or summary statistics. On the other hand, full coalescent or single-step methods integrate Bayesian theory to simultaneously estimate gene trees and species tree (Liu, Anderson, Pearl, & Edwards, 2019).

Single-step methods are computationally more demanding because the posterior probability distribution is calculated using MCMC algorithm. For this reason, in some cases, it is not practical to apply this Bayesian approach to genome-scale sequence data (Liu, Wu, & Yu, 2015). Implementations of this approach are available in BEAST (Bouckaert et al., 2019), BEST (Liu, 2008), BPP (Yang, 2015), and SVD Quartets (Chifman & Kubatko, 2014).

On the other hand, the two-step methods are highly efficient in terms of computational resources and time, with the capacity to analyse thousands of input gene trees in a short time. Several software have been developed based in summary approach; STEM (Kubatko, Carstens, & Knowles, 2009), JIST (O'Meara, 2010), GLASS (Mossel & Roch, 2010), STAR, STEAC (Liu, Yu, Pearl, & Edwards, 2009), Njst (Liu & Yu, 2011), and ASTRAL (Mirarab & Warnow, 2015).

ASTRAL (Accurate Species TRee ALgorithm) implementation has improved since its first release in 2014 (Mirarab et al., 2014) with the implementation of quartet-based support and posterior probabilities as accuracy estimators (Sayyari & Mirarab, 2016), and recently the possibility to include paralogs into analyses (Zhang, Scornavacca, Molloy, & Mirarab, 2020). With all these features ASTRAL is one of the most utilized coalescence based methods for species trees inference.

ASTRAL provides several informative parameters regarding gene tree discordance and species tree accuracy estimation. The species tree is obtained with a quartet score, that is the number of quartets trees that are present in the species tree,

normalised by the total number of quartet trees in input gene trees. It can be used as a measure of gene tree discordance. Branch lengths represent coalescent units (CU) and are proportional to the number of generations spanned by the branch and inversely proportional to the population size. The branch lengths are a function of discordance, the reason why they are very sensitive to gene tree error. In addition, each branch has associated a posterior probability (PP) of the branch being correct. ASTRAL also provides a measure of discordance around each branch based on branch quartet scores. The quartet support of a branch is the proportion of times that quartets around the branch are resolved identically to the species tree in the gene trees. Quartet scores values close to 1/3 indicate high discordance levels. Also, under ILS the quartet scores of the second and third topologies must be identical (Mirarab, 2019).

Although concatenation can be more accurate than summary methods under low levels of ILS, summary methods are generally more accurate than concatenation when there are an adequate number of sufficiently accurate gene trees (Mirarab, Bayzid, & Warnow, 2016). Specifically, ASTRAL implementation outperforms concatenate methods under ILS scenery (Chou et al., 2015). Thereby, coalescent-based methods may be key to estimating highly accurate species trees from multiple loci under high gene tree discordance.

2.3.6. Support measures of phylogenetic trees

The confidence of the estimated phylogenetic trees has been based principally on two estimators of statistical support; the bootstrap value (BV) for distance, parsimony and ML frameworks and the posterior probability (PP) for BI. The bootstrap approach (Felsenstein, 1985b) consists of the resampling of the sites in the sequence alignment with replacement. This process generates a set of pseudo-samples (bootstrap samples) with the same length as the original alignment. Each bootstrap sample is then analysed just like the original alignment. Finally, the bootstrap support value of each node is the proportion of trees obtained from the bootstrap samples that include that node. Although the statistical base of bootstrap support has been widely discussed (Berry & Gascuel, 1996; Holmes, 2003; Soltis & Soltis, 2003), it continues to be one of the most used support measures in phylogenetic inference. Thus, new algorithms for Felsenstein's method correction (Efron, Halloran, & Holmes, 1996) and the improvement of implementation and parallelization of bootstrap calculation have

been developed (Alexandros Stamatakis, Hoover, & Rougemont, 2008; Hoang, Chernomor, Haeseler, Minh, & Vinh, 2018).

On the other hand, the posterior probability of BI approach indicates the probability that a tree or clade is true, given the data and the model, and provides a natural measure of the reliability of the estimated phylogeny (Rannala & Yang, 1996; Yang & Rannala, 1997).

Generally, both methods are used for phylogenetic inference and the results are compared regarding the topology and the bootstrap and posterior probability values. This interpretation equates bootstrap proportion with Bayesian posterior probability. However, the two measures are incomparable. While bootstrap proportion depends on the topology, the posterior probabilities change drastically with the prior for internal branch lengths, suggesting that bootstrap values and posterior probability represent different measures of phylogenetic uncertainty (Yang & Rannala, 2005).

Both bootstrap support and pp values tend to be maximum on large alignment analyses even when the phylogeny is not correct (Lemoine et al., 2018; Yang & Zhu, 2018). Phylogenomic data sets contain thousands of genes evolving under different models that increase systematic errors. Under these conditions, it seems that Bayesian model selection becomes overconfident about a model if it is slightly less bad, conferring high pp to the phylogenetic tree (Yang & Zhu, 2018).

Multispecies coalescent methods based on the summary strategy relieve this problem. For example, the branch posterior probability, implemented in ASTRAL as a branch support measure, can be more accurate (or equal) than the multi-locus bootstrapping (MLBS) method (Sayyari & Mirarab, 2016). In addition, it is much faster and does not require bootstrapping gene trees (Mirarab, 2019).

Nevertheless, the need for new support measures for phylogenomic analyzes is a fact. The discussion is open (Simon, 2020; Thomson & Brown, 2022) and new approaches are beginning to be used (Allman, Mitchell, & Rhodes, 2021; Arcila et al., 2021; Mount & Brown, 2022).

2.4. Orthology inference

Because the orthologs are derived from speciation events, they reflect the phylogenetic relationships between species (Fig. 2A). Therefore, its identification is a

key step in phylogenomic analyses, and errors in this process can lead to high tree discordance. In addition, orthology inference is an important issue in gene ontology annotation and functional evolution studies (Altenhoff & Dessimoz, 2009).

Ortholog genes can be inferred from genomic or transcriptomic data, previous elimination of contaminant sequences as bacterial, parasites, or gut content. The first step must be to obtain a set of gene prediction sequences, generally protein sequences (Altenhoff & Dessimoz, 2009). Although some methods admit nucleotide input sequences.

Methods for *de novo* orthology inference can be classified into two principal groups; tree-based methods and graph-based methods (see Kristensen, Wolf, Mushegian, & Koonin, 2011 for a broader review). The tree-based methods identify orthologues by aligning homologous sequences and reconstructing a tree to find those that are most plausibly related by speciation rather than by duplication or horizontal gene transfer. Genes that coalesce in a speciation node are orthologs, while the paralogs split at a duplication node. As these methods are based on inferred trees, are computationally expensive. In addition, are influenced by the processes that affect the phylogenetic inference; homoplasy, systematic errors, and gene tree discordance sources. Hence, groups of organisms with complex genomes are more likely to suffer orthology prediction errors (Altenhoff & Dessimoz, 2019; Altenhoff et al., 2019).

On the other hand, the graph methods are based on sequence similarity, assuming that a pair of orthologs are more reciprocally similar between them than with another gene. The method uses the bi-directional best hit (BBH) as a symmetric measure of similarity that is accessed by all-against-all pairwise sequence comparisons during the graph construction phase. In the second phase, the pairs are clustered into groups of orthologs genes. This approach is computationally more efficient than tree-based methods and its implementation is quite simple. OrthoFinder (Emms & Kelly, 2015), OrthoDB (Kriventseva et al., 2019) and OrthoMCL (Li, Stoeckert, & Roos, 2003) are the most expanded graph-based methods for orthologs inference. OrthoFinder is very easy to use and not computationally expensive. The general input is protein sequences in fasta files, but recent releases allow nucleotide input data (Emms & Kelly, 2019). It corrects the previously undetected bias related to gene lengths and offers different options for BLAST algorithms, alignments and building gene and species trees, that improve significantly its accuracy concerning other *de*

novo inference methods (Emms & Kelly, 2015). In addition, the output is very friendly to handle and generates an important set of comparative genomic statistics.

One alternative to *de novo* orthology prediction based also on the graph method is to identify the orthologs in the new data using an orthologs reference set. OrthoDB and OMA (Altenhoff et al., 2019) are the most used databases that span all domains of life. These methods are computationally more efficient than *de novo* inference and less susceptible to incomplete gene sampling or gene loss events (Roth, Gonnet, & Dessimoz, 2008).

A broad suite of orthologs inference methods are available (see Nichio, Marchaukoski, & Raittz, 2017 for a broader review), and its performance can be different depending on the analysed data set. In evolutionary scenarios where gene and genome duplications are frequent, the selection of adequate software is vitally important, and testing different methods is highly recommendable (Fernández et al., 2020). Errors in orthology inference can lead to systematic errors that can mimic expected patterns of gene evolution (see Natsidis, Kapli, Schiffer, & Telford, 2021 for a broader review)

Finally, it must be remarkable that although the use of single-copy orthologs with 100% of the gene completeness matrix is the ideal behaviour in phylogenetic inference, it is not always possible. In genomic studies at high taxonomic levels, the number of single-copy genes found in all taxa decreases as the number of species sampled increases. In such cases, the number of single-copy discovered can be low or null. To solve this problem, recent methods to analyse datasets including paralogs have been developed (Smith & Hahn, 2021).

2.5. Phylotranscriptomics

Although the use of entire genomes for phylogenetic reconstruction has been widely extended, it can be very expensive if a high number of taxa have to be analysed. In addition, the *de novo* assembly of non-model organisms can be very challenging in terms of sequencing quality, assembly strategy, and annotation. Several methods have been developed to sample specific regions, capturing large amounts of information without the need to access the entire genome.

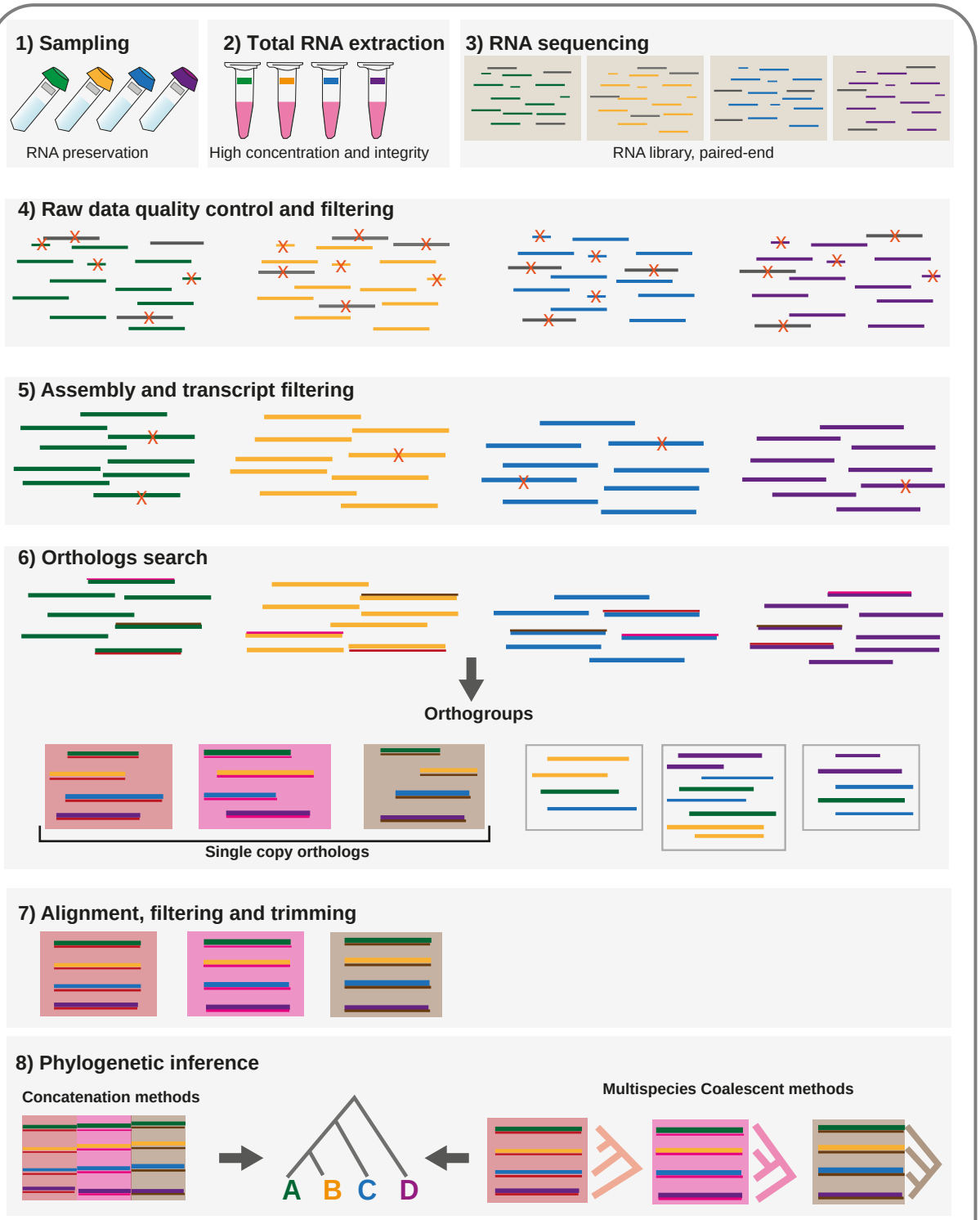


Figure 4. Phylotranscriptomic workflow. 1) The samples must be conserved in conditions that preserve the RNA against degradation. 2) High quality RNA extraction 3) RNA sequencing, paired-end reads are recommended. 4) Raw data quality control and filtering by quality, length and contaminants. 5) Assembly generates a lot of transcripts that must be clustered and filtered. 6) Orthologs prediction. The most recommendable option is to work with single copy orthologs. 7) Obtain multiple sequence alignments by *loci.* The doubtful aligned regions are filtered out and the ends are trimmed. 8) Several methods can be used for phylogenetic inference. Concatenation methods (left) are used with BI and ML frameworks. For large datasets better use mixture models. Multispecies coalescent methods (right) are based on individual gene trees to obtain the species tree and perform better under deep coalescence processes.

The RNA-Seq allows obtaining large data sets of coding genes. Thus, phylotranscriptomics (Fig. 4) is an efficient and cheap alternative for phylogenetic studies and can be as reliable as phylogenomic if an accurate ortholog identification is done (Cheon, Zhang, & Park, 2020). It has been used to resolve ancient evolutionary questions in several taxa (Cunha & Giribet, 2019; Fernández, Edgecombe, & Giribet, 2016; Leebens-Mack et al., 2019; Lemer, Bieler, & Giribet, 2019; Lozano-Fernandez et al., 2019), including Platyhelminthes (Laumer, Hejnol, & Giribet, 2015; Egger et al., 2015). Moreover, transcriptomic data not only allows the inference of phylogenetic relationships but at the same time, combined with genomic data, it can render information on the evolutionary processes acting during speciation or adaptive radiations (Naumenko et al., 2017; Rancilhac et al., 2021; Wickett et al., 2014) as well as on the evolution of genomes (Fernández & Gabaldón, 2020; Guijarro-Clarke, Holland, & Paps, 2020).

In addition, transcriptomic data allow for the detection of genes implicated in metabolic responses under certain conditions. The expression of a transcript/gene can be estimated based on the number of reads mapped against it. Quantification measurements are then integrated into a statistical framework to detect differentially expressed genes (DEGs) between specific conditions (see Costa-Silva, Domingues, & Lopes, 2017 for a broader review). Several tools have been designed for DEG analyses, almost all of them as R packages and very easy to use. The most common are EdgeR (Robinson, McCarthy, & Smyth, 2010) and DEseq2 (Love, Huber, & Anders, 2014), both with high performance (Soneson & Delorenzi, 2013; Zhang et al., 2014). These methods are integrated with Gene Ontology (GO) and enrichment tools to link DGEs with functional annotation analysis (Fradera-Sola et al., 2019; Mahmood et al., 2020; Young, Wakefield, Smyth, & Oshlack, 2010). Although the major number of differential expression studies are carried out under controlled experimental conditions, the application of these techniques to detect genes correlated with the natural adaptation process has spread in recent decades (Akashi, Cádiz Díaz, Shigenobu, Makino, & Kawata, 2016; Balart-García et al., 2021; Li, Zhang, Guan, & Miao, 2013; Santos, Sonoda, Cortez, Coutinho, & Andrade, 2021; Stringlis et al., 2018).

3. The planarians

Planarians are free-living Platyhelminthes belonging to the order Tricladida with representatives in all biogeographical areas of the world (Schockaert et al., 2008). The triclads house approximately 1782 species (Tyler, Schilling, Hooge, & Bush, 2022) whose size varies from a few millimetres to 1m in length in some terrestrial planarians.

The main synapomorphy of the group and which gives its name is the intestine divided into three branches; one simple anterior and two posteriors. They are also characterised by having a muscular pharynx, and several nephridiopores arranged in series. The position of the female gonads closes to the brain and marginal adhesive zones have also been pointed out as distinctive characters, although these show wide variation within the group as a result of secondary changes (Sluys, 1989a).

Planarians are structurally simple acoelomate animals. The sensory system is made up of photoreceptors and chemoreceptors. The nerve network is diffuse and consists of a bilobed brain connected to two ventral nerve cords. Gas exchange occurs through the body wall and nutrients are transported by diffusion. A muscular pharynx opens to the outside through an opening in the medial area of the body. Waste substances are excreted to the outside through nephridiopores connected to two lines of protonephridia, which run dorsoventrally below the epidermis. The planarians are generally hermaphrodites, with only two dioic marine species described; *Sabussowia dioica* Claparède, 1863 and *Cercyra teissieri* Steinmann, 1930 (Charbagi-Barbirou et al., 2011). The female reproductive system consists of two ovaries located near the brain in the anterior region of the body, while the male reproductive system is composed of numerous follicular testis that runs through the entire body. They reproduce sexually by cross-fertilization through sperm exchange using a copulatory organ that opens to the outside below the pharynx. However, some freshwater species can reproduce asexually by fission or by parthenogenesis. The body wall is constituted by circular, longitudinal, and diagonal muscle fibres. A layer of mucus with protective, adhesive and predatory functions is secreted (Ruppert, Fox, & Barnes, 2004). The locomotion is through gliding by the action of ventral cilia and muscle contractions (Stringer, 1917; Talbot & Schötz, 2011).

Despite their structural simplicity, planarians occupy high levels in the trophic webs of their habitats. Their diet is highly varied, consisting mainly of invertebrates,

even preying on other species of planarians (Cuevas-Caballé, Riutort, & Álvarez-Presas, 2019 and references therein). Although some species can become specialised in specific habitats, in general planarians are generalists, which could lead to competition in niches with limited food availability.

Planarians are not tolerant of desiccation. Even the terrestrial ones need highly humid habitats to survive. Therefore, biogeographical hypotheses have been proposed based on their limited dispersion capability (Ball, 1974a).

3.1. Cellular turnover and regeneration

Regenerative capability is spread along most known animal phyla with different potential. Planarians and hydra show the higher regenerative potential, with the capability to regenerate an entire individual from one piece of another one. Fishes and amphibians can regrow sections of the body. While, in mammals only portions of major organs remain regenerative. On the other hand, birds, nematodes, and leeches have lost this ability. Although, the cellular mechanisms of generation also change in different groups; the basic model is common to all and implicates adult stem cell activation and cell proliferation (Li, Yang, & Zhong, 2015; Poss, 2010; Zhao, Qin, & Fu, 2016).

Planarians, specifically some freshwater species, have been used as model organisms for regeneration studies due to their high regenerative capabilities. Since the regenerative capability is highly variable, from whole body regeneration in the Dugesiidae family to the total absence of regeneration in several marine planarians (Ivankovic et al., 2019), almost all studies are limited to the Dugesiidae species models.

Adult individuals show abundant populations of neoblasts, which constitute the 35% of the cellular types (Plass et al., 2018). Neoblasts are pluripotent adult stem cells with the capacity to proliferate and differentiate to maintain a continuous somatic cell turnover (Baguñà, 2012). While in other organisms, the stem cells are restricted to specific lineages and the pluripotency is only shown in the early stages of embryonic development, the neoblasts continuously differentiate into all adult cell types (Pellettieri & Alvarado, 2007).

The differentiated cells do not divide mitotically, and the neoblasts are the only dividing cells in planarians (Morita & Best, 1974; Phillip A Newmark & Sánchez Alvarado, 2000). Thus, the cellular turnover driven by neoblasts differentiation is the only way to maintain tissue homeostasis, cell number and body size (Fig. 5A). This

characteristic allows the regulation of body growth depending on environmental conditions. Under limited food availability, the planarians degrowth by changes in cell number, but they keep a normal metabolic activity. This process is completely reversible, and they can grow again when the food is available (González-Estévez, Felix, Rodríguez-Esteban, & Aziz Aboobaker, 2012; González-Estévez & Saló, 2010). In addition, the response to starvation could also be a mechanism of rejuvenation in planarians, increasing regenerative activity (Felix, Gutiérrez-Gutiérrez, Espada, Thems, & González-Estévez, 2019).

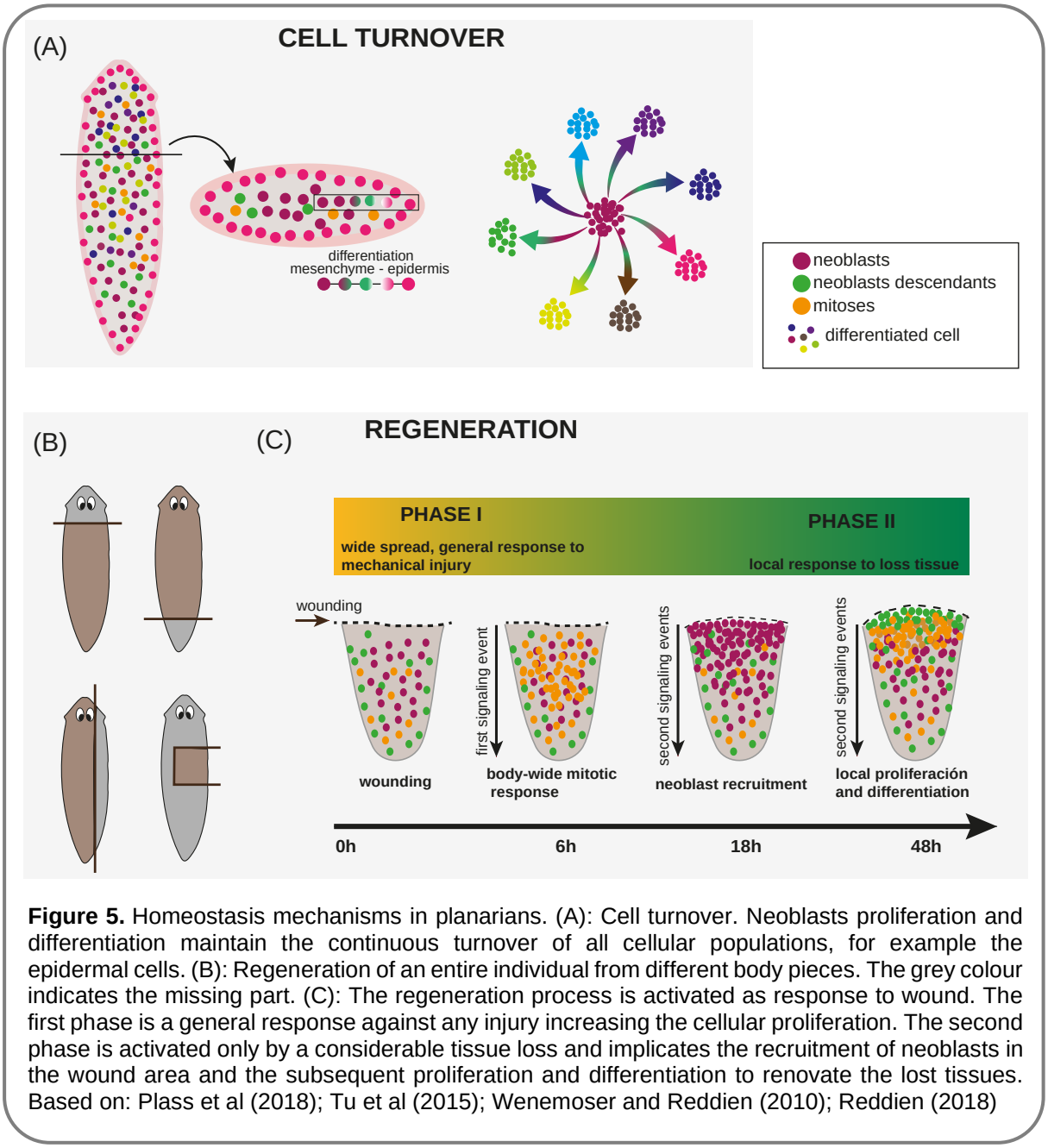


Figure 5. Homeostasis mechanisms in planarians. (A): Cell turnover. Neoblasts proliferation and differentiation maintain the continuous turnover of all cellular populations, for example the epidermal cells. (B): Regeneration of an entire individual from different body pieces. The grey colour indicates the missing part. (C): The regeneration process is activated as response to wound. The first phase is a general response against any injury increasing the cellular proliferation. The second phase is activated only by a considerable tissue loss and implicates the recruitment of neoblasts in the wound area and the subsequent proliferation and differentiation to renovate the lost tissues. Based on: Plass et al (2018); Tu et al (2015); Wenemoser and Reddien (2010); Reddien (2018)

The continuous cell turnover implicates the periodic elimination of selected differentiated cells and their replacement by the differentiated descendant of adult stem cells that become functional (Pellettieri & Alvarado, 2007). The differentiation process implicates a series of cellular changes under strong genetic regulation that guarantees the proper maintenance of the planarian body plan. Thus, it is possible to draw the differentiation path of each specialised cell population based on the transcriptional changes shown by neoblasts and descendants to create an atlas of all cell types present in planarians (Eisenhoffer, Kang, & Alvarado, 2008; Plass et al., 2018; Fincher, Wurtzel, de Hoog, Kravarik, & Reddien, 2018).

In addition to cell turnover, neoblasts allow the regeneration of missing parts or even a new entire individual from a small fragment of the body (Fig. 5B). Neoblast generate new clonal neoblasts (cNeoblasts) with pluripotent capacity, that keep the regenerative capability even after transplantation from one individual to another individual lacking neoblasts (Wagner, Wang, & Reddien, 2011). The neoblast population is not homogeneous. The cNeoblasts differentiate into specialised neoblasts that, although they keep high pluripotent activity, are under strong genetic regulation by distinct transcription factors, which determine their cellular fate (Reddien, 2018 and references therein).

Neoblasts are widely distributed throughout the mesenchyme of the animal, except in the pharynx and the region anterior to the photoreceptors. These are the only areas incapable to regenerate in planarians (Orii, Sakurai, & Watanabe, 2005; Reddien & Alvarado, 2004). The regeneration process implicates two proliferative phases (Fig. 5C). The first is involved in the response to any type of injury. But, the second is only noted as a response to wounds that provoke considerable tissue loss. It is associated with the accumulation of neoblasts in the wound region forming the blastema, from where the missing tissue is regenerated by neoblast proliferation and differentiation (Wenemoser & Reddien, 2010). Under laboratory conditions, the total regeneration of two new individuals, from an individual cut in half, takes approximately 14 days (Saló et al., 2009).

This high regenerative capability of freshwater planarians confers also the capability of asexual reproduction by fission. Using the same machinery for cell turnover and regeneration, asexual lineages reproduce through spontaneous fission

of the individuals in the post-pharyngeal region, and the posterior regeneration of the missing structures in each part, producing two clonal individuals (Saló et al., 2009) .

3.2. Systematics of Tricladida

The Order Tricladida Lang, 1884 is phylogenetically closest to Prolecithophora and Fecampiida (Fig. 6 A). Although this relationship is supported by molecular data (Laumer et al., 2015; Riutort, Álvarez-Presas, Lázaro, Solà, & Paps, 2012), there are no morphological synapomorphies grouping them. Thus, the major knowledge regarding the evolutionary history of Tricladida taxa is based on genetic data.

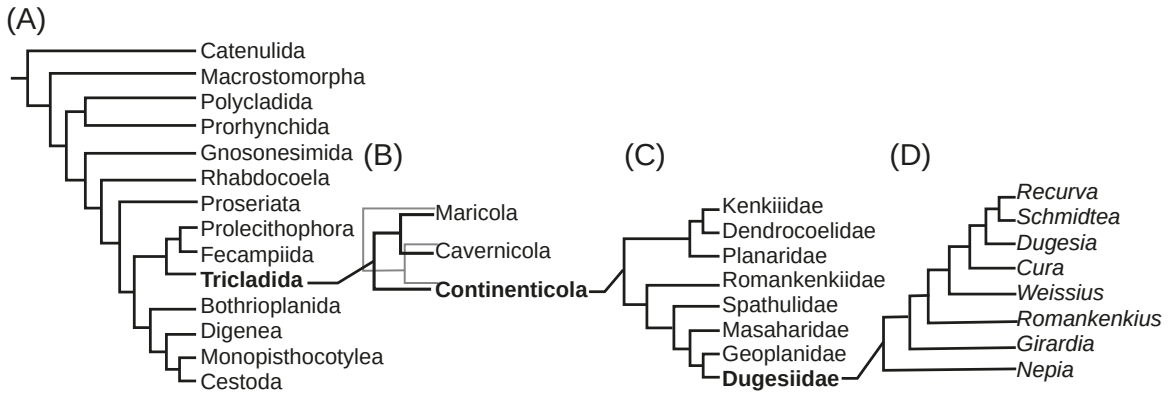


Figure 6. Schematic relationships of Tricladida at different taxonomic levels. A): Relationships for Platyhelminthes orders (Laumer et al., 2015). B): Tricladida relationships, the grey lines represent a different phylogenetic hypothesis based on morphological data (Harrath et al., 2016; Sluys et al 2009). C) Continenticola are formed by two superfamilies, Planriodiea and Geoplanoidea that includes the Dugesiidae family (Riutort et al 2012). D) Hypothesis on the relationships among the main Dugesiidae genera (Alvarez-Presas & Riutort 2014)

Currently, the order is integrated by three suborders; Maricola Hallez, 1892, Cavernicola Sluys, 1990 and Continenticola Carranza et al., 1898 (Fig. 6 B). Maricola species are distributed across all oceans, whose phylogeny is not fully resolved. Its taxonomic classification is based on a morphological phylogeny (Sluys, 1989b; Sluys & Kawakatsu, 2007) and few restricted molecular studies have been carried out (Charbagi-Barbirou et al., 2011; Li et al., 2019; Yang, Sluys, Kawakatsu, & Min, 2018). On the other side, the suborder Cavernicola was created to house five species with discontinuous distribution around the world and unclear relationships among them (Sluys, 1990). Previous phylogenetic hypotheses place Maricola as the sister group of the clade Cavernicola plus Continenticola (Sluys, Kawakatsu, Riutort, & Baguña, 2009).

However, recent molecular studies show the early divergence of Continenticola and the monophyly of Cavernicola and Maricola with high support, but with a low taxon representation (Harrath et al., 2016; Laumer & Giribet, 2014). Therefore, the high-level relationship inside triclads and the evolutionary history of Cavernicola and Maricola suborders remain unclear.

On the other hand, Continenticola has been the most studied clade. This suborder houses two superfamilies; Planarioidea Stimpson, 1857 and Geoplanoidea Stimpson, 1857. The first, groups freshwater planarians from three families; Dendrocoelidae Hallez, 1892, Kenkiidate Hyman, 1937, and Planariidae Stimpson, 1857. Geoplanoidea is constituted by two families; Dugesiidae Ball, 1974, also freshwater, and Geoplanidae Stimpson, 1857; which groups all the known terrestrial planarians (Fig. 6C). The clade Dugesiidae plus Geoplanidae has been defined by a molecular synapomorphy; a gene duplication event that has been demonstrated by the existence of two types of 18S rDNA genes that are highly divergent from each other. The duplication event also affected the rest of the ribosomal genes and the two ITS (Internal Transcribed Spacers) regions (Carranza, Baguña, & Riutort, 1999; Carranza et al., 1998). Thus, molecular data support that Paludicola (all freshwater planarians) constitutes an ecological group but not a natural clade. In addition, the phylogenetic relationships of these groups indicate that the colonisation of freshwater environments has occurred several times and independently within Continenticola (Marta Álvarez-Presas, Baguñà, & Riutort, 2008).

The traditional diagnostic characters used in the identification of triclads species are related to the morphology of their copulatory apparatus (Ball, 1971, 1974b; Sluys, 1989a, 2001; Sluys, Kawakatsu, & Ponce De León, 2005), which implicate a delicate and specialised histological work. However, many triclads reproduce asexually by fission (Stocchino & Manconi, 2013 and references therein), in which case the reproductive system does not develop and identification of fissiparous populations to the species level is very difficult. This aspect constitutes one of the main problems in phylogenetic, biogeographical and ecological studies in Tricladida.

The use of karyological characters has been an alternative to this situation, since the number of chromosomes and their characteristics vary between species and even between populations, and may be correlated with the type of reproduction that they present (D´Souza, Storhas, Schulenburg, Beukeboom, & Michiels, 2004; Leria et

al., 2020). However, even when karyology has given good results in some genera where the karyotypes are species-specific, in others the karyological data are not robust enough for correct identification of asexual populations.

The use of molecular markers offers broad advantages over the absence of decisive diagnostic morphological traits. The first molecular markers used in phylogenetic studies in flatworms were the 18S and 28S ribosomal genes. These genes were used in studies at high levels of taxonomic classification and allowed elucidation of important hypotheses about the evolutionary history of the phylum, its position in the Tree of Life, and internal phylogenies within the group. They have also been used, together with the ITS region, in numerous phylogenetic studies on Tricladida (Baguñà & Riutort, 2004; Riutort et al., 2012 and references therein). However, the use of ribosomal genes is difficult in studies of the Continenticola suborder, because due to the duplication of the ribosomal cluster in this group (Carranza et al 1996), obtaining orthologous sequences of these genes is very difficult.

The mitochondrial gene COI has been widely used in phylogeographic studies in triclads (Álvarez-Presas, Carbayo, Rozas, & Riutort, 2011; Álvarez-Presas et al., 2008; Álvarez-Presas & Riutort, 2014; Dols-Serrate, Leria, Aguilar, Stocchino, & Riutort, 2020; Solà, Sluys, Gritzalis, & Riutort, 2013; Solà, Sluys, Segev, Blaustein, & Riutort, 2015), allowing the identification of asexual populations (Lázaro et al., 2009) and exotic species in European ecosystems (Mazza et al., 2016; Kanana & Riutort, 2019; Álvarez-Presas, Mateos, Tudó, Jones, & Riutort, 2014). In addition, more conserved nuclear markers provide more information and support for phylogenies (Lázaro et al., 2009). The gene Elongation Factor 1 alpha (EF1α) has been used together with COI and ribosomal genes to infer the phylogeny of the subfamily Geoplaninae (Carbayo et al., 2013) and the suborder Continenticola (Álvarez-Presas & Riutort, 2014).

3.2.1. Dugesiidae family

Dugesids are distinguished because the oviducts, together or separately, fall into the bursal canal or very close and posterior to it, in the atrium. The members of the family do not have adhesive organs and the muscles of the pharynx are arranged in two layers (Ball, 1974). Dugesiidae is an ancient family of freshwater triclads (Ball, 1974a; Carranza et al., 1998) and its wide distribution is proof of a complex diversification process. Ball (1974) locates its origin in the Gondwana continental block,

probably in the area that currently corresponds to Antarctica, establishing that the diversification of the family began in the Mesozoic. The biogeographic hypothesis, proposed by Ball, is based on the low dispersal capacity of the group and the current geographic distribution of the genera which, according to Ball, can only be explained by vicariance events (Ball 1974; Ball, 1975). However, another hypothesis based on molecular analyses points to an older origin of the family (Solà et al., *in press*). Based on morphological data, there are 12 described genera of dugesids: *Bopsula* Marcus, 1946; *Cura* Strand, 1942; *Dugesia* Girard, 1850; *Evella* 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 (R Sluys & Riutort, 2018). However, in a molecular study the genera *Romankenkius*, *Spathula*, and *Reynoldsonia* were placed within the land planarians clade, while *Eviella*, and *Weissius* were not included in the analyses (Álvarez-Presas & Riutort, 2014). Therefore, little is known about the phylogenetic relationships within the Dugesiidae (Fig. 6D). However, unpublished data (Grant, 2017) have shown the family Dugesiidae is in need of an urgent taxonomic revision.

Dugesiidae genera are distributed throughout different biogeographic zones of the planet. *Bopsula*, *Romankenkius*, *Girardia*, *Cura* and *Neppia* can be found in America and, except for *Bopsula*, also in Australia together with *Reynoldsonia*, *Weissius*, *Spathula* and *Eviella*. *Spathula* also inhabits in New Zealand along with *Neppia*, which also extends through the Afrotropical region. *Dugesia* genus is spread over Europe, Asia, Africa and Australia, while *Schmidtea* is restricted to Europe and *Recurva* to Greece (Ball, 1974b; De Vries & Sluys, 1991; Kenk, 1974; Sluys, Grant, & Blair, 2007; Sluys & Kawakatsu, 2001; Sluys et al., 2009; Sluys et al., 2013). *Dugesia, Schmidtea* and *Girardia* have been the most studied genera belonging to the Dugesiidae family. A biogeographic hypothesis proposes that during the formation of the current continents, the *Girardia* lineage was restricted to the western landmasses that would later become South America (Ball, 1974) and separated from the ancestor of *Dugesia* and *Schmidtea*. Molecular data have revealed that *Girardia* is sister to the clade that groups *Dugesia* and *Schmidtea* (Álvarez-Presas & Riutort, 2014) (Fig. 6 D).

3.2.2. *Girardia* genus

Giradia was firstly described as a subgenus of *Dugesia* and defined by the following characteristics: a) triangular-shaped head, which can be truncated; b) absent seminal vesicle or non-muscular bifid type; c) absent diaphragm; d) bursal canal formed by internal circular muscles surrounded by longitudinal fibers; e) numerous testes distributed throughout the body and usually ventral (Ball, 1974a) . Later, a review of the phylogeny of *Dugesia* elevated *Girardia* to the genus category based on 17 morphological characters; highlighting the presence of pigmented pharynx as another unique character in *Girardia* (De Vries & Sluys, 1991). However, polymorphic species have been reported in which some individuals may present a depigmented pharynx (Sluys, 2001), or where the degree of pigmentation may vary between populations (Ribas, Riutort, & Baguñà, 1989).

The natural distribution of *Girardia* covers the entire American continent; being the largest representative of Tricladida in terms of number of species in South and Central American regions. It extends until Southern Canada, but with fewer representatives' species (Sluys et al., 2005). Unexpectedly, a new species from China has been described (Chen, Chen, Wu, & Wang, 2015). While, in the case of the "native" Australian populations, the debate remains open pointing out that the Australian species can belong to other lineages (Ball, 1974a; Grant, 2017). The introduction of the genus was reported in Germany in the 1920s (Meinken, 1925), and by the end of the 1960s, it had reached the South of France and the Iberian Peninsula (Saló, Baguñà, & Romero, 1980). It is also present in Australia, Japan, and Hawaii (Sluys et al., 2005; Sluys, Kawakatsu, & Yamamoto, 2010). Although it has always been assumed that *G. tigrina* has been the introduced species in Europe, the report of different morphotypes in the region (Ribas et al., 1989; Stocchino, Sluys, Harrath, Mansour, & Manconi, 2019) makes one suspect the presence of different species in the region.

About 59 *Girardia* species have been described, but despite their wide distribution and diversity, little is known regarding the evolutionary history of the genus. The most complete phylogeny was shown in a cladistic analysis of Dugesiidae family, that included 37 representative species of *Girardia* (Sluys, 2001). However, despite the high number of morphological characters analysed, no resolutive relationships were obtained due to the lack of differential external morphology and the similarities in its internal anatomy. Thus, species identification based on morphological traits is difficult

for no specialist taxonomist. Additionally, reproductive characters are useless for the species assignment of asexual individuals. Therefore, the morphological approach is inefficient for full phylogenetic studies in *Girardia* genus.

Girardia taxa have been used as a representative of Platyhelminthes and Tricladida in molecular phylogenies at high levels of taxonomic classification (Carranza et al., 1998; Ruiz-Trillo, Riutort, Littlewood, Herniou, & Baguña, 1999). However, there is no full information regarding the internal phylogenetic relationships of *Girardia* species or their diversification history, including the expansion out of their natural distribution range.

3.2.3. *Dugesia* genus

Dugesia can be recognized by the presence of a diaphragm between the seminal vesicle and the ejaculatory duct, and because the oviducts emerge from the dorsal surface of the ovaries (Sluys, 2001).

Around 140 species have been described, distributed across Africa, Eurasia and Australasia. Despite the assumed low vagility of planarians, a hypothesis based on molecular analyses point out that the diversification of the *Dugesia* genus has been driven by dispersion, including transoceanic dispersal events, and vicariance. These diversification events have led to the differentiation of seven main clades. A first split separated a Madagascar clade from Africa approximately 160–130 Mya. Followed by the split of a second Madagascar clade and an African clade from the rest. In this remaining group, another African clade is basal to the Asian and European clades (Solà et al., *in press*). A differentiation between Eastern and Western European clades had been envisioned years ago based on molecular data (Lázaro et al., 2009). Currently, a biogeographical hypothesis locates the split of the two sister groups around 30 Mya and proposes that these lineages arrived in Europe from Africa through two different dispersion paths (Solà et al., *in press*).

The Mediterranean region is a recognized biodiversity hot spot and constitutes one of the most sampled regions in evolutionary studies of freshwater planarians. However, while the diversification of *Dugesia* in the Eastern Mediterranean has been fully studied (Solà et al., 2013), the evolutionary history of *Dugesia* in the Western region remains unclear. Molecular studies focused on this group have proposed biogeographical hypotheses to explain the dispersion patterns of *Dugesia* in the region

(Lázaro et al., 2009; Leria, Riutort, Romero, Ferrer, & Vila-Farré, *in press*; Leria, Vila-Farré, Solà, & Riutort, 2019). However, the presence of a high number of asexual populations, with different genetic backgrounds, made it difficult to arrive at supported conclusions. Therefore, a thorough taxon sampling and new molecular approaches are necessary to elucidate the evolutionary history of *Dugesia* in the Western Mediterranean.

3.3. Reproductive strategies of dugesids

Dugesiidae genera are known to present both; sexual and asexual reproduction (Fig. 7) (see Stochino & Manconi, 2013 for a broader review). Sexual planarians are hermaphroditic with a reproductive system composed of testis, ovaries, yolk glands and the copulatory apparatus. All reproductive structures, including the germ cells, are derived from neoblasts. For this reason, and contrary to other animals, germ cells in planarians continue to be generated after embryogenesis and throughout their life span (Issigonis & Newmark, 2019; P. A. Newmark, Wang, & Chong, 2008). During cross-fertilization two individuals interchange sperm that could be stored in the female structures of the partner for months. After fertilisation, several oocytes and multiple yolk cells are enclosed inside a protective shell to be laid under favourable conditions (Benazzi Lentati, 1970; Martín-Durán, Monjo, & Romero, 2012).

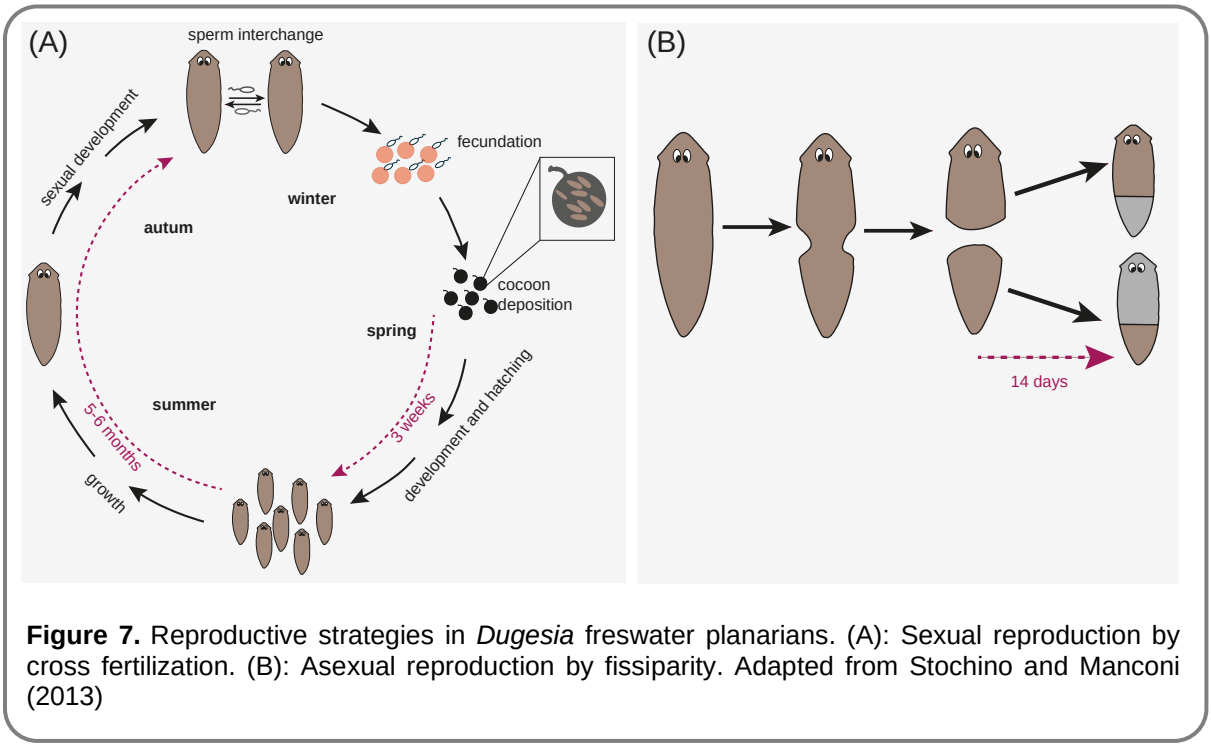


Figure 7. Reproductive strategies in *Dugesia* freshwater planarians. (A): Sexual reproduction by cross fertilization. (B): Asexual reproduction by fissiparity. Adapted from Stochino and Manconi (2013)

On the other hand, asexual reproduction is based on two mechanisms; the high regenerative capability of freshwater planarians which allows the regeneration of two entire organisms from one individual (fission), and parthenogenesis. Parthenogenetic populations have been reported in the *Schmidtea* genus. The parthenogenetic process is sperm-dependent, which means that the zygote division and development is activated by allosperm (D'Souza, Schulte, Schulenburg, & Michiels, 2006; Pongratz, Storhas, Carranza, & Michiels, 2003). However, this type of reproduction is not common in *Dugesia* species.

In *S. mediterranea,* the asexual strain reproduces by fission, and it differentiates from the sexual strain by a chromosomal translocation present in the asexuals (Baguñà et al., 1999; De Vries, Baguñà, & Ball, 1984). However, in *Dugesia* and *Girardia* genera asexuality has appeared in different lineages, with populations strictly sexual or asexual and populations that show both reproductive strategies (Knakievicz, Vieira, Erdtmann, & Bunselmeyer Ferreira, 2006; Puccinelli & Deri, 1991; Stocchino & Manconi, 2013) The genetic differences between sexual and asexual individuals are based principally on chromosome number. Generally, sexual individuals are diploids, but asexual individuals are triploids (Lázaro et al., 2009; Knakievicz, Lau, Prá, & Erdtmann, 2007).

Fissiparous individuals do not have mature reproductive organs, but they become sexual under conditions that stimulate the development of the reproductive system. Triploid asexual individuals belonging to *D. ryukyuensis* species have been sexualized by feeding with sexual planarians. The sexualized individuals developed the complete reproductive system and produced functional gametes. Based on this experiment, it is supposed that a differential meiotic system in germ line cells produces euploid gametes from triploids individuals (Chinone, Nodono, & Matsumoto, 2014). Therefore, this mechanism could allow the sexual reproduction of triploid populations, explaining the shift of reproductive strategies in *Dugesia* species.

3.3.1. Consequences of reproductive strategies

Recombination guarantees the correct segregation of chromosomes during meiosis and the generation of genetic variability due to the formation of recombinant alleles that will be transmitted to the offspring. Further, recombination breaks linked regions allowing natural selection to act more efficiently. Several studies have

demonstrated that the recombination rates increase under strong selection, leading to an increase in adaptation rates. Additionally, the recombination has other implications that affect the heterogeneity in bases composition, substitution rates, and recombination rates across different genomic regions. These effects influence the genome evolution of sexual species, leaving a footprint associated with the recombination processes (see Webster & Hurst, 2012 for a broader review).

Theoretically, the evolutionary advantage of recombination is supported because asexual lineages are rare and incapable of maintaining themselves for long evolutionary times (Engelstädter, 2008; Hartfield, 2016). However, the existence of asexual lineages much more ancient than they should be according to this hypothesis, suggests that there are mechanisms that favour their maintenance over time (Judson & Normark, 1996; Normark, Judson, & Moran, 2003; Brandt et al., 2021).

Nevertheless, the lack of recombination in asexual organisms influences their genetic diversity patterns and genetic selection efficiency. The genetic footprint of asexuality can be different if the asexual reproduction is gametic or agametic. Parthenogenesis or gametic asexual reproduction depends of gamets, while agametic reproduction implicates the development of offspring from a part of the parent's body (De Meeûs, Prugnolle, & Agnew, 2007).

One of the most recognized consequences of no recombination is the irreversible accumulation of deleterious mutations, known as Muller’s ratchet (Felsenstein, 1974; Muller, 1932; Muller, 1964), that theoretically can lead to extinction. However, small amounts of recombination can relieve Muller’s ratchet effect, setting the occasional sex in asexual lineages as a beneficial evolutionary strategy (Hartfield, 2016 and references therein).

During sexual reproduction periods, the genetic interchange between different individuals and the recombination should tend to homogenize the alleles within the population. On the contrary, under long term asexuality, the alleles should accumulate mutations independently of each other, a phenomenon known as the Meselson effect (Fig. 8) (Normark et al., 2003; Schwander, Henry, & Crespi, 2011). This phenomenon predicts that the intraindividual alleles will accumulate high levels of divergence. Therefore, an allele should be more closely related to a homolog allele in another individual than their sister alleles (Birky, 1996; Welch & Meselson, 2000). However, this

consequence of asexuality has not been thoroughly studied, and other processes such as hybridization or gene duplication can lead to a similar genetic pattern. To the present, this effect has been only confirmed in the parthenogenetic *Timema* stick insect (Schwander et al., 2011), the *Trypanosoma* protozoon (Weir et al., 2016) the *Dugesia* fissiparous planarian (Leria et al., 2019), and the oribatid mite *Oppiella nova* (Brandt et al., 2021).

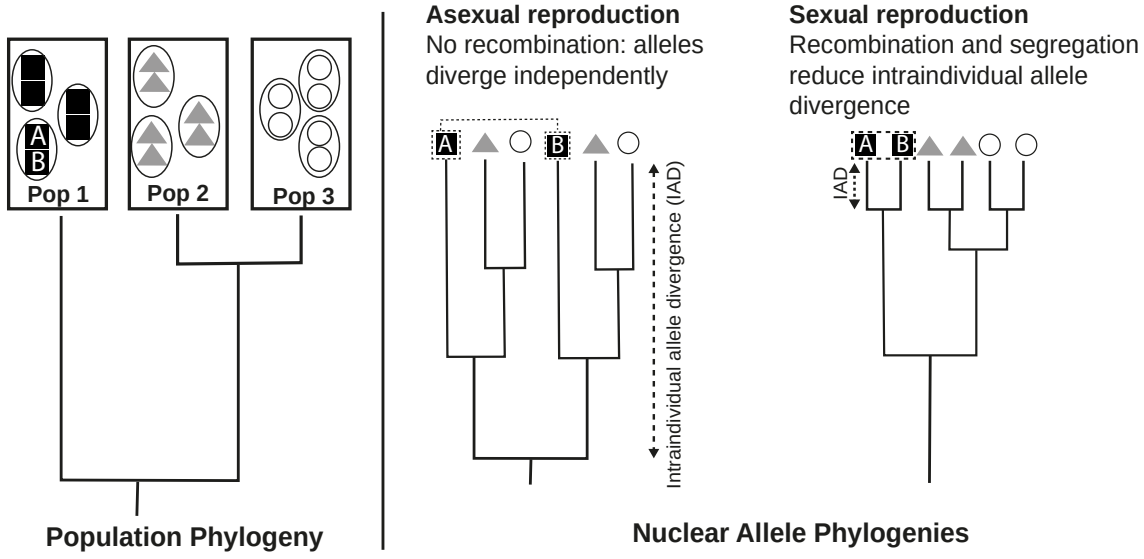


Figure 8. Meselson effect. The intraindividual allele divergence is higher in asexual individuals due to the lack of recombination. Scheme reproduced from Schawander et al., (2011).

In addition, high genetic mosaicism has been reported in clonal plants, associated with their response against pests (Gill, Chao Lin, Perkins, & Wolf, 1995). This genetic mosaicism hypothesis can be applied also to clonal animals (Pineda-Krch & Lehtilä, 2004) and has been proposed as an important adaptive mechanism, especially in coral evolution (Dubé, Planes, Zhou, Berteaux-Lecellier, & Boissin, 2017; Taguchi et al., 2020; Van Oppen, Souter, Howells, Heyward, & Berkelmans, 2011).

3.3.2. Genetic footprint of asexuality in *Dugesia*

The consequences of asexuality in dugesids have been studied at the intraindividual and population-level using *D. subtentacula* and related species as a model (Leria et al., 2019). Using a nuclear marker and COI, Leria and collaborators analysed sexual and asexual populations distributed across the Iberian Peninsula. They detected higher differentiated haplotypes in asexual lineages, consistent with Meselson-effect. Along with this, they found intraindividual mosaicism independently

of the reproductive strategy, although it is more pronounced in the fissiparous individuals as expected. Based on these results the authors define a new process linked to asexuality in planarians; the Mosaic-Meselson-effect. Additionally, no signal of Müller's ratchet was detected in asexual populations because, according to the authors, it is possible that current asexual lineages alternate with sexual periods during their evolution (Leria et al., 2019). This work represents the first attempt to elucidate the effects of asexuality on the evolutionary history of freshwater planarians. Extensive studies analysing other *Dugesia* species with different evolutionary histories are necessary to discern the big picture of the genetic footprint of the asexuality in the evolution of freshwater planarians.

4. Invasive potential of freshwater planarians

In Tricladida several cases of introduced species have been described (Álvarez-Presas et al., 2014; Justine et al., 2015; Justine, Winsor, Gey, Gros, & Thévenot, 2018; Lázaro et al., 2009; Mori et al., 2021; Sluys et al., 2015). For instance, *Schmidtea,* originary from Europe, have some introduced populations of *S. polycroa* in America. On the other hand, *Girardia* is a native genus of America, but it was introduced in Europe at the beginning of the 20th century. Also, within *Dugesia*, the case of *D. sicula*, found all around the Mediterranean basin, has been proposed to be the consequence of human translocations. In the Maricola, it has been described a case of human introduction in the genus *Pentacoelum*. Finaly, many terrestrial planarian species of neotropical origin are being found in multiple countries in Europe most probably as consequence of plant trade.

Biological invasions are an important issue in the conservation of biodiversity. Currently, the introduction of species through anthropogenic actions has become one of the principal causes of the ecological changes in the world (Lowry et al., 2013). Intentional or non-intentional introductions cause important alterations in the evolutionary history of native species through changes in their ecological interactions (see Mooney & Cleland, 2001 for a broader review). In this respect, the freshwater environments are highly sensitive to composition changes and to the introduction of invasive species (Havel, Kovalenko, Thomaz, Amalfitano, & Kats, 2015). For these reasons, monitoring of potentially invader species is vitally important to protect these habitats.

The conceptualization of invasive species as a biological term and its characteristics have been under discussion in the scientific community. Reaching a consensus is complex because the associated terminology is very subjective (Colautti & MacIsaac, 2004). One of the most used and accepted definitions of invasive species is proposed by the IUCN: an alien species is a species introduced outside its natural past or present distribution; if this species becomes problematic, it is termed an invasive alien species (IAS), they may lead to changes in the structure and composition of ecosystems detrimentally affecting ecosystem services, human economy and well-being (IUCN, 2022). This concept makes clear that the status of an alien species as invasive depends on the effects that it causes on the native community and ecosystems. Nevertheless, detractors of this opinion suggest that the invasive term should be used regarding the biogeographic or demographic status of a species without any connotation of impact (Richardson et al., 2000). In addition, other authors propose to focus on the population as an ecological unit, and not the species because the determinant factors of invasion success act at the population level, not species (Colautti & MacIsaac, 2004).

In the case of freshwater planarians there are no specific analyses on their invasiveness and potential detrimental effects. Only few studies have tried to evaluate the effects of introduced *Girardia* on local ecosystems, mainly in the United Kingdom, (Gee, Pickavance, & Young, 1998; Gee & Young, 1993; Pickavance, 1971; Van der Velde, 1975; Wright, 1987). A thorough analysis on the characteristics of the introduced species that make them successful invaders will be of great interest.

In this sense, it is interesting to note that most of the freshwater introduced planarians present fissiparous reproduction. In consequence, one animal accidentally introduced can reproduce by fission generating a clonal line in the new habitat and expand rapidly. Therefore, fissiparity is an important life trait that directly influences the evolutionary history of freshwater planarians. It models the population dynamics and its genetic background. In addition, it favours the dispersion and maintenance of asexual lineages. Therefore, the study of diversification and dispersion patterns of freshwater planarians is not only important to understand the footprint of asexuality in the evolution of this group, but also the biological impact of their expansion outside their natural distribution ranges.

Objectives

The main objective of this thesis is to continue unravelling the evolutionary history of Tricladida, understanding the processes that have shaped the diversification and dispersion of this group across different biogeographical regions and biodiversity hotspots. Using different phylogenetic approaches, depending on the length of genetic data sets, I will infer the evolutionary history of taxonomic groups, which have diversified at different times.

To encounter this general objective, the following specific objectives were proposed:

1. Resolve the internal phylogenetic relationships of Tricladida at the suborder level using ribosomal genes. At the same time, test the monophyly of the Cavernicola suborder, including as many representatives as possible.
2. Obtain the first molecular phylogeny of *Girardia* genus using mitochondrial and nuclear regions as molecular markers. With this phylogeny, test the hypothesis of differentiation between *G. tigrina* from North America and *G. tigrina* from South America. Also Test the hypothesis of multiple human-mediated introductions of *Girardia* out of America and identify the introduced species.
3. Model the potential distribution patterns of *Girardia* introduced species, and their environmental requirements. Analyse the characteristics explaining their invasive success, and its possible impact on the native freshwater fauna. Model their future distribution trends under different climate change scenarios.
4. Prepare and standardize the workflow for sampling, RNA extraction, sequencing, and bioinformatic analyses to carry out comprehensive phylogenetic studies using transcriptomic data for the first time in freshwater planarians.
5. Obtain a supported phylogeny of *Dugesia* genus from the Western Mediterranean using transcriptomic data, including populations never analysed before. Analyse in further detail the internal phylogeny of: *D. subtentaculata* including populations from the Iberian Peninsula and North of Africa; and, *D. etrusca* and *D. liguriensis* species including recently found asexual Iberian populations.

6. Perform a differential expression analysis comparing sexual and asexual individual's conspecific and from different species, and sharing or not the same habitats
7. Analyse the advantages of using large data sets, and the methods associated with them, in evolutionary studies.

Chapter I

Phylogeny and biogeography of the Cavernicola (Platyhelminthes: Tricladida): Relicts of an epigean group sheltering in caves?

LISANDRA BENÍTEZ-ÁLVAREZ, ANA MARIA LEAL-ZANCHET, ALEJANDRO OCEGUERA-FIGUEROA, RODRIGO LOPES FERREIRA, DIEGO DE MEDEIROS BENTO, JOÃO BRACCINI, RONALD SLUYS AND MARTA RIUTORT

Molecular Phylogenetics and Evolution 145 (2020) 106709

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogeny and biogeography of the Cavernicola (Platyhelminthes: Tricladida): Relicts of an epigean group sheltering in caves?

Lisandra Benítez-Álvarez[a], Ana Maria Leal-Zanchet[b], Alejandro Oceguera-Figueroa[c], Rodrigo Lopes Ferreira[d], Diego de Medeiros Bento[e], João Braccini[b], Ronald Sluys[f], Marta Riutort[a,*]

[a] Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Catalonia, Spain
[b] Instituto de Pesquisas de Planárias and Programa de Pós-Graduacão em Biologia, Universidade do Vale do Rio dos Sinos (UNISINOS), 93022-750 São Leopoldo, Rio Grande do Sul, Brazil
[c] Laboratorio de Helmintología, Instituto de Biología, Universidad Nacional Autónoma de México, Tercer Circuito s/n, Ciudad Universitaria, Copilco, Coyoacán, C.P. 04510, A.P. 70–153, Ciudad de México, Mexico
[d] Centro de Estudos em Biologia Subterrânea, Setor de Biodiversidade Subterrânea, Departamento de Biologia, Universidade Federal de Lavras, Campus Universitário, Caixa Postal 3037, Lavras, Minas Gerais, Brazil
[e] ICMBio/CECAV-RN – Centro Nacional de Pesquisa e Conservação de Cavernas, 59015-350 Natal, Rio Grande do Norte, Brazil
[f] Naturalis Biodiversity Center, P.O. Box 9517, 2300 RA Leiden, The Netherlands

ARTICLE INFO

Keywords:
Gondwana
Pangea
Molecular phylogeny
Taxonomy
Troglobitic flatworms

ABSTRACT

The planarian suborder Cavernicola Sluys, 1990 was originally created to house five species of triclad flatworms with special morphological features and a surprisingly discontinuous and broad geographic distribution. These five species could not be accommodated with any degree of certainty in any of the three taxonomic groups existing at that moment, viz., Paludicola Hallez, 1892, Terricola Hallez, 1892, and Maricola Hallez, 1892. The scarce representation of the group and the peculiarities of the morphological features of the species, including several described more recently, have complicated new tests of the monophyly of the Cavernicola, the assessment of its taxonomic status, as well as the resolution of its internal relationships. Here we present the first molecular study including all genera currently known for the group, excepting one. We analysed newly generated 18S and 28S rDNA data for these species, together with a broad representation of other triclad flatworms. The resulting phylogenetic trees supported the monophyly of the Cavernicola, as well as its sister-group relationship to the Maricola. The sister-group relationship to the Maricola and affinities within the Cavernicola falsify the morphology-based phylogeny of the latter that was proposed previously. The relatively high diversity of some cavernicolan genera suggests that the presumed rarity of the group actually may in part be due to a collecting artefact. Ancestral state reconstruction analyses suggest that the ancestral habitat of the group concerned epigean freshwater conditions. Our results point to an evolutionary scenario in which the Cavernicola (a) originated in a freshwater habitat, (b) as the sister clade of the marine triclads, and (c) subsequently radiated and colonized both epigean and hypogean environments. Competition with other planarians, notably members of the Continenticola, or changes in epigean habitat conditions are two possible explanations -still to be tested- for the loss of most epigean diversity of the Cavernicola, which is currently reflected in their highly disjunct distributions.

1. Introduction

Between 1946 and 1983 five species of planarian flatworms (Platyhelminthes, Tricladida) had been described that consistently defied the taxonomic schemes developed by planarian systematists. Four out of these five species (*Opisthobursa mexicana* Benazzi, 1972; *O. josephinae* Benazzi, 1975; *Balliania thetisae* Gourbault, 1978; *Novomitchellia sarawakana* (Kawakatsu & Chapman, 1983)) usually had been assigned to the marine triclads of the Suborder Maricola Hallez, 1892. The fifth species, *Rhodax evelinae* Marcus, 1946, was considered to belong to the freshwater triclads or Paludicola Hallez, 1892. It should be noted that the Suborder Paludicola is no longer valid; its representatives, together with terrestrial planarians –the now obsolete Suborder Terricola Hallez, 1892- are currently classified in the Suborder Continenticola Carranza et al., 1998 (Sluys et al. 2009; Riutort et al., 2012). In all cases, however, some doubt was expressed about the taxonomic assignments of these five species (Ball, 1974; Sluys, 1990). At

* Corresponding author.
E-mail address: mriutort@ub.edu (M. Riutort).

https://doi.org/10.1016/j.ympev.2019.106709
Received 29 July 2019; Received in revised form 5 December 2019; Accepted 11 December 2019
Available online 17 December 2019
1055-7903/

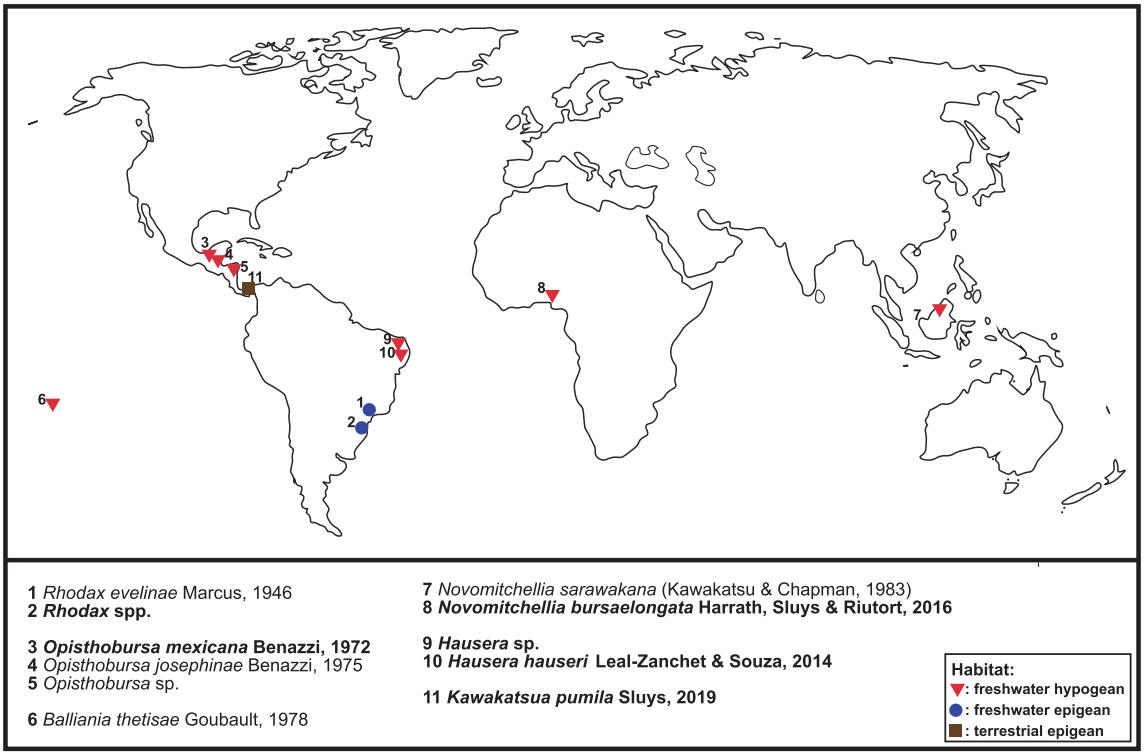


Fig. 1. Distribution of cavernicolan taxa; all known sites are shown. The taxa included in this study are highlighted in boldface.

long last, Sluys (1990) resolved the taxonomic confusion surrounding these five species by showing, on the basis of morphological characters, that they formed a monophyletic group that represented a new and different clade in the phylogenetic tree of the triclad flatworms. At that time three major clades, at the level of suborder or infraorder, were recognized within the Tricladida Lang, 1884, viz., Paludicola, Maricola, and Terricola. For his new, fourth branch on the tree of the planarian flatworms Sluys (1990) erected a new taxon for which he coined the name Cavernicola Sluys, 1990, presently being ranked as a Suborder (Sluys et al., 2009). Although most of its constituent species had a hypogean habitat and exhibited adaptations to life in caves (unpigmented body, lack of eyes), Sluys (1990) stressed the notion that the name of the new taxon had no ecological connotation.

With respect to the phylogenetic position of the new suborder within the Tricladida, Sluys (1990) suggested a possible close relationship between the Cavernicola and the Paludicola, based on the fact that the cavernicolan *Opisthobursa josephinae* exhibits one of the three presumed autapomorphies of the Paludicola, viz., sperm transfer by means of a spermatophore. However, he considered that character distribution as too weak to formally propose presence of a spermatophore as a synapomorphy for the Cavernicola and the Paludicola. Relationships within the Cavernicola were analysed also by Sluys (1990). The fact that the species possess a mixture of primitive features (Marcus, 1946; Sluys, 1990) greatly complicated resolution of their phylogenetic affinities.

After this, it took a long time before the number of species for the Cavernicola started to increase slowly. Two new species and one new genus were described in recent years, viz., *Hausera hauseri* Leal-Zanchet & Souza, 2014 from Brazil, and *Novomitchellia bursaelongata* Harrath, Sluys & Riutort, 2016 from Africa; both species live in a hypogean habitat (Leal-Zanchet et al., 2014; Harrath et al., 2016). In addition, Laumer and Giribet (2014) reported 18S and 28S rRNA sequences for a new, undescribed species of Cavernicola. It was only recently that this new species acquired its proper taxonomic designation when it was described as the new genus and species *Kawakatsua pumila* Sluys, 2019 (Sluys and Laumer, 2019). It is noteworthy that this species was found in a basically terrestrial habitat. Addition of these new species to the Cavernicola made even more evident a conspicuous feature of this small group of species, i.e., their highly disjunct distributions (Fig. 1).

The present study is the first to include molecular data for all cavernicolan taxa, excepting *Balliania* Gourbault, 1978. In our analyses we have incorporated also representatives of 15 genera of triclads belonging to the other two suborders, thus allowing us to test for the first time the previously hypothesized monophyly of the Cavernicola, to analyse its relationships within the Tricladida, as well as the affinities between its constituent taxa.

2. Materials and methods

2.1. Taxon sampling and identification

We obtained samples from six either new or already known localities from South and North America (Southern Mexico), and combined our data with sequences obtained from GenBank, thus including all genera of the Cavernicola presently known, excepting *Balliania* (Table 1, Fig. 1). New specimens of *Opisthobursa mexicana* and *Hausera hauseri* were sampled at the original type localities of these two species, viz., Las Grutas de Coconá, Tabasco, Mexico and Crotes cave, Rio Grande do Norte, Brazil, respectively. In the case of *Rhodax*, the type-locality of *Rhodax evelinae* (the only described species for the genus) no longer exists as it was dramatically transformed due to urbanization, hence representatives of this genus in our study come from other localities. In first instance, we assigned these new representatives to the genus *Rhodax* on the basis of their external features, combined with anatomical and histological features. All new specimens present the following characteristics of the genus *Rhodax*: rounded anterior tip with an adhesive organ; eyes present; pigmented body; short, cylindrical pharynx (Appendix A). Specimens of *Rhodax* spp. 1 and 2, sampled in surface waters located in Tavares and Pinheirinhos, respectively, in southern Brazil, did not have reproductive organs. Specimens of *Rhodax* sp. 2 showed asexual reproduction in the laboratory, similarly to what was described by Marcus (1946) in the original description of the species. Specimens of *Rhodax* sp. 3 from surface water in Tramandaí, southern Brazil, presented a reproductive system, which is characterized by the presence of testes tubes, a large common spermiducal vesicle, and a connection between the

Table 1
Geographic coordinates of the type localities, and habitat as well as stygobiont conditions of cavernicolan taxa.

Taxon	Locality	Geographic coordinates	Habitat	Stygobiont features
Rhodax evelinae Marcus, 1946	São Paulo city, São Paulo, Brazil	–	Lotic and lentic superficial waters	No
Opisthobursa mexicana Benazzi, 1972	Las Grutas de Coconá, Teapa, Tabasco, Mexico	17.616667, −92.966667	Cave	Yes
Opisthobursa josephinae Benazzi, 1975	Pozza Casa Bell, San Cristobal de las Casas, Estado de Chiapas, Mexico	16.716667, −92.666667	Cave	Yes
Ballania thetisae Gourbault, 1978	Maraa, Paea, Tahiti[a]	–	Phreatic layer	Yes
Opisthobursa sp. Kawakatsu & Mitchell, 1983	Grutas de Languín, Alta Verapaz, Guatemala	15.573611, −89.980556	Cave	Yes
Kawakatsua pumila Sluys, 2019	Barro Colorado, Panama	9.15265, −79.85172	Large pile of humid leaf mulch between two buttress roots of an old broadleaf tree	Yes
Hausera hauseri Leal-Zanchet & Souza, 2014	Crotes cave, Felipe Guerra, Rio Grande do Norte, Brazil	−5.592344, −37.686183	Cave	Yes
Novomitchellia sarawakana (Kawakatsu & Chapman, 1983)	Water Polo Cave, Gunung Mulu National Park, Sarawak, Malaysia	4.000000, 114.852778	Cave	Yes
Novomitchellia bursaelongata Harrath, Sluys & Riutort, 2016	Parakou, Benin Republic	9.272972, 2.581861	Waterhole (7.8 m depth and 6.6 m height)	Yes
Hausera sp. 1	Furna Feia cave, Baraúna, Rio Grande do Norte, Brazil	−5.036877, −37.560177	Cave	Yes
Rhodax sp. 1_1	Tavares, Rio Grande do Sul, Brazil	−31.280277, −51.060555	Coastal wetland	No
Rhodax sp. 1_2	Tavares, Rio Grande do Sul, Brazil	−31.317500, −51.122777	Coastal wetland	No
Rhodax sp. 2	Pinheirinhos, Santo Antônio da Patrulha, Rio Grande do Sul, Brazil	−29.71205, −50.638233	Rice field	No
Rhodax sp. 3	Tramandaí, Rio Grande do Sul, Brazil	−30.087777, −50.170833	Coastal wetland	No

[a] Temporary water course close to the Insectarium of the *Institut de Recherches médicales* «Louis Malardé».

copulatory apparatus and the intestine (Leal-Zanchet et al., unpublished results). With respect to their female copulatory organs, these animals show some differences with *R. evelinae*, such as a longer female genital duct, which may be due to intraspecific variation. *Hausera* sp., which was sampled in a cave from northeastern Brazil, is a typical troglobitic animal with an unpigmented body and absence of eyes (Appendix A), similar to *H. hauseri*. *Hausera* sp. also matches other diagnostic features of the genus, such as sperm ducts separately penetrating the penis bulb, the female genital duct communicating with the intestine, ovovitelline ducts without caudal dichotomy, uniting to form a common ovovitelline duct. *Hausera* sp. differs from *H. hauseri* in the shape of the penis papilla and bulbar cavity, the course of the sperm ducts when approaching the penis bulb, and the shape and length of the female genital duct (Hellmann et al., unpublished results).

In order to determine the phylogenetic position of the Cavernicola within Tricladida, we included in our analyses one representative sequence of each genus of the Cavernicola and also sequences of representative taxa of the suborders Maricola and Continenticola. We also included as outgroup species belonging to groups most closely related to the Tricladida, according to previous studies (Laumer et al., 2015; Norén and Jondelius, 2002; Riutort et al., 2012), viz., Fecampiidae and Prolecitophora (Table 2). For determining the relationships within the suborder Cavernicola we used as ingroup all available sequences assigned to this suborder and as outgroup two maricolan taxa. In order to reconstruct ancestral character states related to habitat (epigean / hypogean) and salinity tolerance (freshwater / marine) some recently published sequences of marine triclads that show a tolerance to freshwater were included (Table 2).

2.2. DNA extraction, gene amplification and sequencing

Genomic DNA was extracted from specimens preserved in absolute ethanol by Wizard® Genomic DNA Purification Kit (Promega), according to the manufacturer's instructions. The extraction product was quantified using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA). Genomic DNA was used to amplify a fragment of the nuclear genes 18S rRNA (18S) and 28S rRNA (28S) through a polymerase chain reaction (PCR). For 18S amplification we used the primers 18S1F, 18S4F, 18S5R and 18S9R (Carranza et al., 1996) to amplify two overlapping fragments. For 28S amplification we used the primers 28S1F, 28S2F, 28S4R and 28S6R (Álvarez-Presas et al., 2008). The PCR reactions were performed in a final volume of 25 µl, with final concentrations as follows: $MgCl_2$ 2.5 mM, dNTPs 30 µM, primers 0.4 µM each, 0.75U Go Taq® DNA polymerase enzyme (Promega Madison, Wisconsin, USA) with its corresponding buffer (1X), and approximately 100 ng of template DNA. The amplification program for both fragments of 18S consisted of 30 cycles in the following manner: 30 s at 94 °C, 45 s at 50 °C (AT: annealing temperature), and 1 min at 72 °C, with 2 min for initial denaturation at 95 °C and 4 min for final extension at 72 °C. The program for both fragments of 28S was 35 cycles in the following manner: 45 s at 94 °C, 45 s at 55 °C (AT), 45 s at 72 °C, as well as 1 min of initial denaturation at 94 °C and 3 min of final extension at 72 °C. The PCR products were purified by ultrafiltration in the Merck Millipore MultiScreen System (Darmstadt, Germany). Both chains of purified fragments were sequenced by Macrogen Inc., (Macrogen Europe, Amsterdam). The chromatograms were revised and edited with Geneious v. 10 (https://www.geneious.com) to obtain the final contigs.

2.3. Sequence alignment

Sequences of both genes were independently aligned with MAFFT v7 (Katoh and Standley, 2013) using the web server http://mafft.cbr.jp/alignment/server/ (last visited January 15th, 2019) with the G-INS-i algorithm. The following two principal sets of species were allocated for the phylogenetic analyses. The first (dataset I), was designed to test the monophyly of the Cavernicola, as well as its taxonomic position within the Tricladida. This dataset I included one representative of each cavernicolan genus, as well as one representative per genus for a series of genera belonging to other suborders of the Tricladida and the outgroup

Table 2
Species names, taxonomic classification, and GenBank accession numbers for the Plathyheminthes sequences used in the analyses. Species selected to represent each of the cavernicolan genera are highlighted in boldface.

Species name	Taxon	18S	28S
Protomonotresidae sp.[a]	Rhabditophora: Prolecithophora: Protomonostresidae	KC869820	KC869873
Acanthiella sp.[a]	Rhabditophora: Prolecithophora: Protomonotresidae	KC869786	KC869839
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
Notentera ivanovia[a]	Rhabditophora: Rhabdocoela: Dalyellioida: Fecampiidae	AJ287546.1	AY157167.1
Kronborgia isopodicola[a]	Rhabditophora: Rhabdocoela: Dalyellioida: Fecampiidae	AJ012513.1	AF022862.1
Urastoma cyprinae[a]	Rhabditophora: Mediofusata: Urastomidae	AF065428.2	AY157165.1
Bdelloura candida	Tricladida: Maricola: Bdellouroidea: Bdellouridae	Z99947.1	AJ270167.1
Nerpa fistulata[b]	Tricladida: Maricola: Bdellouroidea: Bdellouridae	MH916614.1	–
Palombiella stephensoni	Tricladida: Maricola: Bdellouroidea: Bdellouridae	DQ666008.2	DQ665988.1
Pentacoelum kasukolinda	Tricladida: Maricola: Bdellouroidea: Bdellouridae	JN009784.1	JN009787.1
Pentacoelum sinensis[b]	Tricladida: Maricola: Bdellouroidea: Bdellouridae	MK140782.1	–
Sluysia triapertura	Tricladida: Maricola: Bdellouroidea: Uteriporidae	MF383119.1	MF383122.1
Paucumara falcata[b]	Tricladida: Maricola; Bdellouroidea; Uteriporidae	MH916612.1	–
Miroplana shenzhensis[b]	Tricladida: Maricola; Bdellouroidea; Uteriporidae	MK140778.1	–
Ectoplana limuli[b]	Tricladida: Maricola; Bdellouroidea; Uteriporidae	D85088.1	–
Obrimoposthia wandeli[b]	Tricladida: Maricola; Bdellouroidea; Uteriporidae	MH108586.1	–
Uteriporus sp	Tricladida: Maricola: Bdellouroidea; Uteriporidae	AF013148.1	–
Cercyra hastata	Tricladida: Maricola: Cercyroidea: Cercyridae	KM200902.1	DQ665962.1
Sabussowia dioica	Tricladida: Maricola: Cercyroidea: Cercyridae	JN009785.1	JN009788.1
Oregoniplana geniculata[b]	Tricladida: Maricola: Cercyroidea: Cercyridae	MH916614.1	–
Procerodes littoralis	Tricladida: Maricola: Procerodoidea: Procerodidae	Z99950.1	DQ665985.1
Hausera hauseri *	**Tricladida: Cavernicola: Dimarcusidae**	MN719501	**MN719494**
Hausera sp. 1*	Tricladida: Cavernicola: Dimarcusidae	MN719502	MN719495
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. 1_1*	Tricladida: Cavernicola: Dimarcusidae	MN719504	MN719497
Rhodax sp. 1_2*	Tricladida: Cavernicola: Dimarcusidae	MN719505	MN719498
Rhodax* sp. 2	**Tricladida: Cavernicola: Dimarcusidae**	MN719506	MN719499
Rhodax sp. 3*	Tricladida: Cavernicola: Dimarcusidae	MN719507	MN719500
Dugesia gonocephala	Tricladida: Continenticola: Geoplanoidea: Dugesiidae	DQ666002.1	DQ665965.1
Girardia sp.	Tricladida: Continenticola: Geoplanoidea: Dugesiidae	AF013156.1	DQ665977.1
Schmidtea polychroa	Tricladida: Continenticola: Geoplanoidea: Dugesiidae	AF013154.1	DQ665993.1
Geoplana quagga	Tricladida: Continenticola: Geoplanoidea: Geoplanidae	KC608497.1	KC608380.1
Cephaloflexa bergi	Tricladida: Continenticola: Geoplanoidea: Geoplanidae	KJ599712.1	KC608355.1
Phagocata vitta	Tricladida: Continenticola: Planarioidea: Planariidae	DQ665998.1	DQ665989.1
Crenobia alpina	Tricladida: Continenticola: Planarioidea: Planariidae	M58345.1	DQ665960.1

* Sequences obtained in this study.
[a] : outgroup sequences.
[b] : Maricola species included only in the Ancestral Reconstruction States analysis.

taxa (Table 2). From dataset I three datasets were generated, viz., two sets for the alignment of the individual genes 18S and 28S (named as dataset I-[18S] and dataset I-[28S], respectively), and another for the concatenation of both genes (named dataset I-[18S + 28S]).

The second major species set, dataset II, comprised all available sequences of cavernicolan representatives, as well as sequences of two maricolan taxa that were used as outgroups (Table 2). This species set was used to infer phylogenetic relationships within the Cavernicola. From dataset II three subsequent datasets were generated, two for the individual genes 18S and 28S (called dataset II-[18S] and dataset II-[28S], respectively), and a third dataset for the concatenation of both genes (called dataset II-[18S + 28S]).

For the ancestral state reconstructions, another major dataset was created on the basis of an 18S alignment, including as ingroup taxa (a) one representative for each cavernicolan genus, and (b) one representative per genus for a series of genera belonging to other suborders of Tricladida; in the following this dataset is abbreviated as dataset III-[18S]. In this dataset some recently published maricolan taxa (Table 2) that live in freshwater or are freshwater-tolerant were included, in order to have a better representation in this group of the feature "salinity tolerance"; representatives of the Fecampiidae were included as outgroup taxa.

Gblocks (Castresana, 2000) was used to remove regions of ambiguous alignment; the parameters used for each alignment are shown in Supplementary Material Table S1. Each alignment was edited by hand to trim the ends and the code N was assigned to sites with missing data. The concatenated alignments were obtained from the alignments of each gene in Mesquite v. 3.04 (Maddison and Maddison, 2008); in these alignments missing sequences were assigned the code N.

We used Xia's method (Xia and Lemey, 2009), implemented in DAMBE6 software (Xia, 2017), to assess the nucleotide substitution saturation. This test is based on the concept of entropy in information theory and calculates an index of substitution saturation (Iss), which is statistically compared to a critical value (Iss.c) that defines a threshold for significant saturation in the data at which the sequences will begin to fail to recover the true tree (Xia and Lemey, 2009). We analyzed each alignment by including all sites and using the proportion of invariant sites previously calculated by the same program.

2.4. Phylogenetic Inference

In order to infer the best sequence evolution model for our datasets we used the jModeltest 2.1 software (Darriba et al., 2012), taking into account the scores of the Bayesian Information Criterion (BIC). The best model for both genes, calculated independently, was GTR + Γ + I (General Time Reversible + Gamma Distribution + Invariable Sites). A gene partition was defined in all the concatenated datasets analyses, so that the estimation of the parameters for each partition was independent. We used two phylogenetic inference methods, viz., Maximum Likelihood (ML) and Bayesian Inference (BI). Both approaches were used to analyze each gene independently, as well as for analyzing the concatenated datasets. ML analyses were performed with RAxML v8.2.4 software (Stamatakis, 2014) under the GTRGAMMA model, taking into account the author's recommendations. A rapid bootstrap analysis with 2000 pseudoreplicates was conducted to obtain bootstrap support values (bs) for the nodes. We ran Bayesian analyses in MrBayes v3.2.2 software (Ronquist et al., 2012) with 5 million generations, sampling frequency every 1000 and 25% burn-in to obtain the consensus tree and posterior probability values (pp). Convergence of the topology and parameter values for the two runs was examined by observing that the average standard deviation of split frequencies was below 0.01. Furthermore, the *.p* file of each run was inspected in Tracer v1.5 software (Rambaut and Drummond, 2007) to ensure that the effective sample size (ESS) values of each parameter were above 200.

2.5. Ancestral states reconstruction

For the Ancestral States Reconstruction analysis (ASR), we obtained a phylogenetic tree with BI using the dataset III-[18S]. Because the Maricola clade shows a good number of species with various degrees of salinity tolerance, we included in this dataset III-[18S] (Table 2) 15 species covering the freshwater tolerance diversity in this group. In the salinity tolerance state reconstruction analysis the state for terrestrial planarians was coded as freshwater, since the animals generally depend on the humidity of forests soils, which usually will be formed by freshwater. This tree was used to independently estimate the ancestral states for habitat (epigean/hypogean) and salinity tolerance (freshwater/freshwater-marine/marine) by using the Phytools package v.0.6.60 of R (Revell, 2012). The posterior probability for each state on the nodes was determined from stochastic character-state maps by applying the empirical Bayes method (Bollback, 2006). For this, we used an all-rates-different (ARD) model and applied the make.simmap function with Markov Chain Monte Carlo (MCMC), and ran 10,000 simulations. In the case of salinity tolerance, which has polymorphic states (freshwater/freshwater-marine/marine), we used the fitpolyMk function integrated with the make.simmap function.

3. Results

3.1. Datasets

The length of the amplified 18S and 28S fragments was approximately 1800 base pairs (bp) and 1500 bp, respectively. For unknown reasons, which may range from problems in the fixation of the specimens, conditions of preservation during transport to intrinsic characteristics of these animals, many of our attempts to obtain good quality DNA for amplification of the genes were unsuccessful. Fortunately, eventually a total of seven new sequences of 18S and of 28S were obtained. After Gblocks processing, the dataset I-[18S] contained 29 sequences with a total length of 1602 sites, while the dataset I-[28S] included 27 sequences with a total length of 1336 sites (Table 2 and Supplementary Material Table S1); these two alignments were concatenated in a dataset I-[18S + 28S] (2939 bp and 29 OTUs) and used to infer phylogenetic relationships within the Tricladida. Another dataset with concatenated alignments (dataset II-[18S + 28S], 3199 positions: 1710 for 18S; 1489 for 28S) with 11 OTUs including only cavernicolan taxa, as well as *Procerodes dohrni* Wilhelmi, 1909 and *Bdelloura candida* (Girard, 1850) as outgroups, was obtained and analyzed to infer the phylogenetic relationships within the Cavernicola. Finally, the dataset III-[18S] included 28 sequences (1709 bp after Gblocks processing; Supplementary material Table S1) and was used for the ancestral character state reconstruction analysis.

Saturation analysis of the alignments for each gene, including outgroups, showed no saturation for our datasets; the proportion of invariant sites was 0.17 and 0.22 for 18S and 28S, respectively.

3.2. Phylogeny

The trees obtained from the three datasets I ([18S], [28S], and [18S + 28S]) by both inference strategies (ML and Bayesian), all group the cavernicolan taxa into a well-supported monophyletic clade that is the sister to Maricola; in turn, Cavernicola + Maricola is the sister-group of Continenticola (Fig. 2, Appendix B). However, the topology of the phylogenetic tree inferred from 28S alone differs in two points from the results obtained with the 18S and concatenated datasets. While the dataset I-[18S] and the dataset I-[18S + 28S] trees (Fig. 2, Appendix B.1) position the Cavernicola as the sister-group of the Maricola with high support (94% bs; 1.00 pp for the 18S tree and 100% bs; 1.00 pp for the concatenated), the dataset I-[28S] tree (Appendix B.2) fails to resolve the relationships between the three triclad suborders, as the Cavernicola groups with Continenticola at a poorly supported node (49% bs; 0.6 pp). Further, in the dataset I-[18S] and the dataset I-[18S + 28S] trees the genus *Opisthobursa* groups with *Novomitchellia* with high support, and together with *Hausera* and *Kawakatsua* they form a monophylum that is highly supported for 18S (87% bs; 1.00 pp) and receives a low or reasonably good support, depending on the method, for the concatenated trees (58% bs; 0.96 pp). In the dataset I-[28S]

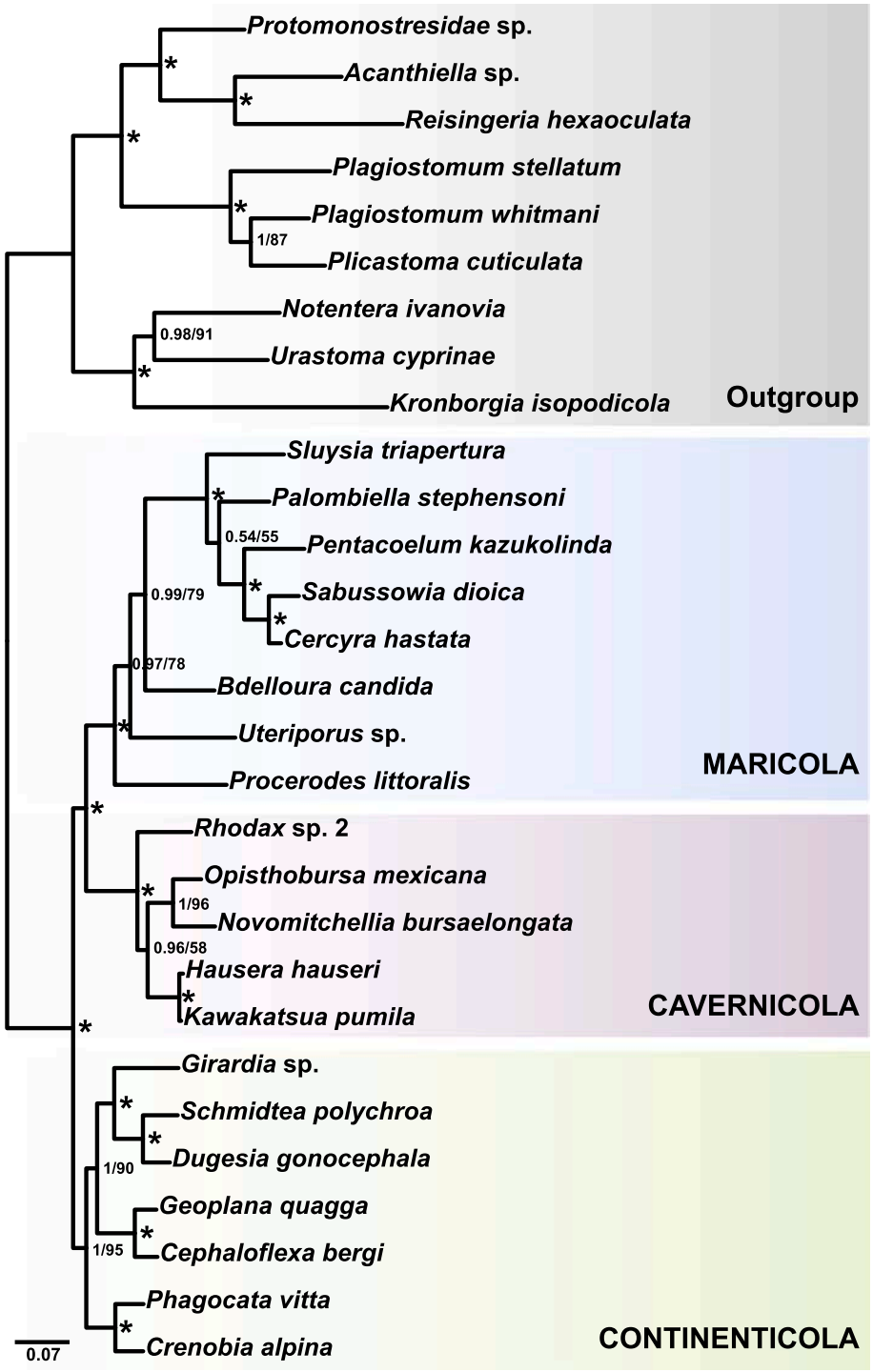


Fig. 2. Bayesian Inference tree inferred from the dataset including 18S and 28S sequences of representatives of the various Suborders of the Tricladida (dataset I-[18S + 28S]). Values at nodes correspond to posterior probability/bootstrap support. *: 1.00 and 100% values for BI and ML, respectively. Scale bar: number of substitutions per nucleotide position.

tree there is no information on *Novomitchellia bursaelongata* (no data in GenBank), while in this phylogeny *Opisthobursa* is the sister genus of *Rhodax* with a reasonably good support (73% bs; 0.96 pp).

The analyses of the three datasets-II ([18S], [28S], and [18S + 28S]) resulted in phylogenetic trees in which *Opisthobursa* and *Novomitchellia* (only *Opisthobursa* in the case of dataset II-[28S]) were positioned as the sister-group of a clade formed by *Hausera* and *Kawakatsua,* with good support in the 18S tree (76% bs; 0.97 pp) and in the concatenated tree (76% bs; 0.99 pp), but with low support in the 28S tree (59% bs; 0.65 pp) (Fig. 3, Appendix C).

In summary, with respect to the phylogenetic position of the Cavernicola within Tricladida, the dataset I-[28S] tree showed a polytomy for the three suborders. In contrast, analyses of the 18S and the concatenated datasets returned a highly supported clade for Cavernicola + Maricola. With respect to relationships within the Cavernicola, analysis of the dataset I-[28S] yielded moderate support for the clade *Rhodax* + *Opisthobursa*, taking into account that *Novomitchelia* is missing from that dataset. The other five trees resulting from analyses of the dataset I and dataset II (Fig. 2, Fig. 3, Appendix B and C) show *Rhodax* as the sister-group of a clade including

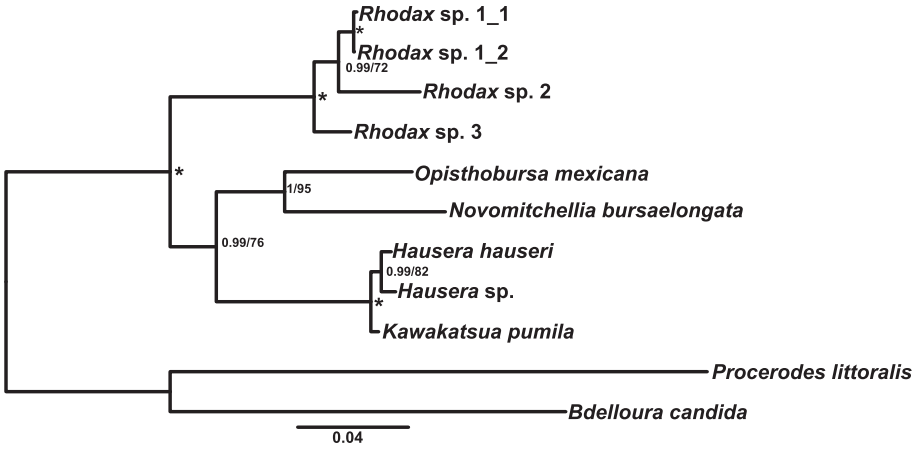


Fig. 3. Relationships within the Cavernicola inferred by Bayesian Inference from the dataset II-[18S + 28S]. Numbers at nodes indicate posterior probability/bootstrap supports for BI and ML, respectively. *: 1.00 and 100% values for BI and ML, respectively. Scale bar indicates number of substitutions per site.

Opisthobursa + *Novomitchelia* and *Hausera* + *Kawakatsua,* generally with high support. These results suggest, in our opinion, that the data at hand strongly support the topology shown in Fig. 2.

3.3. Ancestral habitat

We inferred the ancestral character states for the habitat types (epigean/hypogean) and for salinity tolerance (marine/freshwater) on the nodes of the phylogenetic tree obtained with BI from the dataset III-[18S] (Fig. 4; Appendix D). This dataset was used because it renders the same topology as shown in Fig. 2 and does not contain missing data, so that branch lengths will be more accurate than those in the tree resulting from the concatenated dataset. The hypothesis of an epigean habitat for the ancestor of the clade Maricola + Cavernicola was strongly supported by the ancestral state reconstruction analyses (pp = 0.97; node 10, Fig. 4A, Appendix D.1). In addition, an epigean habitat of the ancestor of the Cavernicola is supported with a high posterior probability value (pp = 0.80, node 11, Fig. 4A, Appendix D.1). Furthermore, a high support value (pp = 0.98; node 11, Fig. 4B, Appendix D.2) suggests that this ancestor lived in a freshwater habitat, while the common ancestor of Maricola + Cavernicola has a nearly equal probability for being either a freshwater animal or exhibiting a tolerance to changes in salinity (pp = 0.59 freshwater; 0.40 freshwater/marine; node 10, Fig. 4B, Appendix D.2).

4. Discussion

4.1. Monophyly of the Cavernicola and its relationship to other suborders of the Tricladida

The phylogenetic trees obtained in the present study corroborate the monophyly of the Cavernicola, as proposed by Sluys (1990). Monophyly of the Cavernicola was proposed on the basis of three apomorphic features: (a) penis bulb with gland cells, (b) horizontal orientation of the bursal canal or female genital duct, combined with the dorsal opening of the common oviduct, or diverticulum, and (c) location of the ovaries at some distance posterior to the brain (Sluys, 1990; Fig. 5). Two of the new species described since Sluys' (1990) monographic study also possess these three features, reinforcing their value as diagnostic characters for the suborder (Leal-Zanchet et al., 2014; Harrath et al., 2016). However, *Kawakatsua pumila* does not possess a penis bulb with gland cells, while *Rhodax* does neither exhibit the character "gland cells in the penis bulb" (character 1 in Sluys 1990; see also Fig. 5). Thus, currently there are two species of cavernicolans, among the eight species known at present, that lack this presumed apomorphic character state of the Cavernicola, one species (*Rhodax*) being positioned as sister to the rest of genera in the phylogenetic tree and the other species (*Kawakatsua*) positioned at one of its tips. Under present conditions absence of this character state in these two taxa is probably best interpreted as being the result of secondary loss.

With respect to the third presumed apomorphic character for the Cavernicola postulated by Sluys (1990), viz., "ovaries situated at some distance posterior to the brain", this character condition is present in *B. thetisae, R. evelinae, O. mexicana, N. bursaelongata,* and *Kawakatsua pumila* (Sluys, 1990; Harrath et al., 2016; Sluys and Laumer 2019), while it is absent in *O. josephinae* and *H. hauseri* (Sluys, 1990; Leal-Zanchet et al., 2014) and ambiguous for *N. sarawakana* (Sluys, 1990). This character state distribution casts some doubt on the presumed synapomorphy for the Cavernicola, as the condition for the cavernicolan ancestor, in view of the topology of our tree (Fig. 4), probably downgrades to being equivocal.

The five currently known cavernicolan genera are housed in a single family, Dimarcusidae Mitchell & Kawakatsu, 1972, which is supported by the fact that these genera together form a monophyletic clade in our analysis (Fig. 2). In addition, its sister-group relationship to the Maricola and the sister-group relationship shared between Dimarcusidae + Maricola and the Continenticola in our phylogenetic trees lends further support to Sluys' (1990) proposal of including all cavernicolan species in a separate suborder. However, in contrast to his hypothesis that the Cavernicola is more closely related to freshwater planarians than to marine triclads, our present phylogeny (Fig. 2) corroborates the conclusions of two previous molecular studies (Laumer and Giribet, 2014; Harrath et al., 2016) that the Cavernicola is most closely related to the Maricola. Our results do not support inclusion of the Dimarcusidae in the suborder Maricola, as suggested by Mitchell and Kawakatsu (1972), but do agree with Sluys' (1990) arguments that the Dimarcusidae does not possess the presumed apomorphous character state of the Maricola, nor the derived features of more restrictive groups of marine triclads.

Our results falsify the presumed close relationship between the Cavernicola and the freshwater planarians. This is in agreement with the fact that cavernicolans lack two out of the three apomorphies hypothesized for the Paludicola (see Sluys, 1989), viz., the probursal condition, and body musculature with an extra outer layer of subepidermal longitudinal fibers. Furthermore, although the third apomorphic feature proposed for the Paludicola, presence of a spermatophore, has been described for *O. josephinae* by Benazzi (1975), it has not been observed in any other cavernicolan species (Sluys 1990; AMLZ, pers. obs.). In point of fact, Sluys (1990, p. 26) himself recognized that "… the data set at hand suggests little else than a sistergroup relationship between the Paludicola and the Dimarcusidae, although this presumed affinity remains poorly supported by apomorphous characters".

In summary, currently we can recognize three suborders within the Tricladida, viz., Cavernicola, Maricola, and Continenticola, which show clear differences in their morphology (Sluys 1990) and are genetically highly differentiated.

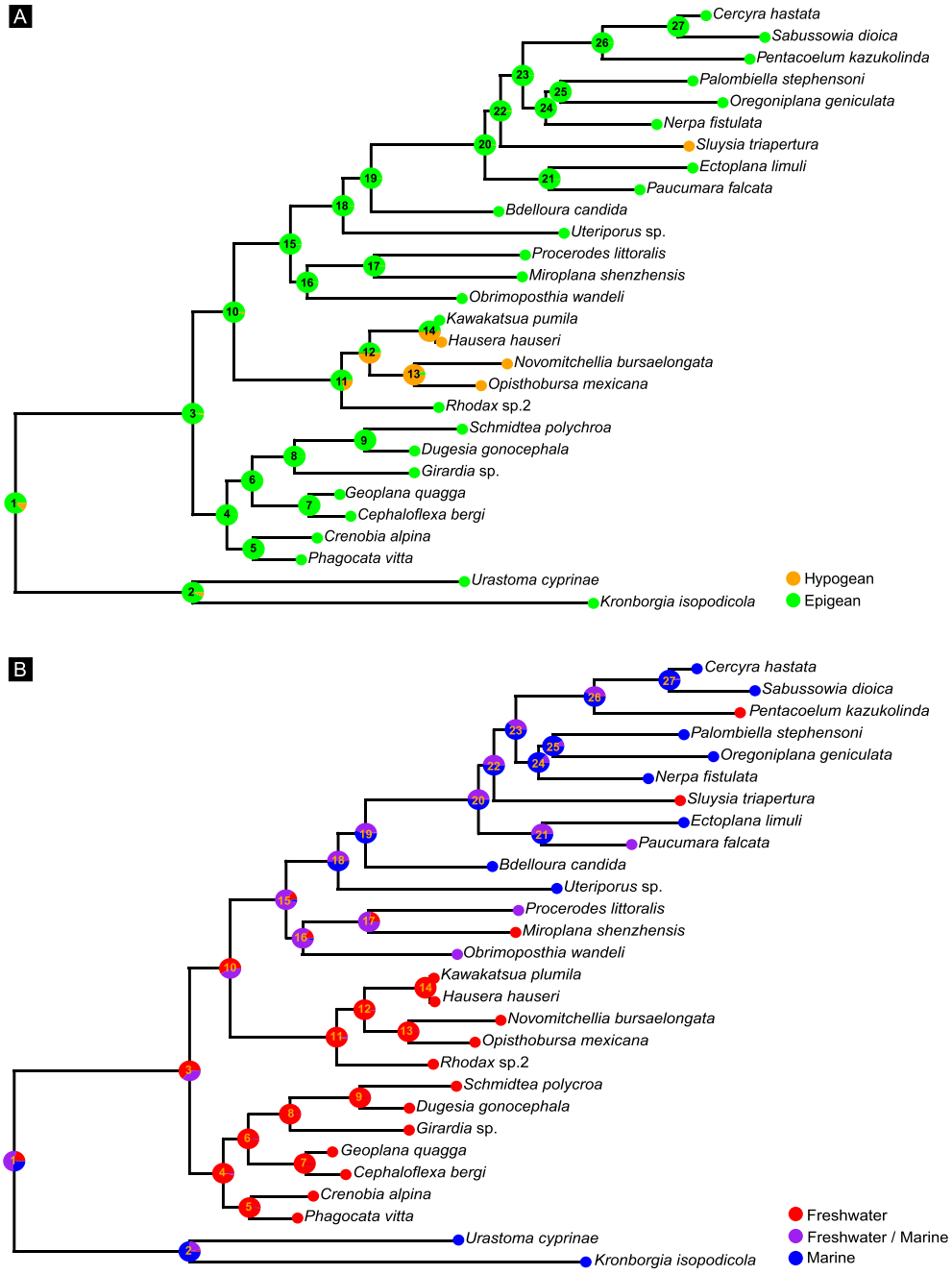


Fig. 4. Results of the Ancestral States Reconstruction analysis based on the Bayesian Inference tree obtained from the dataset III-[18S]. Pie charts at nodes represent the posterior probabilities of ASR analysis for A: epigean (green) and hypogean (yellow) habitat, and B: freshwater (red), freshwater-marine (purple) and marine (blue) salinity tolerance. Terrestrial species were scored for the freshwater condition, as they are only able to survive in a humid habitat. For exact posterior probabilities obtained at each node, see Appendix D. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

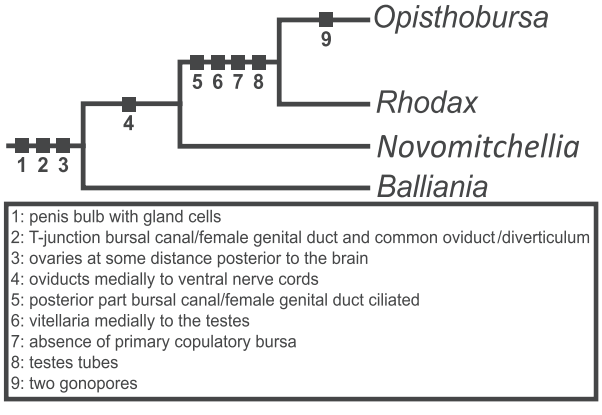


Fig. 5. Morphological phylogeny of Cavernicola based on Sluys (1990). Black rectangles represent morphological apomorphies.

4.2. Relationships and diversity within the Cavernicola

The phylogenetic relationships within the Cavernicola as revealed by molecular data (Figs. 2 and 3) differ from those proposed on the basis of morphological apomorphies (Fig. 5). In the phylogeny presented here, *Rhodax* is sister to all other taxa of the Cavernicola. This implies some changes in our interpretation of the evolution of several morphological features, as will be discussed below.

The clade formed by *Opisthobursa* and *Novomitchellia* is highly supported (Fig. 3), as well as its sister-group relationship to the clade comprising *Hausera* and *Kawakatsua*. This casts some doubt on the four synapomorphies proposed for the sister-group relationship between *Rhodax* and *Opisthobursa* as proposed by Sluys (1990): (a) ciliation being confined to the posterior section of the bursal canal or female genital duct, (b) vitellaria being situated medially to the testes, (c) loss of the primary copulatory bursa, and (d) presence of testes tubes, i.e., fused testicular follicles (characters 5–8 in Fig. 5). In fact, *Opisthobursa*

presents a copulatory bursa, but Sluys (1990) pointed out that it is a secondary bursa and not a primary one. In view of the present phylogenetic results, loss of the primary copulatory bursa in both *Rhodax* and *Opisthobursa* may be interpreted as having evolved independently.

Fused testicular follicles or testis tubes, as present in *R. evelinae* and *O. josephinae* (character 8 in Fig. 5), represent a rare condition among triclad flatworms (see Sluys and Riutort, 2018). In view of the fact that *O. mexicana* has discrete testes follicles (Sluys 1990), presence of testis tubes in *R. evelinae* and *O. josephinae* is presently best interpreted as the result of convergent evolution.

Sluys (1990) hypothesized that oviducts running medially to the ventral nerve cords represent a derived character linking the genera *Rhodax*, *Opisthobursa* and *Novomitchellia*, not shared with *Balliania*, which situated the latter genus as sister to the rest of Cavernicola. Unfortunately, molecular data for *Balliania* is not available and, therefore, we cannot put forward a hypothesis on its position within the Cavernicola. However, for the genus *Hausera* the situation is different, in that the oviducts are exactly dorsal to the nerve cords (Leal-Zanchet et al., 2014); the precise character state for the course of the oviducts in relation to the ventral nerve cords is not known for *Kawakatsua pumila* (Sluys and Laumer, 2019). Under the present phylogeny "oviducts running medially to the ventral nerve cords" may still be postulated as a synapomorphy for all cavernicolans included in our analysis (Fig. 3), under the assumption that at least *Hausera* has evolved another character state in which the oviducts run dorsally to the ventral nerve cords.

The high differentiation found in the tree (Fig. 3) between some of the representatives of *Rhodax* included in our study may point to the presence of more than one species, but this issue may be resolved only by a thorough study that also includes morphological data, which currently is unavailable. Even more surprising is the low genetic differentiation between the genera *Kawakatsua* and *Hausera*, while these are well differentiated at the morphological level (Sluys and Laumer, 2019) and are geographically distant. This situation clearly shows, on the one hand, that a broader sampling most probably will reveal new species for the Cavernicola, and, on the other hand, that within this suborder levels of genetic diversification may vary among different groups.

4.3. Origin and evolution of the Cavernicola

One of the most conspicuous features of this group is its rarity (only a handful of genera, each with merely one or two species, mostly present at a single locality), together with their disjunct distributions (Fig. 1). The fact that currently known species occur at tropical latitudes and that cavernicolans thus far have not been reported from relatively well sampled areas such as Europe and North America suggests that the Cavernicola fauna has a predominantly intertropical distribution. However, it is important to realize that the current distribution, including its disjunctions, may simply reflect a collector's artefact, due to comparatively low investments in research of subterranean habitats in most regions of the world, excepting Europe and North America.

Our phylogenetic trees suggest that the Cavernicola forms an old group. Although no strict time calibration of the entire order Tricladida has been performed, the few molecular timetrees published for *Dugesia* (Solà et al., 2013; Solà, 2014) have situated the diversification of this continenticolan freshwater genus at approximately 150 million years ago (Mya), which implies that the origin of the Continenticola lies even much further back in time. From this perspective, the present phylogenetic tree suggests a great antiquity also for the Cavernicola. This agrees with one of the possibilities for the evolution of the Cavernicola proposed by Sluys (1990), i.e., that the group had differentiated on Gondwana and perhaps already on Pangea. This hypothesis seems plausible given the sister-group relationship between *Opisthobursa* and *Novomitchellia*, genera present in Mexico and Benin, respectively; in turn, this clade is sister to the clade formed by *Hausera* and *Kawakatsua*, from northeastern Brazil and Panama, respectively (Figs. 1 and 3). This suggests that the common ancestor of these four lineages may have lived on Gondwana before this supercontinent started to breakup around 200 Mya (McLoughlin, 2001; Storey, 1995). As Panama and Mexico were not part of Gondwanaland, this scenario presumes that after breakage some descendants of the American lineage dispersed to North America via the Panamanian isthmus when North and South America were eventually connected with each other.

Despite the presumed antiquity of the group, one could argue, alternatively, that it exhibits signs of recent dispersal since both Panama and Tahiti are the result of recent (in the last 5 million years) volcanism, with the latter being a true oceanic island with no connection to any continental landmass. Therefore, an alternative hypothesis might be that there must have been a mechanism for dispersal. But if cavernicolans could disperse, relatively recently, to such habitats, it may similarly have been possible for them to disperse away from Gondwana in ancient times, subsequent to its breakup, or even to Gondwana from elsewhere. One may be tempted to favour such dispersal explanations in view of the situation that an hypothesis about Gondwanan origins of the Cavernicola currently lacks any representation from South Africa, southern South America, Australia, and New Zealand. Generally, much of the evidence for true Gondwanan relicts in other taxa hinges on representation from such areas and even then molecular timetrees may falsify presumed evolution on Gondwana, as was the case with ratite birds (Reilly, 2019 and references therein).

Although calibrated timetrees are as yet not available for the Cavernicola, what is known about the absolute age of triclad flatworms (see above) suggests that the group is ancient. Accepting the antiquity of the group, one may wonder whether the cavernicolans had already evolved their troglobitic adaptations on Pangea or Gondwana. Harrath et al. (2016) put forward two alternative hypotheses for the origin of the Cavernicola. According to their first hypothesis, which was based on the close phylogenetic relationship of the Cavernicola to the Maricola, marine ancestors were forced to invade the underground habitat due to gradual recession of the sea, after which the worms adapted to the hypogean freshwater habitat. A similar scenario was suggested to explain the ecology of *Hausera hauseri* from the karstic Jandaíra formation in northeastern Brazil (Leal-Zanchet et al., 2014). Under their first scenario, Harrath et al. (2016) proposed that epigean *R. evelinae* would have evolved from stygobiont populations and have again acquired the eyes that were lost in its underground ancestors, a possibility that has also been hypothesized for some crustaceans (Humphreys, 2000, and references therein). The second scenario sketched by Harrath et al. (2016) for the evolution of the Cavernicola proposed that an ancestral brackish- and freshwater-tolerant epigean maricolan species led to a brackish water-tolerant *Rhodax* ancestor and to a lineage that invaded the phreatic habitat, possibly to escape presumed competition with continenticolans. In this scenario presence of pigmentation and eyes in *Rhodax* simply reflects retention of the ancestral character states.

Our ancestral states analyses revealed that the character conditions for the ancestor of the Cavernicola are most probably epigean and freshwater (0.80 and 0.98 posterior probability, respectively, Fig. 4, Appendix D), implying diversification of the cavernicolans from worms originally adapted to continental epigean freshwater habitats. In this scenario presence of pigmentation and eyes in *Rhodax* then most probably reflects retention of the ancestral conditions, while for *Kawakatsua pumila* the probability of having retained the ancestral epigean character state is lower (its ancestor with *Hausera* has a 0.51 probability of having been epigean). With respect to salinity tolerance, our analyses show that the ancestor of the Maricola + Cavernicola has a higher probability of being a freshwater animal, or at least being tolerant to freshwater, than that it was a marine species (0.55 and 0.41 vs. 0.03, node 10, Fig. 4A, Appendix D.1); therefore, the Cavernicola may not have had a marine ancestor. This lends support to the second scenario for the evolution of the Cavernicola sketched by Harrath et al. (2016). On the other hand, our results falsify the scenario suggested by Leal-Zanchet et al. (2014) that *Hausera* evolved directly from marine ancestors that entered the caves and then adapted to the freshwater environment.

We hypothesize here that the evolutionary scenario for the Cavernicola, with dispersal of freshwater animals and colonization of

hypogean habitats, resembles cases currently known for the Continenticola, such as representatives of *Girardia* from South America (Souza et al., 2015, 2016; Hellmann et al., 2018), and many species of the Dendrocoelidae in Europe (Stocchino et al., 2017, and references therein) that occur in caves or in surface waters. However, in contrast to *Girardia,* in which epigean species outnumber hypogean taxa, epigean species of the Cavernicola presently represent a minority, as compared with hypogean members of the same suborder (for which at least three more undescribed species occur in the Jandaíra formation; AMLZ, unpublished data), or with epigean species of the Continenticola. This scarcity of cavernicolans in general and that of epigean ones in particular, may be the result of a loss of diversity due to climatic changes or competition with other groups. Specifically, the loss of suitable epigean habitats, and/or competition with continenticolan species, the latter group showing a broad radiation in the same biogeographic regions that also house cavernicolans, could underlie the paucity of epigean cavernicolans. Although these explanations may seem highly speculative with the data at hand, the karstic Jandaíra formation in northeastern Brazil, where *Hausera* species occur (Appendix E.1 and E.2), may be an example of such a loss of suitable epigean habitat. In this area, surface karstification forms recharge zones, favouring the storage and flow of subterranean water (Miranda, 2012), constituting the only water source in most of the region, where epigean streams are scarce (Fernandes et al., 2005). Therefore, no epigean species are expected to be able to survive under these conditions. Thus, under this tentative scenario, caves may have become a refuge for the cavernicolans.

In summary, our results lend support to the hypothesis of a freshwater ancestor of the Cavernicola that colonized continental epigean and phreatic habitats and, subsequently, radiated to form a diverse group with a broad distribution. Under this scenario the evolution of the Cavernicola constitutes a classical example of evolutionary diversification, followed by independent adaptations to hypogean habitats, where caves may have become a refuge habitat for the group for reasons still not fully understood.

Credit authorship contribution statement

Lisandra Benítez-Alvarez: Formal analysis, Writing - original draft, Writing - review & editing. **Ana Maria Leal-Zanchet:** Conceptualization, Funding acquisition , Writing - review & editing. **Alejandro Oceguera-Figueroa:** Funding acquisition, Data curation, Writing - review & editing. **Rodrigo Lop es Ferreira:** Data curation, Writing - review & editing. **Diego de Medeiros Bento:** Data curation, Writing - review & editing. **João Braccini:** Data curation, Writing - review & editing. **Ronald Sluys:** Writing - review & editing. **Marta Riutort:** Conceptualization, Funding acquisition, Writing - original draft, Writing - review & editing.

Acknowledgements

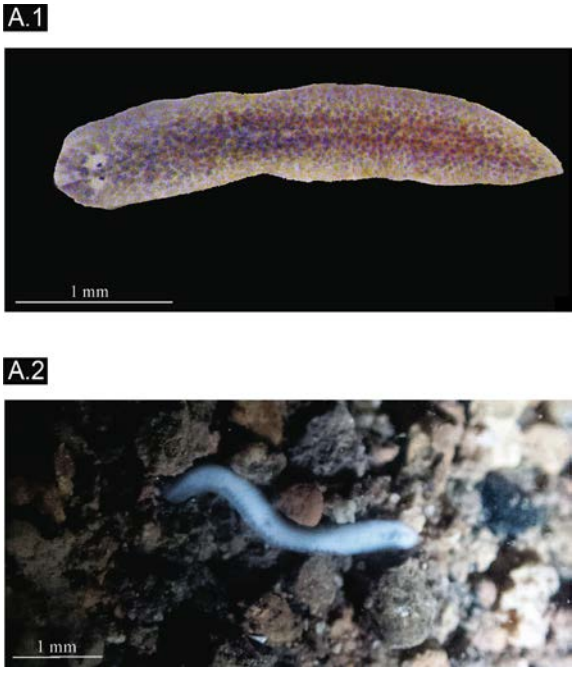
We thank Christopher Laumer for kindly providing information on the locality and some external characteristics of *Kawakatsua pumila* before the species was formally described. We also thank Halim Harrath for making available to us a photograph of the type locality of *Novomitchellia bursaelongata.* We acknowledge Lindsey Hellmann for the Appendix A figure.

Funding

This work was supported by the Brazilian Research Council - CNPq, Brazil [grants 306853/2015-9 and 304682/2014-4]; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil; and Ministerio de Economía y Competitividad, Spain [project CGL2015-63527-1P]. Fieldwork in Mexico was funded by PAPIIT, Mexico [IN210318] and SEP-CONACYT Ciencia Básica, Mexico [220408]. AO-F and MR are grateful for support from the Unión Iberoamericana de Universidades (UIU), Spain [Project BIO02/2017].

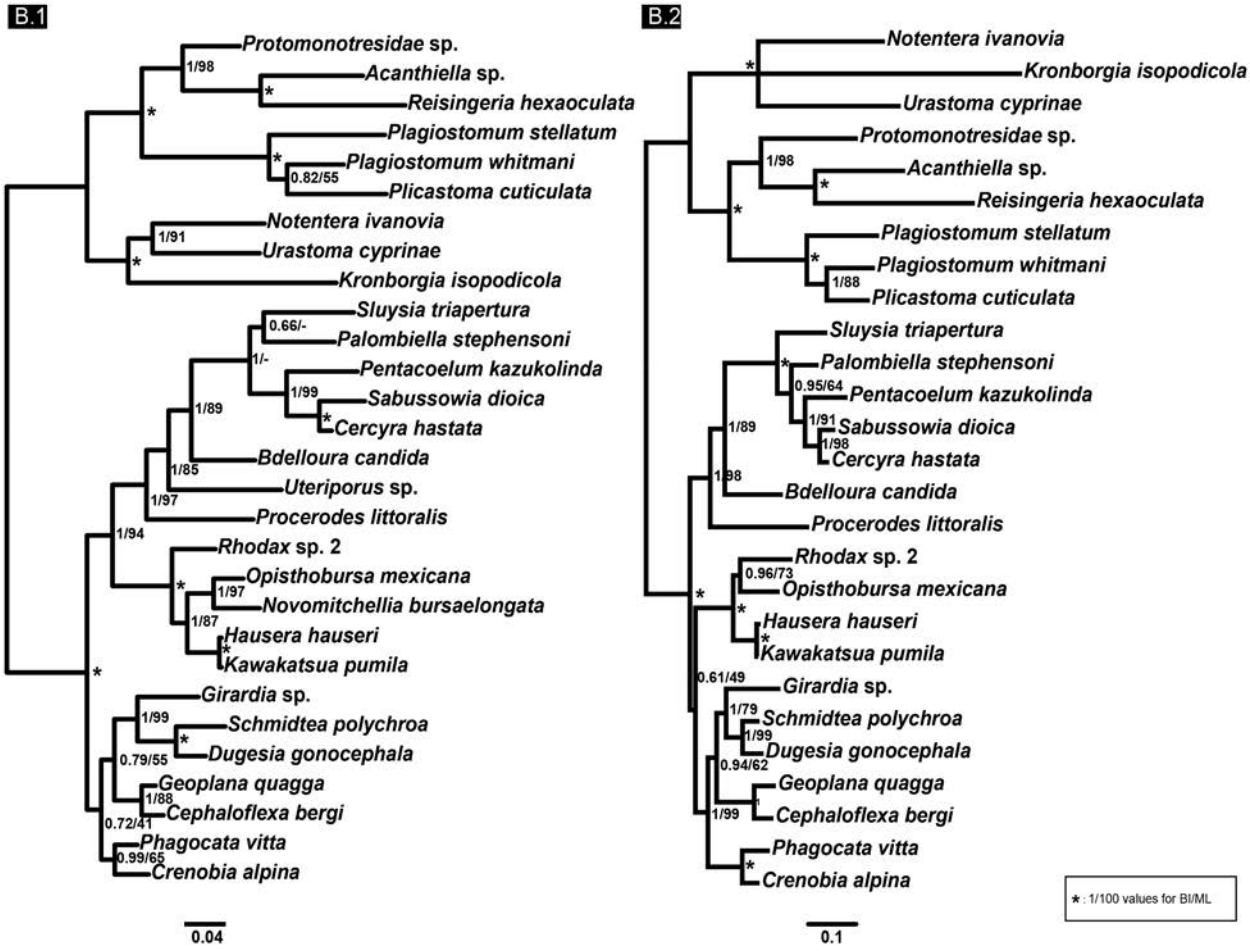
Appendix A

Dorsal view of live specimens of *Rhodax* sp. 3 (A1) from surface water in Tramandaí, southern Brazil, and *Hausera sp.* (A2) from Furna Feia cave, northeastern Brazil. Anterior end to the left.



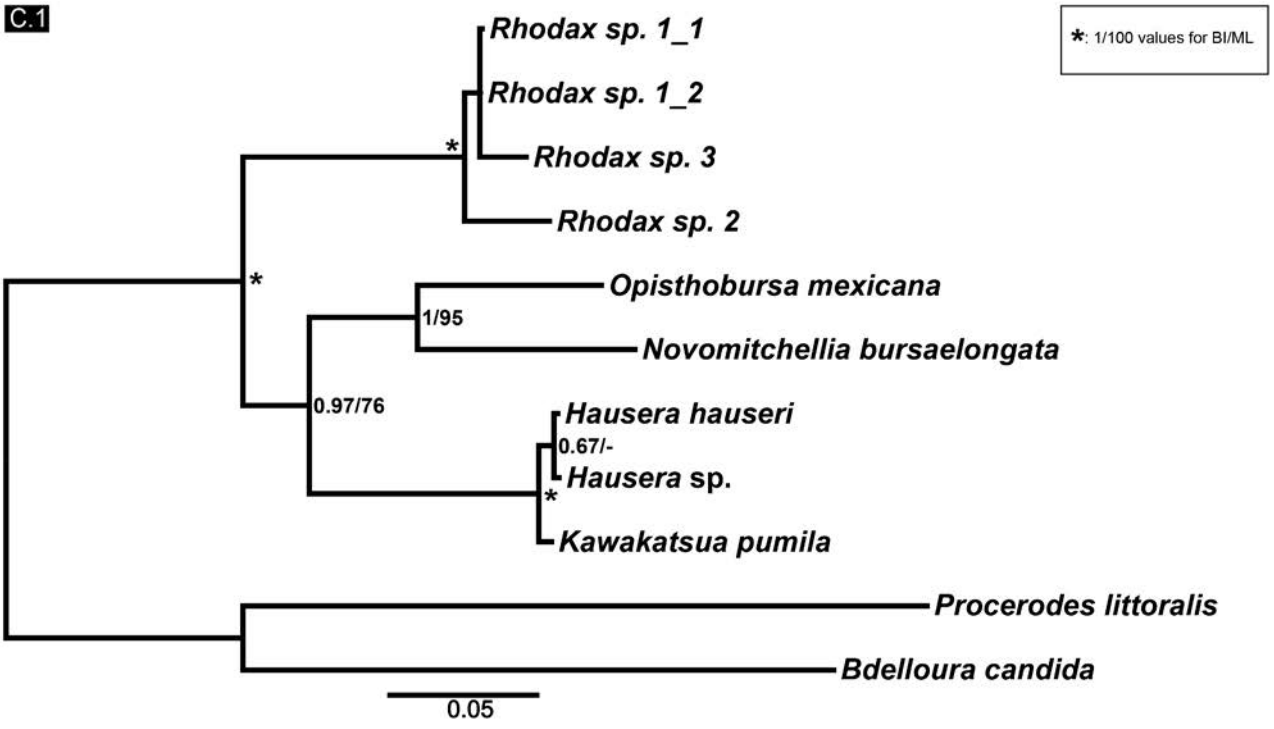
Appendix B

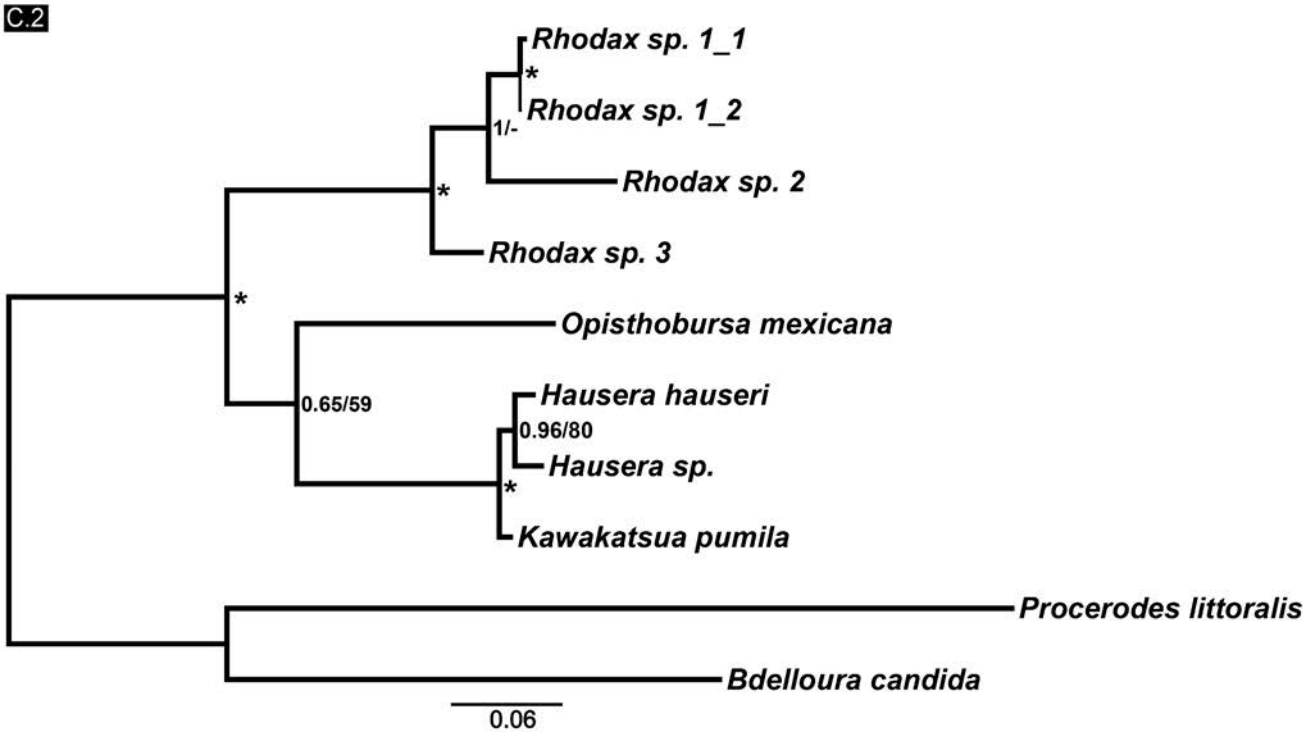
Bayesian Inference tree of Tricladida suborders based on the dataset I-[18S] (B1) and dataset I-[28S] (B2) datasets. Numbers at nodes indicate posterior probability/bootstrap supports for BI and ML, respectively. *: 1.00/100% values for BI/ML. Scale bar indicates number of substitutions per site.



Appendix C

Bayesian Inference tree of Suborder Cavernicola inferred from the dataset II-[18S] (C1) and dataset II-[28S] (C2) data sets. Numbers at nodes indicate posterior probability/bootstrap supports for BI and ML, respectively. *: 1.00/100% values for BI/ML. Scale bar indicates number of substitutions per site.





Appendix D

Posterior probability values at each node in Fig. 4 for the conditions habitat (epigean/hypogean) (D1) and salinity tolerance (freshwater/freshwater-marine/marine) (D2), obtained in the Ancestral States Reconstruction analysis using the make.simmap fuction from the R package Phytools.

D1 Node	Probability	
	Epigean	Hypogean
1	0,88	0,13
2	0,93	0,07
3	0,98	0,02
4	1,00	0,00
5	1,00	0,00
6	1,00	0,00
7	1,00	0,00
8	1,00	0,00
9	1,00	0,00
10	0,97	0,03
11	0,80	0,20
12	0,50	0,50
13	0,05	0,95
14	0,51	0,49
15	0,99	0,01
16	0,99	0,01
17	0,99	0,01
18	0,99	0,01
19	0,99	0,01
20	0,99	0,01
21	1,00	0,00
22	0,99	0,01
23	1,00	0,00
24	1,00	0,00
25	1,00	0,00
26	1,00	0,00
27	1,00	0,00

D2 Node	Probability		
	Freshwater	Freshwater/Marine	Marine
1	0.22	0.52	0.26
2	0.03	0.22	0.75
3	0.64	0.35	0.01
4	0.97	0.03	0.00
5	0.99	0.01	0.00
6	0.99	0.01	0.00
7	1.00	0.00	0.00
8	1.00	0.00	0.00
9	1.00	0.00	0.00
10	0.59	0.40	0.01
11	0.98	0.03	0.00
12	0.99	0.01	0.00
13	1.00	0.00	0.00
14	1.00	0.00	0.00
15	0.15	0.81	0.04
16	0.15	0.82	0.03
17	0.22	0.76	0.01
18	0.02	0.50	0.48
19	0.01	0.45	0.54
20	0.01	0.49	0.51
21	0.00	0.52	0.48
22	0.01	0.48	0.51
23	0.00	0.38	0.62
24	0.00	0.13	0.87
25	0.00	0.09	0.91
26	0.02	0.40	0.58
27	0.00	0.03	0.97

Appendix E

Habitats of cavernicolan taxa included in this study. E1: Crotes cave, northeastern Brazil (type locality of *Hausera hauseri*); E2: Furna Feia cave, northeastern Brazil (*Hausera* sp.); E3: Parakou, Benin (type locality of *Novomitchellia bursaelongata*); E4: Grutas de Coconá cave, México (type locality of *Opisthobursa mexicana*); E5: Mostardas, southern Brazil (*Rhodax* sp.1); E6: Tavares, southern Brazil (*Rhodax* sp.1); E7: Santo Antonio de Patrulha, southern Brazil (*Rhodax* sp.2); E8: Tramandaí, southern Brazil (*Rhodax* sp.3).

Appendix F. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106709.

References

Álvarez-Presas, M., Baguñà, J., Riutort, M., 2008. Molecular phylogeny of land and freshwater planarians (Tricladida, Platyhelminthes): From freshwater to land and back. Mol. Phylogenet. Evol. 47, 555–568.
Ball, I.R., 1974. A contribution to the Phylogeny and Biogeography of the Freshwater Triclads (Platyhelminthes: Turbellaria). In: Riser, N.W., Morse, M.P. (Eds.), Biology of the Turbellaria. McGraw-Hill, NY, New York, pp. 339–401.
Benazzi, M., 1975. Opisthobursa josephinae, a new troglobitic planarian from Chiapas, Mexico. Atti della Accad. Nazi onale dei Lincei LIX, 533–536.
Bollback, J., 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7, 88. https://doi.org/10.1186/1471-2105-7-88.
Carranza, S., Giribet, G., Ribera, C., Riutort, M., 1996. Evidence that two types of 18S rDNA coexist in the genome of *Dugesia* (Schmidtea) *mediterranea* (Platyhelminthes, Turbellaria, Tricladida). Mol. Biol. Evol. 13, 824–832. https://doi.org/10.1093/oxfordjournals.molbev.a025643.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: More models, new heuristics and parallel computing. Nat. Methods 9, 772. https://doi.org/10.1038/nmeth.2109.
Fernandes, M.A.B., Santiago, M.F., Gomes, D.F., Filho, J.M., Frischkorn, H., Lima, J.G., 2005. A origm dos cloretos nas águas subterrâneas na Chapada do apodi-Ceará. Águas subterrâneas 19, 25–34. https://doi.org/10.14295/ras.v19i1.1349.
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. Zool. J. Linn. Soc. 178, 482–491. https://doi.org/10.1111/zoj.12430.
Hellmann, L., Leal-Zanchet, A.M., Lopes Ferreira, R., 2018. Two new sympatric troglobitic freshwater flatworms (platyhelminthes: Dugesiidae) from a hotspot of subterranean biodiversity in the neotropics. Zootaxa 4438 (3), 561–574. https://doi.org/10.11646/zootaxa.4438.3.8.
Humphreys, W.., 2000. Relict fauna and their derivation, in: Wilkens, H., Culver, D., Humphreys, W. (Eds.), Subterranean Ecosystems. Elsevier, Amsterdam, pp. 417–432.
Katoh, K., Standley, D.M., 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.
Laumer, C.E., Giribet, G., 2014. Inclusive taxon sampling suggests a single, stepwise origin of ectolecithality in Platyhelminthes. Biol. J. Linn. Soc. 111, 570–588. https://doi.org/10.1111/bij.12236.
Laumer, C.E., Hejnol, A., Giribet, G., 2015. Nuclear genomic signals of the "microturbellarian" roots of platyhelminth evolutionary innovation. Elife 4, 1–31. https://doi.org/10.7554/eLife. 05503.
Leal-Zanchet, A.M., De Souza, S.T., Ferreira, R.L., 2014. A new genus and species for the first recorded cave-dwelling Cavernicola (Platyhelminthes) from South America. Zookeys 15, 1–15. https://doi.org/10.3897/zookeys.442.8199.
Maddison, W.P., Maddison, D.R., 2008. Mesquite: A modular system for evolutionary analysis. Evolution (N. Y) 62, 1103–1118.
Marcus, E., 1946. Sôbre Turbellaria límnicos brasileiros. Bol. Fac. Filos. Ciênc. Let. Zool. Univ. São Paulo 11, 5–254.
McLoughlin, S., 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. Aust. J. Bot. 49, 271–300. https://doi.org/10.1071/BT00023.
Miranda, T.S. De, 2012. Estudo do comportamento geológico-estrutural dos depósitos Cretácicos do sistema aquífero Apodi, Bacia Potiguar. Ne do Brasil. Estud. Geológicos 22, 3–19.
Mitchell, R.W., Kawakatsu, M., 1972. A new family, genus, and species of cave-adapted planarian from Mexico (Turbellaria, Tricladida, Maricola). Occas. Pap. Museum Texas Tech Univ. 8, 1–16.
Norén, M., Jondelius, U., 2002. The phylogenetic position of the Prolecithophora (Rhabditophora, 'Platyhelminthes'). Zool. Scr. 31, 403–414. https://doi.org/10.1046/j.1463-6409.2002.00082.x.
Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. Disponible en: beast.bio.ed.ac.uk/Tracer.
Reilly, J., 2019. The Ascent of Birds – How Modern Science is Revealing their Story. Pelagic Publishing, Exeter.
Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.
Riutort, M., Álvarez-Presas, M., Lazaro, E.M., Solà, E., Paps, J., 2012. Evolutionary history of the Tricladida and the Platyhelminthes: an up-to-date phylogenetic and systematic account. Int. J. Dev. Biol. 56, 5–17. https://doi.org/10.1387/ijdb.113441mr.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. a, Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–42. https://doi.org/10.1093/sysbio/sys029.
Sluys, R., 1990. A monograph of the Dimarcusidae (Platyhelminthes, Seriata, Tricladida). Zool. Scr. 19, 13–29.
Sluys, R., 1989. Phylogenetic relationships of the triclads (Platyhelminthes, Seriata, Tricladida). Bijdr. Dierk. 59, 3–25.
Sluys, R., Kawakatsu, M., Riutort, M., Baguñà, J., 2009. A new higher classification of planarian flatworms (Platyhelminthes, Tricladida). J. Nat. Hist. 43, 1763–1777. https://doi.org/10.1080/00222930902741669.
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. https://doi.org/10.11646/zootaxa.4586.1.12.
Sluys, R., Riutort, M., 2018. Planarian diversity and phylogeny. In: In: Rink, J.C. (Ed.), Planarian Regeneration: Methods and Protocol. Methods in Molecular Biology, vol. 1774. Springer, New York, pp. 1–56. https://doi.org/10.1007/978-1-4020-6682-5.
Sluys, R., Vila-Farré, M., Álvarez-Presas, M., Riutort, M., Kawakatsu, M., Tulp, A.S., 2015. The diet and distribution of *Pentacoelum kazukolinda* (Kawakatsu & Mitchell, 1984), a maricolan planarian with a freshwater ecology. Zool. Scr. 44, 72–91. https://doi.org/10.1111/zsc.12084.
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). Mol. Phylogenet. Evol. 66, 877–888. https://doi.org/10.1016/j.ympev.2012.11.010.
Solà, E., 2014. Systematics and Historical Biogeography of the Genus Dugesia. Unpublished PhD thesis. Universitat de Barcelona.
Souza, S., Morais, A.L., Bichuette, M.E., Leal-Zanchet, A.M., 2016. Two new species of freshwater flatworms (Platyhelminthes: Tricladida: Continenticola) from South American caves. Zootaxa 4092, 107–121. https://doi.org/10.11646/zootaxa.4092.1.6.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Stocchino, G.A., Sluys, R., Kawakatsu, M., Sarbu, S.M., Manconi, R., 2017. A new species of freshwater flatworm (Platyhelminthes, Tricladida, Dendrocoelidae) inhabiting a chemoautotrophic groundwater ecosystem in Romania. Eur. J. Taxon. 342, 1–21.
Storey, B.C., 1995. The role of mantle plumes in continental breakup: case histories from Gondwanaland. Nature 377, 301–308. https://doi.org/10.1038/377301a0.
Xia, X., 2017. DAMBE6: New tools for microbial genomics, phylogenetics, and molecular evolution. J. Hered. 108, 431–437. https://doi.org/10.1093/jhered/esx033.
Xia, X., Lemey, P., 2009. Assessing substitution saturation with DAMBE. In: Lemey, P., Salemi, M., Vandamme, A.-M. (Eds.), The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny. Cambridge University Press, Cambridge, pp. 615–630. https://doi.org/10.1017/CBO9780511819049.022.

Supplementary Table S1. Data sets characteristics, and parameters used for G-bloks

Alignments	Sequences number	Length in pb	G-blocks parameters					Final lenght
			1	2	3	4	5	
Sp-set I-18S	29	1781	16	16	8	8	with half	1602
Sp-set I-18S	11				8	5	with half	1710
Sp-set I-28S	27	1645	14	14	8	8	with half	1336
Sp-set I-28S	10				8	5	with half	1489
ASR-18S	28	1893	15	15	4	5	with half	694

Chapter II

First molecular phylogeny of the freshwater planarian genus *Girardia* (Platyhelminthes, Tricladida) unveils hidden taxonomic diversity and initiates resolution of its historical biogeography

LISANDRA BENÍTEZ-ÁLVAREZ, RONALD SLUYS, ANA MARIA LEAL-ZANCHET, LAIA LERIA AND MARTA RIUTORT

First molecular phylogeny of the freshwater planarian genus *Girardia* (Platyhelminthes, Tricladida) unveils hidden taxonomic diversity and initiates resolution of its historical biogeography

Lisandra Benítez-Álvarez[1], Ronald Sluys[2], Ana María Leal Zanchet[3], Laia Leria[1], Marta Riutort[1,*]

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

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

[3] Instituto de Pesquisas de Planárias and Programa de Pós-Graduacão em Biologia, Universidade do Vale do Rio dos Sinos (UNISINOS), 93022-750 São Leopoldo, Rio Grande do Sul, Brazil

*corresponding author: mriutort@ub.edu

Under review in: *Zoological Journal of the Linnean Society*

Abstract

The genus *Girardia* (Platyhelminthes: Tricladida) comprises several species of which some have spread from their original areas of distribution in the Americas to other parts of the globe. Due to great anatomical similarities between species, morphology-based phylogenetic analyses struggled to resolve the affinities between species and species-groups. This problem is exacerbated by the fact that populations of *Girardia* may show only asexual reproduction by fissiparity and, thus, do not exhibit a copulatory apparatus, which hampers taxonomic identification and extraction of phylogenetic characters. In the present work this problem has been resolved by constructing a molecular phylogeny of the genus. Although our samples do not include representatives of all known species, they cover a large part of the original distributional range of the genus *Girardia*. Our phylogenetic results suggest the presence of two main clades, which are genetically and karyologically highly differentiated. North and South American nominal *G. tigrina* actually constitute two sibling species that are not even closely related. The South American form is here described as a new species. The phylogenetic tree brings to light that *Girardia* arose on the South American portion of Gondwanaland, from which it, subsequently, dispersed to the Nearctic Region, probably more than once.

Keywords*: Girardia*, evolutionary relationships, historical biogeography, hypogean diversity, introduced species, taxonomy, Tricladida, new species

1. Introduction

The genus *Girardia* comprises about 59 valid species, the natural distribution of which covers the Americas, from Southern Argentina and Chile to Southern Canada, albeit that in North America it is no longer the dominant type of freshwater planarian(Sluys et al. 2005). Furthermore, species of *Girardia* have been introduced into many other regions of the world (see Chapter III.1 and references therein). For Australia, occurrence of introduced *G. tigrina* was established (Sluys et al. 1995 and references therein), apart from three presumed autochthonous species of *Girardia* (cf. Grant et al. 2006; Sluys & Kawakatsu 2001). However, recent molecular work (Grant, 2017) revealed that the latter three species (*G. sphincter* Kawakatsu & Sluys, 2001; *G. graffi* (Weiss, 1909); *G. informis* Sluys & Grant, 2006) do not belong to the genus *Girardia*.

Since the most recent, more comprehensive account on species of *Girardia* from the South American continent and the Caribbean Region by Sluys et al. (2005), 13 new species have been described (Chen et al. 2015; Souza et al. 2016; Souza et al. 2015; Hellmann et al. 2018, 2020; Lenguas-Francavilla et al. 2021; Morais et al. 2021). Phylogenetic analyses of the genus *Girardia* are limited to the study of Sluys (2001), while historical biogeographic studies focusing on the genus are basically absent.

Due to great anatomical similarities between species, morphology-based phylogenetic analyses struggled to resolve the affinities between species and species-groups (cf. Sluys 2001). This problem is exacerbated by the fact that populations of *Girardia* may show only asexual reproduction by fissiparity and, thus, do not exhibit a copulatory apparatus, which hampers taxonomic identification, as well as the extraction of phylogenetic characters. Currently, the use of molecular markers allows overcoming some of the limitations of morphological characters to delimit species and to reconstruct the evolutionary history of triclads. The nuclear gene *Elongation Factor 1 alpha* (*EF1a*) has been used in several phylogenetic and phylogeographic studies in triclads (see Álvarez-Presas & Riutort, 2014), while the mitochondrial gene *Cytochrome Oxidase I* (*COI*) has been used for taxonomic studies as well as for species delimitation in the genus *Dugesia* (Sluys et al. 2013; Solà et al. 2015; Dols-Serrate et al. 2020; Leria et al. 2020). In all of these studies, an integrated approach,

combining molecular and morphological data, proved to be highly successful in furthering our knowledge on the systematics of the various groups.

Here, we present the first molecular phylogeny of the genus *Girardia*, which resulted in several new insights into its taxonomic diversity, particularly in Mexican and South American territories, and into biogeographic history of the genus.

2. Materials and methods

2.1. Taxon sampling

For molecular analyses, samples of *Girardia* were obtained from Asia, Australia, Hawaii, the Americas, and Europe, with greater representation of the two last-mentioned geographical areas. Most of the samples were collected by the authors, while the rest was made available by various colleagues. Individuals were fixed in absolute ethanol. Specimens were identified to species level when both external and internal morphology could be examined (Table S1). When no anatomical information was available, individuals were simply classified as *Girardia* sp. In addition, all available *Girardia* sequences of *Cytochome Oxidase I* (*COI*) and *Elongation factor 1 alpha* (*EF1α*) were downloaded from GenBank. During our analyses some of the latter were excluded because of one or more of the following reasons: (a) low quality or short length of the sequences, (b) uncertain classification of the specimen, (c) avoidance of multiple sequences from a single locality (Table S2).

Figure 1 shows the distribution map of our samples, while an interactive map that allows a better resolution of the information on each locality is available at https://www.ub.edu/planarian-maps/. When available, we used the coordinates of the original samples, otherwise, approximated geographical coordinates were obtained from Google Earth (https://www.google.com/earth/index.html; last visited 12 March 2019) by entering the sampling localities (Table S1). We placed the data points on an open-source map (https://www.openstreetmap.org/) by using a custom script (not available in this publication) of JavaScript (ECMAScript 2015).

2.2. DNA extraction, gene amplification and sequencing

Genomic DNA was extracted using Wizard® Genomic DNA Purification Kit (Promega) and DNAzol® Reagent (Thermo Fisher Scientific, USA), according to the

manufacturer's instructions. The extraction product was quantified using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA).

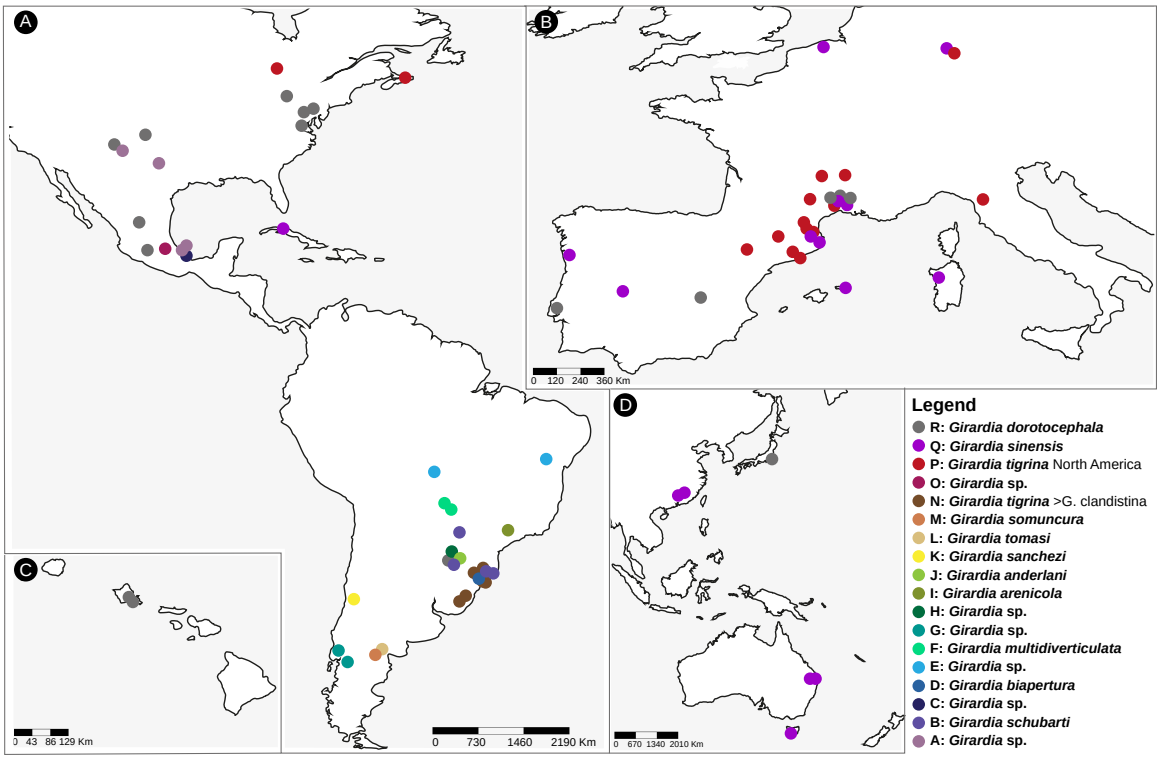


Figure 1. Maps showing the sampling localities for all individuals analyzed in the present study, including those corresponding to GenBank sequences. (A): The Americas; (B): Western Europe; (C): Hawaii; (D): Asia and Oceania. For a finer resolution, visit the interactive map at: https://www.ub.edu/planarian-maps/

A portion of the mitochondrial *COI* and of the nuclear *EF1α* regions were amplified by Polymerase Chain Reaction (PCR), using 100 ng of template DNA and specific primers (Table 1) in 25µl of final reaction volume with $MgCl_2$ (2.5mM), dNTPs (30µM), primers (0.4µM) and 0.75U of Go Taq® DNA polymerase enzyme (Promega Madison, Wisconsin, USA) with its buffer (1x). The amplification program consisted of 2 minutes (m) for initial denaturation at 95°C and 35 cycles of: 50 seconds (sec.) at 94°C, 45 sec. at annealing temperature (Table 1) and 50 sec. at 72°C; with a final extension step of 4 min. at 72°C.

PCR products were run in agarose gels (1%) to check whether the correct band had been amplified. PCR products were purified by ultrafiltration in a Merck Millipore MultiScreen System (Darmstadt, Germany). For those samples that showed a faint PCR band on the electrophoresis, remaining PCR primers and dNTPs were

digested by ExoSAP, a mix of two hydrolytic enzymes (Exonuclease I and Shrimp Alkaline Phosphatase; Thermo Fisher Scientific, USA) in a 3:1 ratio (amplified product: ExoSAP). Both strands of purified fragments were sequenced by Macrogen Inc., (Macrogen Europe, Madrid) with the same primers as used in the amplification. In order to obtain the final contigs, chromatograms were analysed with Genious v.10 (Kearse et al. 2012).

Table 1. Primers used in this study, sequences, references, and annealing temperature (AT).

Primer	Sequence (5´-3´)	Reference	AT
Cytochrome Oxidase I			
BarS (Forward)	GTTATGCCTGTAATGATTG	Alvarez-Presas et al. 2011	43ºC
BarT (Forward)	ATGACDGCSCATGGTTTAATAATGAT	Alvarez-Presas et al. 2011	
COIR (Reverse)	CCWGTYARMCCHCCWAYAGTAAA	Lázaro et al. 2009	
PlatR-Gi (Reverse)	CATCCTGAGGTTTATATWTTGATT	This study	
Elongation Factor 1α			
EF2a (Forward)	GARGCYCARGARATGGGWAAAGGWTC	Barney et al. 2000	54ºC
EF9a (Reverse)	TCNGCRAAYTTGCARGCAATRTGWGC	Barney et al. 2000	
ef1aF (Forward)	ATACGCTTGGGTTTTGG	This study	47ºC
ef1aR (Reverse)	ATGRATTTGACCTGGGTG	This study	
EFGi-2F (Forward)	CCT TCA AAT ACG CTT GGG	This study	51ºC
EFGi-2R (Reverse)	GRATTTGACCTGGRTGATTC	This study	

2.3. Sequence alignment and datasets

Sequences of *COI* and *EF1α*, were aligned independently with ClustalW on the BioEdit Sequence Alignment Editor (Hall 1999). Each gene was translated into amino acids with the corresponding genetic code to check for the absence of stop codons and to produce the alignment, and, thereafter, converted again to nucleotides. Two alignments for each gene were generated, one including only *Girardia* sequences and the other comprising sequences of the closely related outgroup genera *Schmidtea* and *Dugesia* (cf. Álvarez-Presas & Riutort, 2014). We obtained the following four datasets for single gene alignments: (a) *COI* no outgroup (Dataset1), (b) *COI* with outgroup (Dataset2), (c) *EF1α* no outgroup (Dataset3), (d) *EF1α* with outgroup (Dataset4) (Table 2). The number of individuals sequenced for *COI* and *EF1α* differ for the following reasons: (1) *COI* sequences were obtained first from nearly all samples, whereafter the phylogenetic tree was used to select samples for *EF1α* sequencing by

including individuals from different clades, as well as different localities; (2) for some samples, *COI* amplification was impossible and, thus, only *EF1α* was obtained; (3) some species from GenBank have sequences for only one of the two markers (Appendix 1).

A concatenated dataset of both genes without outgroup (Dataset5) (Table 2) was obtained in Mesquite v3.04 (Massidon & Maddison, 2015), including all individuals for which sequences of both genes were available, as well as a few samples lacking one of the sequences. However, the latter sequences had to be included because they concerned the only available representatives for particular clades. Missing data were coded by Ns.

2.4. Phylogenetic Inference

The best sequence evolution model and partition scheme for each gene alignment was estimated independently with PartitionFinder v2.1.1 (Lanfear et al. 2012), thereby considering the score for the Bayesian Information Criterion (BIC). As a preliminary step we hypothesized three partitions, corresponding with the first, second and third codon position for each gene. The results of the PartionFinder program validated this codon partition scheme, both for *COI* and *EF1α*. For each partition the best model was General Time Reversible + Gamma Distribution + Invariable Sites (GTR + Γ + I). This codon partition scheme was then implemented in phylogenetic inference analyses, with the estimations of the parameters for each partition being independent.

Because nucleotide substitution saturation may decrease phylogenetic information contained in the sequences, a saturation test (Xia et al. 2003; Xia and Lemey 2009) was run, using DAMBE (Xia 2017). Third codon positions were analysed alone, while first and second positions were analysed together, including only fully resolved sites. Since the test can only analyse 32 Operational Taxonomic Units (OTUs) at a time, 10,000 replicates of subsets of 4, 8, 16, and 32 OTUs were performed. The proportion of invariant sites was calculated and included in the saturation analysis.

Bayesian Inference (BI) method was applied on the five datasets (Table 2) to infer the best tree and the posterior probabilities (PP), using MrBayes v3.2.2 (Ronquist et al. 2012). The chains were parameterized to 10 million generations, sampling every 1000 generations, and a 25% burn-in (default setting) was applied. Convergence of

parameter values and topologies was examined by checking that the average standard deviation of split frequencies was below 0.01. Estimated sample size values (ESS) of each run were inspected in Tracer v1.5 (Rambaut et al. 2018) to check that the values were over 200.

3. Results

3.1. Sequences and alignments

A total of 124 *Girardia* sequences of *COI* (109 obtained in this study, 15 downloaded from GenBank) and 81 of *EF1α* (of which 80 new) are used in the final analyses, representing localities from all over the range of the genus (Appendix 1, Fig. 1). Several representatives of the genera *Dugesia* and *Schmidtea* are used as outgroups (Appendix 1). Sequences are analysed in individual gene alignments or combined into five datasets (Table 2). The alignments for each gene, including outgroups, show no saturation for any codon position, as determined by the saturation test (in all four tests Iss is significantly lower than Iss.c, thus indicating no saturation).

Table 2. Datasets analysed in this study, with their shorthand description, indication of the phylogenetic trees resulting from the analysis, number of species belonging to either *Dugesia* or *Schmidtea* used as outgroups, number of gene sequences, and the total length of these sequences in nucleotides.

Datasets	Description	Tree figure	Outgroup		Gene sequences		OTUs	Length
			Dugesia	*Schmidtea*	COI	EF1a		
Dataset 1	COI no outgroup	Fig. S2A			124		124	837
Dataset 2	COI with outgroup	Fig. 2	6	3	133		133	840
Dataset 3	EF1a no outgroup	Fig. S2B				81	81	879
Dataset 4	EF1a with outgroup	Fig. S1	4	1		86	86	879
Dataset 5	Concatenated without outgroup	Fig. 3			94	78	98	1716

3.2. Phylogenetic analyses

Phylogenetic trees were obtained from the five datasets (Figs. 2, 3, S1, S2A, S2B). All phylogenetic trees delimit the same major clades and singletons (denoted with letters A to R in the trees), the composition of which does not change between datasets. These clades are fully supported (>0.99 PP) in the concatenated dataset, with the only exception of clade K (0.67 PP support). In the following, we will first describe the composition of the various clades and, where possible, the species assignments, followed by an account on the phylogenetic relationships between the clades.

3.2.1. Species assignment of the terminals and clades

Sequences of individuals of *G. schubarti* (Marcus, 1946) from GenBank and from two localities in southern Brazil constitute a monophyletic group, together with some unidentified specimens from two other localities in southern Brazil (clade B; Fig. 2). This clade was highly differentiated from the rest of the OTUs in the tree, suggesting that it comprises, most likely, a single species, viz., *G. schubarti*. However, given its high diversity, this clade may actually correspond to a complex of species closely related to *G. schubarti*.

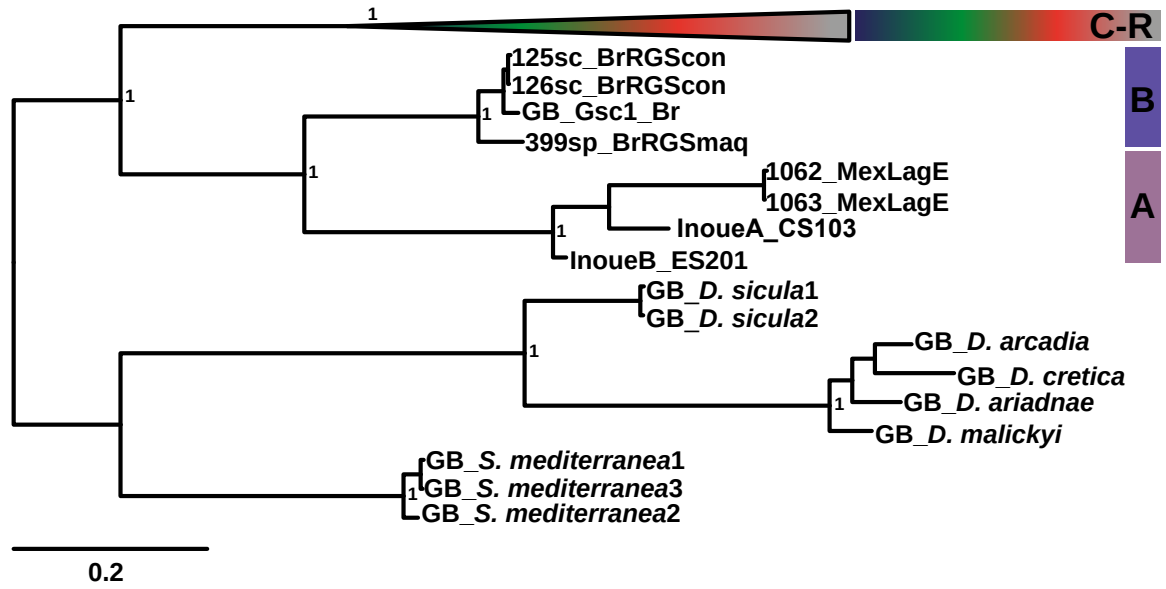


Figure 2. Bayesian Inference tree inferred from Dataset 2 (COI with outgroup). Clades C to R have been collapsed for the sake of clarity. Clade A is constituted by unclassified samples from Mexico and Texas (USA); Clade B includes identified individuals of Girardia schubarti from Brazil and other unidentified Brazilian individuals. Outgroup is composed of several representatives of genera Dugesia and Schmidtea downloaded from GenBank (Appendix 1). Values at nodes correspond to posterior probability. Scale bar: number of substitutions per nucleotide position.

OTUs of *G. multidiverticulata* de Souza et al. 2015 (clade F), *G. biapertura* Sluys, 1997 (clade D), *G. anderlani* (Kawakatsu & Hauser, 1983) (clade J) and *G.* aff. *arenicola* Hellmann & Leal-Zanchet, 2018 (clade I) all group into their own clades, thus representing distinctly separated lineages (Fig. 3A, B). The branch of *G. sanchezi* (Hyman, 1959), represented by two individuals from the type locality in Chile, constitute clade K. Although this is the only clade with rather low support (0.67 PP), it is well-differentiated from all other OTUs, while it is not closely related to any of the other

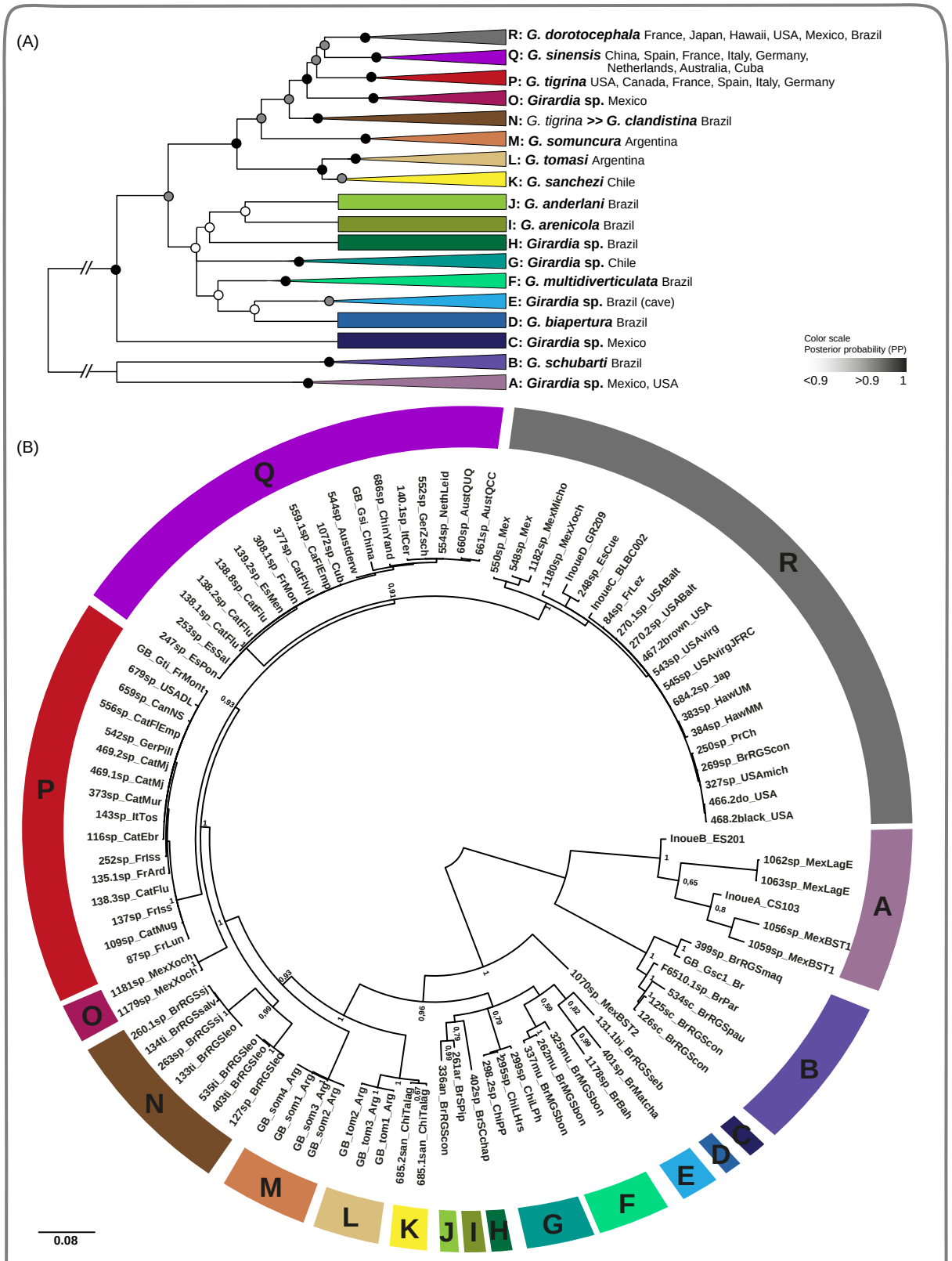


Figure 3. Bayesian Inference tree inferred from Dataset5 (concatenated no outgroup). Different clades indicated by letters and colours. (A): Schematic representation of the tree with collapsed clades, showing species identifications, when available, and countries of origin of the various terminals; shaded circles at nodes indicate posterior probability support, with values below 0.9 considered to be unsupported. (B) circular tree with all terminals; values at nodes correspond to posterior probability support. Scale bar: number of substitutions per nucleotide position.

Chilean individuals that are included in our analyses and that together constitute clade G. OTUs of *G. tomasi* Lenguas-Francavilla et al. 2021 and *G. somuncura* Lenguas-Francavilla et al. 2021 from Argentina group in their respective clades L and M. Among our OTUs, there are five individuals from Brazil that had been identified as *G. tigrina*. Four of these individuals group into clade N, together with three non-identified individuals from Brazil, all from Rio Grande do Sul, thus suggesting that all of these OTUs belong to this species (Fig. 3A, B).

Although we were unable to include into our analyses any *G. tigrina* individual from North America, where the species is native, that was unequivocally identified as a member of this species, clade P comprises OTUs from Michigan, USA (679sp_USADL) and Nova Scotia, Canada (659sp_CanNS) (Fig. 3B). Furthermore, a *G. tigrina* specimen from France, of which sequences were downloaded from GenBank, also falls within clade P. Therefore, we assigned this entire clade to the species *G. tigrina*. All other OTUs in clade P come from outside of the autochthonous area of distribution of *G. tigrina* (Fig. 3A, B).

There are two other species of which taxonomically identified specimens were included in our study, viz., *G. sinensis* Chen & Wang, 2015 from China and *G. dorotocephala* (Woodworth, 1897) from North America. The first-mentioned species was represented by a sequence available from GenBank and the second by specimens purchased from Carolina Biological Supply Company and that were collected from the USA, albeit that exact provenance of this sample was not known. The sequences of these two species group into two separate clades, viz., clade Q (*G. sinensis*; GB_Gsi_China) and clade R (*G. dorotocephala*; 466.2do_USA) (Fig. 3B). However, apart from taxonomically identified *G. dorotocephala,* clade R houses also non-identified OTUs from USA, Mexico, Canada, Europe, Japan, Hawaii, and Brazil. Surprisingly, clade Q not only comprises taxonomically identified *G. sinensis* from China, but also non-identified OTUs from Australia, China, Cuba, and Europe.

Six clades (A, C, E, G, H, and O) can not be assigned to any known species of *Girardia*. Clade A comprises OTUs from very distant localities: Los Tuxtlas, Mexico (1062, 1063, 1056, 1059), and Texas (InoueA_CS103) and New Mexico (InoueB_ES201) in the USA. Clade C is constituted by an unclassified OTU from Los Tuxtlas, Mexico (1070), while clade E is formed by unclassified samples from two very distant (2018 km) Brazilian caves (401, 1178). Clade G comprises unclassified

samples from Huinay Research Station, Chile (295, 298, 299), while clade H is formed by an unclassified sample (402) from Santa Catarina, Brazil. Clade O houses unclassified individuals (1181, 1179) from Xochimilco Mexico (Fig. 3B).

3.2.2. Phylogenetic relationships between clades

The phylogenetic trees based on the individual genes and including the outgroup taxa reveal the presence of two main, well-differentiated lineages within the genus *Girardia* (Figs 2, S1). One of these lineages, with OTUs from Mexico, USA, and Brazil, includes two sister clades (A+B), each of which is highly supported (1 PP), and that are well-separated from each other by long branches. The other main lineage comprises all remaining *Girardia* samples, with OTUs from North, Central and South America, which group in the clades C-R (these clades are collapsed in Figs 2 and S1). Three of the lineages in this second main group (clades P, Q, R), concern OTUs that have been introduced into other parts of the world, outside of the native range of *Girardia*.

In view of these results, we replaced in further phylogenetic analyses the initial outgroup taxa (species of *Dugesia* and *Schmidtea*) by the A+B clade, in order to avoid rooting with outgroup taxa that might be too distantly related to the ingroup. In this way, we attempted to avoid long-branch attraction (Felsenstein 1978) and systematic error due to highly divergent outgroup taxa (Graham et al. 2002).

Phylogenetic trees resulting from analyses of both concatenated and individual-gene datasets, rooted with clade A+B, are shown in Fig. 3 and Fig. S2. Individual-gene analyses (Fig. S2) recovered less internal nodes that are fully supported than the analysis of the concatenated dataset (Fig. 3), probably due to synergetic information in the molecular markers. In the following we describe the relationships and supports found in the concatenated tree (Fig. 3).

With respect to the ingroup C-R, an unclassified OTU from Los Tuxtlas (clade C) is sister to a major branch comprising all remaining clades, with good support (0.96 PP). One branch of this major clade comprises the groups D-J, which does not receive high support (0.79 PP), and concerns several South American lineages with unresolved affinities, such as: *G. biapertura* (D); OTUs from two Brazilian caves (E), and from Santa Catarina (H); *G. anderlani* (J) and the troglobitic *G.* aff. *arenicola* (I); the troglobitic *G. multidiverticulata* (F), and OTUs from Chile (G).

The second major branch on the tree, comprising clades K-R is highly supported (1.0 PP) (Fig. 3A, B). It contains a clade formed by the two sister species *G. sanchezi* (K) and *G. tomasi* (L), as well as clades of the following six well-differentiated taxa: *G. somuncura* (M), *G. tigrina* from Brazil (N), unclassified OTUs from Xochimilco, Mexico (O), *G. tigrina* from North America (P), *G. sinensis* (Q), and *G. dorotocephala* (R). All nodes within the K–R group receive high to maximum support values, ranging between 0.91 and 1.0. Hence, the topology of this portion of the tree (Fig. 3) shows well-supported relationships, in contrast to clade D – J.

3.2.3. Historical biogeographic remarks

Considering the origin of the samples analysed (Fig. 3A), it is possible to comment the biogeographical history of the genus *Girardia*, at least for the taxa included in our study. Therefore, it is important to take into account a number of issues that complicate geographic interpretation of the tree.

For example, *G. sinensis*, although described from China, has a North American origin (see discussion). Moreover, *G. sinensis, G. tigrina*, and *G. dorotocephala* have been introduced from North America into other parts of the world, and, therefore, any country outside of the North American subcontinent should be disregarded in the analysis. However, South America is an exception to this rule, in that the present study shows that presumed *G. tigrina* from this subcontinent actually concerns the new, sibling species *G. clandistina* Sluys & Benítez-Álvarez, sp. nov. (see below). The introduction of species of *Girardia* from their native areas to other parts of the world, and their subsequent settlement and further dispersal, have been analysed in extenso by Benítez-Álvarez et al. (personal communication) and, therefore, shall not be discussed here any further.

From that perspective, it is clear that the ancestral distribution of the clade O-R concerns the North American subcontinent and that of clade D-J concerns the South American subcontinent. When the South American distributions of the clades K-L, M, and N are taken into account, it leaves little doubt that the ancestral distribution on the branch leading to clades D-R must be reconstructed as being South America, as, most likely, is the case also for the most basal branch, leading to all *Girardia* terminals included in the tree. In other words, the ancestral distribution of *Girardia* is South

America, while the North American clade is the result of more recent colonization events.

4. Discussion

4.1. *Girardia*: genetical and chromosomal divergences

Our phylogenetic tree indicated the existence of two major lineages of *Girardia*, one constituted by the sister taxa *G. schubarti* (clade B) and the taxonomically unidentified clade A, and the other comprising all other *Girardia* OTUs and taxa (C–R) (Fig. 3). In a morphological phylogenetic analysis, *G. schubarti* grouped well among the other species of *Girardia* and formed a clade together with *G. arizonensis* (Kenk, 1975) and *G. azteca* (Benazzi & Giannini, 1971) (Sluys 2001). In their study on the phylogeny of continenticolan planarians with the help of molecular markers, Álvarez-Presas & Riutort (2014) included also three species of *Girardia* and found a sistergroup relationship between *G. anderlani* and *G. tigrina*, which together were sister to *G. schubarti*. The analysis of Inoue et al (2020) showed a sister-group relationship between *G. tigrina* and *G. dorotocephala*, which together with *G. anderlani* and two putative new species (InoueC and InoueD in our trees) constituted a monophyletic group that was sister to a clade formed by *G. schubarti* and two other putative new species (InoueA and InoueB in our trees). Lázaro et al. (2011) found also a sister-group relationship between *G. tigrina* and *G. dorotocephala*, which together were sister to *G. schubarti*, being the only three species of *Girardia* included in their analysis. However, our phylogenetic tree indicates a sister-group relationship between *G. schubarti* (clade B) and the taxonomically unidentified clade A (including InoueA and InoueB species), both of these clades together being sister to the major branch comprising all other *Girardia* OTUs and taxa (C–R) in our analysis. Evidently, as our study includes more OTUs than those of Álvarez-Presas & Riutort (2014) and Inoue et al. (2020), a more complex pattern of genealogical affinities is to be expected. In addition to the clear sister-group relationship between clades A+B and C–R, the great genetic distance between these two clades is noteworthy (Fig. 2), since the length of the branches is comparable with the distance between the two sister genera *Schmidtea* and *Dugesia*, which presumably diverged about 135.9 million years ago (Mya) (Solà et al 2022).

In addition to this high genetic differentiation, *G. schubarti* is also differentiated from other *Girardia* species by its number of chromosomes, having a basic haploid complement of n=4 (Kawakatsu et al. 1984; Jorge et al. 2000; Knakievicz et al. 2007), albeit that similar chromosome portraits are found in *G. arizonensis* Kenk , 1975 and *G. jenkinsae* Benazzi & Gourbault, 1977 (Benazzi & Gourbault, 1977; Benazzi, 1982), species not included in the present study (unless they are represented by some of our unidentified specimens from Mexico or the USA). On the other hand, *G. tigrina*, *G. dorotocephala, G. sanchezi, G. anceps*, *G. tahitiensis*, and *G. festae* exhibit haploid complements of n=8 (Gourbault 1977; Puccinelli & Deri 1991), while *G. anderlani*, *G. biapertura* and *G. cubana* (Codreanu & Balcesco, 1973) have n=9 (Benazzi 1982; Jorge et al. 2000; Benya et al. 2007; Knakievicz et al. 2007). For *G. nonatoi,* Marcus (1946) counted in oocytes 10 chromosomes during meiosis in the haploid phase, so that the full complement presumably consists of 20 chromosomes. Unfortunately, the chromosome portraits of other species of *Girardia* are unknown. Despite this paucity of information on chromosome number in the genus *Girardia*, a pattern emerges when the complements are plotted on the phylogenetic tree: clade A+B includes *G. schubarti* with n=4, D– J includes two species (*G. anderlani*, *G. biapertura*) with n=9 , and K– R clade includes three species (*G. tigrina*, *G. dorotocephala, G. sanchezi*) with n =8. If the chromosomal numbers found in the few species within each of these major clades (A+B, D-J, and K-R) are presumed to be common for all species within each of these groups, it may be hypothesized that the origin of the main clades of *Girardia* was associated with events of genomic duplications and/or chromosomal rearrangements.

Differences in chromosome number between closely related species of triclads are relatively common and have been related to speciation events in the genera *Schmidtea* and *Dugesia* (Leria, et al. 2018, 2020). However, with the present information available for *Girardia*, it cannot be excluded that chromosomal changes were not the drivers of the speciation process but accumulated only after speciation had taken place. Therefore, it is only through future, more comprehensive and integrative studies that we may determine whether the great genetical and chromosomal divergences of the A+B clade, as compared to its congeners, warrant taxonomic recognition in the form of a separate genus, or merely represent highly evolved autapomorphic features for a particular branch within the genus *Girardia*.

4.2. Genetic differentiation within clades of *Girardia*

Although we only have scattered samples from all over the Americas (Appendix 1, Fig. 1), in many cases our molecular-based phylogenetic results revealed a high genetic diversity and structure within *Girardia*, particularly in Mexico and Brazil.

Mexico showed the highest molecular diversity, despite the rather low number of samples. Eleven individuals were analysed, two of unknown origin and the rest coming from five localities (Appendix 1), which exhibited clear structure and genetic differentiation, and comprised four different clades (A, C, O, and R, Fig. 3). From these four clades, particularly clade A is noteworthy because of its high internal diversity, albeit that it does not include any taxonomically identified species. However, genetic structure within clade A strongly suggests that it contains more than one species. This is in accordance with the suggestion made by Inoue et al. (2020), who delimited two putative species on the basis of short fragments of *COI*, coded InoueA and InoueB in the present work (Fig. 3). Clades C and O (sister to D-R clade and P+Q+R clade, respectively) comprise only animals from Mexico, while some Mexican individuals occur also in clade R (*G. dorotocephala*). Evidently, at this moment it remains undecided whether the observed genetic diversity concerns new species of *Girardia* or merely reflects the presence of already known species of which molecular data is still lacking.

This high genetic Mexican diversity is not geographically structured as, perhaps, might be expected. Within clade A, the long branch separating samples from the Biological Station (1056, 1059) and individuals from Laguna Escondida (1062, 1063) suggests that two genetically highly differentiated species are present at these two localities, although the collection sites are only 2 Km apart. Clade C is only formed by individual 1070 from a second collection site at the Biological Station. All of this points to a possible co-occurrence of two highly differentiated species in the same river within the Biological Station. At the Xochimilco locality we found three specimens, two constituting the sister clade (O) of the group including clades P, Q and R, while the third specimen (1180) belongs to clade R (*G. dorotocephala*). This mix of genetically distant species at sites that geographically are in close proximity to each other, suggests a complex history for the diversification and evolution of *Girardia* in the Americas.

Another interesting fact that surfaced in our analyses was the relatively high diversity of cave-dwelling species in Brazil, with *G. multidiverticulata* being the first troglobitic continenticolan to be reported from South America. Although its distinctive characters differ from other species of *Girardia*, it shares with *G. anderlani* the presence of a large and branched bulbar cavity (Souza et al. 2015). Unfortunately, our molecular trees did not show sufficient resolution to support a sister-group relationship between *G. multidiverticulata* and *G. anderlani.* In point of fact, the trees suggested a closer relationship between epigean *G. anderlani* and specimen 261, which probably represents troglobitic *G. aff. arenicola*, both showing dorsal testes and a branched bulbar cavity (Hellmann et al. 2018).

Among our Brazilian samples there are two others that originated from hypogean habitats, viz., OTUs 401 and 1178, together constituting clade E (Fig. 3). Although these two individuals constituted a monophyletic group, they are genetically quite distinct, while their sampling localities are far apart. This suggests that clade E comprises two new cave-dwelling species of *Girardia*. This recently discovered flourishing of hypogean *Girardia* species in Brazil (see Morais et al. 2021) may be an indication that the genus is highly successful in adapting to life in caves and that future studies of those habitats in other regions in the Americas may unveil further diversity.

To date, nine species of *Girardia* have been recorded from Mexico, USA and Canada, with *G. tigrina* and *G. dorotocephala* being the most widely distributed ones (Sluys et al. 2005). The present study adds *G. sinensis*, described from a locality in China (Chen et al. 2015), since we identified it molecularly from Cuba (Figs 1, 3A). Moreover, the close phylogenetic relationship that this species shares with *G. dorotocephala* and *G. tigrina*, both of North American origin, also clearly point to that region as the original area of distribution of *G. sinensis*. In a recent molecular study on freshwater planarians from New Mexico and Texas, Inoue et al. (2020) identified two putative new species (InoueC and InoueD in our trees) that were closely related to *G. tigrina* and *G. dorotocephala*. However, in our analyses, both of their sequences grouped among representatives of *G. dorotocephala* (Fig. 3). Several other sequences from individuals collected outside of the Americas fall into the three clades P, Q, R, thus corroborating the North American origin of the introduced populations. The expansion of these three lineages around the world and its possible impact, are analysed in Chapter III.1.

4.3. Nominal *Girardia tigrina*

A very interesting result of our molecular analysis concerns the positions in the phylogenetic tree of the North and South American *G. tigrina* samples, showing that the Brazilian clade (N) is not even closely related to the North American one (clade P) (Fig. 3). This corroborates the conclusion of Sluys et al. (2005) that the North and South American forms are different species. According to Sluys et al. (2005), the only anatomical difference between them resides in the coat of muscles around the bursal canal. In North American *G. tigrina* this coat of muscles is simple, consisting of a thin subepithelial layer of circular muscle, followed by an equally thin layer of longitudinal muscle fibres. In contrast, the South American form possesses a bursal canal musculature that consists of a well-developed coat of intermingled circular and longitudinal muscle fibres. In other characters the two forms are very similar, but our results clearly show that they are genetically well-differentiated and are not even sister species. Therefore, the South American form is here described and named as the new species *Girardia clandistina* Sluys & Benítez-Álvarez, sp. nov. (for differential diagnosis, see Appendix 2). This taxonomic action is not unimportant, since *G. tigrina* is the type species for the genus *Girardia* (Kenk 1974) and, therefore, it is necessary to know the precise boundaries of the taxon and the extension of the species name.

4.4. Historical biogeography of *Girardia*

Our phylogenetic tree suggested that *Girardia* evolved on the South American subcontinent and from there colonized North America. Previous studies argued that the family Dugesiidae, including *Girardia*, had already diversified on Gondwanaland (Ball 1975) or even at pre-Pangaean times and, thus, must have diversified also already on Pangaea (Sluys et al. 1998). Thus, *Girardia* diversified on the South American portion of Gondwanaland, and, subsequently, ancestors of the O-R clade migrated to North America and diversified there. The relatively short inter-branches among the clades on the North American subcontinent (clades O, P, Q, R; Fig. 3B) implies that diversification of *Girardia* in North America is rather recent, and hence that the northward migration for this group did probably take place only after complete closure of the Isthmus of Panama at about 2.8 Mya (O'Dea et al. 2016).

Interestingly, presence of the basal lineages A in Mexico and the USA, and clade C in Mexico, and of the crown group O-R in Mexico, USA, and Canada, suggests

that the North American subcontinent was populated by at least two independent waves of dispersal from the Neotropics. Although in these cases, the available data do not allow us to infer whether this northward migration was relatively recent or took place in more remote epochs, again, it would have been possible only through freshwater tracks in the intermittent connections during emergence of the Isthmus of Panama or once it was fully established (McGirr et al. 2021). Unfortunately, lack of OTUs from the northern parts of South America prevents further elucidation of the precise routes taken by neotropical *Girardia*'s during their dispersal into the Nearctics.

Data availability

All sequences have been deposited in GenBank.

Acknowledgments

We thank all colleagues mentioned in Table S1 for contributing samples of worms. We are grateful to Oleguer Castillo, Ana Paula Goulart Araujo, and Omar Lagunas for DNA extraction and sequencing of some samples, while the first-mentioned also designed a specific Girardia COI primer. We are also grateful to Eduard Solà for designing primers for EF1a amplification. Alejandro Oceguera and Fernando Carbayo are thanked for support in the sampling campaign in Mexico. Also, we are grateful to Juan Manuel Lentijo Mondéjar for his help in making the online map.

Author Contributions

MR did the initial study design. MR, LBA, RS, AMLZ, and LL contribute with the sampling. LBA processed the samples and analysed the molecular data. RS did the formal morphological description of *G. clandistina sp. nov.* LBA, RS, AMLZ, LL, MR, MR, wrote the manuscript.

Conflicts of interest

All authors declare no conflict of interest to disclose.

Funding

This work was supported by the Brazilian Research Council - CNPq, Brazil [grants 306853/2015-9 and 313691/2018-5] and Ministerio de Economía y Competitividad, Spain [projects CGL2015-63527-1P and PGC2018-093924-B-100]. Fieldwork in Mexico was funded by the Unión Iberoamericana de Universidades (UIU), Spain [Project BIO02/2017]

Appendix 1. Sequences included in this study, with indication of sample codes, sampling localities (for *Girardia* samples), taxonomic assignment before and after analyses, codes used in the text and figures, and GenBank accession numbers for COI and EF1α sequences. Sequence codes in bold concern new sequences reported in this study. See Supplementary Table S1 for exact localities, collectors, and criteria used for taxonomic assignment previously to our analysis.

Sample ID	Locality	Taxonomic Identification before analysis	Taxonomic Identification after analysis	ID in Figures	COI	EF1α
84	France, River Lez	*Girardia* sp.	*G. dorotocephala*	84sp_FrLez	**OM307073**	**OM349486**
86.1	France, River Laderge	*Girardia* sp.	*G. dorotocephala*	86.1sp_FrLad	**OM307074**	
86.2	France, River Laderge	*Girardia* sp.	*G. dorotocephala*	86.2sp_FrLad	**OM307075**	
87	France, Lunaç	*Girardia* sp.	*G. tigrina*	87sp_FrLun	**OM307076**	**OM418671**
109	Spain, Catalonia, St. Llorenç de la Muga, River Muga	*Girardia* sp.	*G. tigrina*	109sp_CatMug	**OM307077**	**OM418675**
116	Spain, Catalonia, Riba-roja d'Ebre, River Ebro	*Girardia* sp.	*G. tigrina*	116sp_CatEbr	**OM307078**	**OM418673**
125	Brazil, Rio Grande do Sul, Constantina	*G. schubarti*	*G. schubarti*	125sc_BrRGScon	**OM307079**	**OM418648**
126	Brazil, Rio Grande do Sul, Constantina	*G. schubarti*	*G. schubarti*	126sc_BrRGScon	**OM307080**	**OM418649**
127	Brazil, Rio Grande do Sul, São Leopoldo	*Girardia* sp.	*G. clandistina*	127sp_BrRGSleo	**OM307081**	**OM418688**
131.1	Brazil, Rio Grande do Sul, São Sebastião de Caí	*G. biapertura*	*G. biapertura*	131.1bi_BrRGSseb	**OM307082**	
131.2	Brazil, Rio Grande do Sul, São Sebastião de Caí	*G. biapertura*	*G. biapertura*	131.2bi_BrRGSseb	**OM307083**	
132	Brazil, Rio Grande do Sul, Gramado	*G. tigrina*	*G. clandistina*	132ti_BrRGSgra	**OM307084**	
133	Brazil, Rio Grande do Sul, São Leopoldo	*G. tigrina*	*G. clandistina*	133ti_BrRGSleo	**OM307085**	**OM418690**

134	Brazil, Rio Grande do Sul, Salvador do Sul	*G. tigrina*	*G. clandistina*	134ti_BrRGSsalv	**OM307086**	**OM418689**
135.1	France, Gorges de l'Ardeche	*Girardia* sp.	*G. tigrina*	135.1sp_FrArd	**OM307087**	**OM418677**
135.2	France, Gorges de l'Ardeche	*Girardia* sp.	*G. tigrina*	135.2sp_FrArd		**OM418679**
137	France, Ispagnac	*Girardia* sp.	*G. tigrina*	137sp_FrIsp	**OM307088**	**OM418682**
138.1	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	*G. sinensis*	138.1sp_CatFlu	**OM307089**	**OM418669**
138.2	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	*G. sinensis*	138.2sp_CatFlu	**OM307090**	**OM418664**
138.3	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	*G. tigrina*	138.3sp_CatFlu	**OM307091**	**OM418681**
138.4	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	*G. tigrina*	138.4sp_CatFlu	**OM307092**	
138.5	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	*G. sinensis*	138.5sp_CatFlu	**OM307093**	
138.6	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	*G. sinensis*	138.6sp_CatFlu	**OM307094**	
138.7	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	*G. sinensis*	138.7sp_CatFlu	**OM307095**	
138.8	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	*G. sinensis*	138.8sp_CatFlu	**OM307096**	**OM418663**
139.1	Spain, Menorca, Algendar	*Girardia* sp.	*G. sinensis*	139.1sp_EsMen	**OM307097**	
139.2	Spain, Menorca, Algendar	*Girardia* sp.	*G. sinensis*	139.2sp_EsMen	**OM307098**	**OM418657**
140.1	Italy, Sardinia, R.Fungarone, Putifigari	*Girardia* sp.	*G. sinensis*	140.1sp_ItCer	**OM307099**	**OM418655**
140.2	Italy, Sardinia, R.Fungarone, Putifigari	*Girardia* sp.	*G. sinensis*	140.2sp_ItCer	**OM307100**	

143	Italy, Toscana, Torr.Vincio	*Girardia* sp.	*G. tigrina*	143sp_ItTos	**OM307101**	**OM418683**
247	Spain, Pontevedra, Gondomar	*Girardia* sp.	*G. sinensis*	247sp_EsPon	**OM307102**	**OM418653**
248	Spain, Cuenca, Reíllo	*Girardia* sp.	*G. dorotocephala*	248sp_EsCue	**OM307103**	**OM349494**
250	Portugal, Cheleiros	*Girardia* sp.	*G. dorotocephala*	250sp_PrCh	**OM307104**	**OM349490**
252	France, Issalès	*Girardia* sp.	*G. tigrina*	252sp_FrIss	**OM307105**	**OM418674**
253	Spain, Salamanca, Ciudad Rodrigo	*Girardia* sp.	*G. sinensis*	253sp_EsSal	**OM307106**	**OM418654**
260.1	Brazil, Rio Grande do Sul, São Jose do Norte	*Girardia* sp.	*G. clandistina*	260.1sp_BrRGSsj	**OM307107**	**OM418691**
260.2	Brazil, Rio Grande do Sul, São Jose do Norte	*Girardia* sp.	*G. clandistina*	260.2sp_BrRGSsj	**OM307108**	
261	Brazil, São Paulo, Iporanga (cave)	*Girardia* aff. *arenicola*	*Girardia* aff. *arenicola*	261ar_BrSPip	**OM264750**	**OM418632**
262	Brazil, Mato Grosso do Sul, Bodoquena (cave)	*G. multidiverticulata*	*G. multidiverticulata*	262mu_BrMGSbod	**OM307109**	**OM418642**
263	Brazil, Rio Grande do Sul, São Jose do Norte	*Girardia sp.*	*G. clandistina*	263sp_BrRGSsj	**OM307110**	**OM418692**
269	Brazil, Rio Grande do Sul, Constantina	*Girardia sp.*	*G. dorotocephala*	269sp_BrRGScon	**OM307111**	**OM349488**
270.1	USA, Baltimore	*Girardia sp.*	*G. dorotocephala*	270.1sp_USAbalt	**OM307112**	**OM349498**
270.2	USA, Baltimore	*Girardia sp.*	*G. dorotocephala*	270.2sp_USAbalt	**OM307113**	**OM349499**
295	Chile, Los Lagos, Huinay Research Station	*Girardia* sp.	*Girardia* sp.	295sp_ChiLHrs	**OM307114**	**OM418636**
296	Chile, Los Lagos, Huinay Research Station	*Girardia* sp.	*Girardia* sp.	296sp_ChiLHrs	**OM307115**	

297	Chile, Los Lagos, Huinay Research Station	*Girardia* sp.	*Girardia* sp.	297sp_ChiLHrs	**OM307116**	
298.1	Chile, Pumalin Park	*Girardia* sp.	*Girardia* sp.	298.1sp_ChiPP	**OM307117**	
298.2	Chile, Pumalin Park	*Girardia* sp.	*Girardia* sp.	298.2sp_ChiPP	**OM307118**	**OM418638**
299	Chile, Los Lagos, Peninsula Huequi	*Girardia* sp.	*Girardia* sp.	299sp_ChiLPh	**OM307119**	**OM418637**
300	Chile, Los Lagos, Port Montt	*Girardia* sp.	*Girardia* sp.	300sp_ChiLPm	**OM307120**	
308.1	France, Montpellier	*Girardia* sp.	*G. sinensis*	308.1sp_FrMon	**OM307121**	**OM418656**
308.2	France, Montpellier	*Girardia* sp.	*G. sinensis*	308.2sp_FrMon	**OM307122**	
325	Brazil, Mato Grosso do Sul, Bonito (cave)	*G. multidiverticulata*	*G. multidiverticulata*	325mu_BrMGSbon	**OM307123**	**OM418641**
327	USA, Michigan, Ann Arbor	*Girardia* sp.	*G. dorotocephala*	327sp_USAmich	**OM307124**	**OM349489**
336	Brazil, Rio Grande do Sul, Severiano de Almeida	*G. anderlani*	*G. anderlani*	336an_BrRGScon	**OM232748**	
337	Brazil, Mato Grosso do Sul, Bodoquena (cave)	*G. multidiverticulata*	*G. multidiverticulata*	337mu_BrMGSbod	**OM307125**	**OM418643**
338	Brazil, Mato Grosso do Sul, Bodoquena (cave)	*G. multidiverticulata*	*G. multidiverticulata*	338mu_BrMGSbod	**OM307126**	
373	Spain, Catalonia, Riera de Mura	*Girardia* sp.	*G. tigrina*	373sp_CatMur	**OM307127**	**OM418680**
377	Spain, Catalonia, Girona, Fluvià, Vilert	*Girardia* sp.	*G. sinensis*	377sp_CatFlvil	**OM307128**	**OM418658**
383	Hawaii, Upper Manoa	*Girardia* sp.	*G. dorotocephala*	383sp_HawUM	**OM307129**	**OM349501**
384	Hawaii, Middle Manoa	*Girardia* sp.	*G. dorotocephala*	384sp_HawMM	**OM307130**	**OM349500**

399	Brazil, Rio Grande do Sul, Maquine	*Girardia* sp.	*G. schubarti*	399sp_BrRGSmaq	**OM307131**	**OM418646**
401	Brazil, Mato Grosso, Chapada Guimaraes (cave)	*Girardia* sp.	*Girardia* sp.	401sp_BrMatcha	**OM307132**	
402	Brazil, Santa Catarina, Chapeco	*Girardia sp.*	*Girardia* sp.	402sp_BrSCchap	**OM307133**	**OM418635**
403	Brazil, Rio Grande do Sul, São Leopoldo	*G. tigrina*	*G. clandistina*	403ti_BrRGSleo	**OM307134**	**OM418686**
466.1	USA, Carolina Enterprise†	*G. dorotocephla* *Dugesia dorotocephala*‡	*G. dorotocephala*	466.1do_USA	**OM307135**	
466.2	USA, Carolina Enterprise†	*G. dorotocephla* *Dugesia dorotocephala*‡	*G. dorotocephala*	466.2do_USA	**OM307136**	**OM349491**
467.1	USA, Carolina Enterprise†	*Girardia* sp Brown planaria‡	*G. dorotocephala*	467.1brown_USA	**OM307137**	
467.2	USA, Carolina Enterprise†	*Girardia* sp Brown planaria‡	*G. dorotocephala*	467.2brown_USA	**OM307138**	**OM349502**
468.1	USA, Carolina Enterprise†	*Girardia* sp Black planaria‡	*G. dorotocephala*	468.1black_USA	**OM307139**	
468.2	USA, Carolina Enterprise†	*Girardia* sp Black planaria‡	*G. dorotocephala*	468.2black_USA	**OM307140**	**OM349492**
469.1	Spain, Catalonia, Montjüic	*Girardia* sp.	*G. tigrina*	469.1sp_CatMj	**OM307141**	**OM418670**
469.2	Spain, Catalonia, Montjüic	*Girardia* sp.	*G. tigrina*	469.2sp_CatMj	**OM307142**	**OM418684**
534	Brazil, Rio Grande do Sul, São Francisco de Paula	*G. schubarti*	*G. schubarti*	534sc_BrRGSpau		**OM418647**
535	Brazil, Rio Grande do Sul, São Leopoldo	*G. tigrina*	*G. clandistina*	535ti_BrRGSleo	**OM307143**	**OM418687**
542	Germany, Pillnitz	*Girardia* sp.	*G. tigrina*	542sp_GerPill	**OM307144**	**OM418678**
543	USA, Virginia	*Girardia* sp.	*G. dorotocephala*	543sp_USAvirg	**OM307145**	**OM349495**

544	Australia, Tasmania, Derwent River	*Girardia* sp.	*G. sinensis*	544sp_Austderw	**OM307146**	**OM418667**
545	USA, Virginia, Ashburn, Janelia Farm Research Campus	*Girardia* sp.	*G. dorotocephala*	545sp_USAvirgJFRC	**OM307147**	**OM349496**
546	Francia, Lez	*Girardia* sp.	*G. dorotocephala*	546sp_FrLez	**OM307148**	
547	Spain, Catalonia, Barcelona, Vallvidrera	*Girardia* sp.	*G. tigrina*	547sp_CatBarcVald	**OM307149**	
548	Mexico	*Girardia* sp.	*G. dorotocephala*	548sp_Mex	**OM307150**	
550	Mexico	*Girardia* sp.	*G. dorotocephala*	550sp_Mex	**OM307151**	**OM349497**
551	France, River Herault	*Girardia* sp.	*G. sinensis*	551sp_FrHer	**OM307152**	
552	Germany, Zschorna	*Girardia* sp.	*G. sinensis*	552sp_GerZsch	**OM307153**	**OM418666**
553	Germany, Pillnitz	*Girardia* sp.	*G. tigrina*	553sp_GerPill	**OM307154**	
554	Netherlands, Leiden	*Girardia* sp.	*Girardia sinensis*	554sp_NethLeid	**OM307155**	**OM418665**
556	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*Girardia* sp.	*G. tigrina*	556sp_CatFlEmp	**OM307156**	**OM418685**
558	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*Girardia* sp.	*G. sinensis*	558sp_CatFlEmp	**OM307157**	
559.1	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*Girardia* sp.	*G. sinensis*	559.1sp_CatFlEmp	**OM307158**	**OM418668**
559.2	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*Girardia* sp.	*G. sinensis*	559.2sp_CatFlEmp	**OM307159**	
659	Canada, Nova Scotia, Ainslie Lake	*Girardia* sp.	*G. tigrina*	659sp_CanNS	**OM307160**	**OM418676**
660	Australia, Queensland, UQ Lakes	*Girardia* sp.	*G. sinensis*	660sp_AustQUQ	**OM307161**	**OM418659**

661	Australia, Queensland	*Girardia* sp.	*G. sinensis*	661sp_AustQCC	**OM307162**	**OM418660**
679	USA, Michigan, Douglas Lake	*Girardia* sp.	*G. tigrina*	679sp_USADL	**OM307163**	**OM418672**
683.1	Japan, Hoshikuki-cho, Mizu-no-sato Park, Miyako River	*Girardia* sp.	*G. dorotocephala*	683.1sp_Jap	**OM307164**	
684.1	Japan, Hoshikuki-cho, Mizu-no-sato Park, Miyako River	*Girardia* sp.	*G. dorotocephala*	684.1sp_Jap	**OM307165**	
684.2	Japan, Hoshikuki-cho, Mizu-no-sato Park, Miyako River	*Girardia* sp.	*G. dorotocephala*	684.2sp_Jap	**OM307166**	**OM349493**
685.1	Chile, Talagante, Mapocho River	*Girardia sanchezi*	*G. sanchezi*	685.1san_ChiTalag	**OM307167**	**OM418644**
685.2	Chile, Talagante, Mapocho River	*Girardia sanchezi*	*G. sanchezi*	685.2san_ChiTalag	**OM307168**	**OM418645**
686	China, Conghua, Yadongxi River	*Girardia* sp.	*G. sinensis*	686sp_ChinYand	**OM307169**	**OM418662**
687	Chile, Los Lagos, Huinay Research Station	*Girardia* sp.	*Girardia* sp.	687sp_ChiLHrs	**OM307170**	
1056	Mexico, Biological Station Los Tuxtlas 1	*Girardia* sp.	*Girardia* sp.	1056sp_MexBST1		**OM418694**
1059	Mexico, Biological Station Los Tuxtlas 1	*Girardia* sp.	*Girardia* sp.	1059sp_MexBST1		**OM418639**
1062	Mexico, Los Tuxtlas, Laguna Escondida	*Girardia* sp.	*Girardia* sp.	1062sp_MexLagE	**OM307171**	**OM418640**
1063	Mexico, Los Tuxtlas, Laguna Escondida	*Girardia* sp.	*Girardia* sp.	1063sp_MexLagE	**OM307172**	
1070	Mexico, Biological Station Los Tuxtlas 2	*Girardia* sp.	*Girardia* sp.	1070sp_MexBST2	**OM307173**	**OM418633**
1072.2	Cuba, Matanzas, Martí, El Huequito	*Girardia* sp.	*G. sinensis*	1072.2sp_Cub	**OM307174**	**OM418661**
1178	Brazil, Bahía, Chapada Diamantina, Vale do Pati (cave)	*Girardia* sp.	*Girardia* sp.	1178sp_BrBah	**OM307175**	**OM418634**

1179	Mexico, Mexico City, Xochimilco, Cuemanco	*Girardia* sp.	*Girardia* sp.	1179sp_MexXoch	**OM307176**	**OM41869**
1180	Mexico, Mexico City, Xochimilco, Cuemanco	*Girardia* sp.	*G. dorotocephala*	1180sp_MexXoch	**OM307177**	**OM349487**
1181	Mexico, Mexico City, Xochimilco, Cuemanco	*Girardia* sp.	*Girardia* sp.	1181sp_MexXoch	**OM307178**	
1182	Mexico, Michoacán	*Girardia* sp.	*G. dorotocephala*	1182sp_MexMicho	**OM307179**	
F6510.1	Brazil, Paraná, Toledo, Cerro da Lola	*Girardia* sp.	*G. schubarti*	F6510.1sp_BrPar		**OM418651**
F6510.2	Brazil, Paraná, Toledo, Cerro da Lola	*Girardia* sp.	*G. schubarti*	F6510.2sp_BrPar		**OM418652**
F6510.3	Brazil, Paraná, Toledo, Cerro da Lola	*Girardia* sp.	*G. schubarti*	F6510.3sp_BrPar		**OM418650**
InoueA	USA, Texas, Caroline Spring, Independence Creek	*Girardia* sp.	*Girardia* sp.	InoueA_CS103	MN652340.1	
InoueB	USA, New Mexico, Palomas Creek, Emrick Spring	*Girardia* sp.	*Girardia* sp.	InoueB_ES201	MN652378.1	
InoueC	USA, Texas, Bitter Lake, Bitter Creek	*Girardia* sp.	*G. dorotocephala*	InoueC_BLBC002	MN652301.1	
InoueD	USA, New Mexico, West Fork of the Gila River	*Girardia sp*.	*G. dorotocephala*	InoueD_GR209	MN652373.1	
GB_G.an1	Brazil	*G. anderlani*	*G. anderlani*	GB_Gan1_Br	DQ666038.1	
GB_G.si	China, Guangdong Province Xinghu Lake in Zhaoqing	*G. sinensis*	*G. sinensis*	GB_Gsi_China	KP091895.1	
GB_G.ti	France, Montpellier	*G. tigrina*	*G. tigrina*	GB_Gti_FrMont	DQ666042.1	
GB_G.sc1	Brazil	*G.schubarti*	*G.schubarti*	GB_Gsc1_Br	DQ666041.1	
GB_G.sc2	Brazil, Rio Grande do Sul, Constantina	*G.schubarti*	*G.schubarti*	GB_Gsc2_BrRGScon		KJ599691.1

GB_G.som1	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	*G. somuncura*	GB_som1_Arg	MW271865
GB_G.som2	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	*G. somuncura*	GB_som2_Arg	MW271866
GB_G.som3	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	*G. somuncura*	GB_som3_Arg	MW271867
GB_G.som4	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	*G. somuncura*	GB_som4_Arg	MW271869
GB_G.tom1	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. tomasi*	*G. tomasi*	GB_tom1_Arg	MW271863
GB_G.tom2	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. tomasi*	*G. tomasi*	GB_tom2_Arg	MW271864
GB_G.tom2	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. tomasi*	*G. tomasi*	GB_tom2_Arg	MW271868
GB_S.med1	Tunisia, Lebna	*Schmidtea mediterranea*	*S. mediterranea*	GB_*S.mediterranea*1	JF837060.1
GB_S.med2	Italy, Sardinia	*S. mediterranea*	*S. mediterranea*	GB_*S.mediterranea*2	JF837061.1
GB_S.med3	Italy, Sicily	*S. mediterranea*	*S. mediterranea*	GB_*S.mediterranea*3	JF837062.1
GB_D.sic1	Spain	*Dugesia sicula*	*Dugesia sicula*	GB_*D.sicula*1	KC577350.1
GB_D.sic2	Italy	*D. sicula*	*D. sicula*	GB_*D.sicula*2	KC577351.1
GB_D.ari	Greece	*D. ariadnae*	*D. ariadnae*	GB_*D.ariadnae*	KF308713.1
GB_D.arc	Greece	*D. arcadia*	*D. arcadia*	GB_*D.arcadia*	KF308723.1
GB_D.mal	Greece	*D. malickyi*	*D. malickyi*	GB_*D.malickyi*	KF308777.1
GB_D.cre	Greece	*D. cretica*	*D. cretica*	GB_*D.cretica*	KF308794.1

GB_S.me43	Europe	*S. mediterranea*	*S. mediterranea*	GB_*S. mediterranea*4	KJ599709.1
GB_D.sic3	Europe	*D. sicula*	*D. sicula*	GB_*D. sicula*3	KJ599689.1
1124	South Africa	*D. afromontana*	*D. afromontana*	*D. afromontana*	**OM460743**
135	Japan	*Dugesia* sp.	*Dugesia* sp.	*Dugesia* sp1_Jap	**OM460745**
1336	Japan	*Dugesia* sp.	*Dugesia* sp.	*Dugesia* sp2_Jap	**OM460746**

†: Carolina Enterprise Word-Class Support for Science & Math

‡: classification according to Carolina Biological Supply Company [https://www.carolina.com/living-organisms/classroom-animals/invertebrates/platyhelminthes/10531.ct?Nr=product.siteId%3A100001]

Appendix 2. Differential diagnosis of *Girardia clandistina* Sluys & Benítez-Álvarez, sp. nov.

***Girardia clandistina* Sluys & Benítez-Álvarez, sp. nov.**

Holotype: Naturalis Biodiversity Center, ZMA V.Pl. 976.4, Arroyo Saves Dept., Canalos BS, Uruguay, 1-3 January 1987, sagittal sections on 6 slides.

Etymology: The specific epithet is based on the Latin adjective *clandistinus*, secret, concealed, and alludes to the fact that it concerns a sibling species.

Differential diagnosis:

A species of *Girardia* with low triangular head with bluntly pointed tip and short, broad auricles. Dorsal body colouration variable, being of a reticulated type with darkish spots and also a pair of dark stripes, separated by a pale mid-dorsal streak, or composed of a dark background interspersed with white splotches and with a pale middorsal line, or variations on these two major patterns. Reproductive complex basically as in *G. tigrina*, the only, but consistent, anatomical difference between the two species residing in the coat of muscles around the bursal canal. In North American *G. tigrina* this coat of muscles is simple, consisting of a thin subepithelial layer of circular muscle, followed by an equally thin layer of longitudinal muscle fibres. In contrast, *G. clandistina* possesses a bursal canal musculature that consists of a well-developed coat of intermingled circular and longitudinal muscle fibres.

SUPPORTING INFORMATION

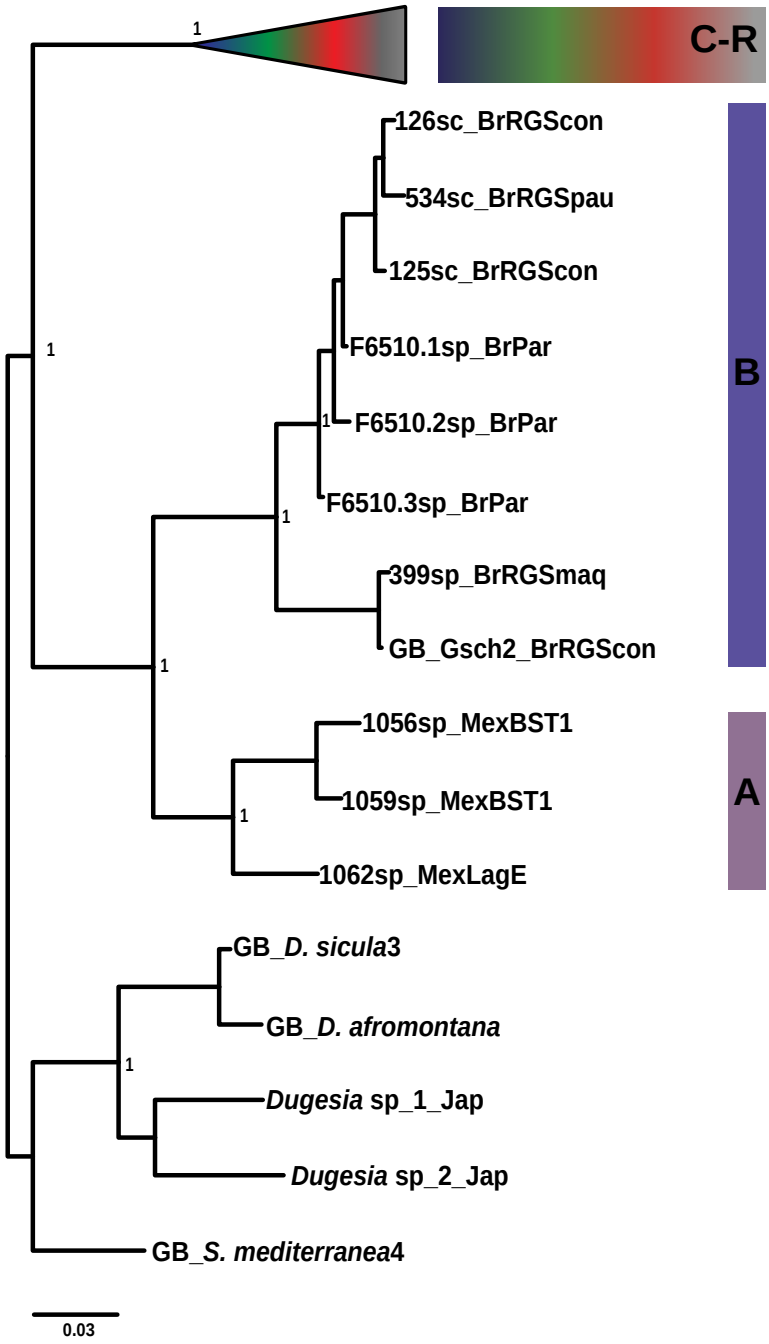
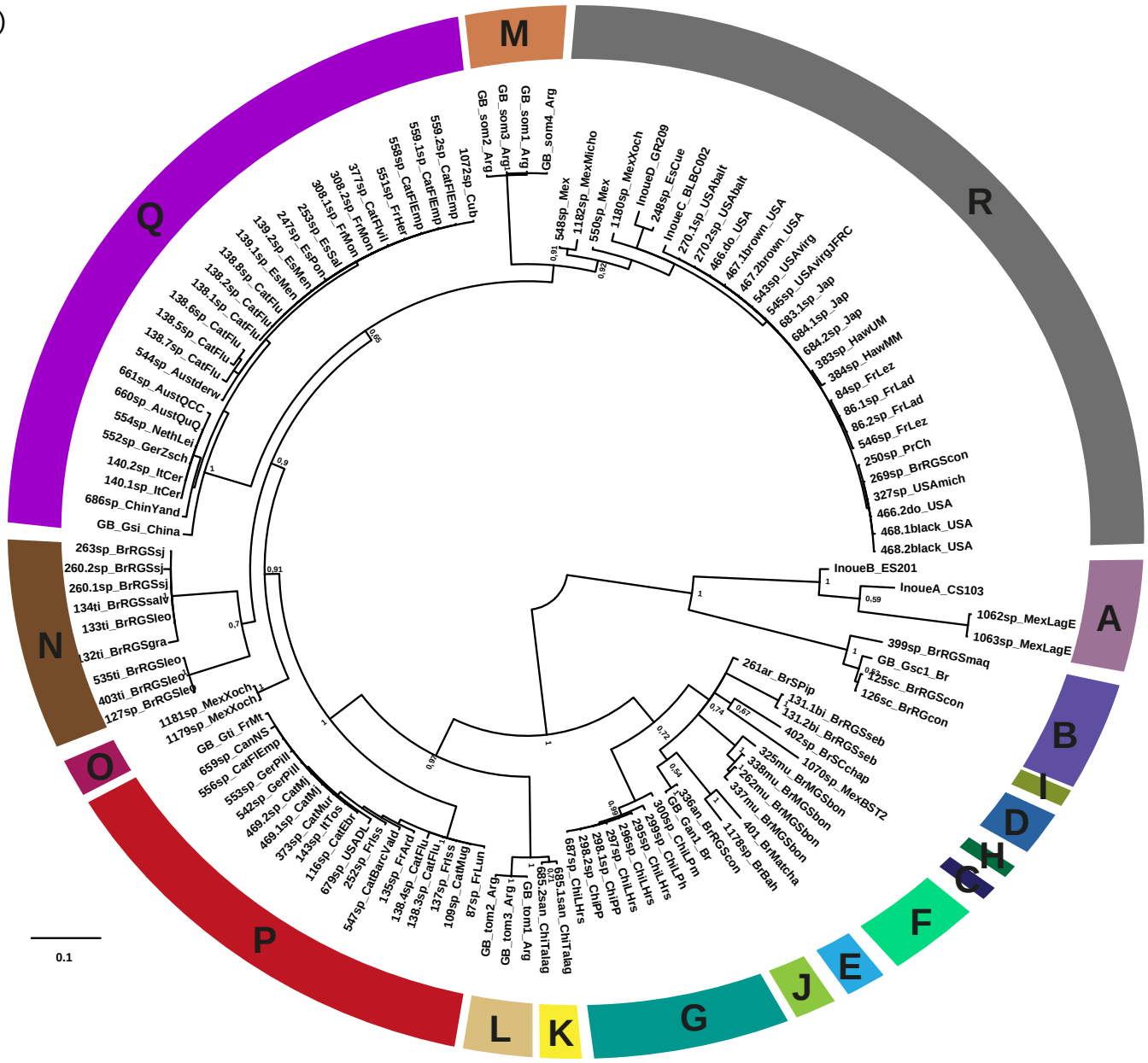


Figure S1. Bayesian Inference tree from Dataset4 (EF1a with outgroup). The sister group of *Girardia schubarti* and unclassified samples from Mexico have been collapsed. Values at nodes correspond to posterior probability support. Scale bar: number of substitutions per nucleotide position.

Figure S2. Bayesian Inference trees from Dataset1 (COI without outgroup) (A) and Dataset3 (EF1α without outgroup) (B). Values at nodes correspond to posterior probability support. Scale bar: number of substitutions per nucleotide position.

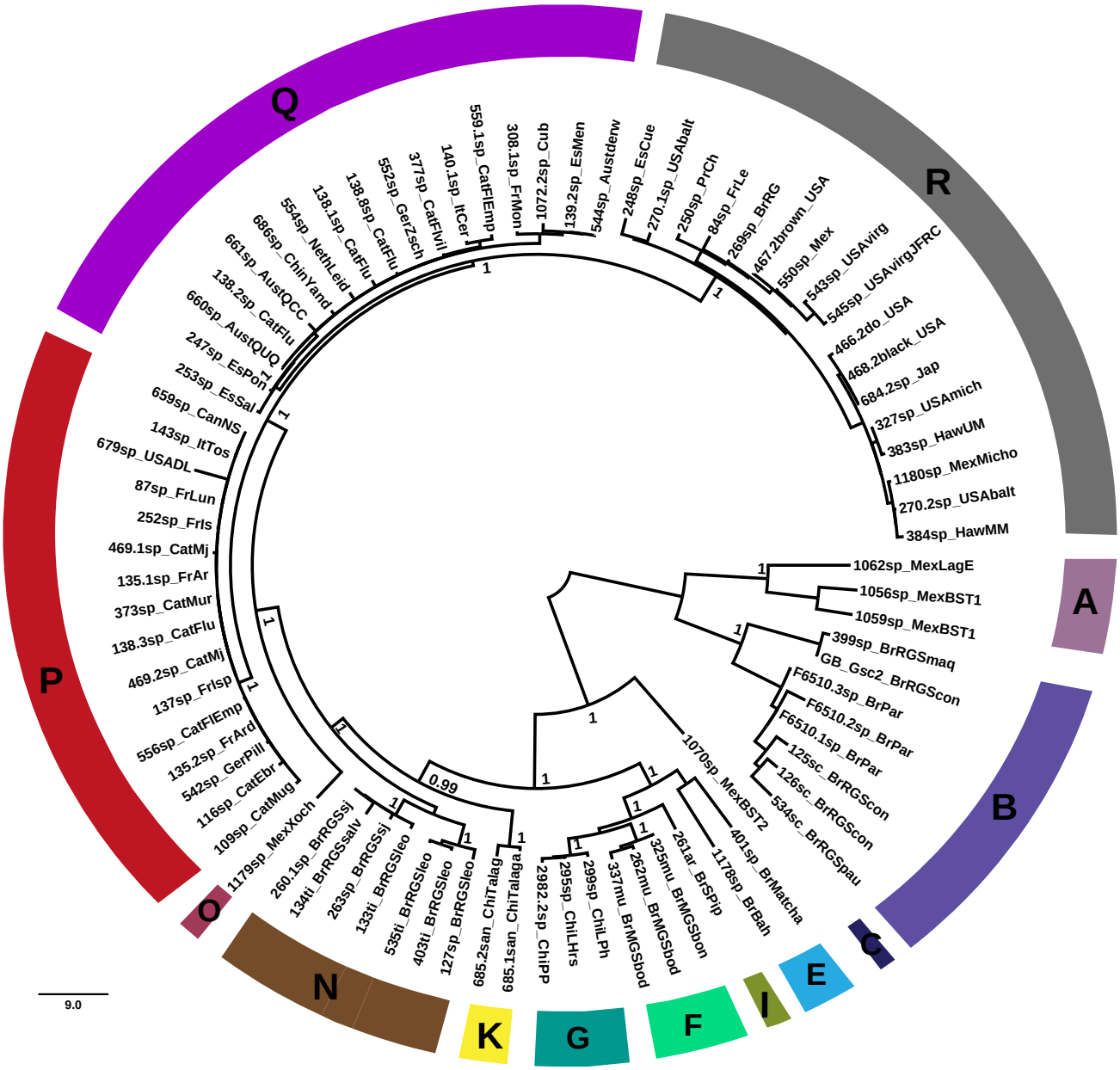
(A)



Phylogenetic clades

- **R:** ***G. dorotocephala*** France, Japan, Hawaii, USA, Mexico, Brazil
- **Q:** ***G. sinensis*** China, Spain, France, Italy, Germany,Netherlands, Australia, Cuba
- **P:** ***G. tigrina*** USA, Canada, France, Spain, Italy, Germany
- **O:** ***Girardia* sp.** Mexico
- **N:** *G. tigrina* **>>** ***G. clandistina*** Brazil
- **M:** ***G. somuncura*** Argentina
- **L:** ***G. tomasi*** Argentina
- **K:** ***G. sanchezi*** Chile
- **J:** ***G. anderlani*** Brazil
- **I:** ***G. arenicola*** Brazil
- **H:** ***Girardia* sp.** Brazil
- **G:** ***Girardia* sp.** Chile
- **F:** ***G. multidiverticulata*** Brazil
- **E:** ***Girardia* sp.** Brazil (cave)
- **D:** ***G. biapertura*** Brazil
- **C:** ***Girardia* sp.** Mexico
- **B:** ***G. schubarti*** Brazil
- **A:** ***Girardia* sp.** Mexico, USA

(B)



Phylogenetic clades

- **R:** ***G. dorotocephala*** France, Japan, Hawaii, USA, Mexico, Brazil
- **Q:** ***G. sinensis*** China, Spain, France, Italy, Germany, Netherlands, Australia, Cuba
- **P:** ***G. tigrina*** USA, Canada, France, Spain, Italy, Germany
- **O:** ***Girardia* sp.** Mexico
- **N:** *G. tigrina* **>>** ***G. clandistina*** Brazil
- **M:** ***G. somuncura*** Argentina
- **L:** ***G. tomasi*** Argentina
- **K:** ***G. sanchezi*** Chile
- **J:** ***G. anderlani*** Brazil
- **I:** ***G. arenicola*** Brazil
- **H:** ***Girardia* sp.** Brazil
- **G:** ***Girardia* sp.** Chile
- **F:** ***G. multidiverticulata*** Brazil
- **E:** ***Girardia* sp.** Brazil (cave)
- **D:** ***G. biapertura*** Brazil
- **C:** ***Girardia* sp.** Mexico
- **B:** ***G. schubarti*** Brazil
- **A:** ***Girardia* sp.** Mexico, USA

Table S1. Samples included in this study, geographical coordinates, collectors, and identification method. The asterisk indicates approximated coordinates based on sampling locality description.

Sample ID	Locality	Taxonomic Identification	Geographic Coordinates		Estimated coordinates	Collector	Identification by
			Latitude	Longitude			
84	France, River Lez	*Girardia* sp.	43,68411110	3,86050000		Laia Leria	External morphology
86	France, River Laderge	*Girardia* sp.	43,72950000	3,32136944		Laia Leria	External morphology
87	France, Lunaç	*Girardia* sp.	43,70821940	3,19550000		Laia Leria	External morphology
109	Spain, Catalonia, St. Llorenç de la Muga, River Muga	*Girardia* sp.	42,32000000	2,79055560		Marta Riutort	External morphology
116	Spain, Catalonia, Riba-roja d'Ebre, River Ebro	*Girardia* sp.	41,24151520	0,43651810		Agustí Munté	External morphology
125	Brazil, Rio Grande do Sul, Constantina	*G. schubarti*	-27,73472220	-52,99222222	*	Ana Leal-Zanchet	External & Internal morphology
126	Brazil, Rio Grande do Sul, Constantina	*G. schubarti*	-27,73472220	-52,99222222	*	Ana Leal-Zanchet	External & Internal morphology
127	Brazil, Rio Grande do Sul, São Leopoldo	*Girardia* sp.	-29,75830830	-51,15338611	*	Ana Leal-Zanchet	External & Internal morphology
131	Brazil, Rio Grande do Sul, São Sebastião de Caí	*G. biapertura*	-29,58797220	-51,38313333	*	Ana Leal-Zanchet	External & Internal morphology
132	Brazil, Rio Grande do Sul, Gramado	*G. tigrina*	-29,39399170	-50,87704444	*	Ana Leal-Zanchet	External & Internal morphology
133	Brazil, Rio Grande do Sul, São Leopoldo	*G. tigrina*	-29,76024170	-51,13541389	*	Ana Leal-Zanchet	External & Internal morphology
134	Brazil, Rio Grande do Sul, Salvador do Sul	*G. tigrina*	-29,45043890	-51,50798333	*	Ana Leal-Zanchet	External & Internal morphology
135	France, Gorges de l'Ardeche	*Girardia* sp.	44,33706940	4,48160833	*	Marta Riutort	External morphology
137	France, Ispagnac	*Girardia* sp.	44,37089170	3,53960556	*	Marta Riutort	External morphology
138	Spain, Catalonia, Girona, Orfes, River Fluvià	*Girardia* sp.	42,17223200	2,87039600	*	Marta Riutort	External morphology
139	Spain, Menorca, Algendar	*Girardia* sp.	39,96622200	3,96615900	*	Sam Pons	External morphology
140	Italy, Sardinia, R.Fungarone, Putifigari	*Girardia* sp.	40,55596000	8,46872400	*	Maria Pala	External morphology
143	Italy, Toscana, Torr.Vincio	*Girardia* sp.	43,94888400	10,83767100	*	Maria Pala	External morphology
247	Spain, Pontevedra, Gondomar	*Girardia* sp.	42,11226110	-8,76286944		Laia Leria	External morphology
248	Spain, Cuenca, Reíllo	*Girardia* sp.	39,90715280	-1,85132222		Laia Leria	External morphology
250	Portugal, Cheleiros	*Girardia* sp.	38,88841000	-9,32990000		Laia Leria	External morphology

252	France, Issalès	*Girardia* sp.	43,51047010	2,39298000		Laia Leria	External morphology
253	Spain, Salamanca, Ciudad Rodrigo	*Girardia* sp.	40,58051000	-6,51324000		Laia Leria	External morphology
260	Brazil, Rio Grande do Sul, São Jose do Norte	*Girardia* sp.	-31,74106600	-51,62754300		João Braccini	External morphology
261	Brazil, São Paulo, Iporanga (cave)	*Girardia* aff. *Arenicola*****	-24,58769800	-48,59291400	*	Rodrigo Ferreira	External morphology
262	Brazil, Mato Grosso do Sul, Bodoquena (cave)	*G. multidiverticulata*	-20,53515800	-56,71542600	*	Lívia Medeiros	External & Internal morphology
263	Brazil, Rio Grande do Sul, São Jose do Norte	*Girardia* sp.	-31,68369900	-51,57208500		João Braccini	External morphology
269	Brazil, Rio Grande do Sul, Constantina	*Girardia* sp.	-27,67584900	-53,03193700	*	Ana Leal-Zanchet	External morphology
270	USA, Baltimore	*Girardia* sp.	39,28576400	-76,61130900	*	?	External morphology
295	Chile, Los Lagos, Huinay Research Station	*Girardia* sp.	-42,37725570	-72,41303561		Ignacio Ribera	External morphology
296	Chile, Los Lagos, Huinay Research Station	*Girardia* sp.	-42,37362400	-72,41374700		Ignacio Ribera	External morphology
297	Chile, Los Lagos, Huinay Research Station	*Girardia* sp.	-42,37204100	-72,40923100		Ignacio Ribera	External morphology
298	Chile, Pumalin Park	*Girardia* sp.	-42,36152780	-72,40422222		Ignacio Ribera	External morphology
299	Chile, Los Lagos, Peninsula Huequi	*Girardia* sp.	-42,45919440	-72,45300000		Ignacio Ribera	External morphology
300	Chile, Los Lagos, Port Montt	*Girardia* sp.	-41,38358330	-73,06541667		Ignacio Ribera	External morphology
308	France, Montpellier	*Girardia* sp.	43,60722800	3,95505300	*	Emili Saló	External morphology
325	Brazil, Mato Grosso do Sul, Bonito (cave)	*G. multidiverticulata*	-21,12961500	-56,49753900	*	Lívia Medeiros	External & Internal morphology
327	USA, Michigan, Ann Arbor	*Girardia* sp.	42,31855000	-83,72958333		Eduard Solà Vázquez	External morphology
336	Brazil, Rio Grande do Sul, Severiano de Almeida	*G. anderlani*	-27,654404	-52,799023	*	Dioneia Vara	External & Internal morphology
337	Brazil, Mato Grosso do Sul, Bodoquena (cave)	*G. multidiverticulata*	-20,53854900	-56,70916800	*	Lívia Medeiros	External & Internal morphology
338	Brazil, Mato Grosso do Sul, Bodoquena (cave)	*G. multidiverticulata*	-20,53644900	-56,70928800	*	Lívia Medeiros	External & Internal morphology
373	Spain, Catalonia, Riera de Mura	*Girardia* sp.	41,72704400	1,91066000	*	Eduardo Mateos Frías	External morphology
377	Spain, Catalonia, Girona, Fluvià, Vilert	*Girardia* sp.	42,17440000	2,82368800		Oleguer Castillo Oliver	External morphology
383	Hawaii, Upper Manoa	*Girardia* sp.	21,28883800	-157,81646000	*	Cory Yap	External morphology
384	Hawaii, Middle Manoa	*Girardia* sp.	21,28480600	-157,82066300	*	Cory Yap	External morphology

399	Brazil, Rio Grande do Sul, Maquine	*Girardia* sp.	-29,62852200	-50,23172700	*	Silvana Amaral	External morphology
401	Brazil, Mato Grosso, Chapada Guimaraes (cave)	*Girardia* sp.	-15,47683700	-55,76426000	*	Lindsey Hellmann	External morphology
402	Brazil, Santa Catarina, Chapeco	*Girardia* sp.	-27,20760100	-52,77371900	*	Ilana Rossi	External morphology
403	Brazil, Rio Grande do Sul, São Leopoldo	*G. tigrina*	-29,75616400	-51,14291400	*	Silvana Amaral	External morphology
466	USA, Carolina Enterprise[a]	*G. dorotocephala*	?	?	*	Lisandra Benítez Álvarez	Carolina Enterprise
467	USA, Carolina Enterprise[a]	*Girardia* sp.	?	?	*	Lisandra Benítez Álvarez	Carolina Enterprise
468	USA, Carolina Enterprise[a]	*Girardia* sp.	?	?	*	Lisandra Benítez Álvarez	Carolina Enterprise
469	Spain, Catalonia, Montjüic	*Girardia* sp.	41,36393500	2,16762200		Lisandra Benítez Álvarez	External morphology
534	Brazil, Rio Grande do Sul, São Francisco de Paula	*G. schubarti*	-29,4422722	-50,57976667	*	Ana Laura Morais	External & Internal morphology
535	Brazil, Rio Grande do Sul, São Leopoldo	*G. tigrina*	-29,7614222	-51,16495278	*	Rafaella Canello	External & Internal morphology
542	Germany, Pillnitz	*Girardia* sp.	51,0068694	13,87168056	*	Miquel Vila	External morphology
543	USA, Virginia	*Girardia* sp.	37,4311778	-78,65413889	*	Miquel Vila	External morphology
544	Australia, Tasmania, Derwent River	*Girardia* sp.	-42,13415	146,2304833	*	Miquel Vila	External morphology
545	USA, Virginia, Ashburn, Janelia Farm Research Campus	*Girardia* sp.	39,0716361	-77,46453333	*	Miquel Vila	External morphology
546	Francia, Lez	*Girardia* sp.	43,6109972	3,896725	*	Miquel Vila	External morphology
547	Spain, Catalonia, Barcelona, Vallvidrera	*Girardia* sp.	41,4170778	2,090980556	*	Miquel Vila	External morphology
548	Mexico	*Girardia* sp.	23,6345	-102,5527833	*	Miquel Vila	External morphology
550	Mexico	*Girardia* sp.	23,6345	-102,5527833	*	Miquel Vila	External morphology
551	France, River Herault	*Girardia* sp.	43,6936556	3,572469444		Miquel Vila	External morphology
552	Germany, Zschorna	*Girardia* sp.	51,2509861	13,74744722	*	Miquel Vila	External morphology
553	Germany, Pillnitz	*Girardia* sp.	51,0068694	13,87168056	*	Miquel Vila	External morphology
554	Netherlands, Leiden	*Girardia* sp.	52,1833861	4,517911111	*	Miquel Vila	External morphology
556	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*Girardia* sp.	42,1631806	2,955611111		Marta Riutort	External morphology
558	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*Girardia* sp.	42,161142	2,958923		Marta Riutort	External morphology

559	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*Girardia* sp.	42,161142	2,958923		Marta Riutort	External morphology
659	Canada, Nova Scotia, Ainslie Lake	*Girardia* sp.	46,117075	-61,21903889		Tobias Boothe	External morphology
660	Australia, Queensland, UQ Lakes	*Girardia* sp.	-27,5048083	153,0194083		James Cleland	External morphology
661	Australia, Queensland	*Girardia* sp.	-27,5459972	153,1837306		James Cleland	External morphology
679	USA, Michigan, Douglas Lake	*Girardia* sp.	45,5801889	-84,66963333		Eduard Solà Vázquez	External morphology
683	Japan, Hoshikuki-cho, Mizu-no-sato Park, Miyako River	*Girardia* sp.	35,6041194	140,1531611		Eduard Solà Vázquez	External morphology
684	Japan, Hoshikuki-cho, Mizu-no-sato Park, Miyako River	*Girardia* sp.	35,6041194	140,1531611		Eduard Solà Vázquez	External morphology
685	Chile, Talagante, Mapocho River	*Girardia sanchezi*	-33,6612444	-70,93283333		Eduardo Ascarruz	Type locality
686	China, Conghua, Yadongxi River	*Girardia sp.*	23,7	113,7166667		Yuan Changjur	External morphology
687	Chile, Los Lagos, Huinay Research Station	*Girardia sp.*	-42,3772557	-72,41303561		Roger Vila	External morphology
1056	Mexico, Biological Station Los Tuxtlas 1	*Girardia sp.*	18,58547	-95,07574		Lisandra Benítez Álvarez	External morphology
1059	Mexico, Biological Station Los Tuxtlas 1	*Girardia sp.*	18,58547	-95,07574		Lisandra Benítez Álvarez	External morphology
1062	Mexico, Los Tuxtlas, Laguna Escondida	*Girardia sp.*	18,59184	-95,0877		Lisandra Benítez Álvarez	External morphology
1063	Mexico, Los Tuxtlas, Laguna Escondida	*Girardia sp.*	18,59184	-95,0877		Lisandra Benítez Álvarez	External morphology
1070	Mexico, Biological Station Los Tuxtlas 2	*Girardia sp.*	18,58208	-95,07276		Lisandra Benítez Álvarez	External morphology
1072	Cuba, Matanzas, Martí, El Huequito	*Girardia sp.*	22,9906889	-80,94398889		Lisandra Benítez Álvarez	External morphology
1178	Brazil, Bahía, Chapada Diamantina, Vale do Pati (cave)	*Girardia sp.*	-12,771346	-41,447793		Pau Balart	External morphology
1179	Mexico, Mexico City, Xochimilco, Cuemanco	*Girardia sp.*	19,2891483	-99,10182917		Omar Lagunas	External morphology
1180	Mexico, Mexico City, Xochimilco, Cuemanco	*Girardia sp.*	19,2891483	-99,10182917		Omar Lagunas	External morphology
1181	Mexico, Mexico City, Xochimilco, Cuemanco	*Girardia sp.*	19,2891483	-99,10182917		Omar Lagunas	External morphology
1182	Mexico, Michoacán	*Girardia sp.*	19,5757278	-101,6678417	*	Omar Lagunas	External morphology
F6510	Brazil	*Girardia sp.*	-24,6815444	-54,19687778		Fernando Carbayo	External morphology
InoueA_CS	USA, Texas, Caroline Spring, Independence Creek	*Girardia sp.* A	30,46912	-101,80352		MN652340.1	Inoue *et al.,* 2020
InoueB_ES	USA, New Mexico, Palomas Creek, Emrick Spring	*Girardia sp.* B	33,150283	-107,671083		MN652378.1	Inoue *et al.,* 2020
InoueC_BLBC	USA, Texas, Bitter Lake, Bitter Creek	*Girardia sp.* C	33,479572	-104,427369		MN652301.1	Inoue *et al.,* 2020
InoueD_GR	USA, New Mexico, West Fork of the Gila River	*Girardia sp.* D	33,248033	-108,300833		MN652373.1	Inoue *et al.,* 2020

GB_Gan1_Br	Brazil	*G. anderlani*			?	DQ666038.1	GenBank Information
GB_Gsi_China	China, Guangdong Province Xinghu Lake in Zhaoqing	*G. sinensis*	23,0758333	112,4786111		KP091895.1	GenBank Information
GB_Gti_FrMont	France, Montpellier	*G. tigrina*			?	DQ666042.1	GenBank Information
GB_Gsc1_Br	Brazil	*G.schubarti*			?	DQ666041.1	GenBank Information
GB_Gsc2_BrRGScon	Brazil, Rio Grande do Sul, Constantina	*G.schubarti*			?	KJ599691.1	GenBank Information
GB_som1_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	-41,0019444	-66,66333333		MW271865	Lenguas *et al.,* 2021
GB_som2_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	-41,0019444	-66,66333333		MW271866	Lenguas *et al.,* 2021
GB_som3_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	-41,0019444	-66,66333333		MW271867	Lenguas *et al.,* 2021
GB_som4_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	-41,0019444	-66,66333333		MW271869	Lenguas *et al.,* 2021
GB_tom1_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. tomasi*	-40,68	-66,17333333		MW271863	Lenguas *et al.,* 2021
GB_tom2_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. tomasi*	-40,68	-66,17333333		MW271864	Lenguas *et al.,* 2021
GB_tom3_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. tomasi*	-40,68	-66,17333333		MW271868	Lenguas *et al.,* 2021

**: the individual sequenced belongs to the same sample of the type-material of *G. arenicola* and *G. paucipunctata*, two syntopic species in a cave. The material was sent for DNA extraction before morphological analysis and taxonomic identification.

Table S2. *COI* and *EF1α* sequences of *Girardia* genus present in GenBank. In bold are indicated the sequences included in the analyses. Reason for exclusion is indicated for certain sequences not included in the present study.

TAXON	Reason of exclusion	COI		EF1a	
		Length	Accession number	Length	Accession number
Girardia anderlani		**315 bp**	**DQ666038.1**		
Girardia anderlani	1	373 bp	AF178313.1		
Girardia dorotocephala	1	777 bp	KM200929.1		
Girardia dorotocephala	4	414 bp	AF178314.1		
Girardia schubarti		**332 bp**	**DQ666041.1**		
Girardia schubarti				**612 bp**	**KJ599691.1**
Girardia sinensis	3	904 bp	KP091891.1		
Girardia sinensis	3	904 bp	KP091892.1		
Girardia sinensis	3	904 bp	KP091893.1		
Girardia sinensis	3	904 bp	KP091894.1		
Girardia sinensis		**904 bp**	**KP091895.1**		
Girardia tigrina		**393 bp**	**DQ666042.1**		
Girardia tigrina	2			991 bp	AJ250913.1
Girardia tigrina	2	777 bp	KM200930.1		
Girardia tigrina	1	435 bp	AF178316.1		
Girardia tigrina	3	519 bp	MN092348.1		
Girardia sp.		**213 bp**	**MN652301.1**		
Girardia sp.		**213 bp**	**MN652340.1**		
Girardia sp.		**213 bp**	**MN652373.1**		
Girardia sp.		**213 bp**	**MN652378.1**		
Girardia sp.	3	213 bp	MN652302-MN652339		
Girardia sp.	3	213 bp	MN652341-MN652372		
Girardia sp.	3	213 bp	MN652374-MN652377		
Girardia sp.	3	213 bp	MN652379-MN652552		
Girardia somuncura		**719 bp**	**MW271865**		
Girardia somuncura		**716 bp**	**MW271866**		
Girardia somuncura		**688 bp**	**MW271867**		
Girardia somuncura		**737 bp**	**MW271869**		
Girardia tomasi		**725 pb**	**MW271863**		
Girardia tomasi		**738 bp**	**MW271864**		
Girardia tomasi		**718 bp**	**MW271868**		

Reasons of exclusion: 1, low quality; 2, incertain classification; 3, not necessary; 4, long branch in trees

References

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

Ball, I. R. (1975). Nature and formulation of biogeographical hypotheses. *Systematic Zoology*, *24*(4), 407–430.

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

Benítez-Álvarez, L. et al. Unraveling a world-wide colonization: molecular phylogenetics and niche modelling reveal the invasion biology of three species of *Girardia* freshwater planarians (Tricladida, Plathyelminthes).

Benya, E. G. F., Leal-Zanchet, A. M., Santos, W. H., Hauser, J., & Erdtmann, B. (2007). Chromosome polymorphism and complements in populations of *Girardia* species (Platyhelminthes, Tricladida, Paludicola) from southern Brazil. *Brazilian journal of biology*, *67*(4 Suppl), 951–5. https://doi.org/10.1590/S1519-69842007000500021

Chen, X., Chen, Y., Wu, C., & Wang, A. (2015). A new species of the genus *Girardia* (Tricladida: Dugesiidae) from China. *Zoological Systematics*, *40*(2), 166–178. https://doi.org/10.11865/zs.20150202

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology*, *53*(1), 401–410. https://doi.org/10.1080/10635150490265012

Gourbault, N. (1977). Etude descriptive et cytotaxonomique d'une planaire polynesienne *Dugesia tahitiensis* n. sp. (Turbellarie, Triclade). *Annales de Limnologie*, *13*(2), 211–220.

Graham, S. W., Olmstead, R. G., & Barrett, S. C. H. (2002). Rooting Phylogenetic Trees with Distant Outgroups: A Case Study from the Commelinoid Monocots. *Molecular Biology and Evolution*, *19*(10), 1769–1781. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A003999

Grant, L. J., Sluys, R., & Blair, D. (2006). Biodiversity of Australian freshwater planarians (Platyhelminthes: Tricladida: Paludicola): New species and localities, and a review of paludicolan distribution in Australia. *Systematics and Biodiversity*, *4*(4), 435–471. https://doi.org/10.1017/S1477200006002064

Grant, L. J. (2017). *Biodiversity, Biogeography and Phylogeny of Australian Freshwater Triclads (Platyhelminthes: Tricladida: Continenticola)*. PhD Thesis, James Cook University. Retrieved from https://researchonline.jcu.edu.au/51512/1/51512-grant-2016-thesis.pdf

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, *41*, 95–98.

Hellmann, L., Ferreira, R. L., Rabelo, L., & Leal-Zanchet, A. M. (2020). Enhancing the still scattered knowledge on the taxonomic diversity of freshwater triclads (Platyhelminthes: Dugesiidae) in caves from two Brazilian Biomes. *Studies on Neotropical Fauna and Environment*, *00*(00), 1–16. https://doi.org/10.1080/01650521.2020.1829901

Hellmann, L., Leal-Zanchet, A. M., & Lopes Ferreira, R. (2018). Two new sympatric troglobitic freshwater flatworms (platyhelminthes: Dugesiidae) from a hotspot of subterranean biodiversity in the neotropics. *Zootaxa*, *4438*(3), 561–574. https://doi.org/10.11646/zootaxa.4438.3.8

Inoue, K., Pohl, A. L., Sei, M., Lang, B. K., & Berg, D. J. (2020). Use of species delimitation approaches to assess biodiversity in freshwater planaria (Platyhelminthes, Tricladida) from desert springs. *Aquatic Conservation: Marine and Freshwater Ecosystems*, aqc.3273. https://doi.org/10.1002/aqc.3273

Jorge, M. A. P., Alvarez, L. E., & Almeida, E. J. C. (2000). Occurrència de mixoploidia em exemplares de Girardia schubarti (Platyhelminthes, Tricladida, Paludicola). *Genetics & Molecular Biology*, *20*(3), Suppl.: 5-6.

Kawakatsu, M., Oki, I., Tamura, S., Yamayoshi, T., Hauser, J., & Friedrich, S. M. G. (1984). Morphological, karyological and taxonomic studies of freshwater planarians from South Brazil. VI. Dugesia schubarti (Marcus, 1946) from the vicinity of São Leopoldo, Estado de Rio Grande do Sul (Turbellaria, Tricladida, Paludicola). *Bulletin of Fuji Women's College*, *No. 22, se*(ser.II), 45–62.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kenk, R. (1974). Index of the genera and species of the freshwater triclads (Turbellaria) of the world. *Smithsonian Contributions to Zoology*, *183*, 1–90. https://doi.org/10.5479/si.00810282.183

Knakievicz, T., Lau, A. H., Prá, D., & Erdtmann, B. (2007). Biogeography and karyotypes of freshwater planarians (Platyhelminthes, Tricladida, Paludicola) in southern Brazil. *Zoological Science*, *24*(2), 123–129. https://doi.org/10.2108/zsj.24.123

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695–1701. https://doi.org/10.1093/molbev/mss020

Lázaro, E. M., 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*(1), 274. https://doi.org/10.1186/1471-2148-11-274

Lenguas-Francavilla, M., Negrete, L., Martínez-Aquino, A., Damborenea, C., & Brusa, F. (2021). Two new freshwater planarian species (Platyhelminthes: Tricladida: Dugesiidae) partially sympatric in Argentinean Patagonia. *Canadian Journal of Zoology*, *99*, 269–278. https://doi.org/10.1139/cjz-2020-0169

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

Maddison, W. P., & Maddison, D. R. (2015). Mesquite: A modular system for evolutionary analysis. Version 3.04. *http://www.mesquiteproject.org*.

Marcus, E. (1946). Sôbre Turbellaria límnicos brasileiros. *Boletim da Faculdade de Filosofia, Ciências e Letras. Zoologia. Universidade de São Paulo*, *11*, 5–254.

McGirr, R., Seton, M., & Williams, S. (2021). Kinematic and geodynamic evolution of the Isthmus of Panama region: Implications for Central American Seaway closure. *GSA Bulletin*, *133*(3–4), 867–884. https://doi.org/10.1130/B35595.1

Morais, A. L., Bichuette, M. E., Chagas-Júnior, A., & Leal-Zanchet, A. (2021). Under threat: Two new troglobitic species of Girardia (Platyhelminthes: Tricladida) from sandstone and limestone caves in Brazil. *Zoologischer Anzeiger*, *293*, 292–302. https://doi.org/10.1016/J.JCZ.2021.06.015

O'Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., et al. (2016). Formation of the Isthmus of Panama. *Science Advances*, *2*(8). https://doi.org/10.1126/sciadv.1600883

Puccinelli, I., & Deri, P. (1991). Comparative karyological analysis of some American planarians belonging to the genus *Dugesia* (subgenus *Girardia*) (Platyhelminthes, Tricladida). *Caryologia*, *44*(3–4), 225–232. https://doi.org/10.1080/00087114.1991.10797189

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/SYSBIO/SYY032

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

Sluys, R. (2001). Towards a phylogenetic classification and characterization of dugesiid genera (Platyhelminthes, Tricladida, Dugesiidae): a morphological perspective. In D. T. J. Littlewood & R. D. Bray (Eds.), *Interrelationships of the*

Platyhelminthes. Syst. Ass. Spec. Vol. Ser. 60 (pp. 57–73). London, England: Taylor & Francis.

Sluys, R., Joffe, B., & Cannon, L. R. G. (1995). An alien flatworm in Australian waters. *Memoirs of the Queensland Museum*, *38*(2), 642.

Sluys, R., & Kawakatsu, M. (2001). Contribution to an inventory of the freshwater planarians of Australia and New Zealand (Platyhelminthes, Tricladida, Dugesiidae), with distribution maps of the species examined. *Beaufortia/Bull Zool Mus Univ Amsterdam*, *51*(10), 163–198.

Sluys, R., Kawakatsu, M., & Ponce de León, R. (2005). Morphological stasis in an old and widespread group of species: Contribution to the taxonomy and biogeography of the genus *Girardia* (Platyhelminthes, Tricladida, Paludicola). *Studies on Neotropical Fauna and Environment*, *40*, 155–180.

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

Solà, E., & et al. (2022). Three dispersal routes out of Africa: The puzzling biogeographical history of the freshwater planarian *Dugesia* (Platyhelminthes: Tricladida) revealed by molecular data. *Journal of Biogeography*.

Souza, S., Morais, A. L., Bichuette, M. E., & Leal-Zanchet, A. M. (2016). Two new species of freshwater flatworms (Platyhelminthes: Tricladida: Continenticola) from South American caves. *Zootaxa*, *4092*(1), 107–121.

Souza, S. T., Morais, A. L. N., Cordeiro, L. M., & Leal-Zanchet, A. M. (2015). The first troglobitic species of freshwater flatworm of the suborder continenticola (Platyhelminthes) from South America. *ZooKeys*, *16*(470), 1–16. https://doi.org/10.3897/zookeys.470.8728

Xia, X. (2017). DAMBE6: New tools for microbial genomics, phylogenetics, and molecular evolution. *Journal of Heredity*, *108*(4), 431–437. https://doi.org/10.1093/jhered/esx033

Xia, X., & Lemey, P. (2009). Assessing substitution saturation with DAMBE. In P. Lemey, M. Salemi, & A.-M. Vandamme (Eds.), *The phylogenetic handbook: a*

practical approach to DNA and protein phylogeny (2nd ed., pp. 615–630). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511819049.022

Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular phylogenetics and evolution*, *26*, 1–7. https://doi.org/10.1016/s1055-7903(02)00326-3

Chapter III

Section 1

Niche modelling and molecular phylogenetics unravel the invasion biology and worldwide colonization of three species of the freshwater planarian genus *Girardia* (Platyhelminthes, Tricladida)

LISANDRA BENÍTEZ-ÁLVAREZ, LAIA LERIA, RONALD SLUYS, ANA MARIA LEAL-ZANCHET, AND MARTA RIUTORT

Niche modelling and molecular phylogenetics unravel the invasion biology and worldwide colonization of three species of the freshwater planarian genus *Girardia* (Platyhelminthes, Tricladida)

Lisandra Benítez-Álvarez[1,*], Laia Leria[1,*], Ronald Sluys[2], Ana María Leal Zanchet[3], Marta Riutort[1,♦]

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

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

[3] Instituto de Pesquisas de Planárias and Programa de Pós-Graduacão em Biologia, Universidade do Vale do Rio dos Sinos (UNISINOS), 93022-750 São Leopoldo, Rio Grande do Sul, Brazil

*Lisandra Benítez-Álvarez and Laia Leria contributed equally to this work

♦corresponding author: mriutort@ub.edu

Abstract

Several species of the freshwater planarian genus *Girardia* have been introduced into freshwater ecosystems all over the world, but little is known about the actual number of species involved and about possible detrimental effects on autochthonous ecosystems. This is due to a paucity of systematic studies focusing on the native region of the genus and to the fact that most introduced *Girardia* show asexual reproduction and do not possess a copulatory apparatus, thus hampering taxonomic identification. In the present study we used molecular phylogenetics and niche modelling under present and future climatic scenarios, in order to examine human-induced dispersal and spread of alien species of *Girardia* from their original areas of distribution in the Americas to other parts of the globe. Samples of *Girardia* were analysed from all along its present distribution. Our results show that the *Girardia* populations spreading over the world belong to three species of North American origin: *G. dorotocephala*, *G. sinensis*, and *G. tigrina*. Our study brings to light that *G. sinensis* is native to North America, from where it colonised China, as well as other regions of the world, including Europe and Australia. It also shows that *G. dorotocephala* has a broader range of localities where it was introduced than previously known, including Europe and Brazil. Niche modelling revealed that the three colonising species have a broad range of potential distribution in extensive regions of the Northern Hemisphere. Their environmental requirements, being generalists with a high suitability for human-

modified habitats, favour colonization. In the Iberian Peninsula, introduced *G. tigrina* and *G. sinensis* have extensive areas of high suitability, overlapping with the more limited suitable areas of autochthonous freshwater planarians. Colonization modelling of future situations, regardless of the climatic scenario, points to an extension of their distributional range towards northern Europe, albeit without diminishing high suitability of regions in the south.

Keywords

Alien species, Dugesiidae, distribution, environmental requirements, colonization, climate scenarios, freshwater ecosystems

1. Introduction

Species distributions may differ considerably in terms of range size, with some species being restricted to small geographic regions and others ranging all over the world. Generally, distributional ranges covering large parts of the globe are not the result of natural causes but frequently have been achieved through human-induced dispersal, with species having been transferred, either intentionally or accidentally, from their original areas to non-native regions (Hulme, 2009). When introduced species manage to successfully survive and reproduce in allochthonous habitats, there is always the risk that they may become invasive and then strongly affect the dynamics of the local ecosystem, which may result in irreversible loss of autochthonous biodiversity (Vitousek et al., 1997; Velde et al., 2006). Such invasions of alien species are considered to form one of the major threats to global biodiversity and to form the second major cause of animal extinctions (Gherardi et al., 2008, and references therein).

The Order Tricladida represents one of the best-known groups of free-living flatworms, with representatives in all biogeographical regions of the world (Schockaert et al., 2008), of which the Suborder Continenticola (including both terrestrial and freshwater representatives) houses the largest number of species. Recent studies have reported the introduction of many species of tropical terrestrial planarians all over the world (Álvarez-Presas et al., 2014; Justine et al., 2014; Jones & Sluys, 2016; Mateos et al., 2020; Mazza et al., 2016; Negrete et al., 2020), with some species almost being naturalized in their non-native habitats (Álvarez-Presas et al., 2014) and some representing a problem for agriculture (Murchie & Gordon, 2013). These studies not only have unravelled the country of origin of these introduced species but have pointed out also the putative causes of these introductions, as well as potential areas for future range extension (Álvarez-Presas et al., 2014; Negrete et al., 2020), which concern vital knowledge for the development of control measures. In contrast, introductions of invasive freshwater planarians have received much less attention, of which the genus *Girardia* Ball, 1974 is of special importance.

The genus *Girardia* includes about 52 valid species, the natural distribution of which extends from the southern regions of Argentina to Southern Canada. Furthermore, species of *Girardia* have been introduced into many other regions of the world. The species *Girardia tigrina* (Girard, 1850) was first reported in Germany in the

1920's, after which it was recorded also from other parts of mainland Western Europe, as well as the Balearic Islands and the Azores, its spread most likely effectuated through the international trade in aquatic plants and the activity of aquarists (Stocchino et al., 2019 and references therein). Representatives of the genus have been reported also from Australia, Japan, and Hawaii (Kawakatsu, Mitchell, Hirao, & Tanaka, 1984; Sluys, Kawakatsu, & Ponce de León, 2005; Sluys, Kawakatsu, & Yamamoto, 2010). In addition, records of *Girardia* in China, Serbia and Eastern Europe have been published in the past few years (Chen, Chen, Wu, & Wang, 2015; Ilić et al., 2018; Kanana & Riutort, 2019). For Australia, occurrence of introduced *G. tigrina* was established (Sluys et al., 1995 and references therein), apart from three presumed autochthonous species of *Girardia* (cf. Grant et al., 2006; Sluys & Kawakatsu, 2001). However, recent molecular work (Grant, 2017) has revealed that the latter three species (*G. sphincter* Kawakatsu & Sluys, 2001; *G. graffi* (Weiss, 1909); *G. informis* Sluys & Grant, 2006) do not belong to the genus *Girardia*.

For long time it has been assumed that only *G. tigrina* was introduced in Europe, although some studies have pointed to the possibility that other alien *Girardia* species may be present, in view of the diversity in morphology, karyology, and reproductive biology of the populations analysed from the Iberian Peninsula (Ribas, Riutort, & Baguñà, 1989) and Italy (Benazzi, 1993; Stocchino et al., 2019). However, due to the fact that, generally, the introduced populations of *Girardia* show only asexual reproduction and do not possess a copulatory apparatus, their identification to species level is greatly hampered, since the main diagnostic characters reside in the reproductive structures (Sluys & Riutort, 2018). This problem of identification of asexual strains is alleviated by using molecular markers, a procedure successfully applied in various studies on planarians, both for identification of species and for resolution of their phylogenetic relationships (see Álvarez-Presas & Riutort, 2014). Therefore, a recent molecular study has pointed out that three species belonging to North American lineages are spread trough Europe and other regions out of America (Chapter II).

An aspect of considerable importance and concern regarding introduced alien species involves their potential invasiveness and detrimental effects on autochthonous ecosystems, an issue that is particularly important for freshwater habitats, since these communities are highly sensitive to compositional changes due to the introduction of

invasive species (Havel, Kovalenko, Thomaz, Amalfitano, & Kats, 2015; Gallardo, Clavero, Sánchez, & Vilà, 2016). Unfortunately, only few studies have tried to evaluate the effects of introduced *Girardia* on local ecosystems, mainly in the United Kingdom, (Gee & Young, 1993; Pickavance, 1971; Velde, 1975; Wright, 1987; Gee, Pickavance, & Young, 1998). Therefore, at present we lack adequate knowledge on the potential danger of introduced *Girardia* to autochthonous species and possible cascading negative effects on the freshwater environment. Ecological niche modelling offers the possibility to evaluate the potential distribution of species and to estimate at the same time the environmental parameters underlying the distributional spread. By analysing and comparing the environmental requirements of introduced, as well as native species, an indirect evaluation is provided of the potential influence exerted by the introduced species on autochthonous ones (Warren & Seifert, 2011; Radosavljevic & Anderson, 2014; Puchałka et al., 2020).

The main objectives of the present study were (a) showing the origin and current diversity of introduced European populations of *Girardia,* (b) identifying their potential area of distribution and environmental requirements, and (c) predicting future tendencies in their potential distribution. To this end, we constructed a molecular phylogeny of the genus *Girardia*, including both native and introduced representatives from all over the world. In addition, a distribution modelling analysis was performed on the introduced species, in order to determine their potential area of distribution, with special focus on the Iberian Peninsula, where the distribution of autochthonous species is well established and equivalent studies have already been performed. Finally, we ran a niche modelling analysis under two different climatic scenarios, in order to predict future, potentially suitable areas for the three introduced species.

2. Materials and methods

2.1. Data acquisition

*Girardia´*s sequences of *Cytochome Oxidase I* (COI) and *Elongation factor 1 alpha* (EF1α) genes from representative samples from Asia, Australia, Hawaii, the Americas, and Europe were downloaded from GenBank. We analysed a total of 98 *Girardia* representatives, originating from 76 localities that are distributed all over the native range of the genus and from several regions in which species were introduced; Europe constituted the most densely represented region (Table S1, Fig. 1).

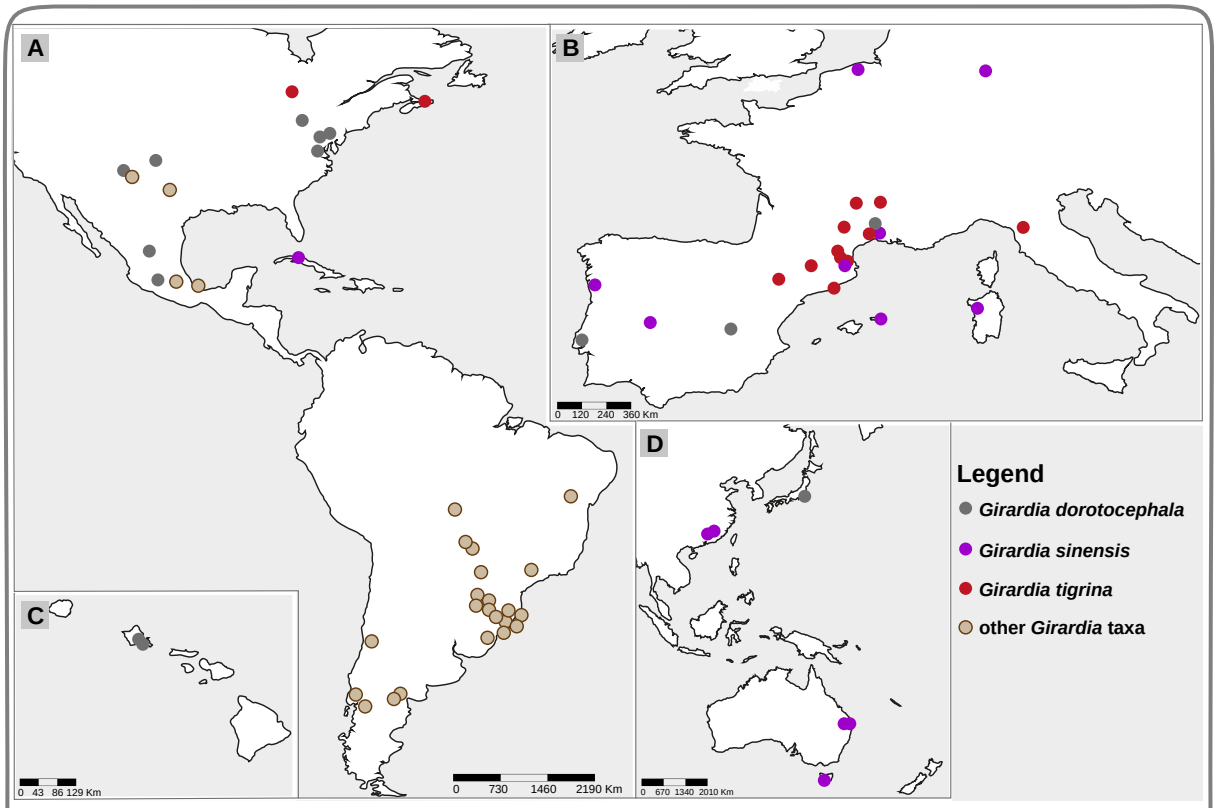


Figure 1. Maps showing the sampling localities for all sequences analysed in the present study. A: The Americas; B: Western Europe; C: Hawaii; D: Asia and Oceania. For a finer resolution, see the interactive map at https://www.ub.edu/planarian-maps/

2.2. Sequence alignment

Sequences of COI and EF1α were aligned independently with ClustalW on the BioEdit Sequence Alignment Editor (Hall, 1999). A concatenated dataset of both genes was obtained in Mesquite v3.04 (Maddison & Maddison, 2015). Missing data were coded by Ns. PartitionFinder v2.1.1 (Lanfear, Calcott, Ho, & Guindon, 2012) was used to find the best partition scheme for each gene, considering the score for the Bayesian Information Criterion (BIC). The results of the PartionFinder program validated a partition scheme considering each codon site independently, both for COI and EF1α. For each partition the best model was General Time Reversible + Gamma Distribution + Invariable Sites (GTR + Γ + I).

2.3. Phylogenetic Inference

Bayesian Inference (BI) method was used to infer a phylogenetic tree from the concatenated dataset. Bayesian analyses were run in MrBayes v3.2.2 (Ronquist et al., 2012) with 10 million generations, sampling every 1000 generations, and 25% burn-in (default setting) to obtain the best tree and posterior probabilities. The best evolutionary model and partition scheme obtained in Partitionfinder was applied with parameters unlinked. Convergence of the parameters' values and topologies was examined by checking that the average standard deviation of split frequencies was below 0.01. Estimated sample size values (ESS) of each run were inspected in Tracer v1.5 (Rambaut et al., 2018) and only runs with values for all parameters above 200 were accepted.

2.4. Species distribution modelling

Based on the molecular analysis, we identified the lineages of *Girardia* that had been introduced into different regions of the world, corresponding to three species, viz., *Girardia tigrina*, *G. sinensis* Chen & Wang, 2015 and *G. dorotocephala* (Woodworth, 1897) (see Results). The software Maxent 3.4.0 (Phillips & Dudík, 2008; Phillips, Anderson, Dudík, Schapire, & Blair, 2017) was used to infer the potential area of distribution of these three species, both at a global scale and for the Iberian Peninsula in particular. In this analysis, worldwide distributional presence data was included for each of the three introduced *Girardia* species, including localities in both native and non-native regions. Six hydro-environmental variables were downloaded from the RiverATLAS database (Linke et al., 2019), at a spatial resolution of either 15 or 30 arc-seconds, for the variables Temperature, Natural Discharge, Precipitation, Land Cover Classes, Lithological Classes, and Slope. In a recent study on planarians of the genus *Dugesia* Girard, 1850 from the Western Mediterranean region it was shown that these variables are not correlated (Leria et al. in press). Future potential distribution of these three *Girardia* species was inferred under two different scenarios of climate change. The bioclimatic variables Annual Mean Temperature (Bio1) and Annual Precipitation (Bio12) were downloaded from the WorldClim2 database (Fick & Hijmans, 2017) at a spatial resolution of 2.5 arc-minutes for the time periods corresponding to Current, 2021-2040, and 2081-2100. For each of the future time periods, two scenarios of greenhouse gas emissions were considered, viz., (1) highest

decrease of emissions (SSP1-2.6,) and (2) highest increase of emissions (SSP5-8.5) (see https://www.carbonbrief.org/cmip6-the-next-generation-of-climate-models-explained for the detailed characteristics of each scenario; last date visited: 24 March 2022). All Maxent analyses were performed by using 10 cross-validated replicates and setting the output to cloglog format. For each analysis, performance of the model was evaluated by checking the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) (Fielding & Bell, 1997). In total, the following number of localities were used for each species: 16 (*G. tigrina*), 17 (*G. sinensis*), 18 (*G. dorotocephala*) (Table S1).

3. Results

3.1. Sequences, alignment, phylogenetic analysis, and distribution

A total of 98 *Girardia* sequences were used for the present work. The concatenated dataset included 94 sequences of COI (837 base pairs - bp) and 78 of EF1α (879 bp) (Table S1). In order to avoid rooting with outgroup taxa that might be too distantly related to the ingroup, the tree was rooted on *G. schubarti* + *Girardia* sp. 1 (below indicated as the *G. schubarti* group). A recent work on the phylogenetic relationships within the genus *Girardia* supports this rooting by the *G. schubarti* group (Chapter II).

Our phylogenetic tree (Fig. 2) recognized monophyletic groups for the individuals of each of the identified species and delimited also seven clades for the unidentified individuals originating from Mexico, Brazil, and Chile (*Girardia* sp. 1 to sp. 7; Fig 2, Table S1). All of these clades were fully supported (>0.99 PP), excepting only *G. sanchezi* (0.67 PP).

In the phylogenetic tree (Fig. 2), all non-American specimens are grouped into three clades, corresponding to the species *G. tigrina, G. sinensis,* and *G. dorotocephala*. The topology shows that the three species are closely related, being connected by relatively short branches. The former two species form a clade, which shares a sister-group relationship with the last-mentioned species. Together, the clade of these three species is sister to an unidentified species from Mexico (*Girardia* sp. 7, 1.00 PP).

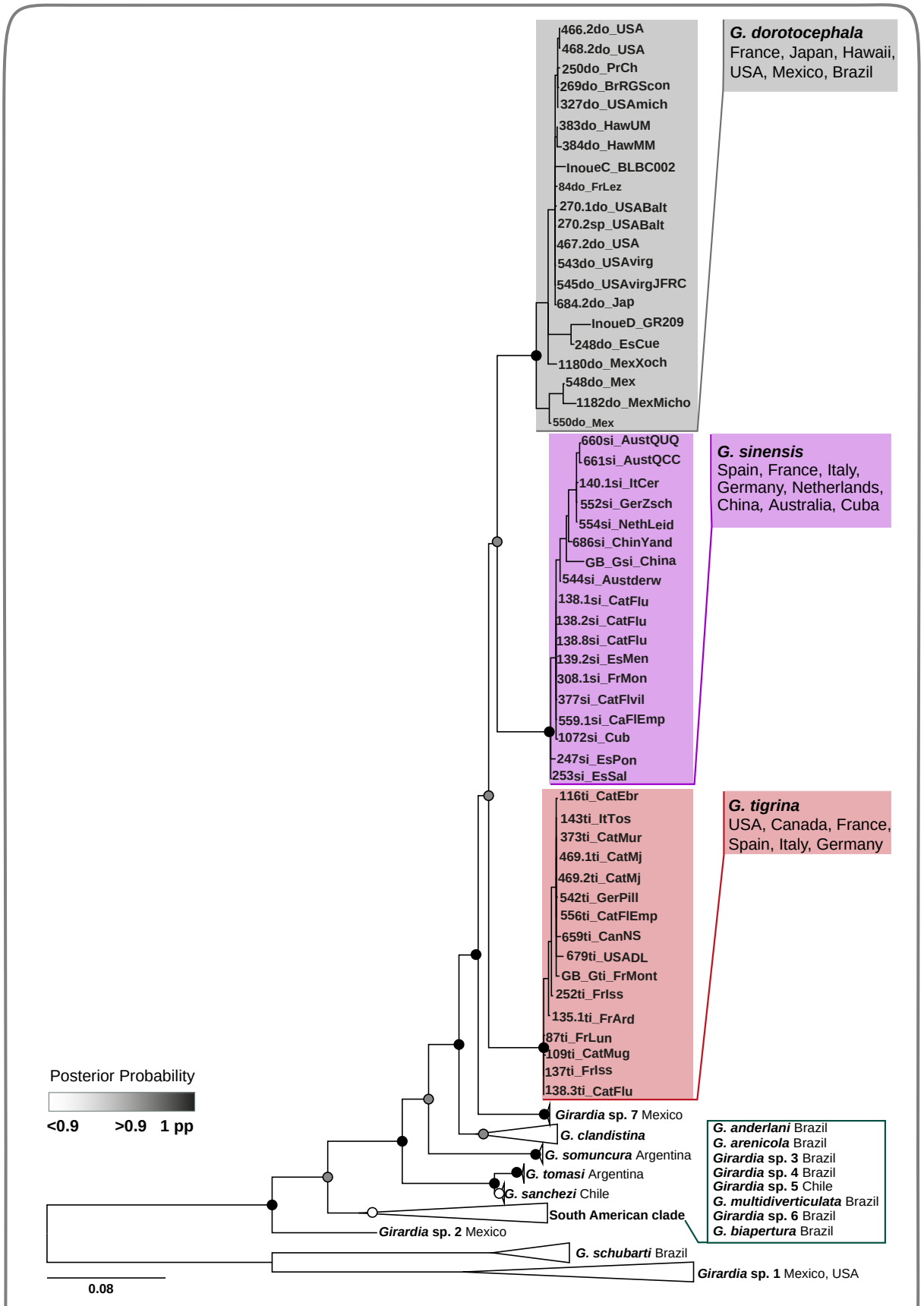


Figure 2. Bayesian Inference tree inferred from the concatenated dataset (COI + EF1α). Most clades have been collapsed for the sake of clarity, excepting those of the three introduced species of *Girardia*. Circles at nodes indicate posterior probability. Scale bar: number of substitutions per nucleotide position

The *G. tigrina* clade comprises Operational Taxonomic Units (OTUs) from Michigan, USA (code 679sp_USADL), and Nova Scotia, Canada (code 659sp_CanNS). All other OTUs in this clade originate from various European countries, viz., France, Spain, Germany, and Italy (Figs 2, S1). It is highly noteworthy that the *G. dorotocephala* clade comprises OTUs from widely separated areas, such as USA, Mexico, Japan, Hawaii, Brazil, Portugal, Spain and France (Figs 2, S1). In similar vein, it is noteworthy and interesting that the *G. sinensis* clade comprises individuals from Australia, China, Cuba, Spain (mainland, as well as the island of Menorca), Italy (Sardinia Island), France, Germany, and The Netherlands (Figs 2, S1).

3.2. Niche modelling

3.2.1. Potential distribution of *Girardia*

The distribution models inferred with Maxent for *G. tigrina*, *G. sinensis* and *G. dorotocephala* gave AUC values close to 1.00 (Table 1), indicating that all models had high predictive power for the potential distribution of these species. Two of the environmental variables that contributed the most to the current models were Temperature (T) and Land Cover Classes (LCC), which in all cases showed a combined percentage of contribution >60%, while the variables that contributed less were Slope (S) and Natural Discharge (ND) of rivers (Table 1).

Table 1. Contribution, in percentages, of the various variables used in the niche modelling for predicting the potential distribution of *Girardia tigrina*, *G. sinensis*, and *G. dorotocephala*. AUC: Area Under the Curve of the Receiver Operating Characteristic.

	G. tigrina	***G. sinensis***	***G. dorotocephala***
AUC	0.918 ± 0.069	0.857 ± 0.091	0.925 ± 0.063
Land Cover Classes	55.9 %	39.4 %	21.3 %
Temperature	33.8 %	22.3 %	42.2 %
LithologicalClasses	4.2 %	17.1 %	28.2 %
Precipitation	4.5 %	12.6 %	3.2 %
Slope	1.6 %	3.3 %	4.3 %
Natural Discharge	0 %	5.3 %	0.8 %

In the following we have considered as being of “low suitability” values below 0.25, as “moderately suitable” values equal or above 0.25 but below 0.6, and “highly suitable” being represented by values equal or above 0.6. Regarding LCC, all three

clades showed a very high suitability for Artificial Surfaces and Associated Areas (e.g., canals and dams), and a high suitability for different types of tree cover (with differences in their preferences, depending on the species), and for cultivated and managed areas (Fig 3). Furthermore, *G. dorotocephala* also showed a high suitability for mosaic regions of croplands and tree cover (Fig 3). The most suitable Temperature for *G. sinensis* and *G. dorotocephala* was ~15 ºC, while for *G. tigrina* it was slightly lower (~11ºC), in all cases with a broad range of moderately suitable temperatures (Fig 3). Regarding the Lithological Classes (LC), the only terrain that was determined as being suitable for all three species was the one composed of carbonated sedimentary rocks, while the rest of lithological preferences were different for each species (or shared between two of the three species) (Fig 3).

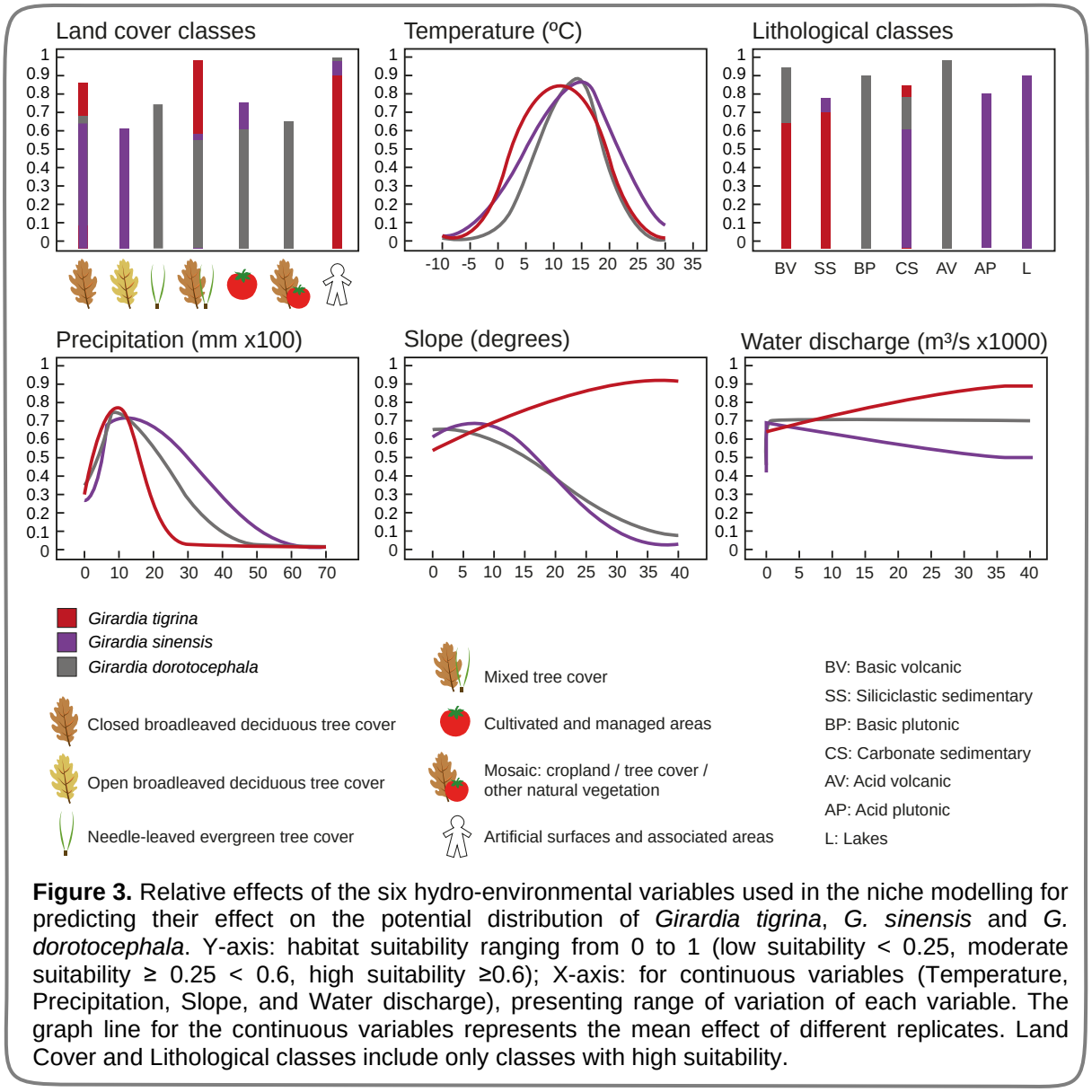


Figure 3. Relative effects of the six hydro-environmental variables used in the niche modelling for predicting their effect on the potential distribution of *Girardia tigrina*, *G. sinensis* and *G. dorotocephala*. Y-axis: habitat suitability ranging from 0 to 1 (low suitability < 0.25, moderate suitability ≥ 0.25 < 0.6, high suitability ≥0.6); X-axis: for continuous variables (Temperature, Precipitation, Slope, and Water discharge), presenting range of variation of each variable. The graph line for the continuous variables represents the mean effect of different replicates. Land Cover and Lithological classes include only classes with high suitability.

All three clades showed the highest suitability for regions with moderate precipitation regimes with an annual average of approximately 1000 mm (1000 litres/square meter). However, the three species show moderate suitability for a wide range of rainfall regimes, going from scarcely any rain during the year (close to 0 mm/year) to regimes reaching accumulated precipitations of 3000 mm/year (*G. dorotocephala*) and 4000 mm/year (*G. sinensis*) (Fig 3), with *G. tigrina* showing the steepest decay in suitability over 1000 mm/year. No clear preference was detected for Water Discharge (WD) of rivers in any of the three species. *Girardia sinensis* and *G. dorotocephala* were found to be highly suited to regions with a moderate terrain slope (between 0 to 15 degrees), thus differing from *G. tigrina*, which turned out to be highly suited to a wide range of terrain slopes (between 5 to 40 degrees).

Predicted current potential distribution of the three species revealed a predominant suitability for the Northern Hemisphere, especially for regions such as Eastern USA, Europe, and Eastern Asia (Fig S1). Additionally, these three species also showed some potentially suitable regions in the southern parts of South America, Africa, and Australia, with *G. sinensis* presenting the broader continuous potential distribution in these southern areas.

On the Iberian Peninsula, *G. tigrina* and *G. sinensis* showed extensive regions of high suitability, with the former having its most extensive area of high suitability being located in the northern region of the Peninsula, particularly in the area ranging from the Basque Country to Asturias (Fig. 4). In contrast, maximally suitable conditions for *G. sinensis* were found in the main rivers in the western region of the Iberian Peninsula, especially in the hydrological basin of the river Tajo. However, it should be borne in mind that the rest of regions of the Iberian Peninsula showed a high suitability for these two species. In contrast to *G. tigrina* and *G. sinensis*, Maxent did not predict any extensive regions of high suitability in the Iberian Peninsula for *G. dorotocephala*. Nonetheless, there were some areas, such as the surroundings of Lisbon or some scattered spots in Catalonia, Andalusia, and the northern part of the Iberian Peninsula, where potentially highly suitable areas were predicted (between 0.8 and 1.0) for this species.

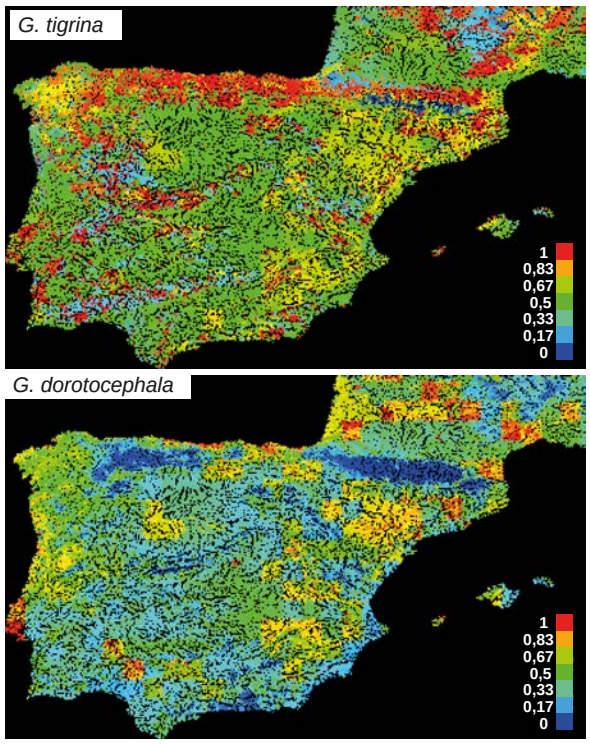


Figure 4. Potential geographic distribution of *Girardia tigrina, G. sinensis*, and *G. dorotocephala* in the Iberian Peninsula, estimated through niche modelling by using six hydro-environmental variables. Colour scale from blue to red indicates habitat suitability, ranging from 0 to 1 (low suitability < 0.25, moderate suitability ≥ 0.25 < 0.6, high suitability ≥0.6). Black pixels: no hydro-environmental data.

The predicted future potential distribution of the three introduced species revealed a gradual northwards expansion into suitable areas in the Northern Hemisphere and a gradual reduction of their suitable areas in the Southern Hemisphere (Fig. S2). Although this pattern was observed in both of our two scenarios on greenhouse gas emissions, it was much more extreme in the scenario with the highest increase in emissions (SSP5-8.5) than in the scenario with the highest decrease in emissions (SSP1-2.6). Fig 5 represents a summary of these results focused on our main area of interest, Europe and the Mediterranean region. In this area, *G. tigrina*, currently having the more extensive potentially suitable area of the three species, will be able to extend its distribution to Scandinavia in the best scenario (SSP1-2.6), reaching the northern coast of Norway in the worst-case scenario (SSP5-8.5), while it will have no suitable areas in the North of Africa. For the other two species, the variations are less pronounced in the best scenario (SSP1-2.6), while in the worst-case scenario they will be able to expand also over nearly all of Europe (except high mountains, such as the Alps) and reduce, but do not completely lose, their suitable regions in Northern Africa.

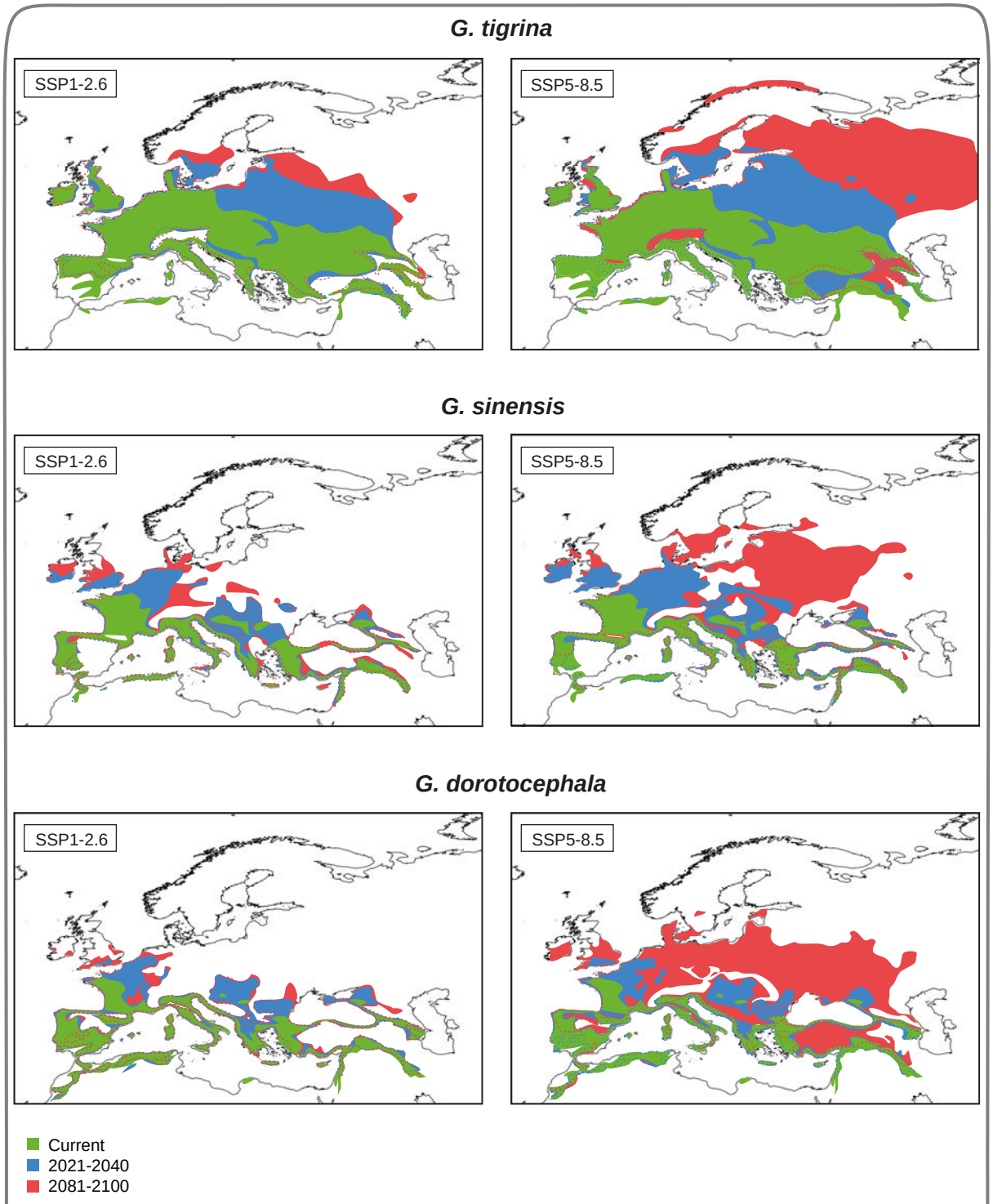


Figure 5. Predicted future geographic distribution of *Girardia tigrina*, *G. sinensis*, and *G. dorotocephala* in Europe, estimated through niche modelling under two different scenarios of climate change (SSP1-2.6: low greenhouse gas emissions; SSP5-8.5: high greenhouse gas emissions) by using the bioclimatic variables Mean Annual Temperature and Annual Precipitation. Colours correspond to three different time periods (green: Current; blue: 2021-2040; red: 2081-2100). Dotted lines denote the boundaries of the overlapping distributions. Detailed figures for each species and scenario are found in Fig. S2

4. Discussion

In the above, we presented a molecular approach to the identification of the several species of *Girardia* that have been introduced from their native ranges into areas all over the world. In addition, we identified the potential environmental drivers that may underlie the spread of alien species of *Girardia,* as well as their potential invasiveness and threat to native species. In the following we will separately discuss these issues.

4.1. Three species of North American *Girardia* colonise the world

Our phylogenetic analysis showed that all non-American *Girardia* samples belonged to either *G. tigrina*, *G. sinensis* or *G. dorotocephala.* The close phylogenetic relationship between *G. tigrina* and *G. dorotocephala* and the fact that they are of North American origin, suggest that they diversified only relatively recently on the North American continent. From the data, it is also apparent that genetic diversity within and among the three introduced *Girardia* species is quite low. In point of fact, it is much lower than, for example, the genetic diversity in freshwater planarian species of the genus *Dugesia* in Western Europe (Dols-Serrate et al., 2020; Leria et al., 2020; Leria et al., press), a region that *Girardia* has amply colonised.

However, *G. sinensis*, which is phylogenetically closely related to *G. tigrina* and *G. dorotocephala* (Fig. 2), was described based on specimens from China, but the tree shows the presence of this species on the island of Cuba, which lies on the North American Plate (Figs 2, S1). Therefore, and on the basis of the phylogenetic relationships and the geographic distribution of these three species, *G. sinensis* is actually a North American species that was introduced into China, but thus far is undetected in North America, just as has been discussed in (Chapter II).

Thus, our results corroborate (1) the earlier hypothesis that *G. tigrina* was introduced from North America, and (2) the introduction of *G. dorotocephala* in Japan and Hawaii. In addition, our results revealed the introduction of *G. dorotocephala* in Europe and Brazil and point to a third introduced species, *G. sinensis,* that has been introduced from North America into China, Europe, and Australia (Figs 1 and 2).

Our present results indicate multiple introductions of *Girardia* into Europe, corroborating previous results of Ribas et al. (1989) and (Stocchino et al., 2019), who had already described the presence of four morphological biotypes of *G. tigrina*: (A)

fissiparous specimens, with a spotted dorsal pattern and pigmented pharynx; (B) fissiparous spotted, with unpigmented pharynx; (C) sexually reproducing animals with a striped dorsal pattern and unpigmented pharynx; (D) fissiparous, striped animals with unpigmented pharynx. On the basis of a protein electrophoretic analysis, Ribas et al. (1989) detected genetic differentiation between animals of class A and those of classes B+C and suggested that this may signal the presence of at least two separate taxa, either at the subspecies or species level. Moreover, both Ribas et al. (1989) and (Stocchino et al., 2019) suggested that presence of the four morphological groups in Europe may be the result of multiple independent introductions of this species or even of the invasion of more than one *Girardia* species. Our present results corroborate this notion of multiple introductions, albeit that now we can establish molecularly that two species are involved in the areas where Ribas et al. (1998) and Stocchino et al. (2019) analysed the presence of *G. tigrina,* viz., *G. tigrina* and *G. sinensis*.

With respect to their external appearance and internal morphology, *G. sinensis* and *G. tigrina* are highly similar. This paucity of morphological differences between both species may be the reason why *G. sinensis* has gone undetected in its presumed original area of distribution, North America. In this sense, it is noteworthy that the copulatory apparatus of the sexually reproducing animals on Menorca analysed by Ribas et al. (1989) conforms to that of *G. tigrina*, and their habitus and pharynx pigmentation pattern to class C (Ribas et al., 1989, Fig. 2). On the other hand, the specimens from Menorca analysed in our study, and which came from a locality close to the B-type locality of Ribas et al. (1989), belonged to *G. sinensis*. All of this provides ample evidence that *G. tigrina* and *G. sinensis* have colonised the same areas, where they may even occur at exactly the same site, exemplified also by the sampling locality #138 in mainland Spain (Table S1).

The morphological similarity of *G. tigrina* and *G. sinensis* and their co-occurrence in Western Europe, implies that records of presumed *G. tigrina* simply based on external appearance can no longer be considered to be reliable. In such cases where the animals only show asexual reproduction through the process of fission, as frequently being the case with introduced populations, discrimination between specimens of *G. tigrina* and *G. sinensis* can be achieved only through molecular markers. In that respect, it is noteworthy that earlier reports mentioned presumed *G. tigrina* for Australia (see Sluys et al., 1995 and references therein), while

the sequences that we analysed from Queensland and Tasmania unequivocally ranked as representatives of *G. sinensis*.

The widely separated geographic localities where all three species were introduced, in combination with the fact that planarians are poor dispersers, strongly suggest that there were multiple introductions and that these were due to human-induced dispersal events. Most likely, the latter were effectuated through the international trade in aquatic plants and the activity of aquarists or through the importation and subsequent culturing of North American strains of *G. tigrina* and *G. dorotocephala* as model organisms in education and scientific studies all around the world. The fact that *G. tigrina* is very sticky, clearly enhances its attachment to all kinds of surfaces and thereby facilitates its passive dispersal (Stocchino et al., 2019). Wright (1987) suggested that multiple occasions of disposal of the contents of aquaria and of small ornamental garden ponds into local rivers might underlie the pattern of distribution of *G. tigrina* in Great Britain. In addition, disposal of such contents into waste-water pipes might facilitate entrance of flatworms into a river via surface water drains (Wright, 1987). The present data does not allow us to infer whether introductions into different countries came directly from North America or may have occurred in a stepping-stone model from one introduction site to the next. Most probably, both types of events have contributed to the spread of alien species of *Girardia*. Future analyses of more variable markers may be necessary, in order to determine the various routes that may have been followed by the three North American species of *Girardia* in their conquest of the world.

4.2. What makes *Girardia* species successful colonisers?

Our niche modelling indicated that *G. tigrina*, *G. dorotocephala* and *G. sinensis* do not only have the ecological potential to successfully survive at the localities where they have been introduced but are also capable of expanding their distributional range. Interestingly, in the Maxent model the responses of the three species to different environmental variables indicated a key feature explaining their high capacity for colonization, viz., shared tolerance for anthropogenic habitats, such as artificial freshwater habitats (like canals) or freshwater habitats in cultivated and managed areas (Fig 3). The importance of this variable has been reported also for many other introduced plant and animal species (Rickart et al., 2011; Salomidi et al., 2013;

Simkanin et al., 2013; Johnston et al., 2017; González-Ortegón & Moreno-Andrés, 2021), including terrestrial planarians (Álvarez-Presas et al., 2014) and is also one of the main factors explaining the distribution of another introduced freshwater planarian, *Dugesia sicula* (Leria et al., in press). In that respect, it is noteworthy that the factor "tolerance for anthropogenic habitats" played no role in the explanation of the autochthonous distribution of *D. subtentaculata* in the Iberian Peninsula (Leria et al., in press). Anthropogenic habitats generally are characterised by low autochthonous species diversity, thus offering empty niches for invasive species that can use these as a springboard to natural habitats (Dietz & Edwards, 2006).

Apart from association with humans and anthropogenic habitats, other traits of successful colonizers that may become invaders, are, for example, high abundance in the native range, short generation time, high genetic variability, wide physiological tolerance, phenotypic plasticity, and asexual reproduction (Kolar & Lodge, 2001). These are general features that may vary or combine in different ways, depending on the intrinsic characteristics of each group of organisms and of the receiver area. Our Maxent analysis allowed us to determine that the introduced species of *Girardia* present some of these traits.

A case of wide physiological tolerance concerns temperature, with the three *Girardia* species presenting a range of highly suitable temperatures that is much broader than in other freshwater planarian species (Vila-Farré & Rink, 2018; Leria et al., in press). For example, *Polycelis felina* exhibits an optimality maximum at 4ºC, with the optimality rapidly decreasing at temperatures above 10ºC, while *Dugesia subtentaculata* shows a narrow temperature range between 12º and 15ºC with a high suitability (Leria et al., in press). In contrast, in *Girardia* high suitability was found for temperatures ranging from around 4ºC to roughly 20ºC. This minimum value of about 4ºC found for *Girardia* in our niche modelling, coincides with ecological studies that found temperature to be a limiting factor in the expansion of *G. tigrina* in the United Kingdom and, curiously, catalogued it as a warm-water species, with a lower tolerance limit of 6ºC for feeding and 10ºC for asexual reproduction (Wright, 1987 and references therein). On the other hand, in the present study, *G. dorotocephala* has been found in Mexico and *G. sinensis* in Cuba in waters with temperatures ranging around 19.5-20.5ºC and 24ºC, respectively.

Our niche modelling also revealed that the three species are suitable to a wide range of lithological classes and water discharge regimes, as well as precipitation regimes (Fig. 3). However, for slope not all of the introduced species presented a flexible suitability, as only *G. tigrina* would be highly suited to an extremely wide range of terrain slopes. This is surprising, since it has been reported that *G. tigrina* generally occurs in the lower reaches of lowland rivers with low slopes (Wright, 1987). All of these results suggest that each of the three introduced *Girardia* species exhibits high tolerances to most of these key environmental variables.

As remarked above, asexual reproduction is considered to be a trait enhancing successful colonization and invasion. For *G. tigrina*, and *G. dorotocephala*, it has been established that they show both sexual and asexual reproduction in their native areas, with populations being either strictly sexual or asexual or showing both reproductive modalities (Kenk, 1937, 1972; Puccinelli & Deri, 1991; Knakievicz, Vieira, Erdtmann, & Ferreira, 2006). For *G. sinensis,* in the laboratory it was established that the species can reproduce both sexually and asexually by fission (Chen et al., 2015). However, most of the introduced populations are constituted by fissiparous individuals. Thus far, only five immigrant sexual populations are known from Western Europe, located in Great Britain, Spain (Menorca Island), Italy, and France (Stocchino et al., 2019 and references therein), while all other known populations reproduce by fission. This prevalence of asexual populations might be advantageous for dispersion and population growth (Gee et al., 1998; Wright, 1987) and has already been proposed as a major factor explaining the successful recent colonization of *Dugesia sicula* in the Mediterranean region (Lázaro & Riutort, 2013).

Another important characteristic of planarians possibly contributing to their colonization capability concerns their top position in the trophic web. Although there is large number of species known to predate on freshwater planarians, the impact of predation on natural populations of flatworms generally is rather low (Vila-Farré & Rink, 2018). Evidently, this may promote rapid population growth, which, in turn, fulfils a basic requirement for high rates of spread.

4.3. Potential ecological impact of introduced *Girardia*

Since freshwater communities are highly sensitive to composition changes through the introduction of invasive species (Havel et al., 2015; Gallardo et al., 2016),

prediction and risk assessment of potential invader species are vitally important for the protection of these habitats. However, risk assessment is distinct from prediction whether a certain species will become an invader, and even more difficult to achieve than the prediction. With respect to freshwater planarians, ecological studies on the specific impact of introduced species on their receiver freshwater communities are (1) limited, (2) focused on only small geographic regions, or (3) are even contradictory. For example, while some studies concluded that *Girardia tigrina* may be a competitive invader (Reynoldson, 1985; Wright, 1987; Gee & Young, 1993), its ecological impact on native communities was downplayed in a recent paper (Ilić et al., 2018).

For making predictions whether certain non-native species will become invasive, it is unfortunate that there is no suite of characteristics that can unequivocally signal successful invaders (Kolar & Lodge, 2001). Nevertheless, a non-native species must go through five phases to become an invader: (1) transport to a new region, (2) release or escapement to the wild, (3) establishment, (4) dispersal or spread, and (5) integration or impact (García-Berthou, 2007). From that perspective, all three species of introduced *Girardia* have already reached phases four and five, albeit that their ecological impact has not been properly analysed.

One of the most often used and accepted definitions of invasive species, is that proposed by the International Union for the Conservation of Nature (IUCN): “An alien species is a species introduced outside its natural past or present distribution; if this species becomes problematic, it is termed an invasive alien species (IAS).” The IUCN website further specifies that IAS “may lead to changes in the structure and composition of ecosystems detrimentally affecting ecosystem services, human economy and wellbeing” (https://www.iucn.org/theme/species/our-work/invasive-species; last visited: 22 January 2021). However, others have suggested that the term “invasive” should refer only to the biogeographic or demographic status of a species, without any connotation of potential or actual ecological impact (Richardson et al., 2000). From the latter perspective, the term “invasive” certainly applies to the introduced species of *Girardia* that are the subject of our present study.

Our preliminary study on the potential impact of introduced species of *Girardia* on autochthonous freshwater planarians was restricted to the Iberian Peninsula, for which we have a relatively good knowledge about its autochthonous species diversity. Most native species in this region occur at high altitudes (*Crenobia alpina* (Dana,

1766), *Polycelis felina* (Dalyell, 1814), various species of *Phagocata* Leidy, 1847), while others, such as *Phagocata ullala* Sluys, 1995, are endemic to very restricted areas in which *Girardia* was never detected (Baguñà, Saló, & Romero, 1980, 1981; Sluys, Ribas, & Baguñà, 1995). Nonetheless, *Dugesia subtentaculata* (Draparnaud, 1801), the most widely distributed autochthonous species, has been found together with *Girardia* representatives. At the locality #247, *D. subtentaculata* coexists with *G. sinensis*, while in the North of the Iberian Peninsula it co-occurs with unidentified *Girardia* individuals (M. Vila-Farré, pers. comm.) and with *G. dorotocephala* (recently identified by DNA barcoding in our laboratory). Moreover, optimal environmental conditions for *D. subtentaculata* were recently estimated in a niche modelling analysis (Leria et al., in press). For this species optimal conditions are present along the northern and western coasts of the Iberian Peninsula, a region with principally deciduous tree cover, an annual average temperature around 14ºC, and relatively high precipitation regimes. These environmental conditions partially overlap with the optimal ecological niche for *G. sinensis,* as identified in the present study (Fig. 4). Moreover, niche modelling showed that the Iberian Peninsula presents extensive unsuitable areas for *D. subtentaculata* (zero suitability values) (Leria et al., in press), while in our modelling analyses of *Girardia*, almost all regions of the entire Peninsula received high suitability values (>0.6), especially for *G. tigrina* and *G. sinensis*. In that light, it is important that it has been shown that introduced species of *Girardia* may exert a negative impact on native planarian species, for instance by competing for the same food. Gee & Young (1993) showed that introduced *G. tigrina* had considerable dietary overlap with the native planarian species *Polycelis tenuis* Ijima, 1884 and *P. nigra* (Müller, 1774). Evidently, ecological impact of introduced *Girardia* may not be restricted to other planarians but may concern also other invertebrates. But such a detailed ecological analysis falls far outside the scope of our present study.

4.5 Future expansion trends of introduced *Girardia*

Our niche modelling indicated that at present the three introduced *Girardia* species have the ecological potential to expand their distributional range (Fig S1), while colonization modelling of future situations, regardless of the particular climatic scenario, pointed to an extension of their potential areas of suitability (Figs 5, S2). Their potential for range expansion was mostly restricted to suitable areas in the Northern Hemisphere, in particular to regions throughout Europe. In this light, and also

considering the generalist characteristics of the species and their demonstrated suitability for human-modified habitats, it is only to be expected that in our current, increasingly anthropogenic world they may have an advantage over native species. In point of fact, our data show clearly that human activities have effectuated important changes in the geographic distribution of the three introduced species of *Girardia* in a relatively short period of time.

Author Contributions

LBA, LL, and MR did the initial study design. LBA and LL processed and analysed the data and wrote the manuscript with input from all authors. All authors read and approved the final manuscript. Authors declare no conflicts of interest.

Conflicts of interest

The authors declare that they have no competing interests.

SUPPORTING INFORMATION

Table S1. Sequences used in the molecular analyses, with indication of their codes, the sampling locality of the individuals, GenBank accession numbers, and the taxonomic identification.

Code in the figure	Locality	Taxonomic Identification	COI	EF1α
84do_FrLez	France, River Lez	*G. dorotocephala*	OM307073	OM349486
87ti_FrLun	France, Lunaç	*G. tigrina*	OM307076	OM418671
109ti_CatMug	Spain, Catalonia, St. Llorenç de la Muga, River Muga	*G. tigrina*	OM307077	OM418675
116ti_CatEbr	Spain, Catalonia, Riba-roja d'Ebre, River Ebro	*G. tigrina*	OM307078	OM418673
125sc_BrRGScon	Brazil, Rio Grande do Sul, Constantina	*G. schubarti*	OM307079	OM418648
126sc_BrRGScon	Brazil, Rio Grande do Sul, Constantina	*G. schubarti*	OM307080	OM418649
127cl_BrRGSleo	Brazil, Rio Grande do Sul, São Leopoldo	*G. clandistina*	OM307081	OM418688
131.1bi_BrRGSseb	Brazil, Rio Grande do Sul, São Sebastião de Caí	*G. biapertura*	OM307082	
133cl_BrRGSleo	Brazil, Rio Grande do Sul, São Leopoldo	*G. clandistina*	OM307085	OM418690
134cl_BrRGSsalv	Brazil, Rio Grande do Sul, Salvador do Sul	*G. clandistina*	OM307086	OM418689
135.1ti_FrArd	France, Gorges de l'Ardeche	*G. tigrina*	OM307087	OM418677
137ti_FrIsp	France, Ispagnac	*G. tigrina*	OM307088	OM418682
138.1si_CatFlu	Spain, Catalonia, Girona, Orfes, River Fluvià	*G. sinensis*	OM307089	OM418669
138.2si_CatFlu	Spain, Catalonia, Girona, Orfes, River Fluvià	*G. sinensis*	OM307090	OM418664
138.3ti_CatFlu	Spain, Catalonia, Girona, Orfes, River Fluvià	*G. tigrina*	OM307091	OM418681
138.8si_CatFlu	Spain, Catalonia, Girona, Orfes, River Fluvià	*G. sinensis*	OM307096	OM418663

139.2si_EsMen	Spain, Menorca, Algendar	*G. sinensis*	OM307098	OM418657
140.1si_ItCer	Italy, Sardinia, R.Fungarone, Putifigari	*G. sinensis*	OM307099	OM418655
143sp_ItTos	Italy, Toscana, Torr.Vincio	*G. tigrina*	OM307101	OM418683
247si_EsPon	Spain, Pontevedra, Gondomar	*G. sinensis*	OM307102	OM418653
248do_EsCue	Spain, Cuenca, Reíllo	*G. dorotocephala*	OM307103	OM349494
250do_PrCh	Portugal, Cheleiros	*G. dorotocephala*	OM307104	OM349490
252ti_FrIss	France, Issalès	*G. tigrina*	OM307105	OM418674
253si_EsSal	Spain, Salamanca, Ciudad Rodrigo	*G. sinensis*	OM307106	OM418654
260.1cl_BrRGSsj	Brazil, Rio Grande do Sul, São Jose do Norte	*G. clandistina*	OM307107	OM418691
261ar_BrSPip	Brazil, São Paulo, Iporanga (cave)	*Girardia* aff. *arenicola*	OM264750	OM418632
262mu_BrMGSbod	Brazil, Mato Grosso do Sul, Bodoquena (cave)	*G. multidiverticulata*	OM307109	OM418642
263sp_BrRGSsj	Brazil, Rio Grande do Sul, São Jose do Norte	*G. clandistina*	OM307110	OM418692
269do_BrRGScon	Brazil, Rio Grande do Sul, Constantina	*G. dorotocephala*	OM307111	OM349488
270.1do_USAbalt	USA, Baltimore	*G. dorotocephala*	OM307112	OM349498
270.2do_USAbalt	USA, Baltimore	*G. dorotocephala*	OM307113	OM349499
295sp_ChiLHrs	Chile, Los Lagos, Huinay Research Station	*Girardia* sp. 5	OM307114	OM418636
298.2sp_ChiPP	Chile, Pumalin Park	*Girardia* sp. 5	OM307118	OM418638
299sp_ChiLPh	Chile, Los Lagos, Peninsula Huequi	*Girardia* sp. 5	OM307119	OM418637
308.1si_FrMon	France, Montpellier	*G. sinensis*	OM307121	OM418656
325mu_BrMGSbon	Brazil, Mato Grosso do Sul, Bonito (cave)	*G. multidiverticulata*	OM307123	OM418641
327do_USAmich	USA, Michigan, Ann Arbor	*G. dorotocephala*	OM307124	OM349489
337mu_BrMGSbod	Brazil, Mato Grosso do Sul, Bodoquena (cave)	*G. multidiverticulata*	OM307125	OM418643

373ti_CatMur	Spain, Catalonia, Riera de Mura	*G. tigrina*	OM307127	OM418680
377si_CatFlvil	Spain, Catalonia, Girona, Fluvià, Vilert	*G. sinensis*	OM307128	OM418658
383do_HawUM	Hawaii, Upper Manoa	*G. dorotocephala*	OM307129	OM349501
384do_HawMM	Hawaii, Middle Manoa	*G. dorotocephala*	OM307130	OM349500
399sc_BrRGSmaq	Brazil, Rio Grande do Sul, Maquine	*G. schubarti*	OM307131	OM418646
401sp_BrMatcha	Brazil, Mato Grosso, Chapada Guimaraes (cave)	*Girardia* sp. 3	OM307132	
402sp_BrSCchap	Brazil, Santa Catarina, Chapeco	*Girardia* sp. 6	OM307133	OM418635
403cl_BrRGSleo	Brazil, Rio Grande do Sul, São Leopoldo	*G. clandistina*	OM307134	OM418686
466.2do_USA	USA, Carolina Enterprise[a]	*G. dorotocephla*	OM307136	OM349491
467.2do_USA	USA, Carolina Enterprise[a]	*G. dorotocephla*	OM307138	OM349502
468.2do_USA	USA, Carolina Enterprise[a]	*G. dorotocephla*	OM307140	OM349492
469.1ti_CatMj	Spain, Catalonia, Montjüic	*G. tigrina*	OM307141	OM418670
469.2ti_CatMj	Spain, Catalonia, Montjüic	*G. tigrina*	OM307142	OM418684
534sc_BrRGSpau	Brazil, Rio Grande do Sul, São Francisco de Paula	*G. schubarti*		OM418647
535ti_BrRGSleo	Brazil, Rio Grande do Sul, São Leopoldo	*G. tigrina*	OM307143	OM418687
542ti_GerPill	Germany, Pillnitz	*G. tigrina*	OM307144	OM418678
543do_USAvirg	USA, Virginia	*G. dorotocephala*	OM307145	OM349495
544si_Austderw	Australia, Tasmania, Derwent River	*G. sinensis*	OM307146	OM418667
545do_USAvirgJFRC	USA, Virginia, Ashburn, Janelia Farm Research Campus	*G. dorotocephala*	OM307147	OM349496
548do_Mex	Mexico	*G. dorotocephala*	OM307150	
550do_Mex	Mexico	*G. dorotocephala*	OM307151	OM349497
552si_GerZsch	Germany, Zschorna	*G. sinensis*	OM307153	OM418666

554si_NethLeid	Netherlands, Leiden	*G. sinensis*	OM307155	OM418665
556ti_CatFlEmp	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*G. tigrina*	OM307156	OM418685
559.1si_CatFlEmp	Spain, Catalonia, Arenys d' Empordà, Fluvià River	*G. sinensis*	OM307158	OM418668
659ti_CanNS	Canada, Nova Scotia, Ainslie Lake	*G. tigrina*	OM307160	OM418676
660si_AustQUQ	Australia, Queensland, UQ Lakes	*G. sinensis*	OM307161	OM418659
661si_AustQCC	Australia, Queensland	*G. sinensis*	OM307162	OM418660
679ti_USADL	USA, Michigan, Douglas Lake	*G. tigrina*	OM307163	OM418672
684.2do_Jap	Japan, Hoshikuki-cho, Mizu-no-sato Park, Miyako River	*G. dorotocephala*	OM307166	OM349493
685.1san_ChiTalag	Chile, Talagante, Mapocho River	*Girardia sanchezi*	OM307167	OM418644
685.2san_ChiTalag	Chile, Talagante, Mapocho River	*Girardia sanchezi*	OM307168	OM418645
686si_ChinYand	China, Conghua, Yadongxi River	*G. sinensis*	OM307169	OM418662
1056sp_MexBST1	Mexico, Biological Station Los Tuxtlas 1	*Girardia* sp. 1		OM418694
1059sp_MexBST1	Mexico, Biological Station Los Tuxtlas 1	*Girardia* sp. 1		OM418639
1062sp_MexLagE	Mexico, Los Tuxtlas, Laguna Escondida	*Girardia* sp. 1	OM307171	OM418640
1063sp_MexLagE	Mexico, Los Tuxtlas, Laguna Escondida	*Girardia* sp. 1	OM307172	
1070sp_MexBST2	Mexico, Biological Station Los Tuxtlas 2	*Girardia* sp. 2	OM307173	OM418633
1072.2si_Cub	Cuba, Matanzas, Martí, El Huequito	*G. sinensis*	OM307174	OM418661
1178sp_BrBah	Brazil, Bahía, Chapada Diamantina, Vale do Pati (cave)	*Girardia* sp. 4	OM307175	OM418634
1179sp_MexXoch	Mexico, Mexico City, Xochimilco, Cuemanco	*Girardia* sp. 7	OM307176	OM418693
1180do_MexXoch	Mexico, Mexico City, Xochimilco, Cuemanco	*G. dorotocephala*	OM307177	OM349487
1181sp_MexXoch	Mexico, Mexico City, Xochimilco, Cuemanco	*Girardia* sp. 7	OM307178	
1182do_MexMicho	Mexico, Michoacán	*G. dorotocephala*	OM307179	

F6510.2sp_BrPar	Brazil, Paraná, Toledo, Cerro da Lola	*G. schubarti*		OM418651
InoueA_CS103	USA, Texas, Caroline Spring, Independence Creek	*Girardia* sp. 1	MN652340.1	
InoueB_ES201	USA, New Mexico, Palomas Creek, Emrick Spring	*Girardia* sp. 1	MN652378.1	
InoueC_BLBC002	USA, Texas, Bitter Lake, Bitter Creek	*Girardia* sp.	MN652301.1	
InoueD_GR209	USA, New Mexico, West Fork of the Gila River	*Girardia sp*.	MN652373.1	
GB_Gsi_China	China, Guangdong Province Xinghu Lake in Zhaoqing	*G. sinensis*	KP091895.1	
GB_Gti_FrMont	France, Montpellier	*G. tigrina*	DQ666042.1	
GB_Gsc_Br	Brazil	*G.schubarti*	DQ666041.1	KJ599691.1
GB_som1_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*Girardia somuncura*	MW271865	
GB_som2_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	MW271866	
GB_som3_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	MW271867	
GB_som4_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. somuncura*	MW271869	
GB_tom1_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*Girardia tomasi*	MW271863	
GB_tom2_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. tomasi*	MW271864	
GB_tom2_Arg	Argentina, Somuncurá Plateau, Head of Valcheta Stream	*G. tomasi*	MW271868	

[a]: Carolina Enterprise Word-Class Support for Science & Math

Figure S1. Potential worldwide geographic distribution of *Girardia tigrina*, *G. sinensis*, and *G. dorotocephala*, estimated through niche modelling by using six hydroenvironmental variables. Color scale from blue to red indicates habitat suitability, ranging from 0 to 1 (low suitability < 0.25, moderate suitability ≥ 0.25 < 0.6, high suitability ≥0.6). Filled white circles indicate the localities of each of the clades included in the analyses. Black pixels: no hydro-environmental data.

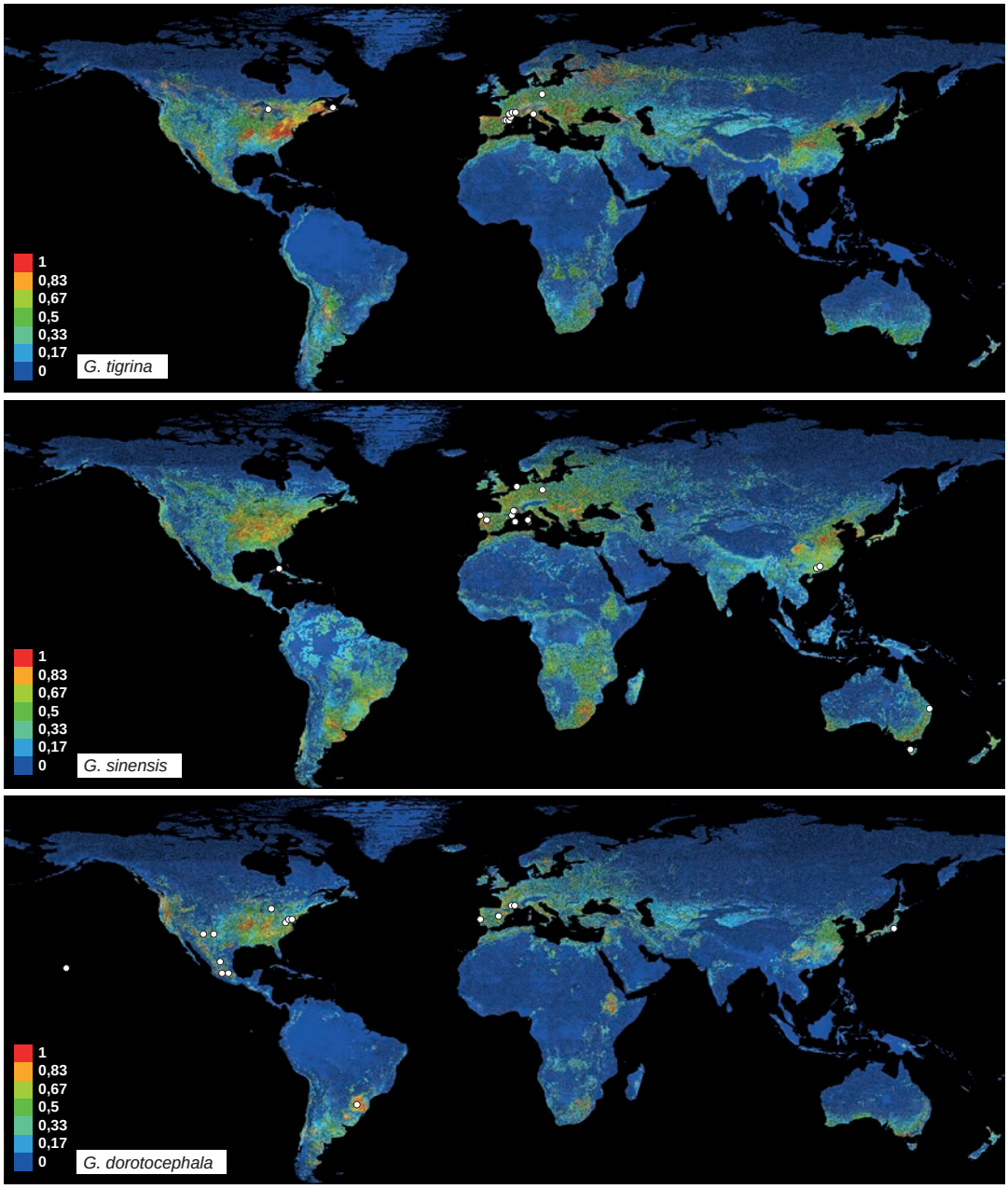
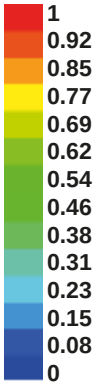
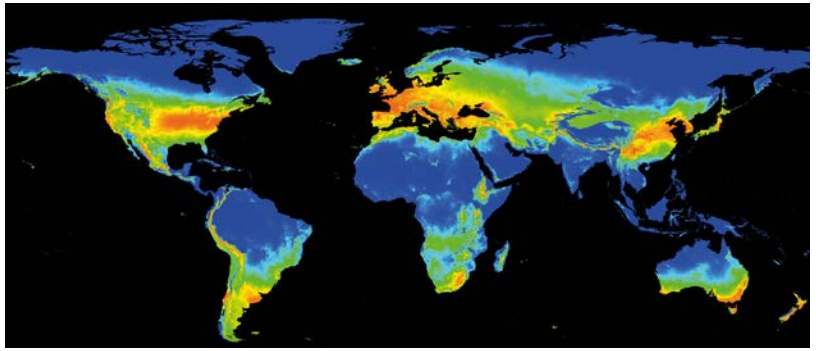


Figure S2. Predicted world-wide current and future geographic distribution of Girardia tigrina, G. sinensis, and G. dorotocephla under two different scenarios of climate change (SSP1-2.6: low greenhouse gas emissions; SSP5-8.5: high greenhouse gas emissions), estimated through niche modelling, using the bioclimatic variables Mean Annual Temperature and annual Precipitation. Future time periods corresponding to 2021-2040 and 2081-2100. Color scale from blue to red indicates habitat suitability, ranging from 0 to 1 (low suitability < 0.25, moderate suitability ≥ 0.25 < 0.6, high suitability ≥0.6). These figures are summarized in Fig. 5.

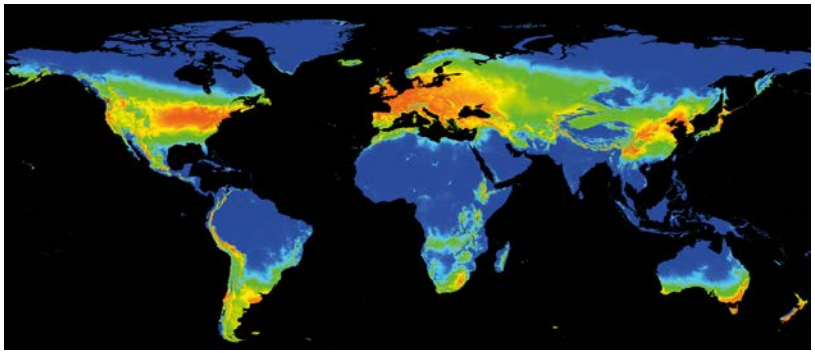


G. tigrina

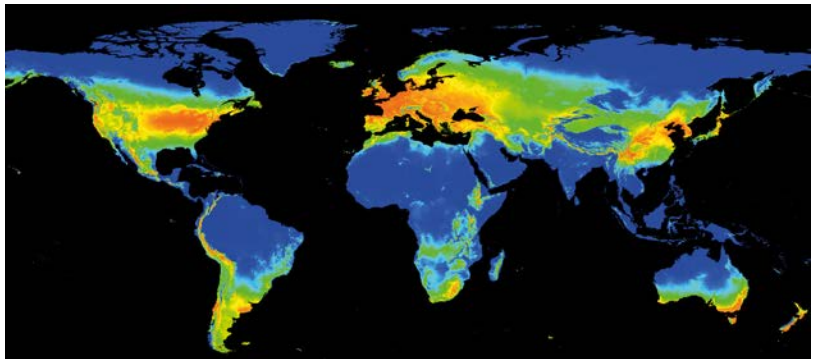
Current



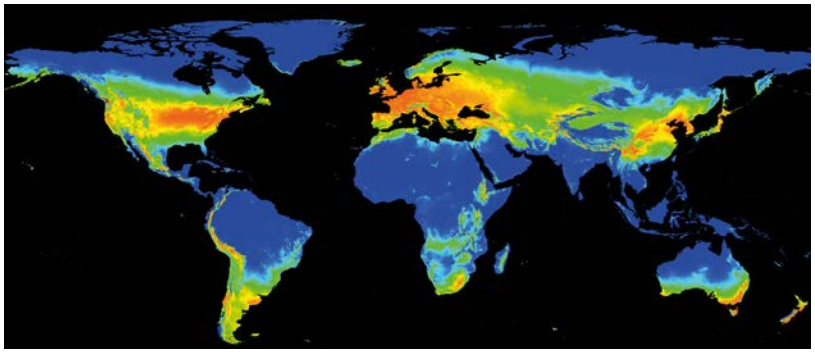
2021-2040 / SSP1-2.6



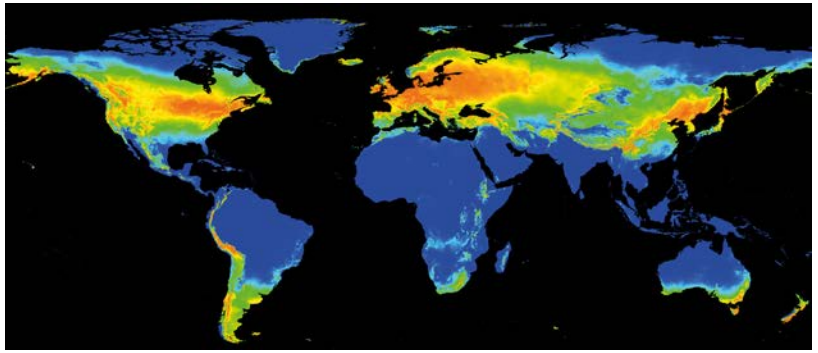
2081-2100 / SSP1-2.6

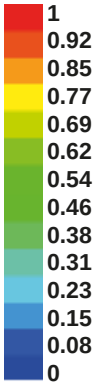


2021-2040 / SSP5-5.8



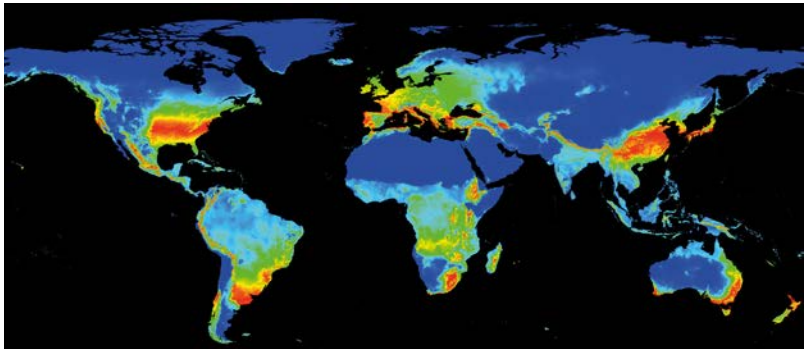
2081-2100 / SSP5-5.8



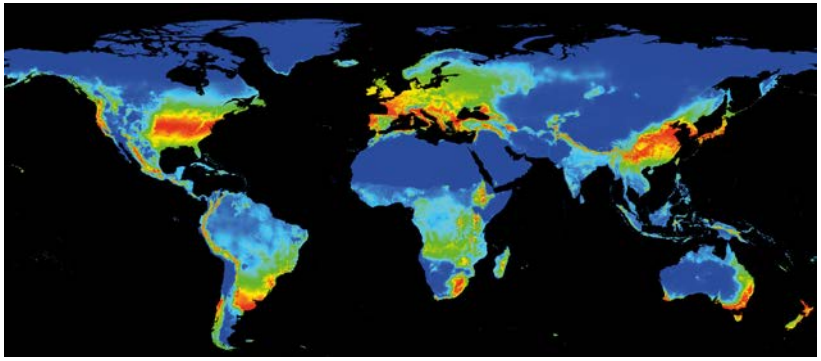


G. sinensis

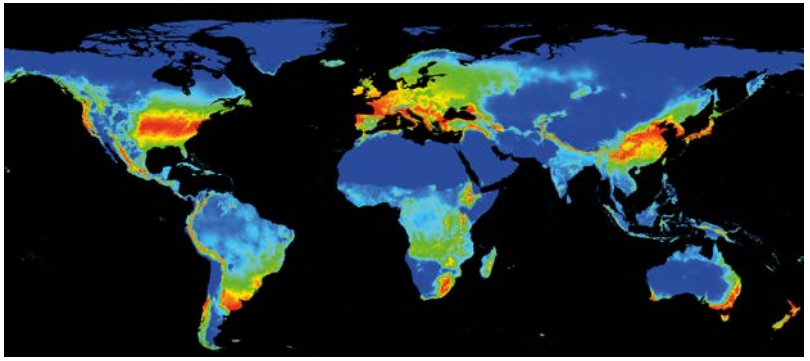
Current



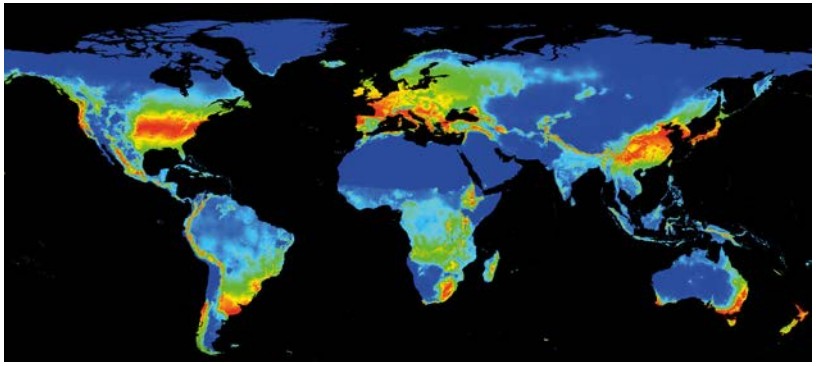
2021-2040 / SSP1-2.6



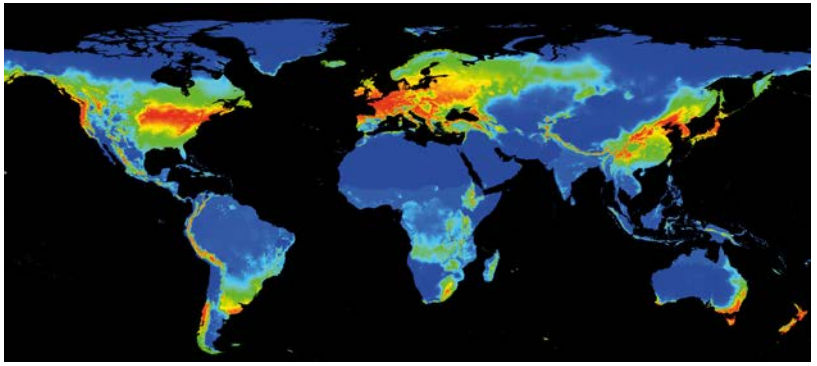
2081-2100 / SSP1-2.6

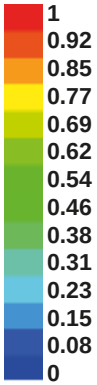


2021-2040 / SSP5-5.8



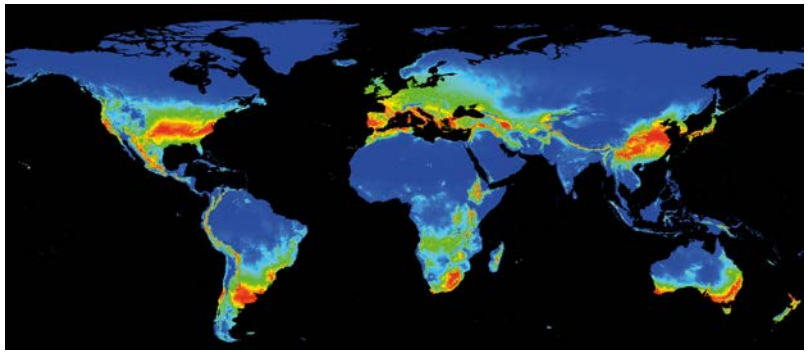
2081-2100 / SSP5-5.8



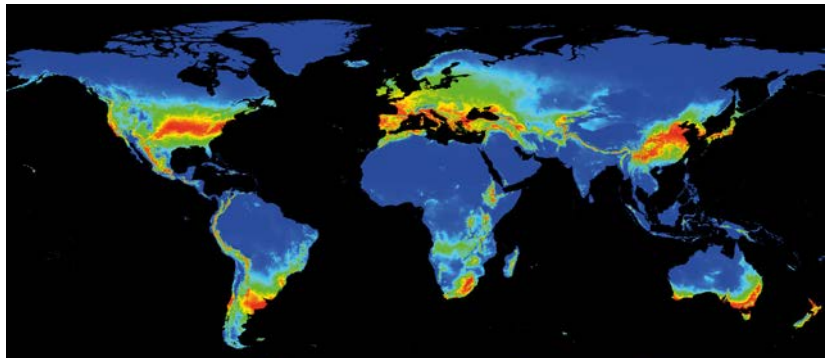


G. dorotocephala

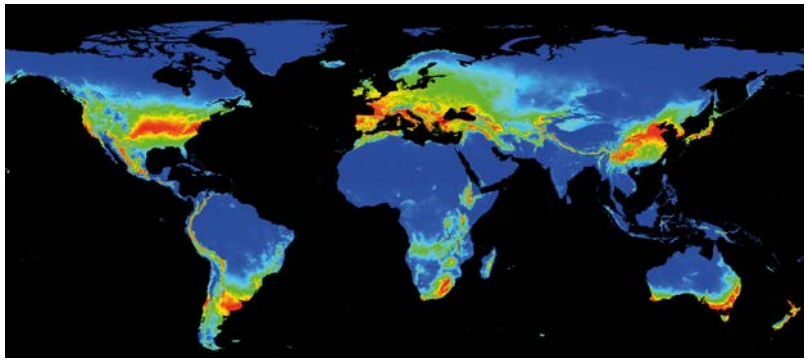
Current



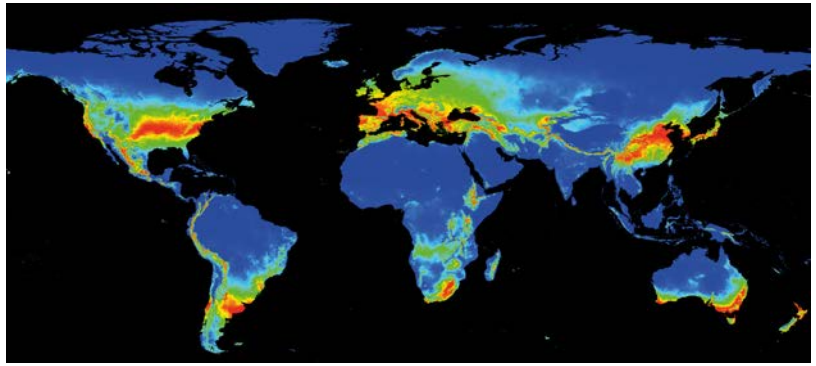
2021-2040 / SSP1-2.6



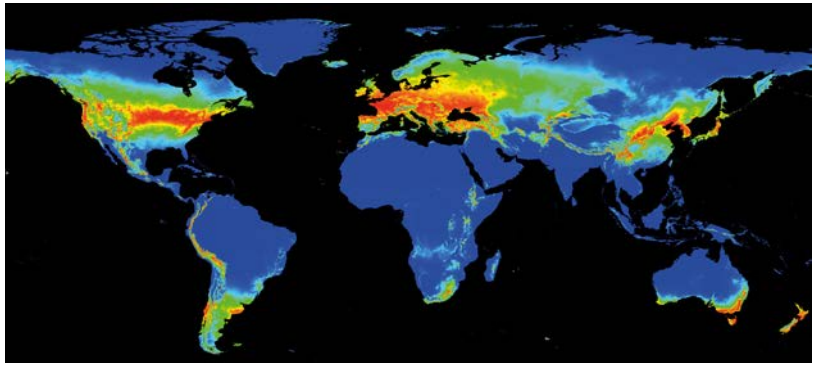
2081-2100 / SSP1-2.6



2021-2040 / SSP5-5.8



2081-2100 / SSP5-5.8



References

Álvarez-Presas, M., Mateos, E., Tudó, A., Jones, H. D., & Riutort, M. (2014). Diversity of introduced terrestrial flatworms in the Iberian Peninsula: a cautionary tale. *PeerJ*, *2014*, e430. https://doi.org/10.7717/peerj.430

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

Baguñà, J., Saló, E., & Romero, R. (1980). Les planàries d'aigües dolces a Catalunya i les Illes Balears. I. Clau sistemàtica i distribució geogràfica. *Butlletí de La Institució Catalana d'història Natural*, *45*, 15–30.

Baguñà, J., Saló, E., & Romero, R. (1981). Biogeografía de las planarias de aguas dulces (Platyhelminthes; Turbellaria; Tricladida; Paludicola) en España. Datos preliminares. *Actas Del Primer Congreso Español de Limnología*, 265–280.

Benazzi, M. (1993). Occurrence of a sexual population of *Dugesia (Girardia) tigrina,* a freshwater planarian native to America, in a lake of southern Italy. *Bolletino Di Zoologia*, *60*(1), 129–130. https://doi.org/10.1080/11250009309355799

Chen, X., Chen, Y., Wu, C., & Wang, A. (2015). A new species of the genus *Girardia* (Tricladida: Dugesiidae) from China. *Zoological Systematics*, *40*(2), 166–178. https://doi.org/10.11865/zs.20150202

Dietz, H., & Edwards, P. J. (2006). Recognition that causal processes change during plant invasion helps explain conflicts in evidence. *Ecology*, *87*(6), 1359–1367. https://doi.org/https://doi.org/10.1890/0012-9658(2006)87[1359:RTCPCD]2.0.CO;2

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/https://doi.org/10.1007/s13127-020-00438-z

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence / absence models. *Environmental Conservation*, *24*(1), 38–49.

Gallardo, B., Clavero, M., Sánchez, M. I., & Vilà, M. (2016). Global ecological impacts of invasive species in aquatic ecosystems. *Global Change Biology*, *22*(1), 151–163. https://doi.org/https://doi.org/10.1111/gcb.13004

García-Berthou, E. (2007). The characteristics of invasive fishes: What has been learned so far? *Journal of Fish Biology*, *71*(SUPPL. D), 33–55. https://doi.org/10.1111/j.1095-8649.2007.01668.x

Gee, H., Pickavance, J. R., & Young, J. O. (1998). A comparative study of the population biology of the American immigrant triclad *Dugesia tigrina* (Girard) in two British lakes. *Hydrobiologia*, *361*, 135–143.

Gee, H., & Young, J. O. (1993). The food niches of the invasive *Dugesia tigrina* (Girard) and indigenous *Polycelis tenuis* Ijima and *P. nigra* (Müller) (Turbellaria; Tricladida) in a Welsh lake. *Hydrobiologia*, *254*(2), 99–106. https://doi.org/10.1007/BF00014313

Gherardi, F., Bertolino, S., Bodon, M., Casellato, S., Cianfanelli, S., Ferraguti, M., … Tricarico, E. (2008). Animal xenodiversity in Italian inland waters: Distribution, modes of arrival, and pathways. *Biological Invasions*, *10*(4), 435–454. https://doi.org/10.1007/s10530-007-9142-9

González-Ortegón, E., & Moreno-Andrés, J. (2021). Anthropogenic Modifications to Estuaries Facilitate the Invasion of Non-Native Species. *Processes*, Vol. 9. https://doi.org/10.3390/pr9050740

Grant, L. J., Sluys, R., & Blair, D. (2006). Biodiversity of Australian freshwater planarians (Platyhelminthes: Tricladida: Paludicola): New species and localities, and a review of paludicolan distribution in Australia. *Systematics and Biodiversity*, *4*(4), 435–471. https://doi.org/10.1017/S1477200006002064

Grant, L. J. (2017). *Biodiversity, Biogeography and Phylogeny of Australian Freshwater Triclads (Platyhelminthes: Tricladida: Continenticola)*. PhD Thesis, James Cook University. Retrieved from https://researchonline.jcu.edu.au/51512/1/51512-grant-2016-thesis.pdf

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, *41*, 95–98.

Havel, J. E., Kovalenko, K. E., Thomaz, S. M., Amalfitano, S., & Kats, L. B. (2015). Aquatic invasive species: challenges for the future. *Hydrobiologia*, *750*(1), 147–170. https://doi.org/10.1007/s10750-014-2166-0

Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, *46*(1), 10–18. https://doi.org/10.1111/j.1365-2664.2008.01600.x

Ilić, M. D., Tubić, B. P., Marinković, N. S., Marković, V. M., Popović, N. Z., Zorić, K. S., … Paunović, M. M. (2018). First report on the non-indigenous triclad *Girardia tigrina* (Girard, 1850) (Tricladida, Dugesiidae) in Serbia, with Notes on its Ecology and Distribution. *Acta Zoologica Bulgarica*, *70*(1), 39–43.

Johnston, E. L., Dafforn, K. A., Clark, G. F., Rius, M., & Floerl, O. (2017). How Anthropogenic Activities Affect the Establishment and Spread of Non- Indigenous Species Post-Arrival. In S. J. Hawkins, A. J. Evans, A. C. Dale, L. B. Firth, D. J. Hughes, & I. P. Smith (Eds.), *Oceanography and Marine Biology* (pp. 389–419). Boca Raton FL.: CRC Press.

Jones, H. D., & Sluys, R. (2016). A new terrestrial planarian species of the genus Marionfyfea (Platyhelminthes: Tricladida) found in Europe. *Journal of Natural History*, *50*(41–42), 2673–2690. https://doi.org/10.1080/00222933.2016.1208907

Justine, J.-L., Winsor, L., Gey, D., Gros, P., & Thévenot, J. (2014). The invasive New Guinea flatworm *Platydemus manokwari* in France, the first record for Europe: time for action is now. *PeerJ*, *2*, e297. https://doi.org/10.7717/peerj.297

Kanana, Y., & Riutort, M. (2019). First record of freshwater planarian *Girardia tigrina* (Platyhelminthes, Tricladida, Continenticola) in Eastern Europe. *Zootaxa*, *4624*(4), 597–600. https://doi.org/10.11646/zootaxa.4624.4.13

Kawakatsu, M., Mitchell, R. W., Hirao, Y., & Tanaka, I. (1984). Ocurrence of *Dugesia dorotocephala* (Woodworth, 1897) (Turbellaria, Tricladida, Paludicola) in Honolulu, Hawaii. *Biological Magazine*, *22*, 1–9.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop

software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kenk, R. (1937). Sexual and Asexual Reproduction in *Euplanaria tigrina* (Girard). *Biological Bulletin*, *73*(2), 280–294. https://doi.org/DOI: 10.2307/1537589

Kenk, R. (1972). Freshwater planarians (Turbellaria) of North America. In *Biota of Freshwater Ecosystems, Identification Manual. Number 1.* Environmental Protection Agency, Washington, U.S.

Knakievicz, T., Vieira, S. M., Erdtmann, B., & Ferreira, H. B. (2006). Reproductive modes and life cycles of freshwater planarians (Platyhelminthes, Tricladida, Paludicula) from southern Brazil. *Invertebrate Biology*, *125*(3), 212–221. https://doi.org/10.1111/j.1744-7410.2006.00054.x

Kolar, C. S., & Lodge, D. M. (2001). Progress in invasion biology: predicting invaders. *Ecology & Evolution*, *16*(4), 199–204. https://doi.org/10.1016/S0169-5347(01)02101-2

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695–1701. https://doi.org/10.1093/molbev/mss020

Lázaro, E. M., & 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

Leria, L., Riutort, M., Romero, R., Ferrer, X., & Vila-Farré, M. Microplate tectonics and the environment as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography*, *In press*.

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

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., … Thieme, M. (2019). Global hydro-environmental sub-basin and river reach characteristics at high

spatial resolution. *Scientific Data*, *6*(1), 283. https://doi.org/10.1038/s41597-019-0300-6

Maddison, W. P., & Maddison, D. R. (2015). Mesquite: A modular system for evolutionary analysis. Version 3.04. Retrieved from http://www.mesquiteproject.org

Mateos, E., Jones, H. D., Riutort, M., & Álvarez-Presas, M. (2020). A new species of alien terrestrial planarian in Spain: Caenoplana decolorata. *PeerJ*, *8*, peerj.10013. https://doi.org/10.7717/peerj.10013

Mazza, G., Menchetti, M., Sluys, R., Solà, E., Riutort, M., Tricarico, E., … Mori, E. (2016). First report of the land planarian *Diversibipalium multilineatum* (Makino & Shirasawa, 1983) (Platyhelminthes, Tricladida, Continenticola) in Europe. *Zootaxa*, *4067*(5). https://doi.org/10.11646/zootaxa.4067.5.4

Murchie, A. K., & Gordon, A. W. (2013). The impact of the “New Zealand flatworm”, *Arthurdendyus triangulatus*, on earthworm populations in the field. *Biological Invasions*, *15*(3), 569–586. https://doi.org/10.1007/s10530-012-0309-7

Negrete, L., Lenguas Francavilla, M., Damborenea, C., & Brusa, F. (2020). Trying to take over the world: potential distribution of *Obama nungara* (Platyhelminthes: Geoplanidae), the Neotropical land planarian that has reached Europe. *Global Change Biology*, (April), 1–12. https://doi.org/10.1111/gcb.15208

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*(7), 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*(2), 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

Pickavance, J. R. (1971). The diet of the immigrant planarian *Dugesia tigrina* (Girard). Food in the wild and comparison with some British species. *Journal of Animal Ecology*, *40*, 637–650.

Puccinelli, I., & Deri, P. (1991). Comparative karyological analysis of some American planarians belonging to the genus *Dugesia* (subgenus *Girardia*) (Platyhelminthes, Tricladida). *Caryologia*, *44*(3–4), 225–232. https://doi.org/10.1080/00087114.1991.10797189

Puchałka, R., Dyderski, M. K., Vítková, M., Sádlo, J., Klisz, M., Netsvetov, M., … Jagodziński, A. M. (2020). Black locust (*Robinia pseudoacacia* L.) range contraction and expansion in Europe under changing climate. *Global Change Biology*, gcb.15486. https://doi.org/10.1111/gcb.15486

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, *41*(4), 629–643. https://doi.org/10.1111/jbi.12227

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/SYSBIO/SYY032

Reynoldson, T. B. (1985). Take-over of an Anglesey lake by an American species of triclad - the potential threat to the native triclad fauna. *Bulletin of the British Ecological Society*, *16*, 80–86.

Ribas, M., Riutort, M., & Baguñà, J. (1989). Morphological and biochemical variation in populations of *Dugesia (G.) tigrina* (Turbellaria, Tricladida, Paludicola) from the western Mediterranean: biogeographical and taxonomical implications. *Journal of Zoology (London)*, *218*, 608–626.

Richardson, D. M., Pyšek, P., Rejmánek, M., Barbour, M. G., Dane Panetta, F., & West, C. J. (2000). Naturalization and invasion of alien plants: Concepts and definitions. *Diversity and Distributions*, *6*(2), 93–107. https://doi.org/10.1046/j.1472-4642.2000.00083.x

Rickart, E. A., Balete, D. S., Rowe, R. J., & Heaney, L. R. (2011). Mammals of the northern Philippines: tolerance for habitat disturbance and resistance to invasive species in an endemic insular fauna. *Diversity and Distributions*, *17*(3), 530–541. https://doi.org/https://doi.org/10.1111/j.1472-4642.2011.00758.x

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (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

Salomidi, M., Katsanevakis, S., Issaris, Y., Tsiamis, K., & Katsiaras, N. (2013). Anthropogenic disturbance of coastal habitats promotes the spread of the introduced

scleractinian coral *Oculina patagonica* in the Mediterranean Sea. *Biological Invasions*, *15*(9), 1961–1971. https://doi.org/10.1007/s10530-013-0424-0

Schockaert, E. R., Hooge, M., Sluys, R., Schilling, S., Tyler, S., & Artois, T. J. (2008). Global diversity of free living flatworms (Platyhelminthes, “Turbellaria") in freshwater. *Hydrobiologia*, *595*(1), 41–48. Retrieved from http://dx.doi.org/10.1007/s10750-007-9002-8

Simkanin, C., Dower, J. F., Filip, N., Jamieson, G., & Therriault, T. W. (2013). Biotic resistance to the infiltration of natural benthic habitats: Examining the role of predation in the distribution of the invasive ascidian *Botrylloides violaceus*. *Journal of Experimental Marine Biology and Ecology*, *439*, 76–83. https://doi.org/https://doi.org/10.1016/j.jembe.2012.10.004

Sluys, R., Joffe, B., & Cannon, L. R. G. (1995). An alien flatworm in Australian waters. *Memoirs of the Queensland Museum*, *38*(2), 642.

Sluys, R., & Kawakatsu, M. (2001). Contribution to an inventory of the freshwater planarians of Australia and New Zealand (Platyhelminthes, Tricladida, Dugesiidae), with distribution maps of the species examined. *Beaufortia/Bull Zool Mus Univ Amsterdam*, *51*(10), 163–198.

Sluys, R., Kawakatsu, M., & Ponce de León, R. (2005). Morphological stasis in an old and widespread group of species: Contribution to the taxonomy and biogeography of the genus *Girardia* (Platyhelminthes, Tricladida, Paludicola). *Studies on Neotropical Fauna and Environment*, *40*, 155–180.

Sluys, R., Kawakatsu, M., & Yamamoto, K. (2010). Exotic freshwater planarians currently known from Japan. *Belgian Journal of Zoology*, *140*(Suppl.), 103–109. https://doi.org/10.3739/rikusui.68.461

Sluys, R., Ribas, M., & Baguñà, J. (1995). Taxonomy, ecology, and karyology of a new species of Phagocata from Spain, with a discussion on the phylogenetic systematics of the genus *Phagocata* s.l. (Platyhelminthes, Tricladida, Paludicola). *Canadian Journal of Zoology*, *73*, 557–568.

Sluys, R., & Riutort, M. (2018). Planarian Diversity and Phylogeny. In J. C. Rink (Ed.), *Planarian Regeneration. Methods and protocols. Methods in Molecular Biology Series.*

(1st ed., Vol. 1774, pp. 1–56). New York: Springer. https://doi.org/10.1007/978-1-4020-6682-5

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

Velde, G. van der. (1975). The immigrant triclad flatworm *Dugesia tigrina* (Girard) (Plalyhelminthes, Turbellaria). Range extension and ecological position in the Netherlands. *Hydrobiological Bulletin*, *9*, 123–130. https://doi.org/https://doi.org/10.1007/BF02263331

Velde, G. van der, Rajagopal, S., Kuyper-Kollenaar, M., Bij de Vaate, A., Thieltges, D. W., & MacIsaac, H. J. (2006). *Biological Invasions: Concepts to Understand and Predict a Global Threat*. https://doi.org/10.1007/978-3-540-33189-6_4

Vila-Farré, M., & Rink, J. C. (2018). The Ecology of Freshwater Planarians. In J. C. Rink (Ed.), *Planarian Regeneration. Methods and protocols. Methods in Molecular Biology Series.* (1st ed., Vol. 1774, pp. 173–205). New York: Springer. https://doi.org/10.1007/978-1-4939-7802-1_3

Vitousek, P. M., D'Antonio, C. M., Loope, L. L., Rejmánek, M., & Westbrooks, R. (1997). Introduced species: A significant component of human-caused global change. *New Zealand Journal of Ecology*, *21*(1), 1–16.

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*(2), 335–342. https://doi.org/10.1890/10-1171.1

Wright, J. F. (1987). Colonization of rivers and canals in Great Britain by *Dugesia tigrina* (Girard) (Platyhelminthes: Tricladida). *Freshwater Biology*, *17*(1), 69–78. https://doi.org/10.1111/j.1365-2427.1987.tb01029.x

Chapter III

Section 2

The expansion continues: Girardia arrives in Africa. First record of *Girardia sinensis* (Platyhelminthes, Tricladida, Continenticola, Dugesiidae) in Morocco.

LISANDRA BENÍTEZ-ÁLVAREZ, EDUARDO MATEOS, YOUNES EL OUANIGHI, NARD BENNAS, MAJIDA EL ALAMI, MARTA RIUTORT

The expansion continues: *Girardia* arrives in Africa. First record of *Girardia sinensis* (Platyhelminthes, Tricladida, Continenticola, Dugesiidae) in Morocco.

Lisandra Benítez-Álvarez[1,*], Eduardo Mateos[2], Younes el Ouanighi[3], Nard Bennas[3], Majida el Alami[3], Marta Riutort[1]

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

[2] Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona. Avinguda Diagonal 643, 08028, Barcelona, Catalonia, Spain.

[3] LESCB URL-CNRST N° 18, FS, Abdelmalek Essaadi University, Tetouan, Morocco. Avenue Sebta, Mhannech II. 93002

*corresponding author: lbenitezalvarez87@gmail.com

Under review in: *Zootaxa*. Format: Correspondence

Biological invasions are an important issue in the conservation of biodiversity. Special interest must deserve the freshwater environments, which are highly sensitive to change in the community composition (Gallardo et al., 2016; Havel, Kovalenko, Thomaz, Amalfitano, & Kats, 2015). The freshwater planarians are active members of aquatic ecosystems with important regulatory function as predators and used as biological indicators of water quality (Knakievicz, 2014; Wu & Li, 2018). The genus *Girardia* Ball, 1974 is the sister group of a clade including *Dugesia* Girard, 1850 and *Schmidtea* Ball, 1974 (Álvarez-Presas & Riutort, 2014). The three genera are the most studied groups in the Dugesiidae family. For *Girardia* a biogeographic hypothesis (Ball, 1974) proposed that during the formation of the current continents, its lineage was restricted to the land masses which would give rise to South America resulting in the Americas as the area of origin for all present-day species. A recent molecular study has given support to that hypothesis (Chapter II). However, in the 20's of last century the presence of *Girardia* genus was reported in Germany (Meinken, 1925), and by the end of the 60's its presence had reached a great part of Western Europe (Gourbault, 1969; Saló, Baguñà, & Romero, 1980). Nowadays, it is also present in Australia, Japan, and Hawaii (Sluys, Kawakatsu, & Ponce De León, 2005; Sluys, Kawakatsu, & Yamamoto, 2010). In addition, new reports of *Girardia tigrina* (Girard, 1850) in the

Balcans region (Ilić et al., 2018) and East Europe (Kanana & Riutort, 2019) have been published recently. Moreover, a new species, *G. sinensis* Chen & Wang, 2015 was described from a locality in China (Chen, Chen, Wu, & Wang, 2015), but molecular data analyses showed that the species have a North American origin (Kanana & Riutort, 2019; Chapter II, Chapter III.1), in consequence their presence in China will also be the result of an introduction. Although it is a valid species, its name is etymologically incorrect.

Here we report for the first time, the presence of *Girardia sinensis* in Morocco and consequently of *Girardia* genus in Africa. Since the specimens found were sexually immature, we identified them to species level using two molecular markers: a fragment of the mitochondrial gene *Cytochrome Oxidase I* (COI) and the nuclear gene *Elongation Factor 1 alpha* (EF1a).

Figure 1. Locality where the three samples were collected in Morocco river Laou

The specimens were collected in Laou river, Morocco, under a bridge on P4104 road, at 199 meters of altitude and the geographical coordinates: 35.26299, -5.26036 in April 2019 (Fig. 1). Three asexual individuals were identified as *Girardia* sp. by their external features (triangular shape of the head, dotted and light pigmentation) and were conserved in absolute ethanol.

Because the individuals were very small (3 mm approximately), the entire animal was used for the DNA extraction with Wizard® Genomic DNA Purification Kit (Promega) according to the manufacturer's instructions. DNA quantification, PCR amplification, purification, sequencing and contig obtaining were performed as described in (Chapter II). Six new sequences were obtained, and 57 sequences of *Girardia* were downloaded from GenBank (Fig. 2). The downloaded sequences are representative of 11 species, including the three North American species: *G. tigrina*, *G. dorotocephala* Woodworth, 1897, and *G. sinensis*; six South American species: *G. schubarti* Marcus, 1946, *G. biapertura* Sluys, 1997, *G. multidiverticulata* Souza, Morais, Cordeiro & Leal-Zanchet, 2015, *G. anderlani* Kawakatsu & Hauser, 1983, *G. sanchezi* Hyman, 1959, *G. clandistina* Sluys & Benítez-Álvarez 2022 (under review), and three non-identified individuals, one from Chile and two from Mexico. The final alignment had 1716 positions (837 for COI and 879 for EF1a). A phylogenetic tree was inferred with the Bayesian Inference method implemented in MrBayes v3.2.2 (Ronquist et al., 2012) using a partitioned scheme by codon and the General Time Reversible model + Gamma Distribution + Invariable Sites (GTR + Γ + I).

The phylogenetic tree (Fig. 2) places the three individuals from Morocco in the *G. sinensis* clade, supporting its classification as belonging to this species. This result also supports the hypothesis that *G. sinensis* is not an autochthonous species from China (Kanana & Riutort, 2019; Benítez-Álvarez et al, under review), but it is a North American lineage described for the first time outside its natural distribution area. At the same time, we can confirm with a broader taxonomic sampling the assignment of Kanana & Riutort (2019) of samples from Ukraine (Gti_ukraine in the tree) to *G. tigrina* species.

Sequence Identifier	Taxonomic Identification	GenBank accession code COI	EF1a
1280.1_Morocco	*G. sinensis**	**OM480692**	**OM487088**
1280.3_Morocco	*G. sinensis**	**OM480693**	**OM487089**
1280.4_Morocco	*G. sinensis**	**OM480694**	**OM487090**
Gsi_China	*G. sinensis*	KP091895.1	
138.8_si_Spain	*G. sinensis*	OM307091	OM418681
139.2_si_Spain	*G. sinensis*	OM307098	OM418657
140.1_si_Italy	*G. sinensis*	OM307099	OM418655
554_si_Netherlands	*G. sinensis*	OM307155	OM418665
660_si_Australia	*G. sinensis*	OM307161	OM418659
686_si_China	*G. sinensis*	OM307169	OM418662
84_do_France	*G. dorotocephala*	OM307073	OM349486
270.1_do_USA	*G. dorotocephala*	OM307112	OM349498
383_do_Hawaii	*G. dorotocephala*	OM307129	OM349501
466.2_do_USA	*G. dorotocephala*	OM307136	OM349491
684.2_do_Japan	*G. dorotocephala*	OM307166	OM349493
Gti_France	*G. tigrina*	DQ666042.1	
Gti_Ukraine	*G. tigrina*	MN092348	
138.3_ti_Spain	*G. tigrina*	OM307091	OM418681
143_ti_Italy	*G. tigrina*	OM307101	OM418683
542_ti_Germany	*G. tigrina*	OM307144	OM418678
659_ti_Canada	*G. tigrina*	OM307160	OM418676
679_ti_USA	*G. tigrina*	OM307163	OM418672
127_cl_Brazil	*G. clandistina*	OM307081	OM418688
133_cl_Brazil	*G. clandistina*	OM307085	OM418690
685.1_san_Chile	*G. sanchezii*	OM307167	OM418644
Gan_Brazil	*G. anderlani*	DQ666038.1	OM232748
262_mu_Brazil	*G. multidiverticulata*	OM307109	OM418642
131.1_bi_Brazil	*G. biapertura*	OM307082	
295_sp_Chile	*Girardia.* sp.2	OM307114	OM418636
Gsc_Brazil	*G. scubarti*	DQ666041.1	KJ599691.1
125_sc_Brazil	*G. scubarti*	OM307079	OM418648
399_sc_Brazil	*G. scubarti*	OM307131	OM418646
1059_sp_Mexico	*Girardia.* sp.1		OM418639
1062_sp_Mexico	*Girardia.* sp.2	OM307171	OM418640

*Identification after results

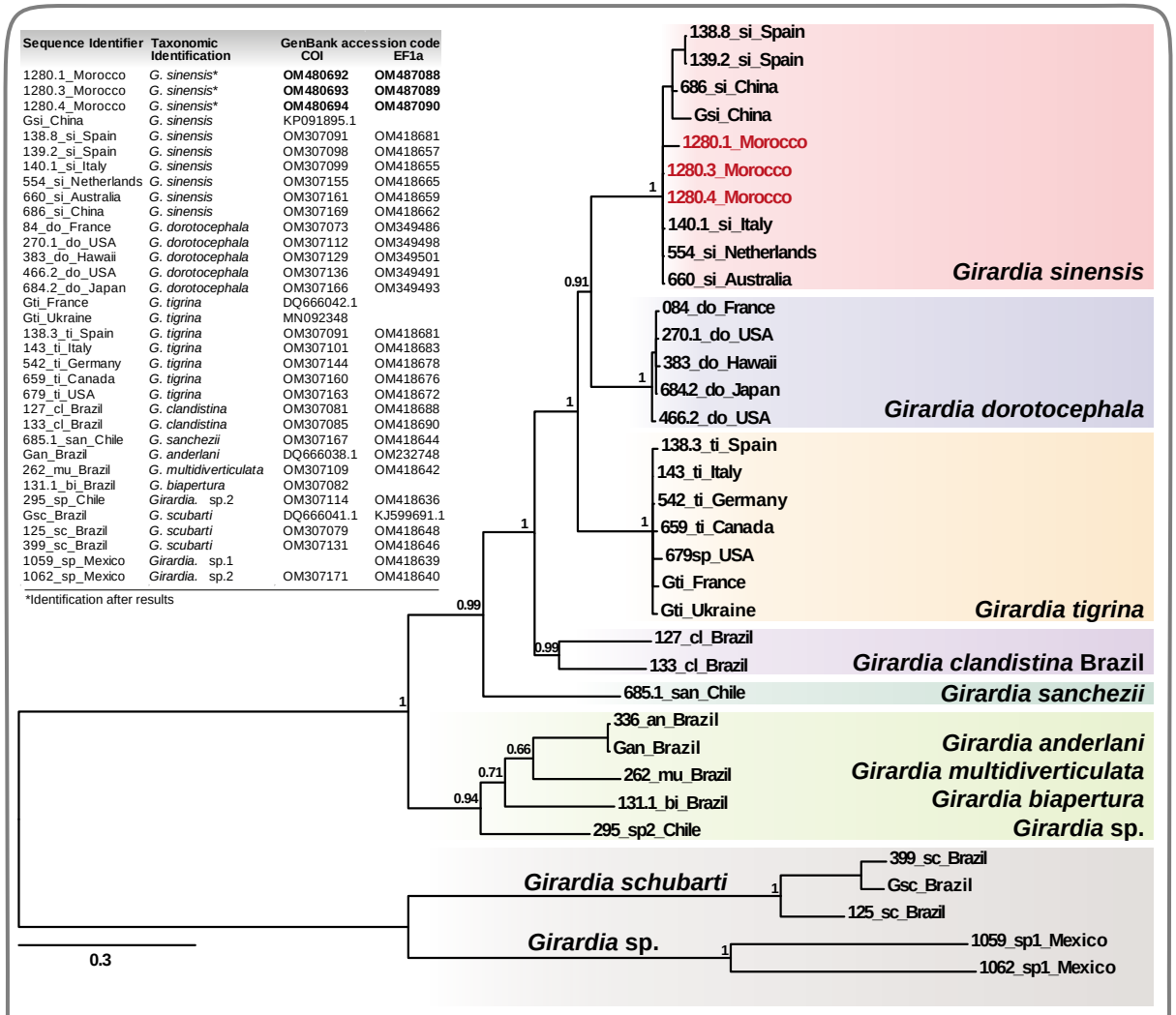


Figure 2. Bayesian Inference tree of concatenated sequences of *Cytochrome Oxidase I* (COI) and *Elongation Factor 1 alpha* (EF1a) genes. The samples from Morocco individuals are highlighted in red and the clades are differentiated by color. Values at nodes: posterior probability. Scale bar: number of substitutions per nucleotide position. The table shows the sequences used in the study, their taxonomic classification and GenBank accession code by gene

In summary, we report for the first time the occurrence of *Girardia* in the African continent, specifically of *G. sinensi* in the Laou river, Morocco. We used molecular data to perform the taxonomic identification of asexual individuals, validating the use of molecular techniques to species identification in this group. Our work supports previous results demonstrating the introduction of three North American species of *Girardia* in Europe, Asia and now, Africa. It is worrisome how the broadening of our samplings of freshwater planarians is taken us to uncover more and more introduced species all around the world. Taking into account the rapid colonization and the success of *Girardia* in Europe, its arrival to Africa deserves special attention for the potential damage that it could represent for African freshwater communities. Other

American invaders as the Nearctic “water boatman” *Trichocorixa verticalis verticalis* (Fieber, 1851), have been reported in Europe and Africa with high colonizing success (Carbonell, Céspedes & Green, 2021; L’Mohdi et al., 2010; Guareschi et al., 2013; Taybi et al., 2020). We want to make a call to the scientific community, about the importance of monitoring the introduced species in freshwater ecosystems, to detect any detrimental consequences to those delicate ecosystems, and to recall the usefulness of genetic data to rapidly identify and study the origin of the alien species.

Data availability

All new sequences have been deposited in GenBank

Acknowledgments

We are grateful to the Department of Water and Forests of Morocco for providing permissions for the study: Decisions 19/2019.

Author Contributions

LBA did the initial study design. LBA, EM, YeO, NB, and MeAl contribute with the sampling. LBA processed the samples and analysed the molecular data. All authors wrote the manuscript.

Conflicts of interest

All authors declare no conflict of interest to disclose.

References

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

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, pp. 339–401.

Carbonell, JA, Céspedes, V, Green, AJ. (2021) Is the spread of the alien water boatman *Trichocorixa verticalis verticalis* (Hemiptera, Corixidae) aided by zoochory and drought resistant eggs? *Freshwater Biology*; 66: 409– 420. https://doi.org/10.1111/fwb.13647

Chen, X., Chen, Y., Wu, C., & Wang, A. (2015) A new species of the genus *Girardia* (Tricladida : Dugesiidae) from China. *Zoological Systematic*, *40*(2), 166–178. https://doi.org/10.11865/zs.20150202

Gallardo, B., Clavero, M., Sánchez, M. I., Vilà, M., Havel, J. E., Kovalenko, K. E., Kats, L. B. (2016) Global ecological impacts of invasive species in aquatic ecosystems. *Global Change Biology*, *22*(1), 151–163. https://doi.org/10.1111/gcb.13004

Gourbault, N. (1969) Expansion de *Dugesia tigrina* (Girard), Planaire americaine introduite en Europe. *Annales de Limnologie*, *t. 5*, *fasc*, 3–7.

Guareschi, S., Coccia, C., Sánchez-Fernández D., Carbonell, J. A., Velasco, J., Boyero, L., Green, A. J., Millán, A. (2013) How Far Could the Alien Boatman *Trichocorixa verticalis verticalis* Spread? Worldwide Estimation of Its Current and Future Potential Distribution. *PLoS ONE*, 8(3): e59757. https://doi.org/10.1371/journal.pone.0059757

Havel, J. E., Kovalenko, K. E., Thomaz, S. M., Amalfitano, S., & Kats, L. B. (2015) Aquatic invasive species: challenges for the future. *Hydrobiologia*, *750*(1), 147–170. https://doi.org/10.1007/s10750-014-2166-0

Ilić, M. D., Tubić, B. P., Marinković, N. S., Marković, V. M., Popović, N. Z., Zorić, K. S., Paunović, M. M. (2018) First report on the non-indigenous triclad *Girardia tigrina* (Girard, 1850) (Tricladida, Dugesiidae) in Serbia, with Notes on its Ecology and Distribution. *Acta Zoologica Bulgarica*, *70*(1), 39–43.

Kanana, Y., & Riutort, M. (2019) First record of freshwater planarian *Girardia tigrina* (Platyhelminthes, Tricladida, Continenticola) in Eastern Europe. *Zootaxa*, *4624*(4), 597–600. https://doi.org/10.11646/zootaxa.4624.4.13

Knakievicz, T. (2014) Planarians as invertebrate bioindicators in freshwater environmental quality: the biomarkers approach. *Ecotoxicology and Environmental Contamination*, *9*(1), 1–12. https://doi.org/10.5132/eec.2014.01.001

L'Mohdi, O., Bennas, N., Himmi, O., Hajji, K., El Haissoufi, M., Hernando, C. Carbonell, J.A., Millán, A. (2010) *Trichocorixa verticalis verticalis* (Fieber, 1851) (Hemiptera, Corixidae): une nouvelle especies exotique au Maroc. *Boletin de la Sociedad Entomológica Aragonesa*, 46: 395–400

Meinken, H. (1925) Ein verkannter Laihruher. *Wochesher, Aquarien Terrarienkunde*, *22*, 94–96.

Taybi, A.F., Mabrouki, Y., Chavanon, G., Berrahou, A., Millán, A. (2020) The alien boatman *Trichocorixa verticalis verticalis* (Hemiptera: Corixidae) is expanding in Morocco. *Limnetica*, 39(1): 49–59. https://doi.org/ 10.23818/limn.39.04

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (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

Saló, E., Baguñà, J., & Romero, R. (1980) Les Planàries d 'aigües dolces a Catalunya i les illes Balears . I . Clau sistemàtica i distribució geogràfica. *Butlletí de La Institució Catalana d'Història Natural*, *45*, 15–30.

Sluys, R., Kawakatsu, M., & Ponce De León, R. (2005). Morphological stasis in an old and widespread group of species : Contribution to the taxonomy and biogeography of the genus *Girardia* (Platyhelminthes, Tricladida, Paludicola). *Studies on Neotropical Fauna and Environment*, *40*(2), 155–180. https://doi.org/10.1080/01650520500070220

Sluys, R., Kawakatsu, M., & Yamamoto, K. (2010). Exotic freshwater planarians currently known from Japan. *Belgian Journal of Zoology*, *140*, 103–109.

Wu, J. P., & Li, M. H. (2018). The use of freshwater planarians in environmental toxicology studies: Advantages and potential. *Ecotoxicology and Environmental Safety*, *161*(May), 45–56. https://doi.org/10.1016/j.ecoenv.2018.05.057

Chapter IV

Section 1

Transcriptomic data uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean

LISANDRA BENÍTEZ-ÁLVAREZ, LAIA LERIA, ROSA FERNÁNDEZ, EDUARDO MATEOS, NARD BENNAS, YOUNES EL OUANIGHI, MAJIDA EL ALAMI, MOHAMED YACOUBI-KHEBIZA, HOUSSAM AYT OUGOUGDAL & MARTA RIUTORT

Transcriptomic data uncovers a complex evolutionary history for the planarian genus *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean

Lisandra Benítez-Álvarez[1], Laia Leria[1], Rosa Fernández[2], Eduardo Mateos[3], Nard Bennas[4], Younes el Ouanighi[4], Majida el Alami[4], Mohamed Yacoubi-Khebiza[5], Houssam Ayt Ougougdal[5] & Marta Riutort[1,*]

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

[2]Metazoa Phylogenomics Lab, Biodiversity Program, Institut de Biologia Evolutiva (CSIC- Universitat Pompeu Fabra). Passeig marítim de la Barceloneta 37-49. 08003 Barcelona, Catalonia, Spain.

[3]Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona. Avinguda Diagonal 643, 08028, Barcelona, Catalonia, Spain.

[4]Laboratoire Ecologie, Biodiversité et Environnement; Département de Biologie; Faculté des Sciences; Université Abdelmalek Essaâdi. Avenue Sebta, Mhannech II. 93002 - Tetouan. Morocco

[d]Water, Biodiversity and Climate Change Laboratory, Semlalia Faculty of Sciences, Cadi Ayyad University. BP 2390, Avenue Le Prince Moulay Abdellah, 400 10 Marrakech, Morocco

*corresponding author: mriutort@ub.edu, (34) 934035432

Highlights:

- Evolutionary history of *Dugesia* (Platyhelminthes, Tricladida) in the Western Mediterranean revisited. New approaches for phylogenetic studies in freshwater planarians
- Hundreds of single copy orthologs extracted from transcriptomic data are used for phylogenetic inference.
- Unexpected phylogenetic position of asexual fissiparous populations unveils putative ancient asexuals and the effect of long-term asexuality in populations genetic diversity.
- The biogeographic history of *Dugesia* in the Western Mediterranean is shaped by paleogeological and paleoclimatic related processes and the diversity of reproductive strategies.

Abstract

The Mediterranean is one of the most biodiverse areas of the Paleartic region. Here, we elucidate the evolutionary history of freshwater planarians belonging to the genus *Dugesia* in the Western Mediterranean basin. Based on large datasets of single copy orthologs obtained from transcriptomic data, we investigated the evolutionary history of the genus in this area, shaped by the paleogeological history of the region, and speciation in islands. These processes led to the diversification of three main biogeographic clades. The internal relationships of these major clades were analysed with several representative samples per species. The use of large datasets regarding the number of *loci* and samples, as well as state-of-the-art phylogenomic inference methods allowed us to answer different unresolved questions about the evolution of particular groups, such as the diversification path of *D. subtentaculata* in the Iberian Peninsula and its colonization of Africa. Additionally, we analysed here for the first time with a comprehensive number of samples several asexual Iberian populations whose assignment at the species level has been an enigma through the years. Our data bring to the light interesting information on the evolutionary history of these populations and the diversification of the major clade to which they belong. We hypothesize that the unexpected topology showed by these asexual individuals in the phylogenetic tree branching at the base of sexual clades, is related to long term asexuality. This work represents the first phylotranscriptomic analysis of Tricladida, laying the first stome of the genomic era in phylogenetic studies on this taxonomic group.

Keywords: *Dugesia, Western Mediterranean, transcriptome, phylogeny, evolution*

1. Introduction

Freshwater planarians constitute one of the most diverse and broadly distributed groups of free-living flatworms. Due to their low dispersion capability and specific ecosystem requirements (Vila-Farré & Rink, 2018), the evolutionary history of these animals has been strongly shaped by the geological changes (Leria, Riutort, Romero, Ferrer, & Vila-Farré, *in press*; Solà et al., *in press*; Solà, Sluys, Gritzalis, & Riutort, 2013) .

The genus *Dugesia* has a broad distribution covering Eurasian, African and Australasian regions. It has been proposed on the basis of molecular data and biogeographic analysis that *Dugesia* arrived to Western Europe from Africa, through terrestrial connections in the Eocene (55-33 mya) splitting, around 30 mya from the Eastern lineages that arrived in another wave from North Africa through the Arabian Peninsula and the ancient Aegean region (Solà et al., *in press*) .

Thirteen species of *Dugesia* are described as endemic from the Western Mediterranean (Leria et al., *in press*); being *D. gonocephala*, the only one that extends its distribution out of the region to the East and the North of Europe (Fig. 1). Moreover, two new lineages suspected to be new species have been recently found in Morocco in previous studies (Leria et al., *in press*) and in the present one. On the other hand, *D. sicula* and *D. maghrebiana,* also present in the Mediterranean region, belong to a different and very distant African lineage (Solà et al., *in press*).

Despite its diversity, the evolutionary history of *Dugesia* in the Western Mediterranean remains poorly studied. Morphological studies have been limited principally because there are many asexual populations in the region (Fig. 1). Asexual reproduction prevents the assignment of populations to specific level due to their lack of reproductive structures, which are the principal source of evidence for taxonomic assignment. Thus, the net result for *Dugesia* in the Mediterranean is the presence of a large number of asexual triploid populations that were classified as *Dugesia gonocephala sensu lato*. This problem began to be solved when molecular data was applied to analyse the *Dugesia* populations present in the Mediterranean area (Lázaro et al., 2009), demonstrating that DNA sequences facilitated the assignment of individuals to its species, and also could yield information to envision their phylogenetic relationships. In that initial work, a basic scheme of major relationships within the group

was found, but the use of only two molecular markers left many relationships poorly resolved. More recently, a study analysed representative samples of a large part of the species from the region using six markers. They obtained a dated phylogeny to put the evolution of the genus in this area under a temporal frame, and with the help of niche modelling and ancestral areas reconstruction analyses, the authors proposed a complex and interesting biogeographic hypothesis (Leria et al., *in press*). However, in their phylogeny still non supported nodes remained, and only one or two representatives per species were included, so that internal relationships and evolutionary history whithin species was not analysed.

In this respect, one of the most diverse and extended lineages in the Western Mediterranean *Dugesia* group is *D. subtentaculata*. This species was considered one unique species in all Iberian Peninsula, South of France and Balearic Island for a long time (De Vries, 1986). However, an integrative taxonomic study divided it in four species: *D. vilafarrei, D. corbata* and *D. aurea*, all sexual populations restricted to their type localities in the South of the Iberian Peninsula and two localities in the Mallorca Island respectively, and *D. subtentaculata sensu stricto,* with sexual, asexual, and facultative (sexual and asexual individuals in the same locality) populations distributed in all the Iberian Peninsula, South of France and the North of Africa (Leria et al., 2020) . The broad distribution of *D. subtentaculata* makes the study of its population structure essential to understand the evolutionary processes that drove its present distribution and evolution. However, Leria et al., (2020) found it was not possible to reconstruct a phylogeny of the populations, possibly due to the small number of markers used for that and the noise that introduces the Mosaic-Meselson effect, a genetic consequence of asexuality described just in this species (Leria, Vila-Farré, Solà, & Riutort, 2019).

Another interesting question related to the species from the Western Mediterranean region has arisen recently. *D. etrusca* and *D. liguriensis* are endemic from the Tuscany and Liguria regions respectively (Benazzi, 1946; De Vries, 1988) and have been described as strictly sexual. Even so, in the last decades new fissiparous reproducing populations from Catalonian region in the Iberian Peninsula have been assigned to this clade using molecular data, but without a precise assignment to species level (Baguñà et al., 1999; Lázaro et al., 2009). These populations are interesting since apparently, they represent a restricted asexual lineage geographically isolated from the sexual populations. The Southern region of France situated between

both groups of organisms has been sampled, without finding any other population that could be assigned to one of these species.

A new approach is now necessary to solve the remaining uncertainties in the evolutionary history of *Dugesia* in the Western Mediterranean. Nowadays, the access to whole genome information has opened the door to a new era in the phylogenetic studies with the substantial increase of the informative regions to analyse. Phylogenomics has demonstrated to be a powerful approach to reconstruct molecular phylogenies and have helped to resolve old questions about the evolution of life (Fernández & Gabaldón, 2020; Guijarro-Clarke, Holland, & Paps, 2020; Y. Li, Shen, Evans, Dunn, & Rokas, 2021). However, genomes are not always accessible, more if the studied group has never been sequenced or no references are available. In those cases, phylotranscriptomics arises as a good and cheaper alternative. Transcriptomic data has been used to resolve several phylogenetic questions in non-model organisms (Feng et al., 2021; Fernández, Edgecombe, & Giribet, 2016; Fernández, Sharma, Tourinho, & Giribet, 2017; Foley et al., 2019; Lemer, Bieler, & Giribet, 2019) and its correct performance compared to genomic data has been demonstrated (Cheon, Zhang, & Park, 2020).

Here, for the first time, we use transcriptomic data to carry out a phylogenetic study in freshwater planarians, focusing on *Dugesia* species belonging to the West Mediterranean clade. We include representatives of most species known to date from the area (11 out of 13 described species, plus two outgroups. Additionally, representatives of not formally described new candidate species were analysed. To obtain and analyse this data, we designed a strategy from the sampling to the phylogenetic inference process, new for the freshwater planarians. Our aims are (1) to obtain a better resolved and more comprehensive phylogeny of *Dugesia* in the Western Mediterranean, (2) to solve the evolutionary history within *D. subtentaculata*, and (3) to understand the origin of the Iberian Peninsula asexual populations and their assignment to either *D. etrusca* or *D. liguriensis* species.

2. Material and methods

2.1. Taxon sampling

A thorough sampling effort was carried out throughout the Western Mediterranean with the objective of covering all the area. We visited known localities

of *Dugesia* species from the Western Mediterranean clade as well as some new localities from April 2018 to March 2020 (Fig. 1). Two species; *D. malickyi* from Mexiates and a candidate new species from Eleonas (Sluys et al., 2013), were collected in Greece to be used as outgroup. The two taxa have been previously anlysed in a biogeographic framework and shown to belong to the Eastern Mediterranean clade sister group of the Western Mediterranean *Dugesia* species clade (Solà et al., *in press*, 2013). In addition, different localities from Morocco were sampled, looking for *D. tubqalis* and the new species reported in Leria et al., (2020). Unfortunately, no representatives of these taxa were found, but other samples collected in the region were included in the analyses (Table 1, Fig. 1, Table S1).

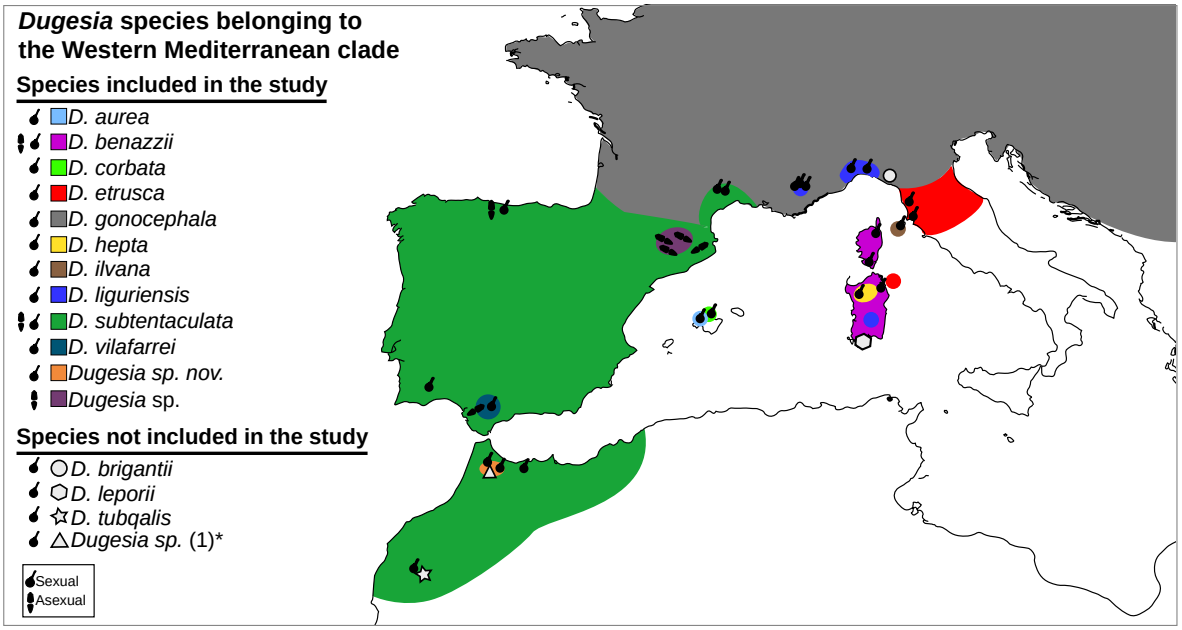


Figure 1. Distribution map of all *Dugesia* species belonging to the Western Mediterranean clade and their reproductive strategies. The map shows their known distribution area. The location of the sampling points included in this study are indicated with the icon that also indicates the reproductive strategy of the sampled population. Species list and distribution are extracted from Leria et al., (2022). *: candidate new species in Leria et al., (2020, *in press*).

Samples were preliminary assigned to species level according to their locality based on information from previous studies (Lázaro et al., 2009; Leria et al., *in press*). Depending on the distance from the sampling point to the laboratory, animals were preserved *in situ* or in the laboratory. All the material used for sampling and handling the animals was cleaned with RNAase away, and the autoclavable material was sterilized twice at 120ºC for 20 min. For the transport of live animals, we used a

portable refrigerator and the tubes with water from the river and the planarians were opened twice a day to aerate them. The fixations to preserve the RNA, *in situ* or in the laboratory, were done with RNAlater (SIGMA) following the recommendations of the manufacturer. When the size of the animal allowed it, a small portion of the posterior end was cut and fixed in absolute ethanol. In the case of very small animals, some individuals from the same sampling point were conserved in absolute alcohol and the rest in RNAlater. DNA for species identification was extracted from absolute ethanol fixed samples.

Table 1. Samples analysed in this study. Are detailed the code used in the trees, taxonomic classification, and the locality. For more information see Supplementary Table S1

Sample Code	Species	Reproductive Strategy	Region, Country	Locality
Daur_1	*D. aurea*	Sexual	Majorca, Spain	Soller
Daur_2	*D. aurea*	Sexual	Majorca, Spain	Soller
Daur_3	*D. aurea*	Sexual	Majorca, Spain	Soller
DbenSard_6	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo
DbenSard_7	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo
DbenSard_9	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo
DbenCors_North_1	*D. benazzii*	Sexual	Corsica, Italy	Campile
DbenCors_North_2	*D. benazzii*	Sexual	Corsica, Italy	Campile
DbenCors_North_3	*D. benazzii*	Sexual	Corsica, Italy	Campile
DbenCors_South_5	*D. benazzii*	Sexual	Corsica, Italy	Monacia-d'Aullène
DbenCors_South_6	*D. benazzii*	Sexual	Corsica, Italy	Monacia-d'Aullène
Dcorb_1	*D. corbata*	Sexual	Majorca, Spain	Sa Calobra
Dcorb_2	*D. corbata*	Sexual	Majorca, Spain	Sa Calobra
Dcorb_3	*D. corbata*	Fissiparous	Majorca, Spain	Sa Calobra
DetruParr_1	*D. etrusca*	Sexual	Italy	Parrana
DetruParr_2	*D. etrusca*	Sexual	Italy	Parrana
DetruPie_2	*D. etrusca*	Sexual	Italy	Pieve
DetruPie_3	*D. etrusca*	Sexual	Italy	Pieve
DetruPie_4	*D. etrusca*	Sexual	Italy	Pieve
Dgono_1	*D. gonocephala*	Sexual	France	Montpellier
Dgono_7	*D. gonocephala*	Sexual	France	Montpellier
Dgono_8	*D. gonocephala*	Sexual	France	Montpellier
Dhept_1	*D. hepta*	Sexual	Sardinia, Italy	Logulento
Dhept_2	*D. hepta*	Sexual	Sardinia, Italy	Logulento
Dhept_5	*D. hepta*	Sexual	Sardinia, Italy	Logulento
Dilv_1	*D. ilvana*	Sexual	Italy	Elba
Dilv_2	*D. ilvana*	Sexual	Italy	Elba
Dilv_4	*D. ilvana*	Sexual	Italy	Elba

DliguBis_1	*D. liguriensis*	Sexual	Italy	Bisagno
DliguBis_2	*D. liguriensis*	Sexual	Italy	Bisagno
DliguBis_3	*D. liguriensis*	Sexual	Italy	Bisagno
DliguAlp_1	*D. liguriensis*	Sexual	France	Alps Marìtims
DliguAlp_3	*D. liguriensis*	Sexual	France	Alps Marìtims
DliguAlp_4	*D. liguriensis*	Sexual	France	Alps Marìtims
DliguGarda_1	*D. liguriensis*	Sexual	France	La Garda
DliguSas_2	*D. liguriensis*	Sexual	Italy	Sassello
DliguSas_3	*D. liguriensis*	Sexual	Italy	Sassello
DliguSas_4	*D. liguriensis*	Sexual	Italy	Sassello
DliguTriga_1	*D. liguriensis*	Sexual	France	Trigance
DliguTriga_2	*D. liguriensis*	Sexual	France	Trigance
DsubMont	*D. subtentaculata*	Sexual	France	Montpellier
DsubBosq_1	*D. subtentaculata*	Fissiparous	Andalusia, Spain	El Bosque
DsubBosq_2	*D. subtentaculata*	Fissiparous	Andalusia, Spain	El Bosque
DsubCangAsex_5	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas
DsubCangAsex_6	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas
DsubCangAsex_7	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas
DsubCangSex_2	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas
DsubCangSex_3	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas
DsubCangSex_4	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas
DsubMor_North_1	*D. subtentaculata*	Sexual	Morocco	Magoo Timriouen
DsubMor_North_2	*D. subtentaculata*	Sexual	Morocco	Magoo Timriouen
DsubMor_North_3	*D. subtentaculata*	Sexual	Morocco	Beni H´amed
DsubMor_South_1	*D. subtentaculata*	Sexual	Morocco	Imlil
DsubMor_South_2	*D. subtentaculata*	Sexual	Morocco	Imlil
DsubMch_1	*D. subtentaculata*	Sexual	Portugal	Monchique
DsubMch_2	*D. subtentaculata*	Sexual	Portugal	Monchique
DsubMch_4	*D. subtentaculata*	Sexual	Portugal	Monchique
DsubStFe_1	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe
DsubStFe_2	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe
DsubStFe_3	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe
Dvila_1	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque
Dvila_2	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque
Dvila_3	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque
Dsp_nov_MorNorth_1	*Dugesia* sp. *nov*	Sexual	Morocco	Beni H´amed
Dsp_nov_MorNorth_6	*Dugesia* sp. *nov*	Sexual	Morocco	Beni H´amed
Dsp_nov_MorNorth_7	*Dugesia* sp. *nov*	Sexual	Morocco	Beni H'amed
DspF.US_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de l´Us
DspF.US_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de l´Us
DspTrilla_1	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspTrilla_2	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspTrilla_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspTrilla_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla

DspTrilla_5	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspTrilla_6	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspBerga_1	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
DspBerga_2	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
DspBerga_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
DspBerga_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
DspBerga_5	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
Outgroup				
Dma_1	*D. malickyi*	Sexual	Greece	Mexiates
Dma_2	*D. malickyi*	Sexual	Greece	Mexiates
DspEast_1	*Dugesia* sp.	Sexual	Greece	Eleonas
DspEast_2	*Dugesia* sp.	Sexual	Greece	Eleonas

The sexuality or asexuality of the individuals was assessed observing them under the stereomicroscope and taking into account previous information about known populations. Sexual individuals were recognized by the presence of the gonopore, the external aperture of the copulatory apparatus, and asexual individuals, by the presence of the blastema, the regenerating bud formed where the fission of the individual has taken place. Each population was assigned to one or both reproductive strategies.

Remaining samples and Nucleic acid extractions are stored in the freezers at the Department of Genetics, Microbiology and Statistics (Universitat de Barcelona) (Table 1). The methodology has been summarized in a diagram highlighting the principal steps of all the procedure: 1) data collection, 2) ortholog search, and 3) phylogenetic inference (Fig. 2).

2.2. Nucleic acids extraction and RNA library preparation

RNA was extracted using Trizol (Thermo Fisher Scientific, USA) following the manufacturer's instructions. The total RNA quantification and the integrity was assessed with Qubit and Bioanalyzer in the Centres Científics i Tecnològics, Universitat de Barcelona (CciTUB). Truseq stranded and ribo-zero libraries were constructed in Macrogen Inc., (Macrogen Europe, Madrid) to obtain Illumina paired-end reads.

DNA was extracted using Wizard® Genomic DNA Purification Kit (Promega) following the manufacturer's instructions and quantified using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA).

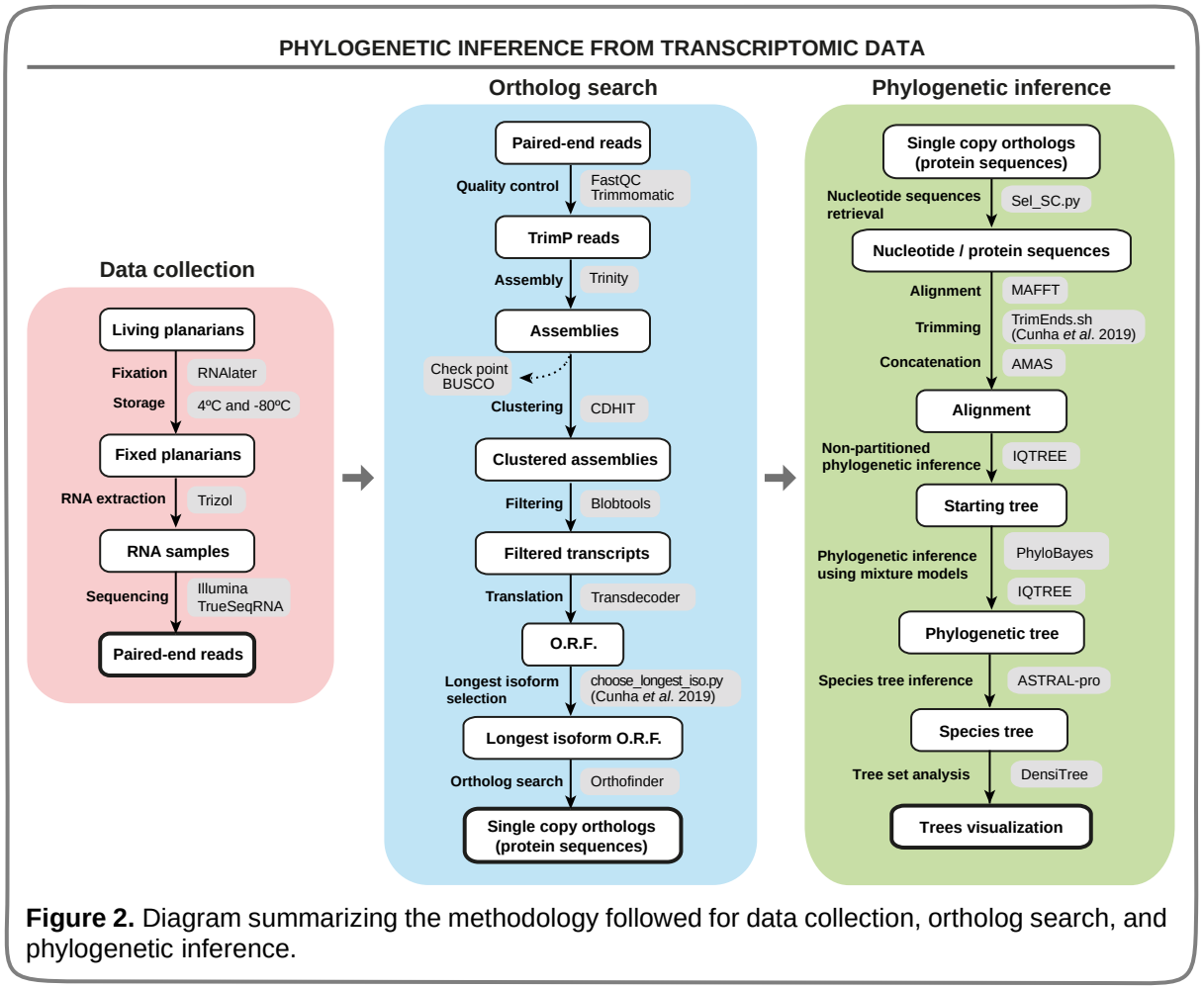


Figure 2. Diagram summarizing the methodology followed for data collection, ortholog search, and phylogenetic inference.

2.3. DNA-Barcoding identification

A fragment of mitochondrial *Cytochrome Oxidase I* (COI) was used as a marker to corroborate the species assignment done in the field. A set of sequences downloaded from GenBank database (Table S2), were used as reference of all *Dugesia* species reported in the area of study. The positions of the new sequences with respect to the references in a phylogenetic tree was used as evidence for the assignment.

A fragment of approximately 800 bp was amplified by Polymerase Chain Reaction (PCR), using 0.4 µM of the BarT (M Álvarez-Presas, Carbayo, Rozas, & Riutort, 2011) and COIR (Lázaro et al., 2009) primers in 25 µl of final reaction volume with $MgCl_2$ (2.5 mM), dNTPs (30 µM), and 0.75 U of Go Taq® DNA polymerase enzyme (Promega Madison, Wisconsin, USA) with its buffer (1X). The amplification conditions

were as following: 1) 2” at 95°C, 2) 50’ at 94°C, 3) 45’ at 43°C, 4) 50’ at 72°C, 5) 4” at 72°C, with 35 cycles of steps 2, 3, and 4.

The amplification was checked in agarose gels (1%) and the PCR products were purified by ultrafiltration in a Merck Millipore MultiScreen System (Darmstadt, Germany). The purified fragments were sequenced by Macrogen Inc., (Macrogen Europe, Madrid) using only COIR. In order to obtain the final contigs, chromatograms were analysed with Genious v.10 (Kearse et al., 2012).

The sequences were aligned with ClustalW Multiple Alignment on the BioEdit Sequence Alignment Editor (Hall, 1999). A Bayesian Inference tree was obtained using MrBayes v3.2.2 (Ronquist et al., 2012) with 10 million generations, sampling every 1000 generations, 25% burn-in, and using three partitions by codon position. Individuals were assigned to the species with which they formed a monophyletic group.

2.4. Bioinformatic Workflow for transcriptomic analysis

The detailed bioinformatic workflow used is available at https://github.com/lisy87/dugesia-transcriptome with all necessary scripts and commands to perform every step.

2.4.1 Quality control and trimming

To explore the quality of the RNA-seq reads, FastQC (Andrews, 2010) was used using default parameters. The raw reads were filtered with Trimmomatic (Bolger, Lohse, & Usadel, 2014) to remove low quality reads and universal Illumina adapters, as well as reads with low quality bases, and reads shorter than 36 bp.

2.4.2. Assembly and clustering

Paired reads were *de novo* assembled using Trinity v2.9.1 (Grabherr et al., 2011; B. J. Haas et al., 2013) following the default options. Some samples were selected for a completeness assessment with BUSCO v5.2.2 (Manni, Berkeley, Seppey, Simão, & Zdobnov, 2021; Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015), averaging a completeness value close to 90% using the metazoan database (OB10). Transcripts were clustered using CD-HIT EST (Fu, Niu, Zhu, Wu, & Weizhong, 2012; W. Li & Godzik, 2006), applying a sequence identity threshold of 0.99, and retaining an average of 97% of the transcripts (Table S3).

2.4.3. Transcript filtering

Transcripts were filtered using a strategy based on the results of Blobtools v 3.6 (Challis & Paulini, 2021; Laetsch & Blaxter, 2017). The assembly (retained transcripts after clustering with CD-HIT), the mapped reads against the assembly (obtained with BWA (H. Li & Durbin, 2009), and the blast of the assembly against the nucleotide database of the NCBI (performed with BLAST+ (Camacho et al., 2009) were used for the individual analysis of every sample with Blobtools. Transcripts with hits against Platyhelminthes and no-hits were captured. No-hits were also captured, as there is not much information about Platyhelminthes in databases. This way, all transcripts matched against other groups were dropped out. An average of 98.1% of all the transcripts were retained, except for the sample Dsp_nov_MorNorth_6 (Table 1) which was eliminated from posterior analysis because of its high content of contaminants transcripts (Table S3).

2.4.4. Ortholog search

Filtered transcripts were translated to proteins using TransDecoder v 5.5.0 (B. Haas & Papanicolaou, 2019) and the longest isoforms were selected using the script “choose_longest_iso.py” (Cunha & Giribet, 2019; Fernández et al., 2014). An average of 21,215 longest isoforms by sample was obtained.

To accomplish our three aims, we used three main groups of samples. The first one included representatives of all taxa to obtain the phylogeny for the whole Western clade. A second group of samples was defined to study the phylogenetic relationships inside *D. subtentaculata.* This group included all samples assigned to this species and samples classified as *D. vilafarrei*, used as outgroup. We refer to this group as subtentaculata samples group. The last group, used to assign the unclassified samples from Font de l’Us, Font de la Trilla, and Berga to species level, includes these samples and the representatives of *D. etrusca, D. liguriensis, D. ilvana, and D. gonocephala* (the last, used as outgroup). We refer to this group as etrusca-liguriensis samples group.

OrthoFinder v 3.6 (Emms & Kelly, 2019) was used to perform the ortholog search using the longest isoforms. Three searches were performed, one for every samples group (Tables S4-5). A total of 717, 4175 and 1984 Single Copy orthologs (SC) were obtained for all samples; subtentaculata samples group, and etrusca-

liguriensis samples group respectively (Table S4). The protein sequences of SC were extracted from the Orthofinder output and the corresponding nucleotide sequences were extracted from Orthofinder and Transdecoder outputs, using two custom python scripts.

2.4.5. Alignment and concatenation

The protein and nucleotide datasets were analysed independently following the same steps and modifying the specific options for each data type.

The individual SC obtained for each of the three orthologs' searches were aligned with MAFFT v 7.487 (Katoh & Standley, 2013) using the --auto and --maxiterate options, with 1000 iterations. The script trimEnds.sh (Cunha & Giribet, 2019) was used to trim the ends of the alignments, and poorly aligned regions were removed afterwards with the software TrimAl v 1.2 (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009) using the automated trimming heuristic option, which is an optimised option for Maximum Likelihood phylogenetic tree reconstructions. The concatenated files were obtained with the program AMAS (Borowiec, 2016).

To perform the analyses, six datasets were built using the output from the three ortholog searches with different compositions of samples and 100% of gene occupancy (Table 2, Table S5). These datasets were designed to respond the main questions in our study: 1) datasets 1-3 constructed from ortholog search 1, the phylogeny of *Dugesia* in the Western Mediterranean, 2) dataset 4 constructed from ortholog search 2, the relationship between populations of *D. subtentaculata* , and 3) datasets 5 and 6 constructed from ortholog search 3, the species assignment of Iberian populations to *D. etrusca* or *D. luguriensis* species. The datasets 2 and 3 were obtained after extracting samples from dataset 1, and the dataset 6 was obtained after extracting samples from dataset 5 (Table S5). These reduced datasets were re-aligned and re-processed after removing samples. For all datasets the protein and nucleotide information were analysed (Table 2, Table S6).

2.4.6. Phylogenetic Inference

Several methods of phylogenetic inference were used to analyse the different datasets. Maximum Likelihood (ML), Bayesian Inference (BI), and Multispecies Coalescence Model (MSC) were performed in IQ-TREE (Minh et al., 2020),

Table 2. Datasets used in the study, and analysis performed with every one of them. Samples included in datasets in bold font were used for ortholog search. See Supplementary Table S5 for further information about the data set composition.

Datasets		Number of samples	Number of SC	Number of Positions		Analyses					
Abbreviation	**Explanation***			**Protein**	**Nucleotide**	**ML Exp. Prot/Nuc**	**ML-Mixt Prot/Nuc**	**BI-Mixt Prot/Nuc**	**ASTRAL Prot/Nuc**	**DensiTree Prot/Nuc**	**ARC Nuc**
Dataset 1	All samples retained after filtering	82	717	274960	839555	X	X				
Dataset 2	Reduced Dataset 1 to infer tree 1	29	717	276745	841425	X	X	X	X	X	
Dataset 3	Reduced Dataset 1 to infer tree 2	13	717	-	831918						X
Dataset 4	Subtentaculata group	23	4175	1764320	5376224	X	X				
Dataset 5	Etrusca-liguriensis group	36	1984	742966	2283579	X	X		X		
Dataset 6	Dataset 5 without samples from Berga	31	1984	744252	2255464	X	X		X		

Abbreviation	***Further Explanation**
ML Exp.	Prot/Nuc Maximum Likelihood without partitions. Protein and nucleotide data
ML-Mixt	Prot/Nuc Maximum Likelihood using Mixture Models. Protein and nucleotide data
BI-Mixt	Prot/Nuc Bayesian Inference using Mixture Models. Protein and nucleotide data
ASTRAL	Prot/Nuc Species Tree using individual gene trees. Protein and nucleotide data
DensiTree	Prot/Nuc Visualization of individual gene trees. Protein and nucleotide data
ARC Nuc	Ancestral Reconstruction Character. Nucleotide data
SC	Single Copy Orthologs

PhyloBayes (Lartillot & Philippe, 2004), and ASTRAL-pro (Zhang, Scornavacca, Molloy, & Mirarab, 2020) respectively. The parameters to carry out every analysis are detailed in Table S6, and a summary of performed analyses by dataset is shown in Table 2.

Maximum Likelihood Trees: one partition

To explore the data and obtain starting trees for posterior analyses, we used the ML approximation implemented in IQ-TREE without defining partitions. These analyses were run using the ModelFinder Plus (MFP) option for the -m parameter, thus looking for the best-fit model (Kalyaanamoorthy, Minh, Wong, Von Haeseler, & Jermiin, 2017) and 1000 to 10,000 replicates of ultrafast bootstrap (Hoang, Chernomor, Haeseler, Minh, & Vinh, 2018) depending on the analysis (Table S6).

Maximum Likelihood Trees: Mixture models

We obtained ML trees with nucleotide and protein data using mixture models in IQ-TREE (Wang, Minh, Susko, & Roger, 2018). For the protein data, we used the non-partitioned tree obtained previously as starting tree, and the following parameters: LG model (Le & Gascuel, 2008) with 20 categories (C20), Gamma rate heterogeneity calculation (+G), site-specific frequency profile inference (+F), and 1,000,000 ultrafast bootstrap replicates. For the nucleotide analyses, we used the MIX option with three components; JC (Jukes & Cantor, 1969), HKY (Hasegawa, Kishino, & Yano, 1985), and GTR (Tavaré, 1986), four Gamma categories (+G4), and 1,000,000 ultrafast bootstrap replicates.

Bayesian Inference Trees: Mixture models

Only the dataset 2, used to infer the phylogeny of all *Dugesia* species included in the study, was analysed with BI methods (Table 2); using 20 categories and the LG model for protein sequences, as well as the CAT GTR for nucleotide data. Two chains were launched with protein data, running until 10368 and 20807 iterations respectively and applying a burnin of 20%. The effective sampling size (ESS) was over 1000 and the discrepancy values were below 0.1 for all parameters. Additionally, the discrepancy observed across all bipartitions was equal to zero. With nucleotide data we launched two chains that ran until 18926 and 18771 iterations. After applying a burnin of 10% taking into account the visualisation of tracer files in the Tracer program (Rambaut & Drummond, 2007), some values were slightly low for a few parameters (ESS between 100 and 300, discrepancy between 0.1 and 0.2) (Table S6). Obtaining the optimal values of ESS and discrepancy is very difficult when large datasets are analysed. If we take into account that the largest and mean discrepancy observed across all

bipartitions are zero (maxdiff = 0, meandiff = 0) we can consider this run pretty acceptable (Lanfear, Hua, & Warren, 2016; Schrempf, Lartillot, & Szöllősi, 2020).

Whereas IQ-TREE has an excellent implementation of mixture models and the ultrafast bootstrap approximation allows to run the analysis relatively easily without excessive consumption of time or computer resources, PhyloBayes is much more needy in computational requirements and time and the analyses may take much longer. For this reason and taking into account the congruent results obtained with both methods after the analysis of dataset 2, we decided to use only Maximum Likelihood approximation in the next analyses.

Reconciling gene trees with species tree: Multispecies Coalescent Model

The species tree was estimated from individual trees using the MSC implemented in ASTRAL-pro, analysing the datasets 2, 5 and 6. Individual trees were obtained with IQ-TREE for both protein and nucleotide single copy orthologs (we refer to these trees as gene trees onward) following the same methodology described before for ML trees using mixture models. Those gene trees were used as input in ASTRAL-pro with default parameters.

2.4.7. Individual gene trees visualisation

To visualize the gene tree discrepancy the individual gene trees were visualised in DensiTree (Bouckaert & Heled, 2014). For that, every tree was independently rooted and ordered using newick-utils (Junier & Zdobnov, 2010) and forced ultrametric and dichotomous using phytools package v.0.7.9 (Revell, 2012) in R (R Team, 2021).

2.4.8. Ancestral character reconstruction (ACR)

Considering the diversity of reproductive strategies of the species included in the study, we inferred the probability of ancestral character states for the reproduction mode on the internal nodes of the ML tree obtained from the dataset 3. In this case 717 SC and only 12 terminals were used, every one representing one species and selecting only a sample of *D. malickyi* as outgroup (Table 2). We set three states for the reproduction mode: Sexual (species with strictly sexual populations), Sexual+Asexual (species with sexual and asexual populations), and Sexual (species with strictly sexual populations). The current states were assigned to the terminals

taking into account the reproductive strategies present in the whole species. Thereby, strictly asexual species have not been included in the analysis, since species with only asexual populations described are very rare and not present in the Mediterranean region. We estimated the ancestral states using the phytools package v.0.7.9 in R. The posterior probability for each state at nodes was determined from stochastic character-state mapping analysis, using the fitpolyMk function with the transient option as model, integrated in the make.simmap function and 10000 simulations on Markov Chain Monte Carlo (MCMC).

3. Results

3.1. Sampling

We were able to obtain specimens from most of the species known from the Western Mediterranean clade (Fig. 1, Table1, Table S1). *D. brigantii* and *D. leporii* have not been found since their initial description. In the case of *D. tubqalis*, from Morocco, its type locality was in a very bad condition and no animals were found, inspection of close localities did not reveal new locations for this species. In addition, due to weather conditions was not possible to sample in Afaska, Morocco, the locality of the candidate new species *Dugesia* sp. (1) reported in Leria et al. (2020; *in press*). However, in a close locality, Beni H´amed, *Dugesia* specimens were found. The DNA-Barcode analysis revealed that the new specimens show a great genetic distance with *Dugesia* sp (1) *sensu* Leria et al (2020; *in press*), and probably represent a sister species, that here we name as *Dugesia* sp. nov, pending a thorough species delimitation study and the description of the new species.

The species assignment of the rest of specimens based on locality was corroborated using the barcoding analysis (Table S1). Specimens from localities Font de l'Us, Font de la Trilla and Berga (voucher ID MR1263, MR1265, MR1361 and MR1360) were left as *Dugesia* sp. since their adscription to either *D. liguriensis* or *D. etrusca* or any other alternative was not conclusive on view of their groupings in the COI based tree, so it was left pending on the analyses of transcriptomes.

3.2. Ortholog searches and datasets

We obtained transcriptome data from 83 specimens (Tables S1 and S3) representatives of 13 species (including the two outgroups from Greece) and

undescribed new candidate species. After the transcript filtering step, 82 samples were retained, and three sample groups were analysed with OrthoFinder. The results of the three orthologs' searches performed are shown in Table S4. A total of 717, 4175 and 1984 Single Copy orthologs (SC) were obtained with all samples (search 1), subtentaculata samples group (search 2), and etrusca-liguriensis samples group (search 3) respectively.

Six datasets were built from the three ortholog searches (Table 2). Dataset 1 included all specimens retained after filtering (82) and 717 SC. Dataset 2 is a reduced version of dataset 1 including 29 specimens to have 2 representatives for each species and perform the phylogenetic analyses for the whole Western Mediterranean clade. Dataset 3 is a reduced version of dataset 2 to include only one representative per species to run the Ancestral Reconstruction of Characters. Dataset 4 is made of the concatenation of SC obtained in the ortholog search 2 to perform the phylogenetic analyses of subtentaculata samples group. Datasets 5 and 6 are made from ortholog search 3 to analyse the relationships of the etrusca-liguriensis samples group and the adscription of the Iberian Peninsula asexual populations, in dataset 6 the specimens from Berga were removed.

3.3. Phylogeny of *Dugesia* in the Western Mediterranean

The phylogenetic trees obtained from dataset 1 showed three main clades, but with incongruence between protein and nucleotide data regarding the topology and the support values for some clades (Fig. S1). Based on protein data, the clade that groups *D. hepta* and *D. benazzii* is the first to diverge, while in the nucleotide-based tree the group including *D. gonocephala, D. ilvana, D. etrusca,* and *D. liguriensis* splits first. However, in both trees the composition of the three major clades does not vary. *D. subtentaculata* is monophyletic, but the internal branches are very short to deduce a supported internal topology. The clade that groups *D. etrusca*, *D. liguriensis*, *D. ilvana, D. gonocephala,* and *Dugesia sp.* from Iberian Peninsula is also monophyletic. However, the grouping of samples from the Iberian Peninsula are atypical. In addition, it is remarkable the position of the samples from Berga, that do not group together in either of the two trees. Of the five analysed samples from this locality; two samples group with *D. etrusca* and three with *D. liguriensis.*

To eliminate the possible noise introduced by the intraspecific diversity and relationships, two representative samples by species were selected to build the dataset 2, using the same 717 SC obtained before. In the case of *D. etrusca* and *D.liguriensis*, a sample from Font de la Trilla was selected as representative of the Iberian localities.

ML, BI and MSC analyses of dataset 2 yielded the same topology for nucleotide and protein data. However, regarding the information content, some support values were lower when using protein data in ML and MSC methods, but not for BI, for which the support values were maximum (pp = 1) for nucleotide and protein data (Fig. 3, Fig. S2, Appendix A).

To visualize the gene tree discordances MSC were visualized in DensiTree (Appendix A). The obtained pattern showed protein data is less informative than the nucleotide data, showing a more blurred pattern. Interestingly, 31 protein trees failed in the rooting process, while only one failed with nucleotides, since those trees were not resolved, reflecting the lower informativeness of proteins for our species group. In addition, the values of the final normalized quartet score of MSC analysis; 0.81 and 0.89 for proteins and nucleotides respectively (Appendix A), indicate that around 81 and 89 percent of quartet trees in the input gene trees agree with the output trees obtained with these datasets. For this reason, from here, we will talk on base to the results obtained from analyses done with nucleotide data.

The results from the three phylogenetic inference methods based on dataset 2 have been summarized on the ML tree (Fig. 3). Three main clades have been differentiated. The first divergent clade including *D. gonocephala*, *D. ilvana*, *D. etrusca*, D. *liguriensis,* and Iberian populations from Catalonia region is mainly continental, except for *D. ilvana*, which is endemic from Elba island in the Tuscan archipelago. Taking into account the unknown assignment of Iberian populations to *D. etrusca* or *D. liguriensis*, we decided to denote these branches as *D. etrusca sensu lato* (*s.l*) and *D. liguriensis s.l* and use *sensu stricto* (*s.s*) when sexual population are referred. We have named this branch the Iberia-Apennines-Alps-plus clade since it includes populations from these geographic regions plus *D. ilvana* from Elba island and *D. gonocephala,* distributed in almost all continental Europe. Next, two sister clades are defined, one endemic from the Corsica and Sardinia islands, Corsica-Sardinia clade, groups *D. hepta* and *D. benazzi* species. The second clade, Iberia-Africa clade, includes a complex species group formed by: a) *D. aurea* and *D. corbata*, endemic from the

Balearic Islands, b) *D. subtentaculata* broadly distributed in all the Iberian Peninsula and the North of Africa, c) *D. vilafarrei* restricted to one locality in the South of the Iberian Peninsula, and c) a new lineage from the Rift, in Africa (Fig. 3).

All methods showed high support values for all nodes, except for the node grouping *D. subtentaculata* and *D. vilafarrei*, where the branch support of MSC is relatively low (0.88), but the final normalised quartet score (a measure of support for the entire topology) is high (Fig. 3, Appendix A).

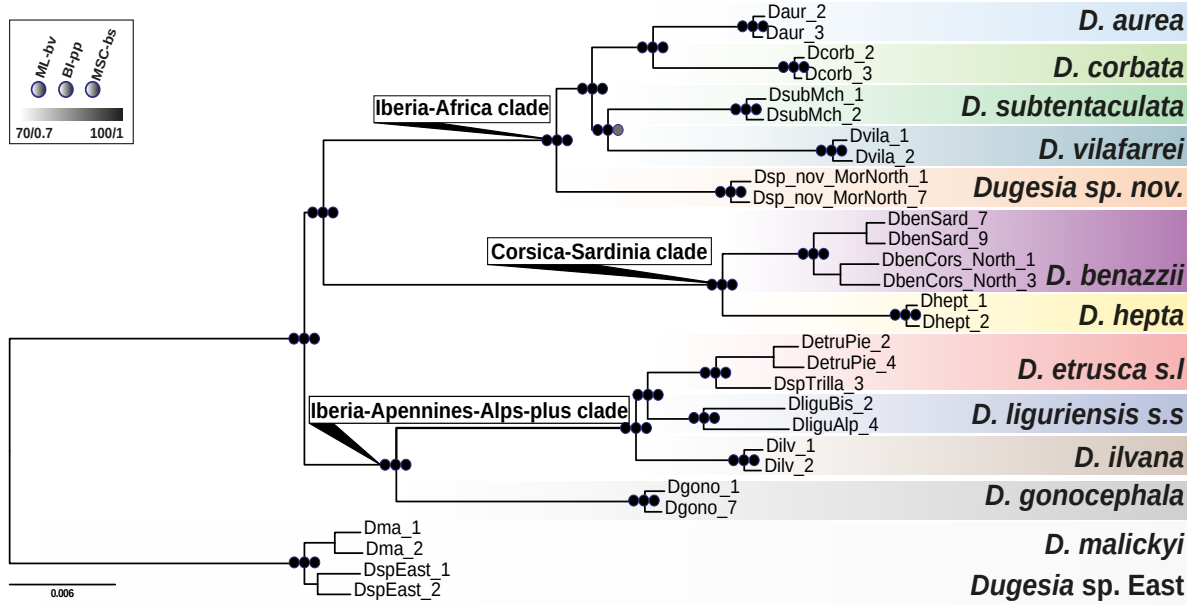


Figure 3. Phylogenetic tree of *Dugesia* from the Western Mediterranean obtained with transcriptomic data. The tree summarizes the results obtained with Maximum Likelihood (ML, IQ-TREE), Bayesian Inference (BI, PhyloBayes), and Multispecies Coalescence Model (MSC, ASTRAL-pro) analysing the nucleotide information of dataset 2 (13 species, 29 samples, and 717 single copy orthologs). All approaches yield the same topology, shown here with the ML tree. At the nodes are shown thw bootstrap value (ML-bv), the posterior probability (BI-pp), and the branch support (MSC-bs) in a color scale from minimum (white) to maximum (black). Scale bar: substitutions per site, *s.s: sensu stricto, s.l: sensu lato*

3.4. Phylogeny of *D. subtentaculata* in the Iberian Peninsula

The ML tree obtained from the analysis of dataset 4 (Table 2) shows a clear structure in this species (Fig. 4). The population of Cangas, in the North of the Iberian Peninsula is the first to diverge. Although this population is represented by sexual and asexual individuals, no differentiation by reproductive strategy was shown. The next group to diverge includes samples from North of Catalonia (Santa Fe) and South of France (Montpellier) to constitute the Northeast clade, followed by a Southwestern

clade, including a population from South of Portugal (Monchique), that is the sister group of the Southern clade formed by samples coming from the South of Iberian Peninsula (El Bosque) and two differentiated populations from Africa (Magoo in the North and Imlil in the South of Morocco) (Fig. 4).

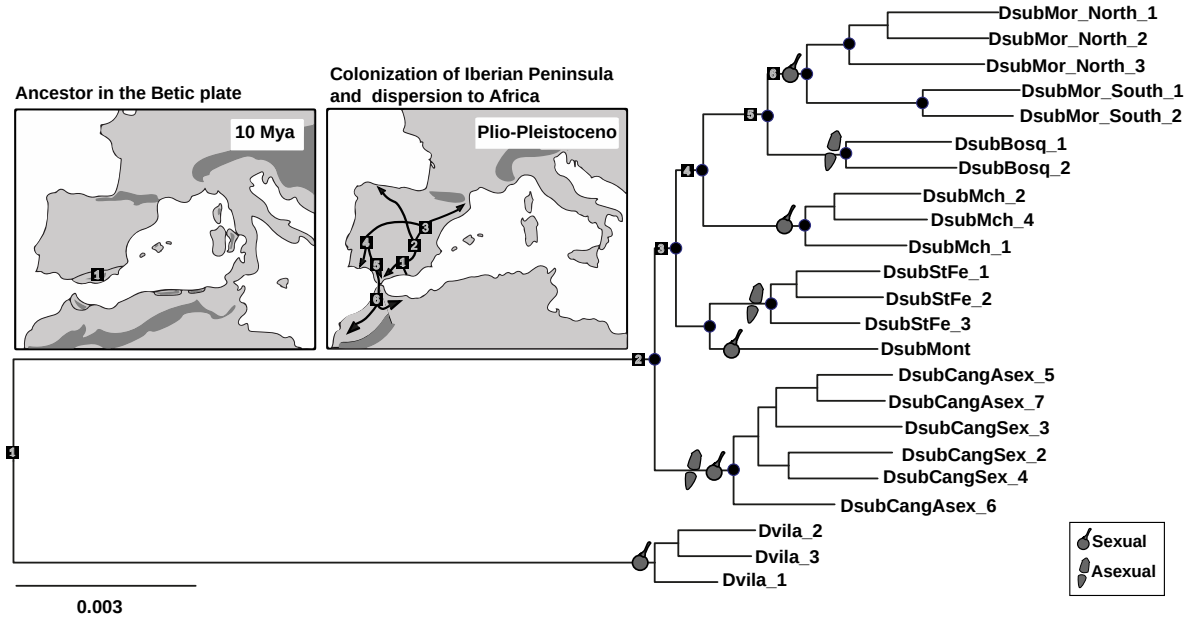


Figure 4. ML tree of *D. subtentaculata* obtained with IQT-REE from Dataset 4 (subtentaculata samples group: 2 species, 23 samples, 4175 SC). The black circles on the nodes indicate bs = 100 and the icons on the branches the reproductive strategy of populations. Scale bar: substitutions per site. The scheme represents the hypotheses proposed to explain the colonization process of *D. subtentaculata* in the Iberian Peninsula based on this data.

3.5. Species assignment to *D. etrusca* and *D. liguriensis*

The ML tree obtained from dataset 5 (Table 2) shows an unexpected topology (Fig. 5A). *D. etrusca s.s* is monophyletic (red in Fig. 5). However, *D. liguriensis s.s* is divided in two clades and results in a paraphyletic group, since samples from Sasello and Bisagno group with *D. etrusca s.l*. The unclassified samples from Font de l'Us, Font de La Trilla, and Berga occupy an intermediate position within this group, showing a very strange branching pattern, where nearly several individuals constitute one lineage resulting in a ladder-like pattern. In addition, the bootstrap supports are low for many nodes (bs < 97, grey circles in Fig. 5A).

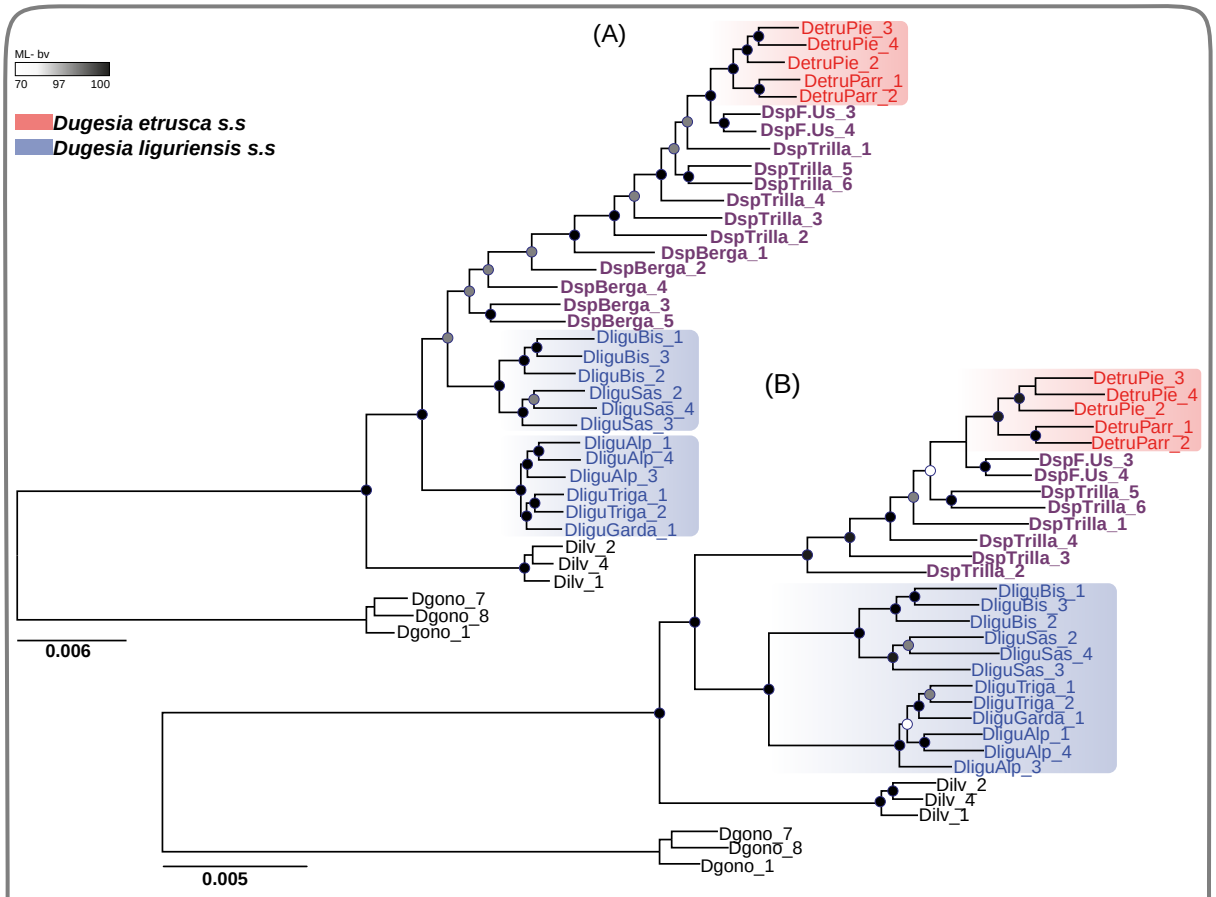


Figure 5. ML trees of Iberian-Apennines-Alps-plus clade obtained with IQ-TREE using nucleotide data from (A): Dataset 5 (etrusca-liguriensis samples group: 4 species, 36 samples, 1984 SC) and (B): Dataset 6 (Dataset 5 without Berga representatives: 4 species, 23 samples, 1984 SC). The bootstrap values (ML-bv) are shown on the nodes in a color scale from minimum (white) to maximum (black). Scale bar: substitutions per site, *s.s: sensu stricto*

The species trees obtained with Astral-pro using the MSC method show a different topology (Fig. 6A). In this case, *D. liguriensis s.s* and *D. etrusca s.s* form both monophyletic groups highly supported. The samples from Font de l'Us group with high support with *D. etrusca s.s.* Although some internal nodes have low support, the samples from Font the la Trilla and two samples from Berga also group with high support in this clade, showing the ladder-like pattern. The rest of samples from the Berga group with *D. liguriensis s.s,* also with a high support.

Taking into account that the samples from Font de l'ÚS and Font de La Trilla grouped with *D. etrusca s.s* in all analyses (Fig. S1. Fig. 5, and Fig. 6), while the samples from Berga in some cases splitted in the two species; *D. etrusca s.l* and *D. liguriensis s.s* (Fig. S1, Fig. 6A) we considered the samples from Berga were problematic and may be causing some artifactual groupings. To explore their effect in the phylogenetic inference, we extracted the representatives from Berga to obtain the

dataset 6 (Table 2). From this dataset, resolved trees were obtained, showing two monophyletic groups (Fig. 5B, Fig. 6B). Nonetheless, although the samples from Font de La Trilla grouped in *D. etrusca s.l*, they continued showing the stranger ladder-like pattern and its grouping has low support (Fig. 5B and Fig. 6B). With these results, the populations from Font de l'Us and Font de la Trilla could be classified as *D. etrusca*. The population from Berga remains uncertain since both *D. etrusca s.l* and *D. liguriensis s.l* include representatives from Berga population (Fig. S1, Fig. 6A).

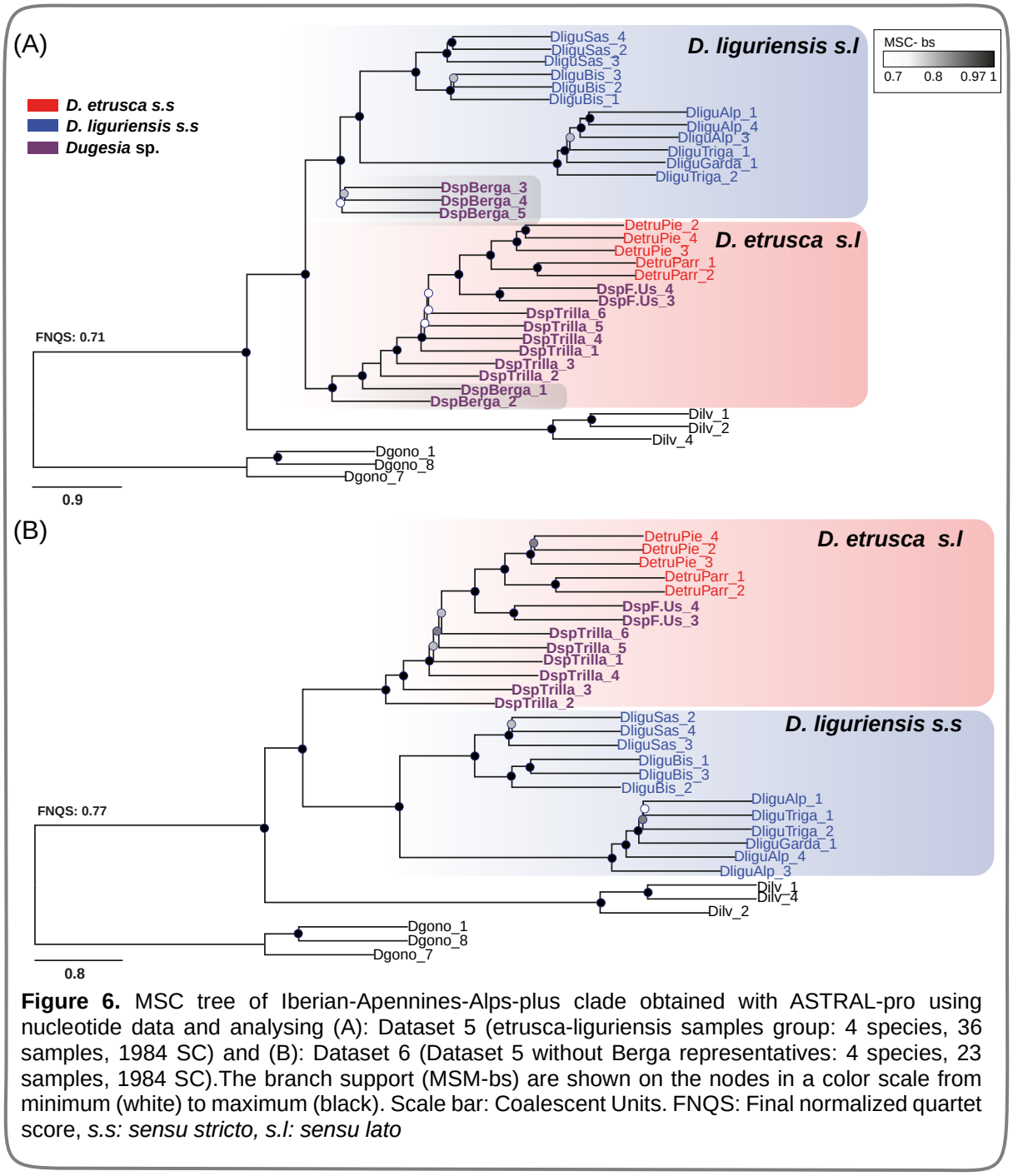


Figure 6. MSC tree of Iberian-Apennines-Alps-plus clade obtained with ASTRAL-pro using nucleotide data and analysing (A): Dataset 5 (etrusca-liguriensis samples group: 4 species, 36 samples, 1984 SC) and (B): Dataset 6 (Dataset 5 without Berga representatives: 4 species, 23 samples, 1984 SC).The branch support (MSM-bs) are shown on the nodes in a color scale from minimum (white) to maximum (black). Scale bar: Coalescent Units. FNQS: Final normalized quartet score, *s.s: sensu stricto, s.l: sensu lato*

3.6. Ancestral character reconstruction

For the assignment of reproductive strategies mode to *D. etrusca* s.l and *D. liguriensis* s.l, given the problematic groupings of the Iberian asexual populations, we took into account the topology showed by the ML trees of dataset 1 (Fig. S1) and the species tree obtained with ASTRAL-pro (Fig. 6) that groups some samples from Berga population with *D. liguriensis s.s*. In this scenario, we assumed that the two lineages have both types or reproduction, but this assumption must be taken with caution, since the evolutionary history of the Iberia-Apennines-Alps-plus clade seems to be complex and may not be fully resolved.

The ancestral character reconstructed for the reproductive mode of ancestors along the evolutionary tree of the Western Mediterranea *Dugesia* is shown in Fig. 7. The hypothesis of a strictly sexual ancestor was strongly supported in almost all nodes (pp > 0.9), except for the node joining *D. etrusca s.l* and *D. liguriensis s.l*, where the ancestor shows a high probability to be sexual and/or asexual (pp = 0.76) suggesting, in this case, the possibility of facultative ancestor for this lineage (Fig. 3, Fig. S3).

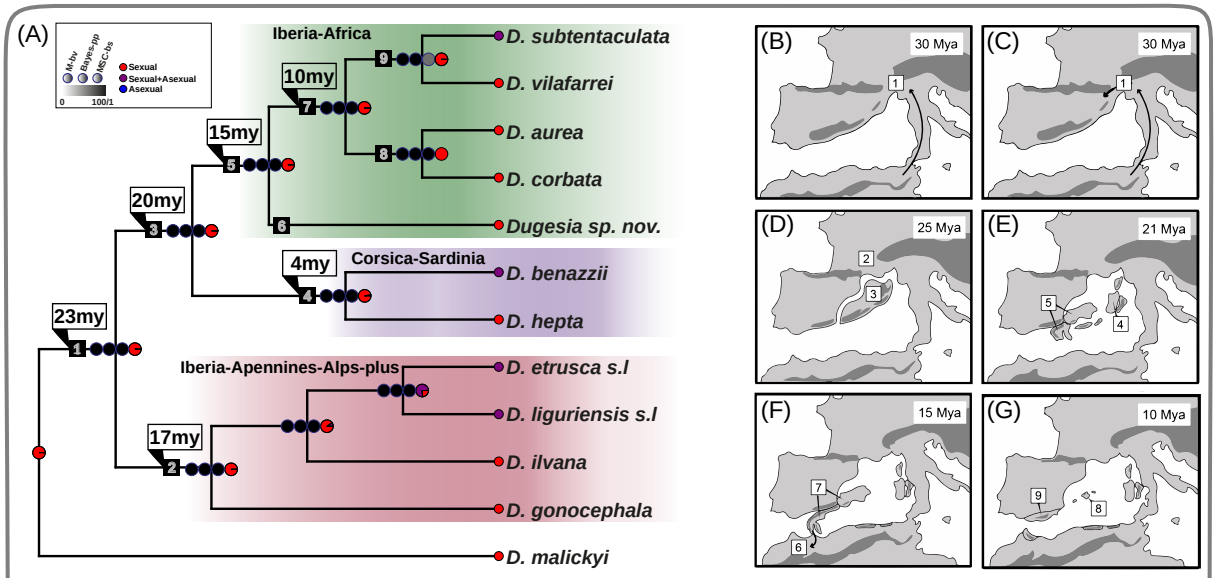


Figure 7. Scheme summarizing the evolutionary history of *Dugesia* from the Western Mediterranean. (A): a forced ultrametric tree obtained from dataset 3 (12 species, 12 samples, and 717 SC) highlighting the three main clades. The divergence times shown on the nodes have extracted from Leria et al., *(in press)* based on a calibrated phylogeny obtained using 6 molecular marker and different taxon composition. B-F: schemes showing the geological events that shaped the diversification of *Dugesia* in the Western Mediterranean modified from Leria et al., (*in press*).

4. Discussion

4.1. New methodologies applied to the phylogenetic studies in *Dugesia*

Two previous studies based on molecular data have tried to ascertain the phylogenetic relationships of *Dugesia* from the Western Mediterranean. The first (Lázaro et al., 2009) was based on only two markers while the second increased the markers to six (Leria et al., *in press*) resulting in a substantial change in topology and resolution, validating not only the importance of the number of markers but also their nature. Lázaro and collaborators used fragments of the nuclear gene Internal Transcribed Spacer (ITS-1) and the mitochondrial region Cytochrome Oxidase I (COI). Although these markers have been widely used in taxonomic and phylogenetic studies in diverse taxa (Chen, Chen, Wu, & Wang, 2015; DeSalle & Goldstein, 2019; Phillips, Hyde, Alves, & Liu, 2018; Vu et al., 2019) and particularly in planarians (For review articles see Marta Álvarez-Presas & Riutort, 2014; Baguñà et al., 1999; Riutort, Álvarez-Presas, Lázaro, Solà, & Paps, 2012), for Western Mediterranean *Dugesia* the use of 1039 positions in a concatenated alignment did not yield enough information to obtain a completely resolved phylogeny. The addition of four markers (Leria et al., *in press*) increased the compared positions to 5,439 combining ribosomal and protein coding genes from the nucleus and mitochondrial regions under different selective pressure. This broader data yielded a new topology that nonetheless still showed a few unsupported nodes, indicating that the molecular evolutionary rate of the regions sequenced were still insufficient to reflect the diversification process of the group in specific cases.

In the present work, we moved to a strategy based on phylotranscriptomics and developed the pipeline of programs and scripts needed to perform all the analyses for the first time in freshwater planarians. With this new strategy we obtained a strongly supported topology for all nodes using 717 single copy orthologs (>800,000 bp, Table S6) identified from coding data. The transcriptomic strategy has the advantage of increasing to hundreds of thousands the positions analysed, while the price and the time needed to obtain them are equal or even lower than the PCR amplification of only a few markers. On the other hand, the bioinformatic processing of the data increases the time spent in the analyses, especially the initial quality controls, cleaning, and the search of orthologs. But with the advantage that once the pipeline has been

established, it can be repeated including any new data. The development of these large datasets has moreover stimulated the improvement of the ways to implement the evolutionary models needed in the probabilistic based methods (Holder & Lewis, 2003) that may also result in the improvement observed in our case (Table S6) with respect to previous studies. From the application of the same model for the entire alignment to the use of different models for partitions by gene or codon, diverse strategies have been developed. Since the upraise of genomic datasets, the use of Mixture Models has expanded in the field of phylogenomic analyses, replacing the traditional partition schemes. 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 (Quang, Gascuel, & Lartillot, 2008) making it unnecessary to give beforehand any partition scheme. Overall, the good performance of this method when applied to phylogenetic inference has been demonstrated (Chris Venditti, Andrew Meade, 2008; Schrempf et al., 2020).

Having a genome wide representation of genes also allowed us to apply, in addition to the traditionally used inference methods with mixture models, an approach based on MSC to infer the species tree from individual gene trees. This method has been broadly used in phylogenomics demostrating its high performance against the traditional concatenation methodology (Liu, Anderson, Pearl, & Edwards, 2019), even using transcriptomic data (Edwards et al., 2016 and references there in). This methodology is specially recommended when deep coalescence processes take place in the evolutionary process of the studied species group (Mirarab, Nakhleh, & Warnow, 2021).

Regarding the use of different data types, we also have been able to analyse aminoacidic data and compare its performance against nucleic data to answer our questions. The DensiTree graph showed a most blurred pattern in protein trees indicating higher gene tree discordance in protein data (Appendix A at the bottom), which explains that the maximum support values with all methods were obtained with nucleotide sequences. These results indicated that the aminoacidic sequences do not contain enough information to resolve the phylogeny. The protein data is more conserved than nucleotide sequences and its information is more useful to study the evolution in a broad scale of time (Nei & Kumar, 2000), while the diversification of *Dugesia* in the Western Mediterranean is relatively recent (Sola et al., *in press*). For

these reasons we based our main results and discussion on the analyses of nucleotide data.

Our new strategy has helped to support the phylogeny of *Dugesia* in the Western Mediterranean region, to obtain the first resolved phylogeny of *D. subtentaculata* with a large molecular dataset, and the assignment of some Iberian asexual fissiparous populations to geographically distant clades. Nonetheless the results obtained with this new dataset opens new questions about the complex evolutionary history of *Dugesia* in the Western Mediterranean.

4.2. Evolutionary history of *Dugesia* in the Western Mediterranean

In the first and very preliminar molecular approximation to study the evolutionary history of *Dugesia* (Lázaro et al., 2009) we included a few individuals of several species belonging to the Western Mediterranean area. Those included four populations of *D. subtentaculata sensu stricto* from Iberian Peninsula, as well as the populations currently considered different species (*D. aurea* and *D. corbata*), also *D. gonocephala*, *D. etrusca, D. liguriensis, D. ilvana*, *D. hepta,* and *D. benazzii*. The results showed for the first time the differentiation in the three main clades that are clearly defined in our phylogeny but did not resolve the relationships among them nor even the relationships within some of them (for some of which the sampling was very poor).

More recently, a study analysing the drivers of the biogeographic history of Western Mediterranean *Dugesia*, extended the number of markers used to six (Leria et al., *in press*) and also the sampling to include representatives of *Dugesia* from Morocco and a new species from the Iberian Peninsula (*D. vilafarrei*). However, in that work only one or two representatives per species were included, not allowing the analysis of intraspecific genetic structure and evolutionary history. Their results showed the same three main clades of the previous study of Lázaro et al., (2009), but with a different topology. Leria et al., (*in press*) showed the same topology that we obtained here (Fig. 3), but with still low support for the node joining the Iberia-Africa clade with the Corsica-Sardinia clade (names of the clades used in Fig. 3). In the present work, the support values for all the nodes have been high giving full support to that topology (Fig. 7).

Leria et al., (*in press*) also dated their phylogeny and performed ancestral range reconstruction and niche modeling analyses that allowed them to put forward a biogeographical hypothesis. They showed that the evolution of *Dugesia* has been shaped by the paleogeological processes that formed the Western Mediterranean as well as climatic changes. The good support we give to the topology obtained in that study allows us to put our tree in the time frame they proposed. We have summarized Leria et al., (*in press*) hypothesis in Fig. 7B-G over a scheme of our phylogenetic tree (Fig. 7A). Our scheme shows the divergence times obtained by Leria et al., (*in press*) using a different dataset to infer the calibrated phylogeny in BEAST. Therefore, the branch length of our scheme does not always match with their divergence times, and we only used it to show the coincident topology found in both studies, supporting the biogeographical hypothesis proposed by Leria et al (*in press*).

The scheme shows the three main groups Iberia-Apennines-Alps-plus clade, the Iberian-African clade, and the Corsica-Sardinia clade on the temporal frame proposed by Leria et al (*in press*). Their hypothesis locates the ancestor of the Western clade arriving into Europe through the Italian Peninsula 30 Mya (Fig. 7B), matching with the results of a recent biogeographic study of *Dugesia* genus (Sola et al., *in press*). This ancestor, could start dispersing throughout the continental region, passing also to the Iberian Peninsula (Fig. 7C).

The first diversification event of *Dugesia* from the Western Mediterranean clade occurred around 23 Mya, coinciding with the breakage of Eastern Iberia and Southern France from the continent (Rosenbaum, Lister, & Duboz, 2002). This diversification event putatively isolated the ancestor of the Iberia-Apennines-Alps-plus clade, which remained in the continent, from the ancestor of the Corsica-Sardinia and Iberia-Africa clades, which remained in the landmass that would become the Corsica and Sardinia islands, the Balearic Islands, the Betic region and part of the North of Africa (Fig. 7D) (Leria et al. *in press*)

4.3. The Corsica-Sardinia clade

We will not get into much detail within this clade since it is an objective for an ongoing investigation dedicated to the group. However, we can not avoid noticing some important facts for this group in our trees. In the first place, as stated above, it is the first time that the sister relationship of this clade with the Iberia-Africa clade receives

maximum support in a phylogenetic analysis. In the archipelago four species have been described, *D. benazzii, D. hepta, D. brigantii* and *D. leporii* however the latter two have not been found since their initial description, so unluckily have never been included in a molecular study. For *D.benazzii* and *D. hepta*, a recent study based on two molecular markers and including multiple populations of both species (Dols-Serrate, Leria, Aguilar, Stocchino, & Riutort, 2020) along its distribution has shown a very complex evolutionary history. The evolution of the group will be characterized by a putative process of hybridization between both species in the Sardinia island, the combination of different reproductive strategies, and the complex processes of chromosome rearrangements. Moreover, in that work it was put forward the hypothesis that *D. benazzi* could be separated into two species; one from Sardinia and another one from Corsica (Dols-Serrate et al., 2020). Our work shows a clear differentiation within *D. benazzii*, with representatives from the islands of Sardinia and Corsica forming two monophyletic groups well differentiated and with high support. Moreover, we show a clear differentiation between the populations of *D. benazzii* from the South and the North of Corsica indicating a geographic diversification of this clade in these archipelagos. Our data hence renders further support to *D. benazzii* being in fact two species, one distributed in Corsica and the other one in Sardinia.

4.4. The *D. subtentaculata* journey; from the Betic-Riff plate to Iberia and back to Africa

Basing on the biogeographic hypothesis of Leria et al., (*in press*), the ancestor of the Iberian Peninsula and Balearic Island representatives of the Iberia-Africa clade (Fig. 7, node 7) originated in the Betic-Riff plate after it broke separating this clade from their sister group in the North of Africa (Fig. 7G), here represented by a non described species. Later, after the Balearic Islands split, the ancestor of *D. vilaferrei* and *D. subtentaculata* appeared in the Iberian Peninsula. Here we corroborate this biogeographic history and moreover obtain for the first time a resolved phylogeny of *D. subtentaculata* populations (Fig. 4).

Even though not all reported populations have been included, our phylogeny is enough to propose a first hypothesis on the colonization process of Iberian Peninsula by *D. subtentaculata*. Basing on our tree topology (Fig. 4), we hypothesized that the ancestor of this species moved from the ancient Betic plate to the Northeast of the

Iberian Peninsula. We show a first split that separates the Cantabric population from all the rest of populations included in the present work. Followed by the separation of the Northeastern populations; Santa Fe (Catalonia) and Montpellier (Southeastern France). This last may be either the product of one anthropogenic introduction, since by that time, the Pyrenees were fully formed (Dèzes, Schmid, & Ziegler, 2004, 2005) or could represent a natural dispersion through the well-recognised corridor between the Pyrenees and the Mediterranean Sea (Martinez Rica & Montserrat Recoder, 1990). The diversification of the species continued with the split of the Southwestern population of Monchique in Portugal and, the split of the population from El Bosque in Andalusia. A similar biogeographic pattern showing an earlier diversification of the lineages from the Eastern basins of the Iberian Peninsula than the lineages from the Western basins has been pointed out for the native freshwater fishes of this region (Filipe, Araújo, Doadrio, Angermeier, & Collares-Pereira, 2009), which may be related to the hydrographical evolution of the Iberian Peninsula (De Vicente, Cloetingh, Van Wees, & Cunha, 2011). Finally, the last split within the phylogeny of *D. subtentaculata* concerned the populations from Morocco, which probably diversified from the Iberian populations due to its dispersion to North Africa. Leria et al., (*in press*) dated the divergence between one Iberian (Peralejos in the Iberian Peninsula Meseta) and one African populations around 1.6 Mya, indicating that the pass putatively occurred during the Pleistocene, when the sea level in the Gibraltar strait was lower.

The lack of phylogenetic resolution reported in Leria et al., (2020) for this species was a consequence of the mosaicism in most individuals due to their reproductive strategy (fissiparity combined with occasional events of sexual reproduction) affecting both genes analysed, COI and Dunuc12 (Leria et al., 2019). Thereby, analysis including asexual populations and the few markers traditionally used, fails to infer the phylogenetic history of this species due to the intraindividual diversity and the Mosaic-Meselson effect (Leria et al., 2019). In the present study the use of thousands of exonic regions from single copy orthologs has resulted in a resolved and supported phylogeny within *D. subtentaculata*. The advantage of our methodology resides, in the first place, in the use of coding regions under different selective pressure that possibly restrict the emergence of mosaicism. Thus, many conserved sites probably presenting fixed substitutions among populations can counter the effect of intraindividual diversity in a few sites and result in a

phylogenetically informative set of data. In this aspect it is important to notice that most of the intraindividual variability found in the Dunuc12 marker by Leria et al., (2019) was situated in the intronic region. In addition, the methodology used in the present study including new strategies for the application of evolutionary models, possibly allowed to better retrieve most of this information.

In view of the success of this new approach, more analysis, including the great number of populations reported by (Leria et al., 2020) can be foreseen as a good strategy to build a stronger case on the diversification of *D. subtentaculata* in the Iberian Peninsula and North of Africa.

In relation to the diversity of *D. subtentaculata* in the Northwest of Africa, we report two differentiated populations; one from the North and the other from the South of the Atlas landscape in Morocco. The presence of this species in Africa has been previously refereed (Harrath et al., 2012; Giacinta Angela Stocchino, Sluys, & Manconi, 2012), but no phylogenetic analyses had included representatives of this geographical area before. Moreover, it is remarkable that the new species from North of Morocco analysed here (Dsp_nov_Mor North in Fig. 3) is different of the candidate species included in previous analyses (Leria et al., 2020, *in press*), indicating that at least three species belonging to the Western Mediterranean clade are present in the Atlas area (*D. tubqalis*, and the two new candidates species), apart from the representatives of *D. subtentaculata* mentioned above. This suggests there is a high species's richness hidden in the Atlas and the Riff region.

4.5. The Iberia-Apenines-Alps-plus clade: geographically broader than thought

D. etrusca s.s and *D. liguriensis s.s* were initially described from individuals coming from Tuscany (Benazzi, 1946) and Liguria (De Vries, 1988) respectively. In a first attempt to infer a molecular phylogeny for Tricladida (Baguñà et al., 1999) basing on ITS-1 sequences a single individual of *Dugesia*, coming from Northeastern Spain surprisingly grouped with representatives of *D. etrusca* (*D. liguriensis* was not included in that study). Later, in the first molecular analysis of western Mediterranean species (Lázaro et al., 2009) a single individual coming from Sardinia and the previously cited locality from Northeastern Spain were ascribed to *D. liguriensis*. Finally, the presence of putative *D. etrusca* and *D. liguriensis* from Aragon and Catalonia in the Northeastern section of the Iberian Peninsula have been detected by DNA-Barcoding in posterior

samplings of Iberian regions (Riutort *pers. com.*). These findings raised the question to which of both species did the Iberian Peninsula populations belong and how biogeographically we could explain their existence.

In the present work, we include for the first-time multiple representatives of a number of the Iberian Peninsula localities harboring specimens belonging to this clade as well as *D. ilvana,* endemic from Elba island, not analysed in Leria et al. (*in press*). With this broad representation of the species and geographic diversity within the group we expected to be able to set a supported hypothesis on the relationships, species assignations and biogeographical history for this group. Nonetheless, the special characteristics of the Iberian representatives has hampered a part of our aims, opening new and interesting questions on the evolution of this group of *Dugesia* species and of asexual lineages in general, as we develop in the following.

Our phylogenetic tree shows *D. ilvana* as the first split within the clade. Elba island is formed by the westernmost outcrops of the Northern Apennines and is the link with the Alpine region of Corsica (Fig. 7G). Its geological history is complex, and the tectonic activity persisted from the Middle Miocene until the late Miocene-Pliocene (Bortolotti, Fazzuoli, et al., 2001; Bortolotti, Pandeli, & Principi, 2001) setting a geological maximum for the divergence of approximately between 8 and 5 Mya. However, Leria et al (*in press*) set the split of *D. etrusca s.s* + *D. liguriensis s.s* clade around 9 Mya. Therefore, the geological history of Elba Island might seem slightly too young to have preceded the *D. liguriensis-D. etrusca* split. Since the evolutionary history of Elba Island fauna has been linked to connections with Corsica regions and Tuscany coast (Fattorini, 2009 and references therein; Dapporto & Cini, 2007; Di Nicola & Vaccaro, 2020) it is difficult to drag a biogeographic scenery for the split of *D. ilvana* from the ancestor of *D. etrusca s.l* and *D. liguriensis s.l*. Therefore, the phylogenetic position of *D. ilvana* shown here, must be reviewed under more detailed analyses, taking into account the complex diversification process of Iberian-Apennines-Alps-plus clade.

According to the topology (Figs. 3 and 7) the clade including *D. etrusca s.l* and *D. liguriensis s.l* would have diversified occupying the area from the Apennines region to the Northeastern in the Iberian Peninsula. It has been hypothesized that the Alps prevented the expansion of the Apennines-Alps lineage to Central Europe, while the Pyrenees possibly limited its entry to the Iberian Peninsula (Leria et al., *in press*).

However, we found populations belonging to this clade in the Catalonian region in the Northeastern on the Peninsula. We propose that this pass was possible because the Eastern Pyrenees orogenesis encompassed two stages; a first stage from Early Cretaceous to middle Lutetian time (99 to 47 Mya) and the second one from middle Lutetian to late Oligocene (47 to 23 Mya). The first stage was characterized by a low topography, because the increase in relief was partially compensated by downward flexion of the Iberian plate, as well as high levels of mountain erosion. While the last stage, with more orogenic activity, was not concluded until approximately 23 Mya (Vergés et al., 1995). These complex geological history lead to the occurrence of several fauna corridors along the Pyrenees landscape (Ninot, Carrillo, & Ferré, 2017). Taking into account these orogenic process, their geological dates, and the broad uncertainty around divergence times estimated by Leria et al (*in press*), the Eastern Pyrenees could have acted as the corridor for the ancestors of the populations found in the North of Iberian Peninsula (Fig. 7C). Why these species did not expand further in the Iberian Peninsula and remained restricted to the localities described here, is a question we can not resolve with the present data.

In summary, we detect some incongruencies and a certain difficulty to explain the evolutionary history of the group and its biogeography. When we analysed with more detail the relationships within and among the *D. etrusca s.l* and the *D. liguriensi s.l* clades (Fig. 5, 6, and S1) to try to have a better understanding of the relationships among the Iberian populations and the rest, and to assign them to one or the other species, the situation became much more complex.

4.6. The Iberia-Apenines-Alps-plus clade: a complex evolutionary history driven by asexuality?

In the species identification analysis with COI (Materials and Methods section 2.3) individuals from Font de la Trilla and Berga already showed an anomalous ladder-like pattern as the one we later found in the transcriptome-based trees (Figs. 5, 6 and S1). The populations did not constitute monophyletic groups as happened for all the rest of analysed localities. For this reason, a broader number of samples from these populations were included in the transcriptomic analyses. Despite the strange branching pattern, all our trees show that the Iberian fissiparous populations belong to Iberia-Apennines-Alps-plus clade. However, regarding their species assignment to *D.*

etrusca or *D. liguriensis* species, our results are more than disquieting, exciting. Based on our phylogenetic trees, populations from Font de l'Us and Font de la Trilla can be assigned to *D. etrusca*. But, from these trees three questions arise: i) the species assignment of individuals from Berga, ii) what is causing the strange ladder pattern of individuals from Font de la Trilla and Berga, and iii) whether this anomalous branching pattern could affect the phylogenetic inference.

Two putative explanations for these issues could be related with a case of hybridization between both species or alternatively to the asexuality of the Iberian populations. We discard the hybrid origin hypothesis for the Iberian specimens since no hybrid populations have been found in the Apennine-Alps region, where the two sexual lineages inhabit. In fact, there is no knowledge of both species duelling in the same water course or even the same area anywhere. There is neither knowledge of sexual populations of the species in the Iberian Peninsula, a region that has been thoroughly sampled by Leria et al (2020). In case of hybridization between the two species, this would have had to take place very long ago, so that the hybrid lineage could pass through the Pyrenees, with the subsequent extinction of all its populations in France and Italy, leaving only the Iberian hybrid populations.

In the alternative scenario, that we see more plausible, the ancestor of both clades will have crossed to the Iberian Peninsula, this ancestor will have been asexual or the lineage will have become asexual shortly after the crossing. The asexual reproduction by fission is very common in *Dugesia* genus (Baguñà et al., 1999; Kobayashi, Maezawa, Nakagawa, & Hoshi, 2012; Lázaro et al., 2009; Nishimura et al., 2015; G. A. Stocchino & Manconi, 2013). It has been proposed that in other species of planarians as *D. sicula* and some *Girardia* species the asexuality have been an advantage for the colonization of a broad territory (Lázaro & Riutort, 2013; Chapter III). In addition, in the case of *D. subtentaculata* the alternation of sexual and asexual reproduction could be an adaptive strategy that guarantees the evolutionary success of this lineage in the Iberian Peninsula (Leria et al., 2019).

The ACR analysis in the present study results in a high probability for the existence of ancestral populations with both, sexual and asexual reproduction for the node joining *D. etrusca s.l* and *D.liguriensis s.l* (Fig. 7, Fig. S3), which will support the idea of an ancestral fissiparous lineage arriving to the Iberian Peninsula. These results, nonetheless, are based on the assumption that both, *D. etrusca s.l* and *D.liguriensis*

s.l clades present both types of reproductive strategies taking into account the grouping of samples from Berga in both clades (Figs. 1 and 6A).

The asexuality of these ancestral populations could be a factor explaining the ladder topology showed by their descendant specimens in our trees (Figs. 1, 5, and 6). The pattern observed looks alike the described for nuclear alleles in asexual populations under Meselson Effect, when alleles show high divergence by accumulation of mutations independently of each other (Schwander, Henry, & Crespi, 2011). Under long term fissiparous reproduction different clonal lines are stablished in the population, and present day individuals can belong to independent lineages of fissiparity inside the populations, with their alleles being related far away in time. However, *D. subtentaculata* also presenting fissiparous reproduction, show the typical monophyletic pattern expected for populations geographically separated (Fig. 4). This difference could be explained by a more recent diversification of *D. subtentacula* (Leria et al., *in press*) and the alternation of fissiparity with sexual periods described in this species. Thus, we suspect that the Iberian populations of the Iberia-Apennines-Alps-plus clade show haplotypes derived independently from the ancestral haplotypes of *D. etrusca s.l* and *D.liguriensis s.l* species. This ancestral information would have been maintained by asexuality since no recombination events take place. The grouping of Berga individuals at the base of both *D. etrusca* and *D.liguriensis* lineages, could indicate the presence of haplotypes from the two species ancestors in the Berga population.

This takes us to the third question, on the potential effect that ancestral lineages may have on the tree inference method, which is again difficult to respond to. Further researches are necessary to demonstrate that the ladder-like branching pattern is real and not an artifact. However, in any case, this “artifact” could be showing an underlying evolutionary process different to the forces that lead the phylogenetic history of the other main clades, possibly related to long term asexuality. Analyses focused on this particular clade are necessary to elucidate the evolutionary process that underlies their diversification, and to demonstrate or reject the hypothesis that their fissiparity may be in the base of their strange topology in the phylogenetic trees. Maybe, a new theoretical framework deserves to be established to explain the effect of ancient asexual populations in the phylogenetic inference.

Final remarks

The study of the processes that shaped the biodiversity is key to understanding the evolution of life on our planet. This knowledge is useful not only to understand basic evolutionary processes, but to apply this information to conserve the diversity and their functions in the ecosystems. The Mediterranean region constitutes a hotspot of diversity, and its paleogeographical history is one of the most complex in the world. Large mountain systems, unstable climate periods, volcanic activity and plate fragmentation events act as modellers of the biodiversity in this region. Here, we contribute to the study of one important component of the Mediterranean freshwater ecosystem, the free-living planarians. Focused in the Western Mediterranean region, we help to elucidate the evolutionary history of *Dugesia* genus using transcriptomic data. Our work represents a step forward in the phylogenetic studies in this group, passing from the analysis of few markers to hundreds of them, supporting previous information, and contributing with new valuable data to the knowledge on these species. We corroborated a biogeographic hypothesis that explains the diversification of *Dugesia* in the Western Mediterranean; affected by the tectonic dynamics of the region during the Cenozoic. In addition, we bring to light new questions about the evolution of the asexual populations of different species, and how this disparity in reproductive modes can affect the phylogenetic inference. It is necessary to use species as *D. subtentaculata, D. benazzii, D. etrusca s.l, D. liguriensis s.l,* and *D. sicula*, which integrate the asexuality in different evolutionary scenarios to understand the effect of asexuality in the natural process of resistance, resilience, and diversification of life.

Data accessibility

All reads generated for this study are deposited in the NCBI-SRA repository under the BioProject accession code: PRJNA797284. All used scripts and detailed bioinformatic methodology can be found in https://github.com/lisy87/dugesia-transcriptome

Credit authorship contribution statement

MR did the initial study design. LBA, LL, EM, NB, YeO, MeA, MY-K, HAO, and

MR contributed to sampling. LBA performed RNA extractions and processed and analysed the transcriptomic data with the input from RF. LBA, LL, and MR wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Acknowledgements

Thanks to Julio Rozas for sharing with us the computational cluster of the Evolutionary Genomics & Bioinformatics research group. Thanks also to Silvia Hinojosa for guiding the first bioinformatics steps and Juan Manuel Lentijo Mondejar for his advice with the python scripts. In addition, we thanks to Ignacio Tenaguillo Arriola for the help launching some analysis to the cluster. Also thanks to Eduard Solà for the sampling campaign in Greece. Our thanks also to the Laboratoire Ecologie, Biodiversité et Environnement, Université Abdelmalek Essaâdi and the Water, Biodiversity and Climate Change Laboratory, Cadi Ayyad University by the support with the sampling and transportation permissions in Morocco. Sampling Permission in Morocco: Decision Nº 19/2019 HCEFLCD/DLCDPN/DPRN/CFF

Funding

This research was supported by the Ministerio de Economía y Competitividad of Spain (project PGC2018-093924-B-100). This research was also supported by a FI Grant from the Generalitat de Catalunya (to Lisandra Benítez-Álvarez).

Supplementary Tables

Table S1. Detailed information for every sample, including the voucher and the accession codes in the NCBI-SRA repository

Sample Code	Species	Reproductive Strategy	Region, Country	Locality	Geographic Coordinates (DMS)		Voucher ID	Biosample acession code in NCBI-SRA repository BioProject PRJNA797284.	SRA acession code in NCBI-SRA repository
					Latitude	Longitude			
Daur_1	*D. aurea*	Sexual	Majorca, Spain	Soller	39º45'41"	2º42'44"	MR1079.1	SAMN24979312	SRR17642743
Daur_2	*D. aurea*	Sexual	Majorca, Spain	Soller			MR1079.2	SAMN24979313	SRR17642732
Daur_3	*D. aurea*	Sexual	Majorca, Spain	Soller			MR1079.3	SAMN24979314	SRR17642721
DbenSard_6	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo	40º34'30"	9º40'31"	MR0652.6	SAMN24979320	SRR17642753
DbenSard_7	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo			MR0652.7	SAMN24979321	SRR17642752
DbenSard_9	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo			MR0652.9	SAMN24979322	SRR17642751
DbenCors_North_1	*D. benazzii*	Sexual	Corsica, Italy	Campile	42°30'24.7"	9°22'27.8"	MR1257.1	SAMN24979315	SRR17642710
DbenCors_North_2	*D. benazzii*	Sexual	Corsica, Italy	Campile			MR1257.2	SAMN24979316	SRR17642699
DbenCors_North_3	*D. benazzii*	Sexual	Corsica, Italy	Campile			MR1257.3	SAMN24979317	SRR17642688
DbenCors_South_5	*D. benazzii*	Sexual	Corsica, Italy	Monacia-d'Aullène	41° 32' 41.7"	09° 01' 24.1"	MR1260.5	SAMN24979318	SRR17642677
DbenCors_South_6	*D. benazzii*	Sexual	Corsica, Italy	Monacia-d'Aullène			MR1260.6	SAMN24979319	SRR17642673
Dcorb_1	*D. corbata*	Sexual	Majorca, Spain	Sa Calobra	39º49'44.6"	2º48'52.8"	MR1080.1	SAMN24979323	SRR17642750
Dcorb_2	*D. corbata*	Sexual	Majorca, Spain	Sa Calobra			MR1080.2	SAMN24979324	SRR17642749
Dcorb_3	*D. corbata*	Sexual	Majorca, Spain	Sa Calobra			MR1080.3	SAMN24979325	SRR17642748
DetruParr_1	*D. etrusca*	Sexual	Toscana, Italy	Parrana	43º32'23"	10º27'37"	MR1256.1	SAMN24979328	SRR17642745
DetruParr_2	*D. etrusca*	Sexual	Toscana, Italy	Parrana			MR1256.4	SAMN24979329	SRR17642744
DetruPie_2	*D. etrusca*	Sexual	Toscana, Italy	Pieve	43º29'52"	10º37'33"	MR1355.2	SAMN24979330	SRR17642742
DetruPie_3	*D. etrusca*	Sexual	Toscana, Italy	Pieve			MR1355.3	SAMN24979331	SRR17642741
DetruPie_4	*D. etrusca*	Sexual	Toscana, Italy	Pieve			MR1355.4	SAMN24979332	SRR17642740
Dgono_1	*D. gonocephala*	Sexual	France	Montpellier	43º43'22"	3º8'38"	MR1283.1	SAMN24979333	SRR17642739
Dgono_7	*D. gonocephala*	Sexual	France	Montpellier			MR1283.7	SAMN24979334	SRR17642738

Dgono_8	*D. gonocephala*	Sexual	France	Montpellier			MR1283.8	SAMN24979335	SRR17642737
Dhept_1	*D. hepta*	Sexual	Sardinia, Italy	Logulento	40º49'16.682"	8º35'35.213"	MR0980.1	SAMN24979336	SRR17642736
Dhept_2	*D. hepta*	Sexual	Sardinia, Italy	Logulento			MR0980.2	SAMN24979337	SRR17642735
Dhept_5	*D. hepta*	Sexual	Sardinia, Italy	Logulento			MR0980.5	SAMN24979338	SRR17642734
Dilv_1	*D. ilvana*	Sexual	Italy	Elba			MR1357.1	SAMN24979339	SRR17642733
Dilv_2	*D. ilvana*	Sexual	Italy	Elba	42º 47' 9"	10º 10' 3"	MR1357.2	SAMN24979340	SRR17642731
Dilv_4	*D. ilvana*	Sexual	Italy	Elba			MR1357.4	SAMN24979341	SRR17642730
DliguBis_1	*D. liguriensis*	Sexual	Italy	Bisagno			MR1254.1	SAMN24979345	SRR17642726
DliguBis_2	*D. liguriensis*	Sexual	Italy	Bisagno	44º26'40"	9º5'6"	MR1254.2	SAMN24979346	SRR17642725
DliguBis_3	*D. liguriensis*	Sexual	Italy	Bisagno			MR1254.3	SAMN24979347	SRR17642724
DliguAlp_1	*D. liguriensis*	Sexual	France	Alps Marìtims			MR1252.1	SAMN24979342	SRR17642729
DliguAlp_3	*D. liguriensis*	Sexual	France	Alps Marìtims	43º47'9"	6º38'7"	MR1252.3	SAMN24979343	SRR17642728
DliguAlp_4	*D. liguriensis*	Sexual	France	Alps Marìtims			MR1252.4	SAMN24979344	SRR17642727
DliguGarda_1	*D. liguriensis*	Sexual	France	La Garda	43°49’25”	6°34’55”	MR1302.3	SAMN24979348	SRR17642723
DliguSas_2	*D. liguriensis*	Sexual	Italy	Sassello			MR1253.2	SAMN24979349	SRR17642722
DliguSas_3	*D. liguriensis*	Sexual	Italy	Sassello	44°29'19.2"	8º28'2.9"	MR1253.3	SAMN24979350	SRR17642720
DliguSas_4	*D. liguriensis*	Sexual	Italy	Sassello			MR1253.4	SAMN24979351	SRR17642719
DliguTriga_1	*D. liguriensis*	Sexual	France	Trigance	43°47’44”	6°26’40”	MR1301.1	SAMN24979352	SRR17642718
DliguTriga_2	*D. liguriensis*	Sexual	France	Trigance			MR1301.2	SAMN24979353	SRR17642717
DsubMont	*D. subtentaculata*	Sexual	France	Montpellier	43º43'22"	3º8'38"	MR1283.5	SAMN24979381	SRR17642686
DsubBosq_1	*D. subtentaculata*	Fissiparous	Andalusia, Spain	El Bosque	36°45'42"	-5°30'20"	MR1267.1	SAMN24979370	SRR17642698
DsubBosq_2	*D. subtentaculata*	Fissiparous	Andalusia, Spain	El Bosque			MR1267.2	SAMN24979371	SRR17642697
DsubCangAsex_5	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas			MR1297.5	SAMN24979372	SRR17642696
DsubCangAsex_6	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas			MR1297.6	SAMN24979373	SRR17642695
DsubCangAsex_7	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas	43°21'54.65"	-5°9'6.19"	MR1297.7	SAMN24979374	SRR17642694
DsubCangSex_2	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas			MR1297.2	SAMN24979375	SRR17642693
DsubCangSex_3	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas			MR1297.3	SAMN24979376	SRR17642692
DsubCangSex_4	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas			MR1297.4	SAMN24979377	SRR17642691

DsubMor_North_1	*D. subtentaculata*	Sexual	Morocco	Magoo Timriouen	35°06'42.5"	-5°11'19.9"	MR1276.1	SAMN24979382	SRR17642685
DsubMor_North_2	*D. subtentaculata*	Sexual	Morocco	Magoo Timriouen			MR1276.2	SAMN24979383	SRR17642684
DsubMor_North_3	*D. subtentaculata*	Sexual	Morocco	Beni H´amed	35°08'37.3"	-5°06'51.8"	MR1252.2	SAMN24979384	SRR17642683
DsubMor_South_1	*D. subtentaculata*	Sexual	Morocco	Imlil	31°09'35.8"	-7°55'41.1"	MR1278.1	SAMN24979385	SRR17642682
DsubMor_South_2	*D. subtentaculata*	Sexual	Morocco	Imlil			MR1279.3	SAMN24979386	SRR17642681
DsubMch_1	*D. subtentaculata*	Sexual	Portugal	Monchique			MR1358.1	SAMN24979378	SRR17642690
DsubMch_2	*D. subtentaculata*	Sexual	Portugal	Monchique	37º 17' 20"	-8º 33' 15"	MR1358.2	SAMN24979379	SRR17642689
DsubMch_4	*D. subtentaculata*	Sexual	Portugal	Monchique			MR1358.4	SAMN24979380	SRR17642687
DsubStFe_1	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe			MR1082.1	SAMN24979387	SRR17642680
DsubStFe_2	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe	41º 46' 26.35	2º 27' 42.24	MR1082.2	SAMN24979388	SRR17642679
DsubStFe_3	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe			MR1082.3	SAMN24979389	SRR17642678
Dvila_1	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque			MR1266.1	SAMN24979390	SRR17642676
Dvila_2	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque	36°45'42"	-5°30'20"	MR1266.2	SAMN24979391	SRR17642675
Dvila_3	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque			MR1266.3	SAMN24979392	SRR17642674
Dsp_nov_MorNorth_1	*Dugesia sp. nov*	Sexual	Morocco	Beni H´amed			MR1251.1	SAMN24979361	SRR17642708
Dsp_nov_MorNorth__6	*Dugesia sp. nov*	Sexual	Morocco	Beni H´amed	35°08'37.3"	-5°06'51.8"	MR1251.6	SAMN24979362	SRR17642707
Dsp_nov_MorNorth__7	*Dugesia sp. nov*	Sexual	Morocco	Beni H'amed			MR1251.7	SAMN24979363	SRR17642706
DspF.Us_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de l´Us	42º15'50"	0º59'45"	MR1263.3	SAMN24979326	SRR17642747
DspF.Us_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de l´Us			MR1263.4	SAMN24979327	SRR17642746
DspTrilla_1	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1265.1	SAMN24979364	SRR17642705
DspTrilla_2	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1265.2	SAMN24979365	SRR17642704
DspTrilla_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla	41º55'51"	1º1'14"	MR1265.3	SAMN24979366	SRR17642703
DspTrilla_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1361.2	SAMN24979367	SRR17642702
DspTrilla_5	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1361.3	SAMN24979368	SRR17642701
DspTrilla_6	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1361.4	SAMN24979369	SRR17642700
DspBerga_1	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga			MR1360.1	SAMN24979356	SRR17642714
DspBerga_2	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga	42º6'21.97"	1º52'48.05"	MR1360.2	SAMN24979357	SRR17642713
DspBerga_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga			MR1360.3	SAMN24979358	SRR17642712

DspBerga_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga			MR1360.4	SAMN24979359	SRR17642711
DspBerga_5	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga			MR1360.5	SAMN24979360	SRR17642709
Outgroup									
Dma_1	*D. malickyi*	Sexual	*Greece*	Mexiates	38º53'4.09"	22º18'53.16"	MR1261.1	SAMN24979354	SRR17642716
Dma_2	*D. malickyi*	Sexual	*Greece*	Mexiates			MR1261.3	SAMN24979355	SRR17642715
DspEast_1	*Dugesia* sp.	Sexual	Greece	Eleonas	38º34'29.1"	22º23'38.50"	MR1262.1	SAMN24979310	SRR17642755
DspEast_2	*Dugesia* sp.	Sexual	Greece	Eleonas			MR1262.2	SAMN24979311	SRR17642754

Summary

27 localities

83 samples

10 species described for Western Mediterranean

1 not described species

2 species from Greece as outgroup

Table S2. Sequences used in the DNA-Barcoding identification

Species	GenBank accession code
D. aenigma	KC006968
D. aurea 1	MK712631
D. aurea 2	MK712632
D. benazzii	MK385926
D. corbata 1	MK712635
D. corbata 2	MK712636
D. cretica	KC006976
D. damoae	KC006979
D. etrusca	MK712651
D. gonocephala	OL410667
D. hepta	MK385923
D. ilvana	FJ646989
D. improvisa	KC006987
D. liguriensis	OL410632
D. subtentaculata 1	MK712608
D. subtentaculata 2	MK712605
D. tubqalis	OM281843
D. vilafarrei 1	MK712649
D. vilafarrei 2	MK712648
Dugesia sp.	MK712634
Dugesia sp.	KC007021

Table S3. Results of the analysis by sample; from quality control to selection of longer isoforms.

AND

Table S6. Parameters and results of analysis performed with every dataset.

Are available at Riutort´s Lab site

Table S4. Results of the three analysis done with OrthoFinder

Orthofinder results (with samples in dataset 1)		**Orthofinder results (with samples in dataset 4)**		**Orthofinder results (with samples in dataset 5)**	
Number of species	82	Number of species	23	Number of species	36
Number of genes	1739698	Number of genes	493341	Number of genes	781321
Number of genes in orthogroups	1702579	Number of genes in orthogroups	480355	Number of genes in orthogroups	761690
Number of unassigned genes	37119	Number of unassigned genes	12986	Number of unassigned genes	19631
Percentage of genes in orthogroups	97,9	Percentage of genes in orthogroups	97,4	Percentage of genes in orthogroups	97,5
Percentage of unassigned genes	2,1	Percentage of unassigned genes	2,6	Percentage of unassigned genes	2,5
Number of orthogroups	36056	Number of orthogroups	26832	Number of orthogroups	29742
Number of species-specific orthogroups	1144	Number of species-specific orthogroups	426	Number of species-specific orthogroups	745
Number of genes in species-specific orthogroups	3120	Number of genes in species-specific orthogroups	1335	Number of genes in species-specific orthogroups	2021
Percentage of genes in species-specific orthogroups	0,2	Percentage of genes in species-specific orthogroups	0,3	Percentage of genes in species-specific orthogroups	0,3
Mean orthogroup size	47,2	Mean orthogroup size	17,9	Mean orthogroup size	25,6
Median orthogroup size	15	Median orthogroup size	18	Median orthogroup size	20
G50 (assigned genes)	86	G50 (assigned genes)	23	G50 (assigned genes)	38
G50 (all genes)	86	G50 (all genes)	23	G50 (all genes)	37
O50 (assigned genes)	5399	O50 (assigned genes)	6576	O50 (assigned genes)	6271
O50 (all genes)	5615	O50 (all genes)	6858	O50 (all genes)	6531
Number of orthogroups with all species present	**5387**	**Number of orthogroups with all species present**	**8505**	**Number of orthogroups with all species present**	**7576**
Number of single-copy orthogroups	**717**	**Number of single-copy orthogroups**	**4175**	**Number of single-copy orthogroups**	**1984**

Table S5. Sample composition of each Dataset. The outgroups are indicated in bold format. The asterisk indicates that these samples were used for ortholog search. SC: Single Copy Orthologs included

Dataset 1*	**Dataset 2**	**Dataset 3**	**Dataset 4***	**Dataset 5***	**Dataset 6**
717 SC	**717 SC**	**717 SC**	**4175 SC**	**1984**	**1984**
Daur_1					
Daur_2	X				
Daur_3	X	X			
DbenSard_6					
DbenSard_7	X				
DbenSard_9	X				
DbenCors_North_1	X	X			
DbenCors_North_2					
DbenCors_North_3	X				
DbenCors_South_5					
DbenCors_South_6					
Dcorb_1					
Dcorb_2	X	X			
Dcorb_3	X				
DetruParr_1				X	X
DetruParr_2				X	X
DetruPie_2	X			X	X
DetruPie_3				X	X
DetruPie_4	X	X		X	X
Dgono_1	X			**X**	**X**
Dgono_7	X	X		**X**	**X**
Dgono_8				**X**	**X**
Dhept_1	X				
Dhept_2	X	X			
Dhept_5					
Dilv_1	X	X		X	X
Dilv_2	X			X	X
Dilv_4				X	X
DliguBis_1				X	X
DliguBis_2	X			X	X
DliguBis_3				X	X
DliguAlp_1	X			X	X
DliguAlp_3				X	X
DliguAlp_4		X		X	X
DliguGarda_1				X	X
DliguSas_2				X	X
DliguSas_3				X	X
DliguSas_4				X	X
DliguTriga_1				X	X

DliguTriga_2				X	X
DsubMont			X		
DsubBosq_1			X		
DsubBosq_2			X		
DsubCangAsex_5			X		
DsubCangAsex_6			X		
DsubCangAsex_7			X		
DsubCangSex_2			X		
DsubCangSex_3			X		
DsubCangSex_4			X		
DsubMor_North_1			X		
DsubMor_North_2			X		
DsubMor_North_3			X		
DsubMor_South_1			X		
DsubMor_South_2			X		
DsubMch_1	X		X		
DsubMch_2	X	X	X		
DsubMch_4			X		
DsubStFe_1			X		
DsubStFe_2			X		
DsubStFe_3			X		
Dvila_1	X		**X**		
Dvila_2	X	X	**X**		
Dvila_3			**X**		
Dsp_nov_MorNorth_1	X				
Dsp_nov_MorNorth__7	X	X			
DspF.Us_3				X	X
DspF.Us_4				X	X
DspTrilla_1				X	X
DspTrilla_2				X	X
DspTrilla_3	X			X	X
DspTrilla_4				X	X
DspTrilla_5				X	X
DspTrilla_6				X	X
DspBerga_1				X	
DspBerga_2				X	
DspBerga_3				X	
DspBerga_4				X	
DspBerga_5				X	
Dma_1	**X**				
Dma_2	**X**	**X**			
DspEast_1	**X**				
DspEast_2	**X**				

Supplementary Figures

Figure S1. ML trees obtained from dataset 1 (13 species, 82 samples, and 717 single copy orthologs, SC) in IQ-TREE for (A): protein data, and (B): nucleotide data. The dots on the nodes represents bootstrap support = 100. Scale bar: substitutions per site.

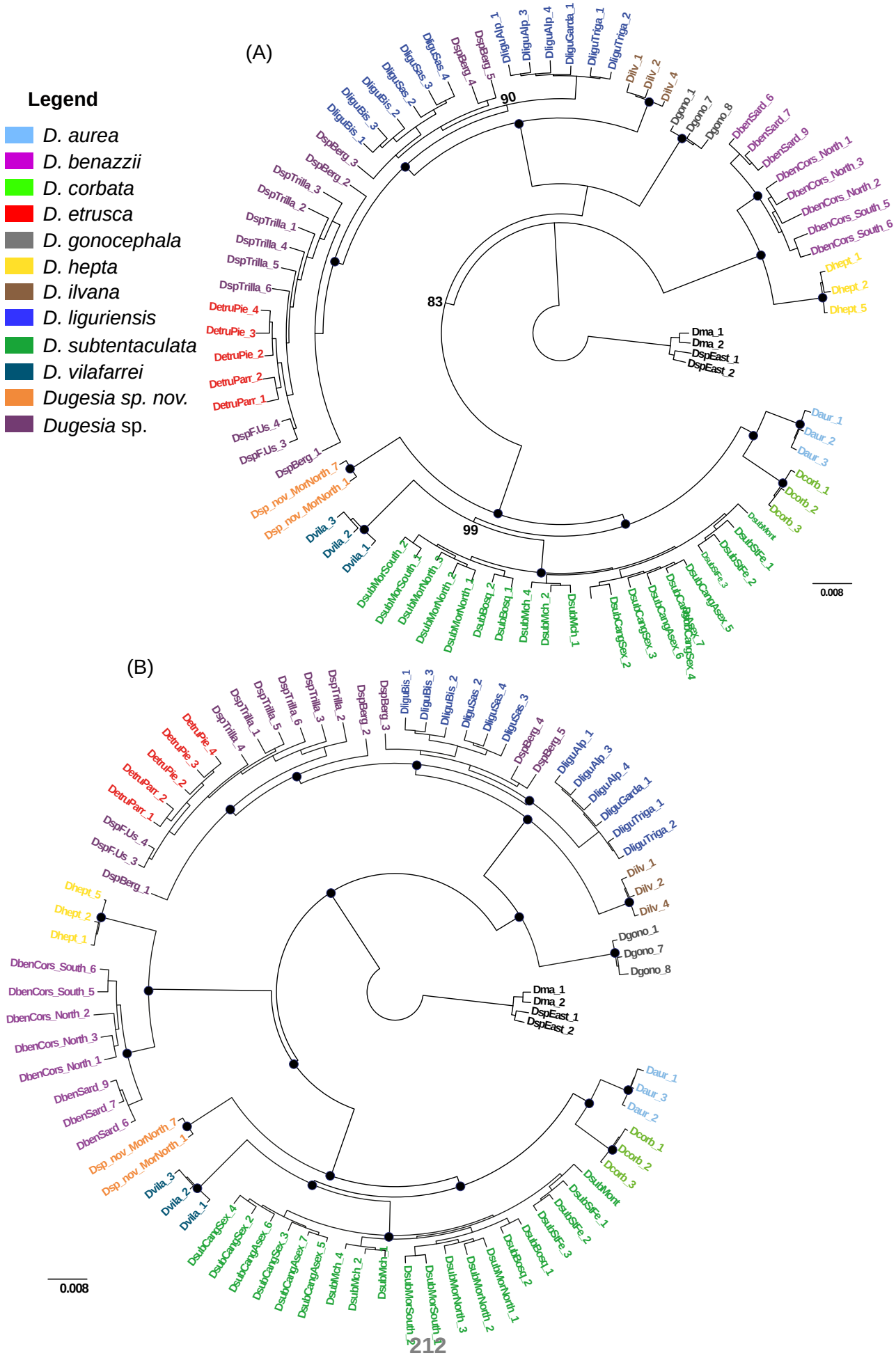


Figure S2. Phylogenetic trees obtained from dataset 2 (13 species, 29 samples, and 717 SC) with (A): ML using protein data, (B): BI using protein data, and (C): BI using nucleotide data. Supporting values are shown on the nodes, the black circles represents the maximum values (bv = 100 and pp = 1). Scale bar: substitutions per site.

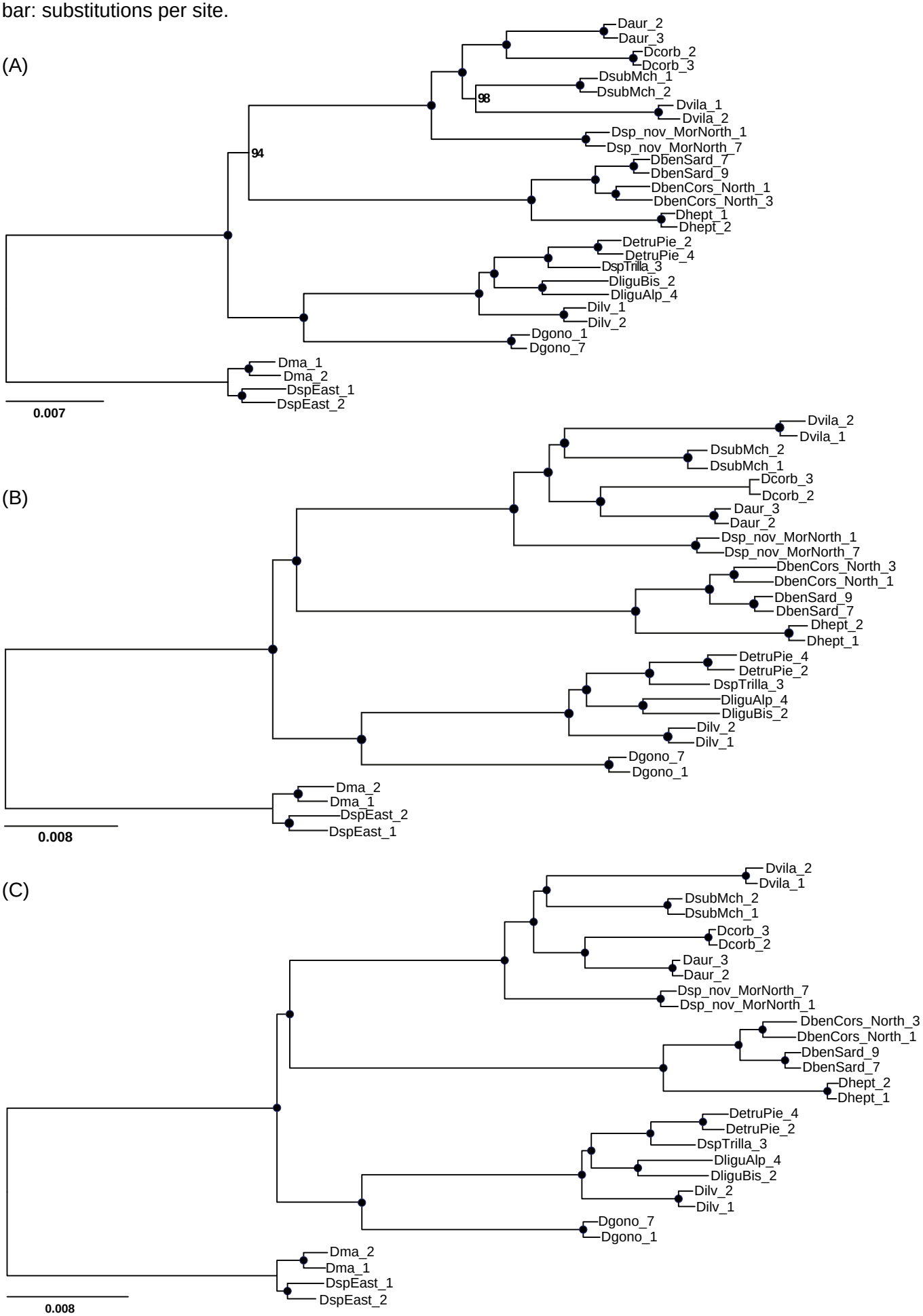
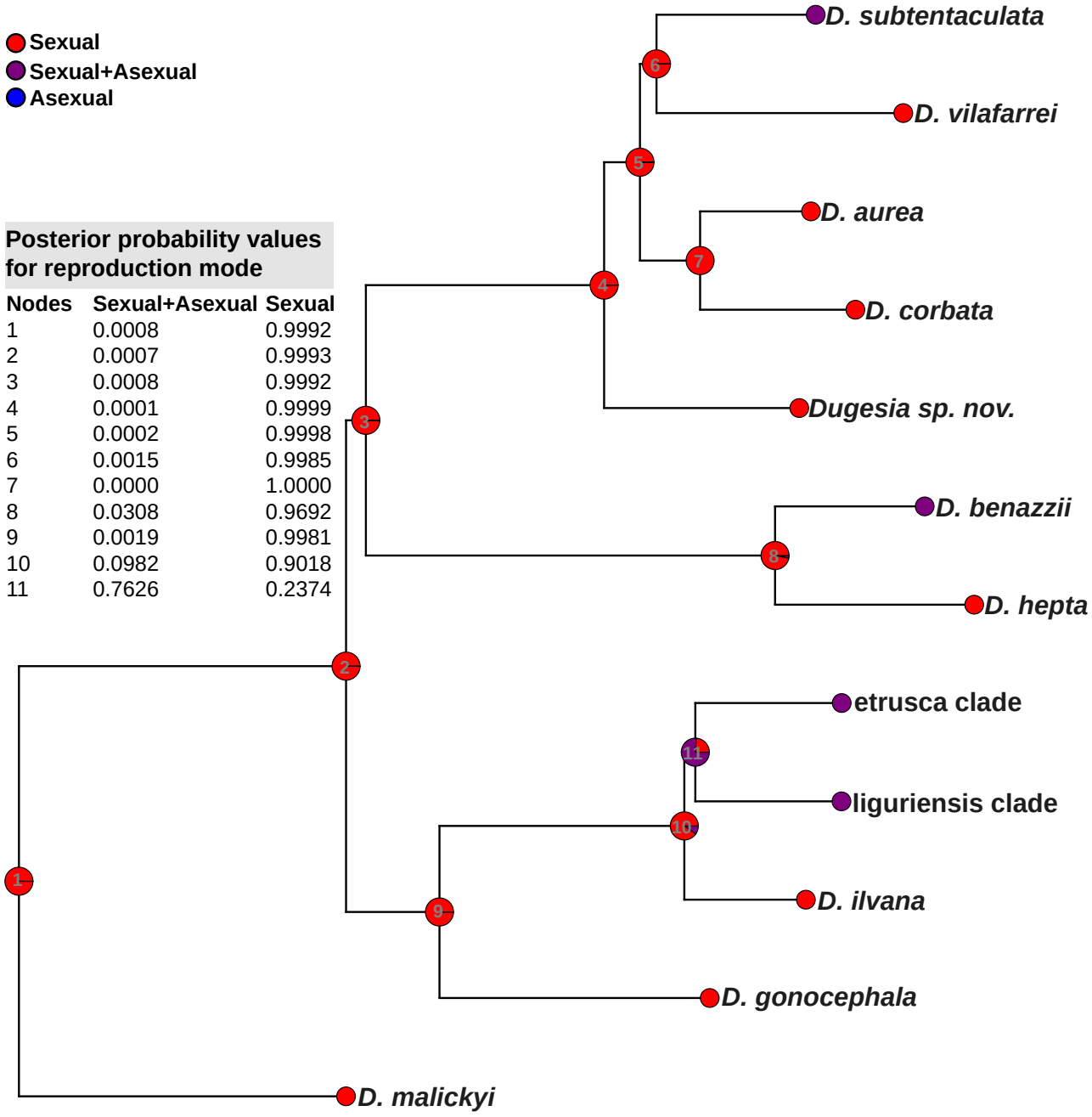


Figure S3. Ancestral character reconstruction using dataset 3 (12 species, 12 samples, and 717 SC). Pie chart of posterior probability (pp) for reproduction mode (sexual, sexual+asexual, and asexual) are shown on the nodes. The pp values are shown on the table



Posterior probability values for reproduction mode

Nodes	Sexual+Asexual	Sexual
1	0.0008	0.9992
2	0.0007	0.9993
3	0.0008	0.9992
4	0.0001	0.9999
5	0.0002	0.9998
6	0.0015	0.9985
7	0.0000	1.0000
8	0.0308	0.9692
9	0.0019	0.9981
10	0.0982	0.9018
11	0.7626	0.2374

References

Álvarez-Presas, M, Carbayo, F., Rozas, J., & Riutort, M. (2011). Land planarians (Platyhelminthes) as a model organism for fine scale phylogeographic studies: understanding patterns of biodiversity in the Brazilian Atlantic Forest hotspot. *Journal of Evolutionary Biology*, *24*, 887–896. https://doi.org/10.1111/j.1420-9101.2010.02220.x

Álvarez-Presas, Marta, & 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

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Baguñà, J., Carranza, S., Pala, M., Ribera, C., 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. (1946). Sopra una nova planaria d'acqua dolce. *Estratto Dall' Archivo Zoologico Italiano*, *XXXI*.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, *4:e1660*. https://doi.org/10.7717/peerj.1660

Bortolotti, V., Fazzuoli, M., Pandeli, E., Principi, G., Babbini, A., & Corti, S. (2001). Geology of central and eastern Elba island, Italy. *Ofioliti*, *26*(2 A), 97–105. https://doi.org/10.4454/ofioliti.v26i2a.137

Bortolotti, V., Pandeli, E., & Principi, G. (2001). The geology of the Elba Island: An historical introduction. *Ofioliti*, *26*(2 A), 79–96. https://doi.org/10.4454/ofioliti.v26i2a.136

Bouckaert, R. R., & Heled, J. (2014). DensiTree 2: Seeing Trees Through the Forest. *BioRxiv*, 012401. https://doi.org/10.1101/012401

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(421). https://doi.org/10.1186/1471-2105-10-421

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scalephylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/doi:10.1093/bioinformatics/btp348

Challis, R., & Paulini, M. (2021). *blobtoolkit/blobtools2: v2.6.1*. https://doi.org/10.5281/zenodo.5032179

Chen, X., Chen, Y., Wu, C., & Wang, A. (2015). A new species of the genus *Girardia* (Tricladida: Dugesiidae) from China. *Zoological Systematic*, *40*(2), 166–178. https://doi.org/10.11865/zs.20150202

Cheon, S., Zhang, J., & Park, C. (2020). Is Phylotranscriptomics as Reliable as Phylogenomics? *Molecular Biology and Evolution*, *37*(12), 3672–3683. https://doi.org/10.1093/molbev/msaa181

Chris Venditti, Andrew Meade, M. P. (2008). Phylogenetic Mixture Models Can Reduce Node-Density Artifacts. *Systematic Biology*, *57*(2), 286–293. https://doi.org/10.1080/10635150802044045

Cunha, T. J., & Giribet, G. (2019). A congruent topology for deep gastropod relationships. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1898). https://doi.org/10.1098/rspb.2018.2776

Dapporto, L., & Cini, A. (2007). Faunal patterns in Tuscan archipelago butterflies: The dominant influence is recent geography not paleogeography. *European Journal of Entomology*, *104*(3), 497–503. https://doi.org/10.14411/eje.2007.070

De Vicente, G., Cloetingh, S., Van Wees, J. D., & Cunha, P. P. (2011). Tectonic classification of Cenozoic Iberian foreland basins. *Tectonophysics*, *502*(1–2), 38–61. https://doi.org/10.1016/j.tecto.2011.02.007

De Vries, E. J. (1986). On the taxonomic status of *Dugesia gonocephala* and *Dugesia subtentaculata* (Turbellaria, Tricladida, Paludicola). *Journal of Zoology*, *209*(1), 43–59. https://doi.org/https://doi-org.sire.ub.edu/10.1111/j.1469-7998.1986.tb03565.x

De Vries, E. J. (1988). Further contribution to the taxonomy and biogeography of the subgenus *Dugesia* (Platyhelminthes: Tricladida: Paludicola) in the Mediterranean region and the Middle East. *Isarael Journal of Zoology*, *35*, 109–136.

DeSalle, R., & Goldstein, P. (2019). Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution*, *7*(September), 1–11. https://doi.org/10.3389/fevo.2019.00302

Dèzes, P., Schmid, S. M., & Ziegler, P. A. (2004). Evolution of the European Cenozoic Rift System: Interaction of the Alpine and Pyrenean orogens with their foreland lithosphere. *Tectonophysics*, *389*(1–2), 1–33. https://doi.org/10.1016/j.tecto.2004.06.011

Dèzes, P., Schmid, S. M., & Ziegler, P. A. (2005). Reply to comments by L. Michon and O. Merle on 'Evolution of the European Cenozoic Rift System: Interaction of the Alpine and Pyrenean orogens with their foreland lithosphere' by P. Dèzes, S.M. Schmid and P.A. Ziegler, Tectonophysics 389 (2004) 1-33. *Tectonophysics*, *401*(3–4), 257–262. https://doi.org/10.1016/j.tecto.2005.02.002

Di Nicola, M. R., & Vaccaro, A. (2020). New data on the presence of the Aesculapian snake *Zamenis longissimus* (Laurenti, 1768) (Serpentes Colubridae) on Elba Island (Tuscany, Italy). *Biodiversity Journal*, *11*(2), 611–614. https://doi.org/10.31396/biodiv.jour.2020.11.2.611.614

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*(3), 369–386. https://doi.org/10.1007/s13127-020-00438-z

Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., … Davis, C. C. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, *94*, 447–462. https://doi.org/10.1016/j.ympev.2015.10.027

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

Fattorini, S. (2009). Both Recent and Pleistocene geography determine animal distributional patterns in the Tuscan Archipelago. *Journal of Zoology*, *277*(4), 291–301. https://doi.org/10.1111/j.1469-7998.2008.00540.x

Feng, Y. Y., Shen, T. T., Shao, C. C., Du, H., Ran, J. H., & Wang, X. Q. (2021). Phylotranscriptomics reveals the complex evolutionary and biogeographic history of the genus *Tsuga* with an East Asian-North American disjunct distribution. *Molecular Phylogenetics and Evolution*, *157*, 107066. https://doi.org/10.1016/J.YMPEV.2020.107066

Fernández, R., Edgecombe, G. D., & Giribet, G. (2016). Exploring Phylogenetic Relationships within Myriapoda and the Effects of Matrix Composition and Occupancy on Phylogenomic Reconstruction. *Systematic Biology*, *65*(5), 871–889. https://doi.org/10.1093/SYSBIO/SYW041

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

Fernández, R., Laumer, C. E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P. P., … Giribet, G. (2014). Evaluating Topological Conflict in Centipede Phylogeny Using Transcriptomic Data Sets. *Molecular Biology and Evolution*, *31*(6), 1500–1513. https://doi.org/10.1093/MOLBEV/MSU108

Fernández, R., Sharma, P. P., Tourinho, A. L., & Giribet, G. (2017). The Opiliones tree of life: shedding light on harvestmen relationships through transcriptomics. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1849). https://doi.org/10.1098/RSPB.2016.2340

Filipe, A. F., Araújo, M. B., Doadrio, I., Angermeier, P. L., & Collares-Pereira, M. J. (2009). Biogeography of Iberian freshwater fishes revisited: The roles of historical versus contemporary constraints. *Journal of Biogeography*, *36*(11), 2096–2110. https://doi.org/10.1111/j.1365-2699.2009.02154.x

Foley, S., Lüddecke, T., Cheng, D. Q., Krehenwinkel, H., Künzel, S., Longhorn, S. J., … Piel, W. H. (2019). Tarantula phylogenomics: A robust phylogeny of deep theraphosid clades inferred from transcriptome data sheds light on the prickly issue of urticating setae evolution. *Molecular Phylogenetics and Evolution*, *140*, 106573. https://doi.org/10.1016/J.YMPEV.2019.106573

Fu, L., Niu, B., Zhu, Z., Wu, S., & Weizhong, L. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, *29*(7), 644. https://doi.org/10.1038/NBT.1883

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

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Haas, B., & Papanicolaou, A. (2019). *TransDecoder 5.5.0*. Retrieved from https://github.com/TransDecoder/TransDecoder/wiki

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, *41*, 95–98.

Harrath, A. H., Sluys, R., Merzoug, D., Yacoubikhebiza, M., Alwasel, S., & Riutort, M. (2012). Freshwater planarians (Platyhelminthes, Tricladida) from the Palearctic section of the African continent: New records, with the description of a new species. *Zootaxa*, *15*(3182), 1–15. https://doi.org/10.11646/zootaxa.3182.1.1

Hasegawa, M., Kishino, H., & Yano, T. aki. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution 1985 22:2*, *22*(2), 160–174. https://doi.org/10.1007/BF02101694

Hoang, D. T., Chernomor, O., Haeseler, A. von, Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, *35*, 518–522. https://doi.org/10.1093/molbev/msx281

Holder, M., & Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*, *4*(4), 275–284. https://doi.org/10.1038/NRG1044

Jukes, T. H., & Cantor, C. R. (1969). Evolution of Protein Molecules. *Mammalian Protein Metabolism*, 21–132. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics*, *26*(13), 1669–1670. https://doi.org/10.1093/bioinformatics/btq243

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*(6), 587–589. https://doi.org/10.1038/nmeth.4285

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kobayashi, K., Maezawa, T., Nakagawa, H., & Hoshi, M. (2012). Existence of Two Sexual Races in the Planarian Species Switching between Asexual and Sexual Reproduction. *Zoological Science*, *29*(4), 265. https://doi.org/10.2108/zsj.29.265

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research*, *6*(1287). https://doi.org/10.12688/f1000research.12232.1

Lanfear, R., Hua, X., & Warren, D. L. (2016). Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. *Genome Biology and Evolution*, *8*(8), 2319–2332. https://doi.org/10.1093/gbe/evw171

Lartillot, N., & Philippe, H. (2004). A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Molecular Biology and Evolution*, *21*(6), 1095–1109. https://doi.org/10.1093/molbev/msh112

Lázaro, E. M., Sluys, R., Pala, M., Stochino, 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

Lázaro, E. M., & 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

Le, S. Q., & Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, *25*(7), 1307–1320. https://doi.org/10.1093/MOLBEV/MSN067

Lemer, S., Bieler, R., & Giribet, G. (2019). Resolving the relationships of clams and cockles: Dense transcriptome sampling drastically improves the bivalve tree of life. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1896). https://doi.org/10.1098/rspb.2018.2684

Leria, L., Riutort, M., Romero, R., Ferrer, X., & Vila-Farré, M. Microplate tectonics and the environment as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography, In Press*.

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). https://doi.org/10.1186/s12862-019-1440-1

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

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li, Y., Shen, X. X., Evans, B., Dunn, C. W., & Rokas, A. (2021). Rooting the Animal Tree of Life. *Molecular Biology and Evolution*, *38*(10), 4322–4333. https://doi.org/10.1093/molbev/msab170

Liu, L., Anderson, C., Pearl, D., & Edwards, S. V. (2019). Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model. In M. Anisimova (Ed.), *Evolutionary Genomics. Methods in molecular biology* (Vol. 1910, pp. 211–239). https://doi.org/10.1007/978-1-4939-9074-0_7

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Martinez Rica, J. P., & Montserrat Recoder, P. (1990). Biogeographic features of the Pyrenean range. *Mountain Research and Development*, *10*(3), 235–240.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Haeseler, A. von, & Lanfear, R. (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

Mirarab, S., Nakhleh, L., & Warnow, T. (2021). Multispecies Coalescent: Theory and Applications in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *52*, 247–268. https://doi.org/10.1146/annurev-ecolsys-012121-095340

Nei, M., & Kumar, S. (2000). *Molecular Evolution and Phylogenetics* (M. Nei & S. Kumar, Eds.). New York: Oxford University Press, Inc.

Ninot, J. M., Carrillo, E., & Ferré, A. (2017). The Pyrenees BT - *The Vegetation of the Iberian Peninsula: Volume 1* (J. Loidi, Ed.). https://doi.org/10.1007/978-3-319-54784-8_8

Nishimura, O., Hosoda, K., Kawaguchi, E., Yazawa, S., Hayashi, T., Inoue, T., … Agata, K. (2015). Unusually Large Number of Mutations in Asexually Reproducing Clonal Planarian *Dugesia japonica*. *PLOS ONE*, *10*(11), e0143525. https://doi.org/10.1371/JOURNAL.PONE.0143525

Phillips, A. J. L., Hyde, K. D., Alves, A., & Liu, J. K. (Jack). (2018). Families in Botryosphaeriales: a phylogenetic, morphological and evolutionary perspective. *Fungal Diversity*, *94*(1), 1–22. https://doi.org/10.1007/S13225-018-0416-6

Quang, L. S., Gascuel, O., & Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*, *24*(20), 2317–2323. https://doi.org/10.1093/bioinformatics/btn445

Rambaut, A., & Drummond, A. J. (2007). *Tracer v1.4.* Available at: beast.bio.ed.ac.uk/Tracer.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/J.2041-210X.2011.00169.X

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. *Int. J. Dev. Biol.*, *56*, 5–17. https://doi.org/10.1387/ijdb.113441mr

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (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

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

Schrempf, D., Lartillot, N., & Szöllősi, G. (2020). Scalable Empirical Mixture Models That Account for Across-Site Compositional Heterogeneity. *Molecular Biology and Evolution*, *37*(12), 3616–3631. https://doi.org/10.1093/molbev/msaa145

Schwander, T., Henry, L., & Crespi, B. J. (2011). Molecular evidence for ancient asexuality in timema stick insects. *Current Biology*, *21*(13), 1129–1134. https://doi.org/10.1016/j.cub.2011.05.026

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with

single copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/BIOINFORMATICS/BTV351

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*, 523–547. https://doi.org/10.1111/zoj.12077

Solà, E., Leria, L., Stocchino, G. A., Bagherzadeh, R., Balke, M., Daniels, S. R., … Riutort, M. Three dispersal routes out of Africa: The puzzling biogeographical history of Dugesia freshwater planarians. *Journal of Biogeography, In Press*.

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

Stocchino, G. A., & Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology*, *80*(3), 319–328. https://doi.org/10.1080/11250003.2013.822025

Stocchino, Giacinta Angela, Sluys, R., & Manconi, R. (2012). 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*, *58*(3551), 43–58. https://doi.org/10.11646/zootaxa.3551.1.3

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, *17*, 57–86.

Team, R. C. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Retrieved from https://www.r-project.org/

Vergés, J., Millán, H., Roca, E., Muñoz, J. A., Marzo, M., Cirés, J., … Cloetingh, S. (1995). Eastern Pyrenees and related foreland basins: pre-, syn- and post-collisional crustal-scale cross-sections. *Marine and Petroleum Geology*, *12*(8), 903–915. https://doi.org/10.1016/0264-8172(95)98854-X

Vila-Farré, M., & Rink, J. C. (2018). The Ecology of Freshwater Planarians. In J. C. Rink (Ed.), *Planarian Regeneration. Methods in Molecular Biology, vol 1774* (pp. 173–205). https://doi.org/10.1007/978-1-4939-7802-1_3

Vu, D., Groenewald, M., de Vries, M., Gehrmann, T., Stielow, B., Eberhardt, U., … Verkley, G. J. M. (2019). Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. *Studies in Mycology*, *92*, 135–154. https://doi.org/10.1016/j.simyco.2018.05.001

Wang, H. C., Minh, B. Q., Susko, S., & Roger, A. J. (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

Zhang, C., Scornavacca, C., Molloy, E. K., & Mirarab, S. (2020). ASTRAL-pro: Quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution*, *37*(11), 3292–3307. https://doi.org/10.1093/molbev/msaa139

Chapter IV

Section 2

Genomic footprints of ancient fissiparity in the evolutionary history of freshwater planarians

LISANDRA BENÍTEZ-ÁLVAREZ, LAIA LERIA, ROSA FERNÁNDEZ, IGNACIO TENAGUILLO, AND MARTA RIUTORT

Genomic footprints of ancient fissiparity in the evolutionary history of freshwater planarians

Lisandra Benítez-Álvarez[1], Laia Leria[1], Rosa Fernández[2], Ignacio Tenaguillo[1], and Marta Riutort[1, *]

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

[2]Metazoa Phylogenomics Lab, Biodiversity Program, Institut de Biologia Evolutiva (CSIC- Universitat Pompeu FabraIBE). Passeig marítim de la Barceloneta 37-49. 08003 Barcelona, Catalonia, Spain.

*corresponding author: mriutort@ub.edu, (34) 934035432

Abstract

Reproductive strategies shape the genetic background of populations. Recombination leads to increase the genetic variability and the higher efficiency of purifying selection, while evolutionary disadvantages have always been attributed to asexuality. However, long term strict asexual lineages defy this rule. Furthermore, except for the typical consequences of asexuality, little is known about the genetic footprint of long term fissiparity. Freshwater planarians are recognized for their outstanding regeneration capability, which allows the group to reproduce asexually by fission. The genus *Dugesia* shows a great diversity of reproductive strategies, with species being strictly sexual, strictly fissiparous or combining both reproductive modes. *Dugesia* species from the Western Mediterranean have been recently studied from a biogeographic point of view using transcriptomic data. In that study, asexual populations belonging to the most ancient clade showed an unexpected phylogenetic pattern putatively associated with long term asexuality. Here, we further investigate the ladder-like topology as an effect of ancient fissiparity and not as an artefact of the phylogenetic inference. Our hypothesis proposes that under long-term asexuality, individuals belonging to ancient divergent fissiparous lineages coalesce very back in time. Then, the phylogenetic trees show the descendants of divergent asexual lineages in a ladder-like pattern. Additionally, ancient asexual populations can act as reservoirs of ancestral polymorphisms, that may result in the attraction of the sexual lineages. This information is misleading for the phylogenetic inference methods that fail to infer the true species

tree topology in this case. We propose that the ladder-like pattern can be considered a genomic footprint of long term fissiparity, which can affect the phylogenetic inference and therefore, it strongly impact the species tree estimation.

Keywords: asexual reproduction, fissiparity, planarians, *Dugesia,* phylotranscriptomics

1. Introduction

The reproductive strategy of organisms determines the genetic background of their populations. Recombination during sexual reproduction is evolutionary advantageous due to the increasing of genetic variability and a higher efficiency of purifying selection, while asexual reproduction tends to decrease the genetic variability of populations, making them clonal (Webster & Hurst, 2012).

The main predicted effect of asexuality is the irreversible accumulation of deleterious mutations, known as Muller's ratchet (Felsenstein, 1974; H. Muller, 1932; Muller, 1964), which theoretically, can lead to the extinction of asexual lineages. Additionally, under an asexual scenario, recombination does not homogenize the alleles within the population and the alleles should accumulate mutations independently of each other (Birky, 1996; Welch & Meselson, 2000). This effect of asexuality is named Meselson effect and is recognized because the intraindividual alleles show high levels of divergence. Therefore, an allele might be more closely related to a homolog allele in another individual than their sister alleles within the same individual (Schwander, Henry, & Crespi, 2011). This effect is detectable in populations under long periods of time without sexual reproduction (Normark, Judson, & Moran, 2003). However, processes such as hybridization or gene duplication can lead to a Meselson effect-like allele divergence pattern (Schwander et al., 2011).

A variant of this effect, the Mosaic-Meselson effect, has been proposed to occur in agametic asexual organisms (Leria, Vila-Farré, Solà, & Riutort, 2019). Agametic asexual species, differing from those gametic, inherit part of the soma of the progenitor during each reproductive event. This situation would allow mutations not only to independently accumulate in the alleles within a cell but also among different cells, generating a Meselson effect at a mosaic level. This effect was investigated based on the information of two molecular markers using the freshwater planarian *Dugesia subtentaculata* as a model (Leria et al., 2019). However, evidence of this effect at the genomic level is still missing.

On the other hand, recombination has other direct effects on genome evolution. One of these effects is the GC-biased gene conversion (gBGC), resulting from biased incorporation of G and C nucleotides during repair of mismatches formed during meiotic homologous recombination. gBGC is thought to be responsible for the

correlation between GC content and recombination (see Webster & Hurst, 2012 for a broader review). Thus, substantial differences in the gBGC strength have been found between low and high recombining regions (Capra, Hubisz, Kostka, Pollard, & Siepel, 2013; Duret & Galtier, 2009; Galtier, 2021).

The percent of GC at the third codon position (GC3%) has been used to characterize the genomic GC-content in different phylogenetic groups, especially the mammals (Romiguier & Roux, 2017). In *Timema* species (stick insects) the GC3% has been used to study the effect of long-term asexuality on gBGC strength. Using transcriptomic data, these authors analysed the GC3% of pairs of sexual/asexual species with different divergence times. They found higher GC3 content and higher variance among genes from sexual species than asexual species within each pair. In addition, the proportion of genes in which GC3% was higher in the sexual than in the asexual species was greater in older pairs than in youngest pairs (Bast et al., 2018).

Freshwater planarians are broadly recognized by its outstanding regeneration capability. Genera such as *Schmidtea* and *Dugesia* (family Dugesiidae) lead the list of favourite model organisms in regeneration studies (Rink, 2013). This regeneration capability allows the group to reproduce asexually by fission, which can be alternated in different combinations with the sexual mode (Stocchino & Manconi, 2013). Sexual reproduction in planarians occurs by cross fertilization between two hermaphrodite individuals. The asexual reproduction by fission happens when one individual divides and each part of the body regenerates the missing structures (Reddien & Alvarado, 2004). Generally, asexual individuals do not develop the reproductive systems and although the sexual ones maintain the regenerative capabilities in response to injuries, they are not able to regenerate the entire body so efficiently as in the case of asexual individuals (Saló et al., 2009).

The genus *Dugesia* constitutes one of the most diverse groups of freshwater planarians in the Paleartic region (Solà et al., *in press*). Moreover, *Dugesia* species show a great diversity of reproductive strategies, with species being strictly sexual, strictly fissiparous or combining both reproductive modes (Leria et al., 2019; Stocchino & Manconi, 2013). Among all *Dugesia* species, those from the Western Mediterranean region have been recently studied from a biogeographic point of view using transcriptomic data (Chapter IV.1). The Western Mediterranean species are divided into three main clades: the Iberia-Apennines-Alps-plus clade, the Corsica-Sardinia

clade and the Iberia-Africa clade. Interestingly, in Chapter IV.1 was found out that the species from the Iberia-Apennines-Alps-plus region show an unexpected phylogenetic pattern, putatively associated to a long term asexuality.

The Iberia-Apennines-Alps-plus clade includes three populations from the Iberian Peninsula still not assigned to a species, *D. etrusca* from the Apennines region, *D. liguriensis* from the Alps, *D. ilvana* from Elba island, and *D. gonocephala* distributed in almost all continental Europe (Fig. 1). One of the objectives of Chapter IV.1 was the species assignation of the Iberian populations from Berga, Font de l'Us and Font de La Trilla to *D. etrusca* or *D. liguriensis* species, but it was not possible because an unexpected topology and several controversial nodes in the phylogenies (Fig. S1). Even though the samples from Font de l'Us and Font de La Trilla could had been assigned to *D. etrusca* with certain doubts, the assignment of samples from Berga was impossible due to their incongruent groupings in the trees obtained with different methods. Therefore, the clades including the asexual individuals are referred as *sensu lato* (*s.l*) species. In addition, the unexpected topology consisting of a ladder-like pattern of samples from Font de La Trilla and Berga in the trees (Fig. S1) raised suspicions on fissiparity affecting the evolutionary history of these populations. The authors outlined the idea that ancient asexuality could be underlying the odd pattern observed, since the Iberians populations constitute the only extant asexual lineages in the Iberia-Apennines-Alps-plus clade. Moreover, in that study the ladder-like pattern was shown only in the Iberia-Apennines-Alps-plus clade, but not in other asexual lineages from the Western Mediterranean, namely *D. subtentaculata* populations. In consequence, the authors proposed the idea that the populations from Berga, Font de la Trilla and Font de l'Us are the remnants of a very ancient asexual lineage, while the other extant asexual lineages belonging to *Dugesia* from Western Mediterranean are more recent and, moreover, it is known they combine fissiparity with occasional sex (Leria et al., 2019).

The main objective of the present study is to validate the ladder-like topology as an effect of ancient asexuality and not an artifact of the phylogenetic inference. For that, we first validated the methodology of ortholog search using a different ortholog inference method and also a set of paralogs. Second, we used recognized effects of the lack of recombination to demonstrate that these populations have maintained asexuality for a long time without alternating with sexual stages. Finally, we analysed

how the inclusion of ancient asexual lineages can affect the inference of the phylogenetic relationships with closest groups.

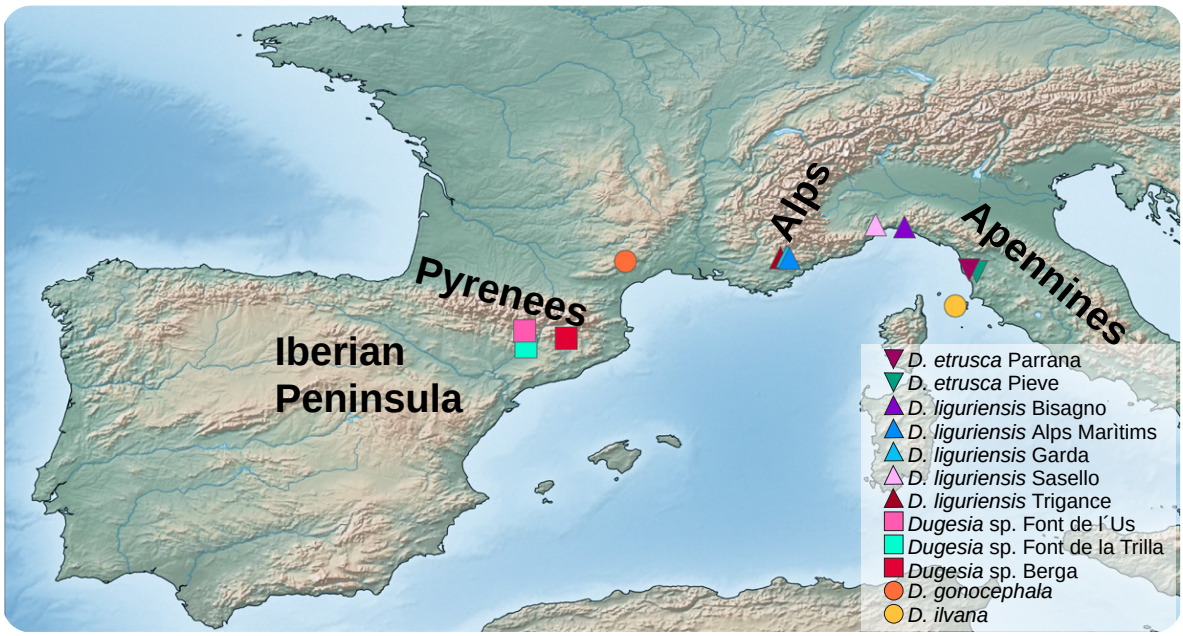


Figure 1. Localization of sampling points

2. Methods

2.1. RNA data acquisition and pre-processing

To carry out the present study, we used data previously processed in Chapter IV.1 (Table S1). This data was obtained from RNAseq sequencing using TruSeq libraries and cured to perform phylogenetic analyses.

In Chapter IV.1 were established three groups of samples to perform different ortholog searches. A first group formed by 82 samples belonging to 12 taxa from the Western Mediterranean and 2 from Eastern Mediterranean clade (outgroup) was used to obtain 717 Single Copy orthologs (SC). The second group of samples was named the subtentaculata group and included 20 samples belonging to *D. subtentaculata* and 3 samples belonging to *D. vilafarrei* (outgroup) (Table S1). Finally, the etrusca-liguriensis group included 36 samples; 5 samples belonging to *D. etrusca sensu stricto* (*s.s)* from The Tuscany, 12 samples belong to *D. liguriensis* (*s.s*) from the Ligurian region, 3 belonging to *D. ilvana* from Elba Island, 3 belonging to *D. gonocephala* from Montpellier (used as outgroup), and the 13 samples from The Iberian Peninsula,

whose classification is uncertain (Table S1; Fig. 1). The ortholog search in this sample group yielded 1984 SC. For a better understanding in the text, here we defined the two groups of orthologs as WM_717 for the ortholog search using all samples from Western Mediterranean, and etru-ligu_1984 for the search using etrusca-liguriensis samples group.

For the present work, we retained the sequences of the two groups of SC mentioned above. In the case of the WM_717, we used the nucleotide sequences belonging to etrusca-liguriensis and subtentaculata sample groups. Further, we used the nucleotide and protein sequences of the etru-ligu_1984.

In addition to the SC, here we extracted a sequence set of paralog candidate genes. We selected orthogroups with more than one sequence per sample, as paralog candidates based on OrthoFinder output. Using the Orthogroups.GeneCount.tsv file, the custom python script select_OG_all_sp.py, and simple bash commands, we generated a list of paralogs candidates (par_158) and extracted their sequences.

2.2. Ortholog search with OMA

A new ortholog search with OMA (Altenhoff et al., 2019) was performed using the same set of longest isoforms previously analyzed in Chapter IV.1 with OrthoFinder. The protein sequences of the SC were extracted from the OMA output using simple bash commands and the nucleotide sequences were extracted from Transdecoder output as is described in Chapter IV.1.

2.3. RNA datasets

Dataset 1 includes all samples belonging to etrusca-liguriensis samples group (Table S1) and the nucleotide sequences of 1984 SC shared by them. To obtain the datasets 2 and 3 we removed the samples from Berga, and all samples from Iberian Peninsula respectively from the dataset 1, keeping the same number of SC (Table 1).

The dataset 4 includes 1984 SC and 158 paralogs, while the dataset 5 only includes the paralogs. For the viability of the work, we avoided the step of nucleotide sequence extraction, and used protein data to build these datasets.

The dataset 6 was built using 717 SC and two groups of samples: the etrusca-liguriensis samples group, and the subtentaculata samples group (Table S1).

Table 1. Datasets used in the study, and analyses performed

Datasets	Composition	Number of samples	Data Type	Species Tree (Astral-pro -t2)	Species Tree (Astral-pro default)	DensiTree	ML tree	Topology test	GC3% content
Dataset 1	etrusca-liguriensis samples group / 1984 SC	36	Nucleotide	X		X			
Dataset 2	etrusca-liguriensis samples group without samples from Berga / 1984 SC	31	Nucleotide	X		X			
Dataset 3	etrusca-liguriensis samples group without samples from Iberian Peninsula / 1984 SC	23	Nucleotide	X	X	X	X	X	
Dataset 4	etrusca-liguriensis samples group / 1984 SC +158 paralogs	36	Protein		X				
Dataset 5	etrusca-liguriensis samples group / 158 paralogs	36	Protein		X				
Dataset 6	etrusca-liguriensis + subtentaculata samples groups / 717 SC	59	Nucleotide						X
Dataset 7	etrusca-liguriensis samples group / 2570 SC from OMA	36	Protein				X		
Dataset 8	etrusca-liguriensis samples group / 2570 SC from OMA	36	Nucleotide				X		

Finally, two datasets were obtained using the SC discovered by OMA; the dataset 7 for protein sequences and the dataset 8 for nucleotide ones.

In summary, 8 datasets were built from RNAseq data combining three groups of SC (WM_717, etru-ligu_1984, and OMA_2570), one group of paralogs (par_158), and different samples composition (Table 1); all of them with 100% of gene occupancy.

2.4. Data analyses

2.4.1. Maximum Likelihood trees

Maximum Likelihood *(*ML) trees were obtained in IQ-TREE (Wang, Minh, Susko, & Roger, 2018). For nucleotide data, we used three substitution models as components; JC (Jukes & Cantor, 1969), HKY (Hasegawa, Kishino, & Yano, 1985), and GTR (Tavaré, 1986), with four Gamma categories (+G4), and 1,000,000 ultrafast bootstrap replicates. For the protein data, we obtained a starting tree from non-partitioned data using MFP (-m option) and 10,000 replicates of ultrafast bootstraps. After, we ran the mixture model analysis with the following parameters: LG model (Le & Gascuel, 2008) with 20 categories (C20), Gamma rate heterogeneity calculation (+G), site-specific frequency profile inference (+F), and 1,000,000 ultrafast bootstrap replicates. ML trees were obtained from datasets 3, 7, and 8 taking into account the nature of the data to follow one or another strategy (Table 1).

2.4.2. Multispecies Coalescent Model trees

We used the Multispecies Coalescent Model (MSC) approach implemented in ASTRAL-pro (Zhang, Scornavacca, Molloy, & Mirarab, 2020) to obtain the species tree based on individual gene trees, leveraging both single copy genes and genes with paralogs. The analyses were performed using -t2 option to recover the full branch annotation of the trees obtained from datasets 1-3. The obtained information was summarized in pie charts using the AstralPlane package (Hutter, 2020) in R. Species trees from datasets 4-5 were obtained with default options.

The individual gene trees were obtained in IQ-TREE using the parameters described for ML trees, but using 10,000 ultrafast bootstrap replicates for nucleotide data, and the model LG+G+F without bootstrap replicates for protein data. The visualization of individual trees in DensiTree (Bouckaert, 2010; Bouckaert & Heled, 2014) was performed for datasets 1-3, just as is described in Benítez-Álvarez et al (in preparation).

2.4.3. Topology test

A topology test was carried on with dataset 3 using IQ-TREE. We tested the topology obtained from ML analysis and the alternative one regarding the position of *D. ilvana* in the tree. The same model (MIX{JC,HKY,GTR}+G4) described previously for nucleotide data was used. Through the -zb and -au options, several topology tests were performed: the Bootstrap proportion (BP), KH (Kishino & Hasegawa, 1989), SH (Shimodaira & Hasegawa, 1999), and expected likelihood weights (ELW) (Strimmer & Rambaut, 2002), and the approximately unbiased (AU) test (Shimodaira, 2002). All tests performed 10,000 resamplings using the RELL method (Kishino, Miyata, & Hasegawa, 1990).

2.4.4. GC content calculation

GC3% was calculated for dataset 6 (Table 1). We used MACSE (Ranwez, Harispe, Delsuc, & Douzery, 2011) to align the nucleotide coding regions without losing the codon structure. The third positions were extracted using the tool extract-codon-alignment (https://github.com/linzhi2013/extract_codon_alignment) and the percent of GC by length calculated with the fx2tab module from seqkit program (Shen, Le, Li, & Hu, 2016). A tabulate file (samples by rows and genes by columns) was analysed in R

(Team, 2021) to obtain and graph the average (AVE) and standard deviation (STD) of GC3% content by sample.

2.4.5. Dunnuc12 validation

To validate Dunuc12 as a marker for intraindividual genetic diversity studies, we extracted the exonic region of this marker from all cloned sequences and blasted (Camacho et al., 2009) them against the corresponding transcriptome assembly. The transcripts were identified through blasting against the NCBI nucleotide database. The coverage of this region in every transcriptome was obtained using samtools (Li et al., 2009) and visualized in IGV (Robinson, Thorvaldsdóttir, Wenger, Zehir, & Mesirov, 2017). We looked for the identified transcripts inside the WM_717 and the etru_ligu_1984 SC groups. Finally, ML trees were obtained from the identified orthogroups (For a more detailed description of the methodology used see Text S1).

2.4.6. Intraindividual genetic diversity

DNA of eight samples (DspBerga_2, DspBerga_3, DspBerga_5, DspTrilla_3, DspTrilla_5, DspTrilla_6, DetruPie_2, and DliguGarda_1) were used to amplify a region of approximately 680pb of the Dunuc12 marker, a fragment of TMED9 (nuclear gene Transmembrane p24 trafficking protein 9). The new specific primers D12LigEt-F2 (CGGCAAATCTACAAAATAT) and D12LigEt-R1 (CATCTTGAAATTCAAACAG), were used at 25 µM with $MgCl_2$ (25 mM), dNTPs (0.5 mM), 0.25 U of Go Taq® DNA polymerase enzyme (Promega Madison, Wisconsin, USA), and the Taq buffer (5X) in a final reaction volume of 25 µl. The amplification conditions were as follows: 1) 2’ at 94 ^{o}C, 2) 45” at 94 ^{o}C, 3) 50” at 54 ^{o}C, 4) 50” at 72 ^{o}C, and 5) 3’ at 72 ^{o}C, with steps 2, 3, and 4 ran for 35 cycles. PCR products purification, cloning procedure, and sequence analyses were performed as described in Leria et al., (2019).

A set of 9 alignments (690 pb) of Dunuc12 cloned sequences were built to construct the haplotype networks; one alignment for every individual (8-14 sequences per individual), plus the alignment including all sequences (89) obtained from all individuals (Table S2). The networks were constructed using the median-joining method (Bandelt, Forster, & Röhl, 1999). In addition, the haplotype diversity (Hd), and the nucleotide diversity (Pi) were calculated in the DnaSP software from a total alignment length of 690 positions (Librado & Rozas, 2009).

To compare the results obtained in asexual populations from Iberia-Apennines-Alps-plus clade with the asexual populations belonging to *D. subtentaculata* clade, we downloaded the data of individuals from Estella (3), Santa Fe (3), and Truchas (3) (Leria et al., 2019) from DRYAD database. We used an alignment of 691 positions including 129 cloned sequences of Dunuc12 from asexual individuals to recalculate the intraindividual diversity parameters in *D. subtentaculata.*

Additionally, proportion of Non-synonymous (*Ka*) and Synonymous (*Ks*) mutations were calculated from the same coding region (120 pb) in both *D. subtentaculata* and etrusca-liguriensis lineages.

2.4.7. Genetic distance calculation

Using the dataset 1 (Table 1), we calculated the genetic distance by pair using the TN93 model (Tamura & Nei, 1993) in the dist.dna function from ape package (Paradis & Schliep, 2019) in R.

3. Results

3.1. Validating the ladder-like topology

3.1.1. Ortholog search with OMA

A total of 548,739 orthogroups were obtained with OMA, of which 2,570 were defined as single copy (only one sequence per sample with all samples present). The ML trees obtained with the protein and nucleotide sequences of OMA_2570 recovered the previous topology, showing the ladder-like pattern of the samples from Berga and Font de la Trilla (Fig. 2).

3.1.2. Paralogs information

We extracted 158 paralog candidates for the Iberia-Apennines-Alps-plus clade. The individual trees from these paralogs were inspected by eye, to confirm that the gene tree showed duplication events. Taking advantage of the capability of ASTRAL-pro to handle paralogs information, we analysed our paralogs set together with the SC (Fig. 3A), and alone (Fig. 3B). These trees showed the same topology previously obtained. However, the values of branch length, branch support, and final normalized quartet score were low, especially in the tree obtained only with the 158 paralogs (Fig. 3B).

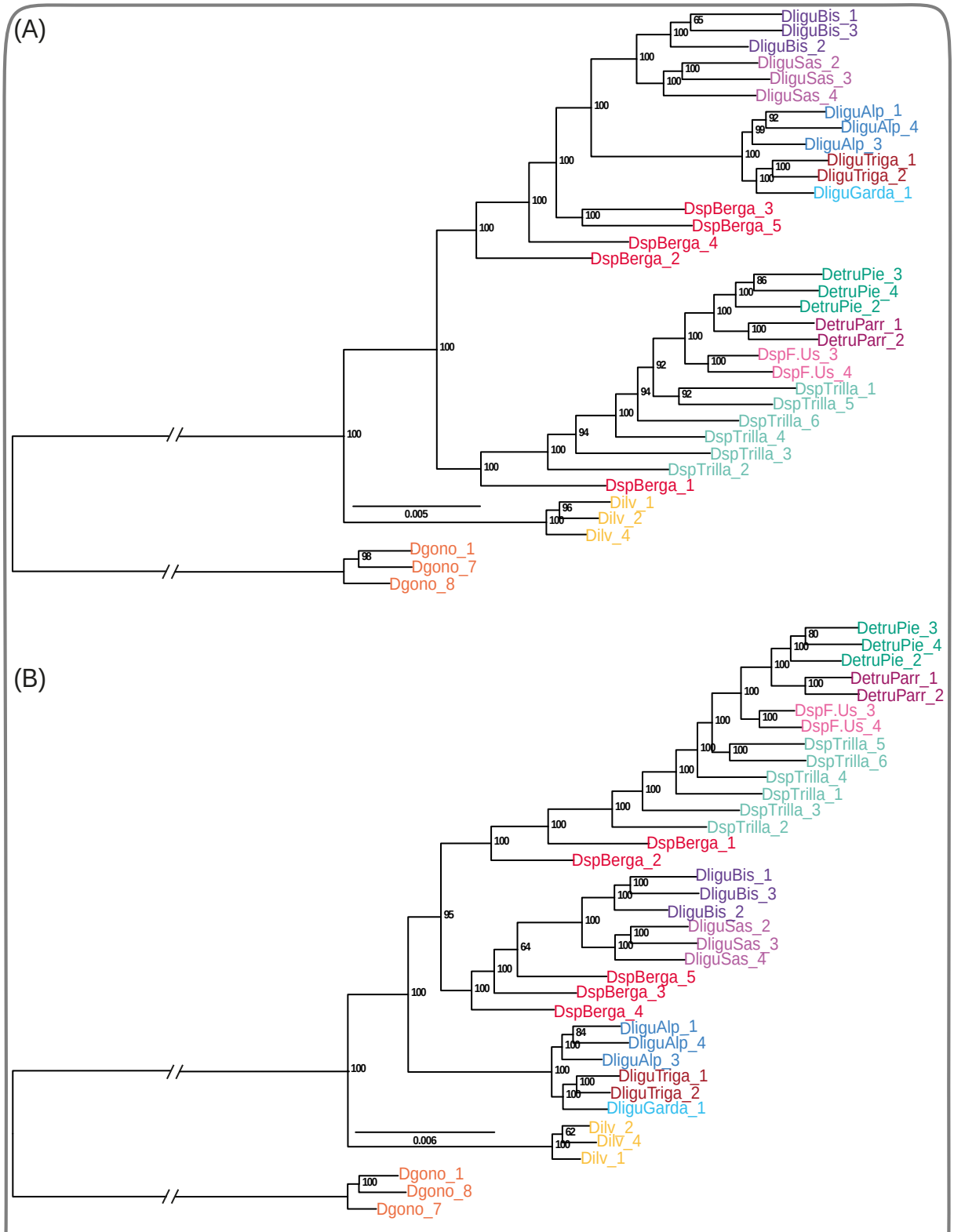


Figure 2. ML tree obtained from 2570 SC discovered with OMA analysing all the samples belonging to the Iberia-Apennines-Alps-plus clade. (A): Using protein data. (B): Using nucleotide data. Scale bar: substitutions per site. The bootstrap values are shown on the nodes.

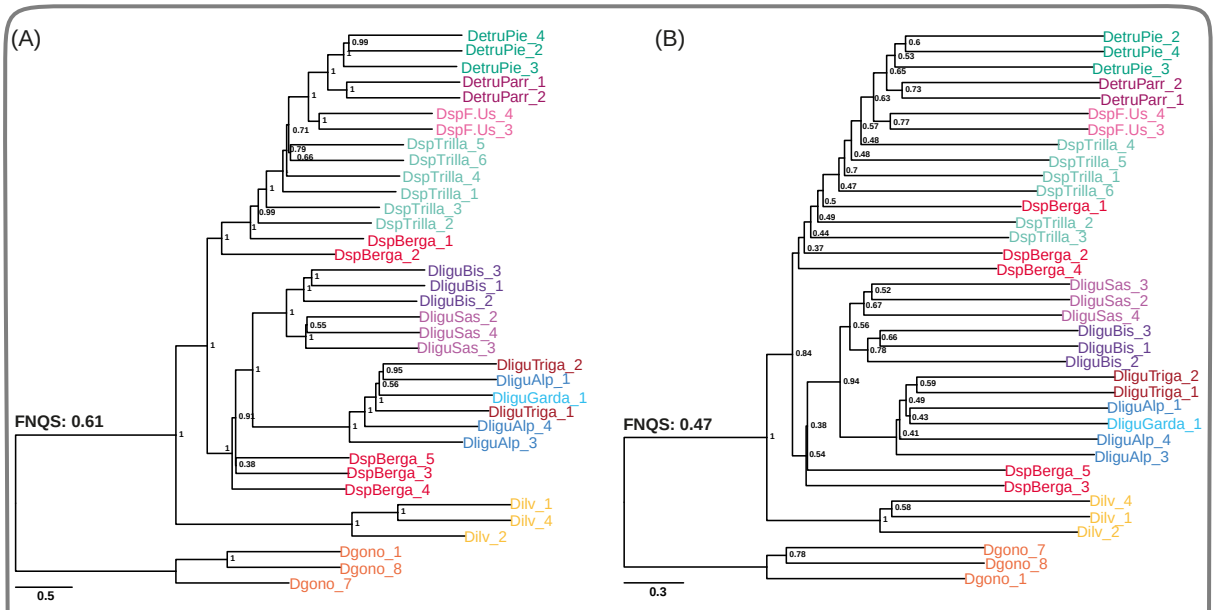


Figure 3. MSC trees obtained with ASTRAL-pro using protein data for the Iberia-Apennines-Alps-plus clade. (A): Using 1984 SC and 158 paralogs. (B): Using 158 paralogs. The branch support values are shown on the nodes. Scale bar: Coalescent Units. FNQS: Final normalized quartet score.

3.2. Ancient asexuality in the Iberia-Appenines-Alps-plus clade

3.2.1. GC content in third positions

The GC3% was calculated in sexual and asexual individuals belonging to the etrusca-liguriensis samples group and the subtentaculata samples group. *D. subtentaculata* was included because despite their great number of asexual populations, the ladder-like pattern has not been seen in this clade ever.

We analyzed the WM_717 SC set. The average of GC3% among all genes by sample was low and not much different between sexual and asexual samples (Fig. S2). In the case of Iberia-Appenines-Alps-plus clade, non-noticeable differentiation is shown when the standard deviation is plotted against the GC3% average. On the other hand, *D. vilafarrei* (sexual) shows a slight differentiation from the *D. subtentaculata,* whose sexual and asexual populations appear mixed. However, this graph has to be analyzed with caution, since the scale is very large and, in fact, the values are very close.

3.2.2. Dunuc12 validation

Using the transcriptomic data, we validated Dunuc12 as SC gene (Text S1). An exonic region of 120 positions (Dunuc12-exo) was identified in every cloned sequence. All the Dunuc12-exo from the same individual, matched against the same transcript in its individual transcriptome assembly. In addition, all identified transcripts (one by individual) mapped against sequences from *Dugesia* identified as TMED9 in the nucleotide database of NCBI. We analysed two groups of SC used here, finding that all the transcripts formed only one orthogroup; OG0007336 in WM_717, and OG0010797 in etru-ligu_1984. Although neither was defined as SC by OrthoFinder (see Supporting Information Text S1 for a broader explanation), the ML trees obtained from OG0007336 and OG0010797 showed the topology of a region that has not suffered any duplication event in the evolutionary history of the group.

3.2.3. Intraindividual genetic diversity

From the eighth selected individuals a total of 89 sequences of Dunuc12 were obtained (Table S2). The length of the sequences was 690bp, with a total of 138 variable sites. After the correction of the sequences eliminating the putative amplification errors, the analysis in DnaSP software revealed a total of 35 different haplotypes. For the two sexual individuals selected, one from Pieve (*D. etrusca*) and other from La Garda (*D. liguriensis*), 6 and 4 different haplotypes were found respectively, while for the 6 fissiparous individuals from Iberian Peninsula the number of intraindividual haplotypes ranged between 3 and 8. All analyzed individuals, except DspTrilla_3, presented a higher number of haplotypes than its ploidy would suggest, an indication of mosaicism (Table S2, S3).

Intraindividual haplotype networks of the asexual individuals (Fig. S3A) showed a pattern in which the haplotypes were highly divergent from one another with similar frequencies. However, the sexual individuals (Fig. S3B) showed a star-like pattern, with a predominant haplotype in the center and some closely related haplotypes at a distance of one or few mutations surrounding it.

The haplotype network with all cloned sequences (Fig. S3C) shows that sexual individuals do not share any haplotype with any other population. However, fissiparous individuals shared haplotypes among populations, forming clusters of closely related

haplotypes, but those clusters are diverging among them, a clear Meselson-effect pattern.

Regarding the intraindividual genetic diversity parameters; the nucleotide diversity was higher in asexual individuals of etrusca-liguriensis group (Fig. 4A, Table S3). However, the haplotype diversity (Hd) was higher in *D. subtentaculata* individuals than in the etrusca-liguriensis ones (Fig. 4B, Table S3).

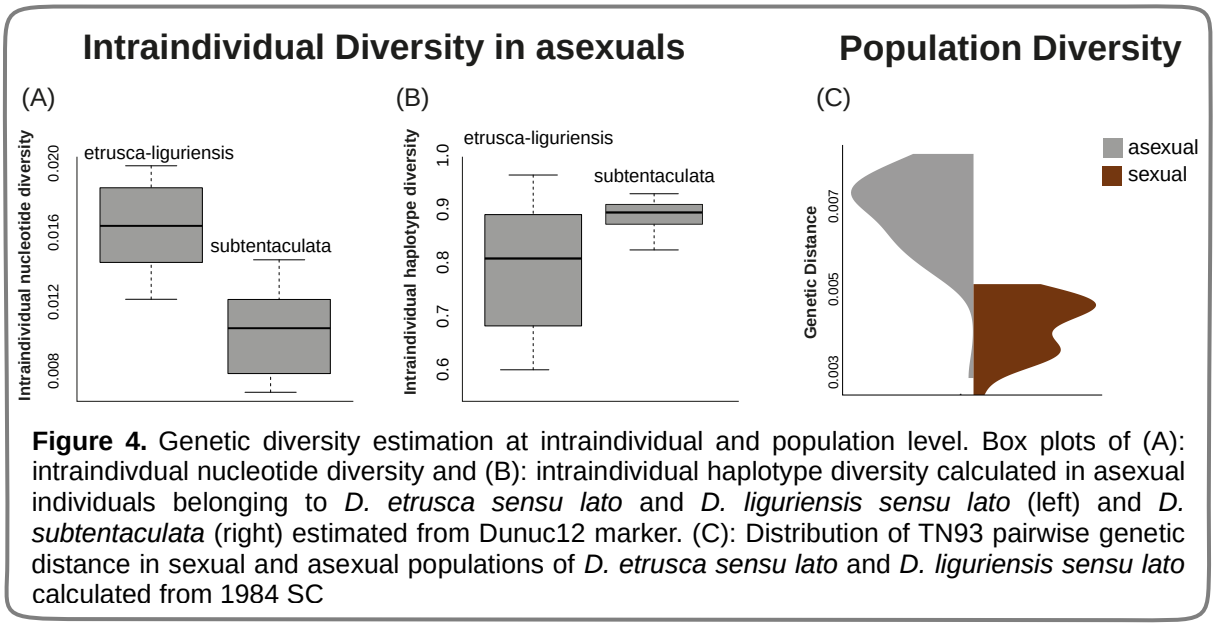


Figure 4. Genetic diversity estimation at intraindividual and population level. Box plots of (A): intraindivdual nucleotide diversity and (B): intraindividual haplotype diversity calculated in asexual individuals belonging to *D. etrusca sensu lato* and *D. liguriensis sensu lato* (left) and *D. subtentaculata* (right) estimated from Dunuc12 marker. (C): Distribution of TN93 pairwise genetic distance in sexual and asexual populations of *D. etrusca sensu lato* and *D. liguriensis sensu lato* calculated from 1984 SC

Surprisingly, the proportion of synonymous mutations (Ks) in the exonic region (120 pb) was nule for all individuals belonging to the etrusca-liguriensis group, except for DspBerga_3 (Ks=0.0107), and only individuals from Berga showed non-synonymous mutations (Ka>0). With this data was impossible to calculate the Ka/Ks ratio and obtain information regarding the Muller's ratchet effect. On the other hand, no signals of this effect were observed in the subtentaculata lineage, since Ka values were lower than Ks (Table S3).

3.2.4. Genetic distance between individuals

A genetic distance matrix between all sexual individuals from Alps and Apennines regions, and asexual individuals from Iberian Peninsula was built using 1984 SC and the TN93 model. The genetic pair-wise distances between asexual individuals from Berga and Font de la Trilla are higher than the genetic pair-wise distances between individuals from sexual population (Fig. 4C; Fig. S4).

3.3. Effect of asexual taxa in species tree inference

We obtained the per branch quartet support and the branch support (local posterior probability) in the species tree obtained with the MSC method from datasets 1-3. Figure 5 shows the quartet support and local posterior probability for the three possible topologies around the internal branches of the Iberia-Apennines-Alps-plus clade. The red arrows indicate the most controversial nodes in the phylogeny. The values of quartet support indicate the amount of gene tree conflict around a branch. Values close to 0.33 indicate a very high level of discordance around the branch (Fig. 5).

In the case of the tree obtained from dataset 1 (Fig. 5A), two discordant nodes are shown; the node that joins samples from Font de la Trilla and Berga (nodes 49 and 60 in Supplementary Figure S5A). In addition, the final normalized quartet support of this tree is low (FNQS=0.71) and the DensiTree is very blurred.

When the samples from Berga were removed, the nodes around the samples from Font de la Trilla continued showing discordance (Fig. 5B; Fig. S5B). The individual trees show a less burred pattern and the FNQS is slightly greater (0.77).

From the dataset 3, a different topology was obtained (Fig. 5C). In this analysis were eliminated all asexual samples from the Iberian Peninsula. Unlike datasets 1 and 2, in this case *D. ilvana* and *D. etrusca s.s* form a monophyletic group. None discordant-nodes are shown in this tree. The quartet supports and the branch supports on the internal branches are high (Fig. S5C). Furthermore, the individual trees show a more defined pattern and the FNQS is higher than in the previous topologies (0.82). In addition, the ML tree obtained from this dataset, shows the same topology and high bootstrap values (Fig. S6, Table S4).

3.3.1. Topology test

Taking into account the discordant results regarding the position of *D. ilvana*, we tested the topologies obtained with dataset 3 [Tree 1 (*D. gonocephala*, (*D. liguriensis sensu stricto*, (*D. ilvana*, *D. etrusca sensu stricto*)));] and the previously ones [Tree 2 (*D. gonocephala*, (*D. ilvana,* (*D. liguriensis sensu stricto*, *D. etrusca sensu stricto*)));]. All tests showed a significantly better adjustment to the data than to the tree 1 (Table 2), supporting the topology where *D. ilvana* is sister group of *D. etrusca sensu stricto* (Fig. 3C, Fig. S4) when the Iberian populations were removed from the analyses.

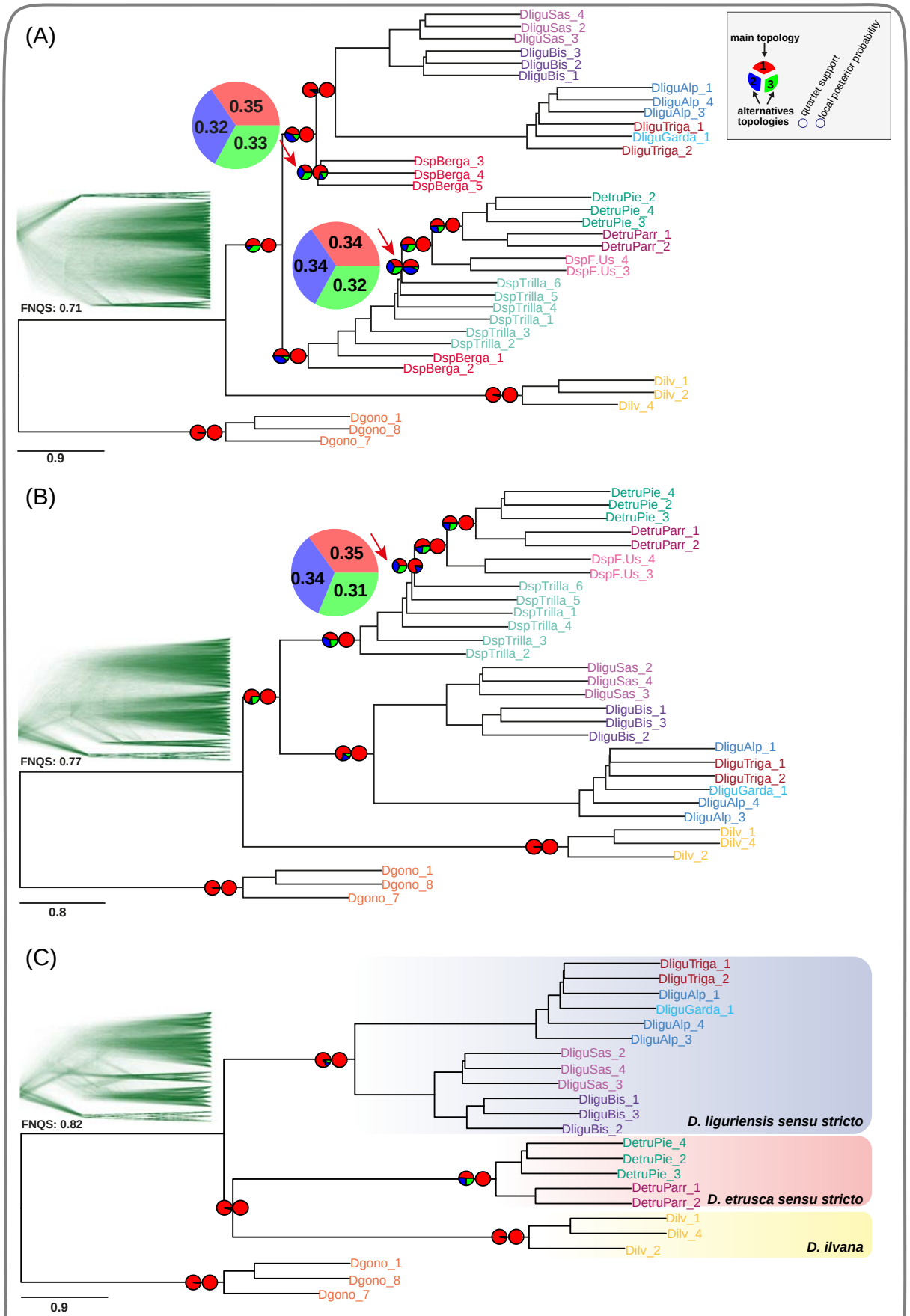


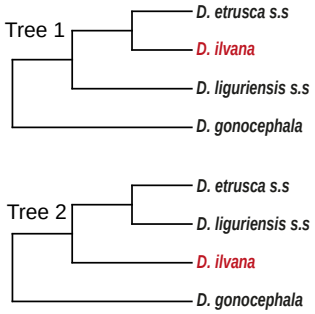
Figure 5. Species trees obtained using Multispecies Colaescent Model (MSC) from three data sets of single copy orthologs and different sample composition. *The figure caption continues in the next page*

Figure 5 cont. (A): MSC tree from data set 1 containing all samples belonging to the Iberian-Apennines-Alps-plus clade. (B): MSC tree from dataset 2 excluding the samples from Berga locality. (C): MSC tree from dataset 3 excluding all samples from the Iberian Peninsula. The per branch quartet support and the branch support (local posterior probability) values for the main and the alternative topologies are shown on the branch in pie charts. The red arrows indicate discordant topologies around a branch and the big pie charts the corresponding quartet support values. Scale bar: Coalescent Units (MCM). FNQS: Final normalized quartet score. See Supplementary Table S4 and Figure S5 for more information about the support values of the branches. The individual trees are represented in DensiTree schemes in green.

Table 2. Results of topology test between Tree 1 and Tree 2, using the Dataset 3.

Tree	LogL	deltaL	Bp-RELL	P-KH	P-SH	C-ELW	P-AU
1	-4466502,48	0	1+	1+	1+	1+	1+
2	-4466713,34	210,86	0-	0-	0-	-1,97E-42	-1,41E-61

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
c-ELW: Expected Likelihood Weight
p-AU: p-value of approximately unbiased (AU) test
Plus signs denote the 95% confidence sets.
Minus signs denote significant exclusion.



4. Discussion

Phylogenies describe the diversification process through bifurcated branches. Thus, dichotomous nodes with high support values are always the expected topology to validate any phylogenetic relationship. In the same way, monophyly constitutes a qualitative measure to validate taxa, and it is expected that taxonomic categories always correspond with monophyletic groups. When this pattern disappears, the first doubts hover around the methodologies used and the systematic errors associated with them.

Here, we studied an unexpected topological pattern associated to the inclusion of asexual individuals in phylogenetic analyses. In addition, we analysed how the genetic information carried by asexual taxa affect the tree inference.

4.1. The ladder-like topology is not an artifact of phylogenetic reconstruction

The species *D. etrusca s.s* and *D. liguriensis s.s* were described for the first time based on morphological characters of the reproductive apparatus, just as is common in freshwater planarians (Benazzi, 1946; De Vries, 1988). Asexual populations from the Iberian Peninsula have been reported without being possible its accurate identification to one species or another using specific molecular markers (Baguñà et al., 1999; Lázaro et al., 2009). Using transcriptomic data, a larger sampling size, and different inference methods were obtained odd topologies in the phylogenies (Chapter IV.1). The trees obtained using ML approach showed an unexpected ladder-like pattern of asexual individuals and a paraphyletic topology for *D. liguriensis s.l* (Fig. S1A). When asexual representatives from Berga population were removed from the dataset, ML tree recovered a monophyletic topology for the two species, but asexual samples grouping with *D. etrusca s.s* continued showing the ladder-like topology (Fig. S1C). On the other hand, the MSC method recovered two monophyletic groups, but it cast doubt on the assignment of Berga's samples, since different samples were placed with different species (Fig. S1B and D). Additionally, all trees showed low support values on the nodes joining the asexual samples from Font de la Trilla and Berga.

Orthologs selection can lead to incorrect topology inference, which is one of the most sensible steps in phylogenomics (Fernández et al., 2020; Kapli, Yang, & Telford, 2020). In Chapter IV.1 was used OrthoFinder (Emms & Kelly, 2019) software for ortholog search from transcriptomic data. This method infers orthogroups across multiple species based on pairwise sequence similarity between all sequences. It defines an orthogroup as the set of genes descended from a single gene in the last common ancestor of all species under consideration (Emms & Kelly, 2015). OrthoFinder currently constitutes one of the most accurate methods for orthology inference (Deutekom, Snel, & Van Dam, 2021; Emms & Kelly, 2019). To corroborate that the ladder-like pattern observed in the previous study is not a product of ortholog error, we conducted a new ortholog search using OMA, which identifies the orthologs using a reference database with high performance (Altenhoff, Schneider, Gonnet, & Dessimoz, 2011). With OMA we recovered 2,570 orthologs, several more than with OrthoFinder, a situation that has been reported in other studies comparing both software (Altenhoff et al., 2019). The ML trees obtained with the SC discovered with OMA recovered the same ladder-like topology of asexual individuals with both protein

and nucleotide sequences (Fig. 2). Also, the paraphyly of *D. liguriensis* was recovered with nucleotide data (Fig. 2B). These results corroborate the topologies obtained with the SC discovered by OrthoFinder, indicating that the ladder-like pattern is not an artifact of ortholog inference.

Additionally, it would be logical to think that an artifact produced by an incorrect orthologs estimation would affect the whole tree, showing anomalous topologies in all groups. However, this pattern has only been shown by asexual individuals belonging to the Iberia-Appenines-Alps-plus clade (Chapter IV.1).

For a second validation, we took advantage of the approaches developed to extract phylogenetic information from paralog genes. The use of paralogs with adequate approaches, specially MSC methods, have been expanded in phylogenomics as another source of phylogenetic information since obtaining single copy gene sets can become more difficult when more divergent species are included in the analyses (Hellmuth et al., 2015; Yan, Smith, Du, Hahn, & Nakhleh, 2022). The species trees inferred including the set of paralogs showed the same ladder-like topology, with a consequent decrease of support due to the use of paralogs (Fig. 3), demonstrating that this topology is as real as ancient and can be recovered even from paralog information.

In conclusion, we have demonstrated that although the ladder-like topology is weird, it is not an artifact of the methodology used. Otherwise, we propose that this topology represents the phylogenetic footprint of ancient fissiparity in these populations, which have been for a long time without recombination.

4.2. Ancient asexuality in the Iberian-Apennines-Alps-plus clade

Using different methodologies, we have shown that the ladder-like pattern is not an effect of systematic errors but the consequence of a real underlying biological process. We suspect that this pattern can be reflecting the effect of ancient asexuality in Iberian populations that belong to the Iberian-Apennines-Alps-plus clade.

Because this pattern was more obvious in the populations from Font de la Trilla and Berga, with more individuals sampled, we only refer to these populations and seem that we leave the population from Font de l'Us in the backstage. The issue with this population is that only two individuals were sampled, and they group with *D. etrusca* with high support values. In this scenario, it is impossible to know if this population has

a “standard” monophyletic grouping, or it can show the ladder-like pattern if more individuals are included in the analyses. If from Font de la Trilla and from Berga, we had only analysed individuals DspTrilla_6, DspTrilla_5, DspBerga_3, and DspBerga_4 maybe this debate wasn't happening since they constitute monophyletic groups per population (Fig. S1A-D). Moreover, the fact is that Berga and Font de la Trilla revealed this unexpected topology, even in the trees done with *Cytochrome Oxidase I* for an initial species assignment (Chapter IV.1).

We hypothesize that contrary to *D. subtentaculata* lineage, in which asexual populations have never shown a ladder-like pattern, the Iberian-Appenies-Alps-plus clade has been under long-term asexuality without recombination events. (Leria, Riutort, Romero, Ferrer, & Vila-Farré, *in press*) dates the diversification of Iberian-Appenies-Alps-plus clade around 17 Mya and the *D. subtentaculata* diversification in Iberian Peninsula approximately 1.69 Mya. In Chapter IV.1 it was proposed that the ancestor of *D. etrusca-D. Liguriensis* lineage could be also asexual and colonized the Iberian Peninsula through fauna corridors in the Eastern Pyrenees, early in its diversification process along the Western Mediterranean. In addition, Leria et al., (2019) propose the shift between sexual and asexual reproduction as a strategy during *D. subtentaculata* evolution.

Here we used several approaches to confirm the ancient asexuality of Berga and Font de la Trilla populations with respect to the *D. subtentaculata* lineage. First, we used the GC content in the third codon position as a recombination measure (Fig. S2). Under higher recombination rates, it is expected an increase of GC content due to the gbGC (Webster and Hurst 2012). This calculation has been used as measure of the lack of recombination in ancient asexual lineages in *Timema* stick insects (Bast et al., 2018). However, here we did not detect a considerable change between sexual and asexual individuals. The low content of GC in freshwater planarians transcriptomes has been reported before (Abril et al., 2010; Nelson Hall et al., 2021; Resch, Palakodeti, Lu, Horowitz, & Graveley, 2012). Nevertheless, although it is suspected that the high variation of GC content in Platyhelminthes can be related to some adaptive trait (Lamolle, Fontenla, Rijo, Tort, & Smircich, 2019), there is no clear explanation for this. We hypothesize that a high AT content in planarian genomes can favour the active DNA replication processes that take place in these organisms regarding the cellular turnover and regeneration, but further studies are necessary to

corroborate it. In consequence, planarians do not show the expected shift in GC content related to recombination events.

Second, we used intraindividual diversity parameters to compare the effect of asexuality in the Iberian-Appenies-Alps-plus clade and in the *D. subtentaculata* lineage. For that, we validated the Dunuc12 marker as a good marker for intraindividual genetic diversity studies. Although OrthoFinder does not classify Dunuc12 as SC, we can corroborate that Dunuc12 is a single copy gene in the *Dugesia* genome. First, all sequences are grouped only in one orthogroup in two ortholog searches; and second, the trees did not show a paralog topology (Text S1). Additionally, these trees corroborated that the coding region of this marker is not informative, and all variability is found in the intron.

Then, we analysed the Meselson effect as a measure of long-term asexuality. We obtained a higher intraindividual nucleotide diversity in asexual populations belonging to Iberian-Apennines-Alps-plus clade than in asexual populations belonging to *D. subtentaculata* lineage (Fig. 4A). This result supports the premise that Berga and Font de La Trilla populations have been evolving without recombination for a longer period of time than the asexual populations of *D. subtentaculata*.

Although these signals also can be found in asexual species originated from recent hybridization (Lunt, 2008), here we discard this option since the hybrid origin of asexual Iberian populations is biogeographically unfeasible, and a facultative sexual ancestor of sexual and asexual lineages is a most plausible hypothesis (Chapter IV.1).

In agametic clonal organisms it is common to find genetic mosaicism due to the inheritance of somatic mutations from the parents (Gill, Chao Lin, Perkins, & Wolf, 1995). Leria et al., (2019) described the Mosaic-Meselson-effect in *D. subtentaculata* by the first time, as a consequence of asexuality that combine the Meselson effect with high mosaicism levels in agametic organisms.

Here, we also found that asexual populations of *D. etrusca s.l* and *D. liguriensis s.l* show high levels of intraindividual haplotype diversity (Fig. 4B, Table S3), indicating that the Mosaic-Meselson effect is also occurring in these lineages. The slightly lower values of this parameter found in *D. etrusca-D.liguriensis s.l* compared to *D. subtentaculata*, likely result from a higher number of intraindividual sequences obtained in the last mentioned species (Leria et al., 2019).

Another way to show the Meselson effect is analyzing the haplotype networks. Our haplotype networks showed the typical Meselson effect pattern and are similar to those shown by Leria and collaborators for sexual and asexual individuals. Highly divergent haplotypes in asexual individuals (Fig. S3A) and star-like pattern in sexual ones (Fig. S3B). Therefore, the accentuated Meselson effect found in Berga and Font de La Trilla populations is a reliable proof of the oldest asexuality of these lineages and could explain why asexual *D. subtentaculata* populations do not show the ladder-like pattern.

Finally, our last way to test the ancient asexuality of the Iberian populations of *D. etrusca* and *D. liguriensis* was to investigate the putative accumulation of deleterious mutations in these populations (i.e., the putative existence of Müller's Ratchet). To do so, we analyzed the proportion of non-synonymous and synonymous mutations (Ka/Ks) in the same coding region of Dunuc12 in these populations and compared it with the asexual populations of *D. subtentaculata*. Surprisingly, we found that the coding region of Dunuc12 was extremely conserved in the asexual populations of *D. etrusca s.l* and *D. liguriensis s.l*, hindering the calculation of this parameter in most individuals. Nevertheless, the very few mutations found in these populations were indeed in most cases non-synonymous (Table S3), pointing to a potential existence of Müller Ratchet.

4.3. The ladder-like topology; a new genomic footprint of ancient fissiparity

Basing on the aforementioned results, here we propose the theoretic framework leading to the ladder-like pylogenetic footprint caused by ancient fissiparous reproduction. The footprint of asexuality in the evolutionary history of lineages lies on the biological characteristics of the organisms. The Meselson effect is the consequence of the lack of recombination. Under this situation the alleles in the population are not homogenized, leading to highly differentiated alleles inside the same individual, that could be more related with homologous alleles from other populations (Birky, 1996; Schwander et al., 2011; Welch & Meselson, 2000) (Fig. 6A). However, in lineages that combine both reproductive strategies, the pattern shown by sexual and asexuals is different. In this case, the intraindividual divergence (ID) of alleles in sexuals will be shorter than the ID of asexual ones, which coalesce back in time (Fig. 6B).

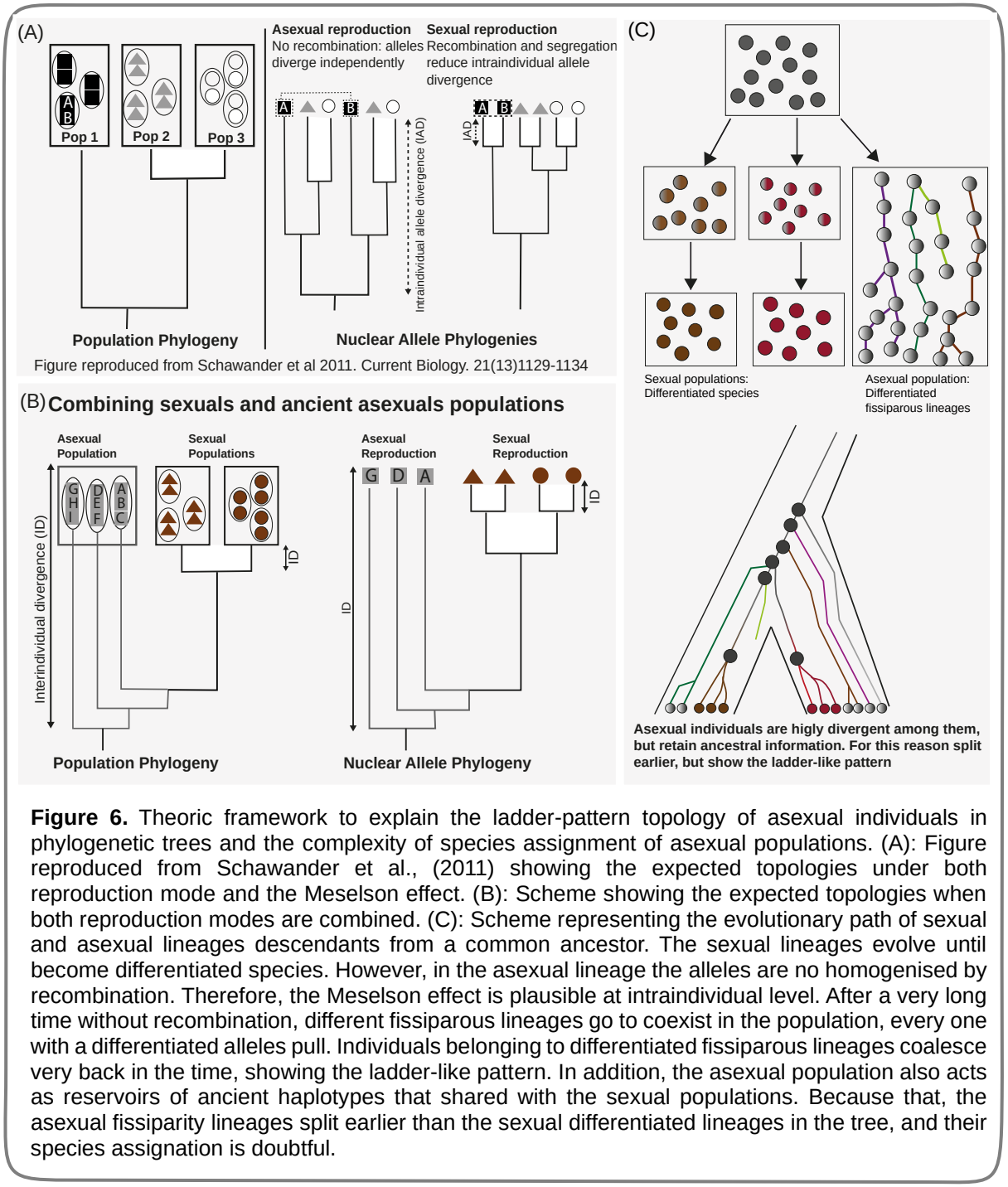


Figure 6. Theoric framework to explain the ladder-pattern topology of asexual individuals in phylogenetic trees and the complexity of species assignment of asexual populations. (A): Figure reproduced from Schawander et al., (2011) showing the expected topologies under both reproduction mode and the Meselson effect. (B): Scheme showing the expected topologies when both reproduction modes are combined. (C): Scheme representing the evolutionary path of sexual and asexual lineages descendants from a common ancestor. The sexual lineages evolve until become differentiated species. However, in the asexual lineage the alleles are no homogenised by recombination. Therefore, the Meselson effect is plausible at intraindividual level. After a very long time without recombination, different fissiparous lineages go to coexist in the population, every one with a differentiated alleles pull. Individuals belonging to differentiated fissiparous lineages coalesce very back in the time, showing the ladder-like pattern. In addition, the asexual population also acts as reservoirs of ancient haplotypes that shared with the sexual populations. Because that, the asexual fissiparity lineages split earlier than the sexual differentiated lineages in the tree, and their species assignation is doubtful.

In the case of planarians, the fissiparity implies the expansion of clonal lineages in the population through time. After long evolutionary time without recombination, every clonal lineage will be differentiated, drawing their own genetic diversity. To demonstrate this effect under long term fissiparity, we calculated the pair-wise distances based on transciptomic data in sexual populations of *D. etrusca* and *D. liguriensis* and asexual population from Berga and Font de La Trilla. The genetic

distances, calculated from 1984 SC, between asexual individuals were higher than between sexual individuals (Fig. 4C; Fig. S4), supporting that under long term asexuality the recombination does not homogenize the pool of alleles within individuals, and neither within the population, leading to the existence of independent lineages of fissiparity.

Under the scenario where sexual and fissiparous lineages evolve from the same ancestor, the species tree will show the footprint of both reproductive strategies (Fig. 6C). While sexual lineages evolve becoming differentiated species, in asexual lineages the lack of recombination produces Meselson effect at an intraindividual level. Under longer asexuality without genetic crossing, distant fissiparous lineages accumulate high amounts of divergence between them. Therefore, individuals belonging to ancient divergent fissiparous lineages coalesce very back in the time. Then, the phylogenetic trees show the descendants of divergent ancient lineages in a ladder-like pattern. The probability to find representatives of these lineages will depend on the population size. Thus, in a random sampling it is possible to find also related individuals belonging to the same lineage, whose coalescence point is more recent in the time (Fig. 6C).

Additionally, these asexual “independent lineages” split earlier than the sexual differentiated lineages in the tree because the asexual populations also act as reservoir of ancestral polymorphisms, attracting the related sexual lineages. This ancestral relationship, conserved by asexuality masks the diversification process of the group and made unachievable the species assignation of asexual populations, since asexual lineages coalesce with those sexual ones far back in time, when the sexual lineages were not differentiated yet.

Systematics of asexual taxa constitute a challenge for taxonomist specialists (Fehrer, Krak, & Chrtek, 2009; Hörandl, 2018; Martens, Rossetti, & Baltanás, 1998), and the taxonomy and systematics of asexual lineages have been belittled due to the evolutionary constrains of asexuality (Fontaneto et al., 2007). Following the definition that establishes the species as separately evolving metapopulation lineages (De Queiroz, 2007), it could be plausible to describe the asexual populations from the Iberian Peninsula as different species. However, this group would not meet the monophyly principle to be considered a clade because the individuals are grouped in a ladder-like topology. In the case of the Berga population the species assignment is

more confused, due to the grouping with both species. Therefore, we decide to continue using *sensu estricto* as reference to sexual populations and *sensu lato* as reference to the two lineages including asexual populations. Since no morphological characters are available for systematic studies in asexual freshwater planarians, and the genetic diversity of asexual populations is shaped by the fissiparity, new theoretical frameworks are necessary to include the diversity of reproductive strategies in our taxonomic system of classification.

4.4. Ancient asexual lineages confound phylogenetic inference. A new challenge for evolutionary studies in freshwater planarians

The inclusion of asexual individuals in phylogenetic analyses also affect the relationship with closer related groups. Figure 5 shows the analysis of different datasets and how the support of species trees increases with the elimination of asexual individuals. Full dataset 1 (Fig. 5A) shows the lowest FNQS, indicating highest discordance among gene trees. Unexpectedly, when all asexual representatives are excluded, the tree shows *D. ilvana* as the sister group of *D. etrusca s.s,* instead of being sister to the clade formed by *D. etrusca* s.s and *D. ligurienses s.s* (Fig. 5C). This topology was not recovered in any analysis performed in Chapter IV, which always included representative samples of the asexual lineages in the datasets. However, Lázaro et al (2009), recovered a low supported relationship between *D. etrusca* and *D. ilvana*. Moreover, this topology is highly supported here by the MSC method (Fig. 5C; Fig. S5 C), ML method (Fig. S6) and topology tests (Table 2).

These results point to new biogeographic hypotheses to explain the diversification of the Iberia-Apennines-Alps-plus clade. Based on the results obtained in Chapter IV was proposed that the ancestor of *D. ilvana* split first from *D. etrusca-D. liguriensis* ancestor. However, they already pointed that the timing for this split based on the dating analysis in Leria et al (*in press*) did not support such an ancient split. Additionally, taking into account that Elba Island was formed by Apennines outcrops (Bortolotti, Fazzuoli, et al., 2001; Bortolotti, Pandeli, & Principi, 2001), it is more plausible that *D. ilvana* shared a common ancestor with the Apennines species *D. etrusca s.s*. Biogeographical studies have demonstrated that Elba Island was connected with Tuscany by a natural land bridge during the Pleistocene (Dapporto & Cini, 2007; Di Nicola & Vaccaro, 2020; Fattorini, 2009). Therefore, several animal

groups in the Island show a phylogenetic relationship with the Tuscan archipelago fauna (Fattorini, 2009 and references therein; Di Nicola & Vaccaro, 2020; Dapporto & Cini, 2007).

We suspect that the inclusion of the asexual lineages introduced ancestral polymorphisms maintained in the different fissiparous lineages shared by *D. liguriensis* and *D. etrusca*, reinforcing the ancestral relationship against the more recent splitting with *D. ilvana*. Ancestral polymorphisms can lead to incongruent phylogenetic hypothesis (Moran & Kornfield, 1993; MacEachern, McEwan, & Goddard, 2009). Moreover, the effect of ancestral shared polymorphisms in phylogeographic reconstructions based on molecular data has been demonstrated in several taxa (Boluda, Rico, Naciri, Hawksworth, & Scheidegger, 2021; Bowie, Fjeldså, Hackett, Bates, & Crowe, 2006; Du et al., 2019; B. Wang et al., 2019). Therefore, ancient asexual populations acting as reservoirs of ancestral polymorphisms will introduce noisy information that the phylogenetic inference methods incorporate as support of alternative phylogenetic hypotheses, which do not reflect the real species tree.

Summarizing, here we demonstrated that ancient asexuality in agametic organism can lead to a ladder-like topology showing the descendants of differentiated clonal lines as independent lineages that split earlier than the sexual clades in the tree. In addition, the asexual populations retain ancestral polymorphisms shared by the sexual ones, reinforcing the ancestral relationship against more recent splittings. Since the reproductive strategies shape the genetic diversity of populations and lead to differentiated genetic backgrounds between sexual and asexual lineages, the discordance between gene trees and species trees can increase and affect the phylogenetic inference. Therefore, we recommend the use of MSC methods against concatenation methods for species tree estimation with large datasets, especially in cases with putative ancient asexuals. These methods offer a framework based on Coalescent Theory that reconstruct in a better way the ancestral relationship between sexual and asexual lineages. Additionally, is highly recommendable to test different phylogenetic hypothesis with different datasets to study the effect of the inclusion of asexual representatives on the trees. Unfortunately, the inference methods and evolutionary models are designed based on sexual reproduction premises, and do not contemplate the effects of the lack of recombination or different ploidies. New

frameworks including these issues are necessary to reconstruct the evolutionary history of asexual lineages in the Tree of Life.

Data accessibility

RNAseq data is accessible under the BioProject accession code PRJNA797284 at the NCBI SRA database and the cloned sequences of Dunuc12 are accessible at the GenBank nucleotide database. Also, the scripts used are accessible at https://github.com/lisy87/dugesia-transcriptome

Acknowledgments

We are grateful to Marina Pérez for the design of Dunuc12 primers, specifics for *D. etrusca* and *D. liguriensis s.l.* We thank Joan Ferrer Obiol and Julio Rozas for their valuable suggestions. We also thank our lab manager Gema Blasco for her support in all our research projects.

Authors Contributions

LBA and MR did the initial study design. LBA processed and analysed the data with the input from RF. LL and IT performed Dunuc12 cloning and analyses. LBA wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Funding

This research was supported by the Ministerio de Economía y Competitividad (project PGC2018-093924-B-100) from Spain. This research was also supported by a FI Grant from the Generalitat de Catalunya (to Lisandra Benítez-Álvarez).

Supplementary Material

Supplementary tables S1, S2, and S4, are available at Riutort´s Lab site

Table S3. Intraindividual diversity parameters calculated in individuals belonging to two different asexual lineages. Haplotype and nucleotide diversity were calculated from a Dunnuc12 region of 691 positions in subtentaculata lineage and 690 positions in etrusca-liguriensis lineage. The proportions of non-synonymous and synonymous mutations were calculated from the same coding region of 120 positions in both lineages.

Individual	Lineage	Haplotype Diversity (Hd)	Nucleotide Diversity (Pi)	Proportion of Non-synonymous mutations (Ka)	Proportion of Synonymous mutations (Ks)	Ka/Ks
Estella1A	subtentaculata	0,909	0,0086	0,0018	0,0281	0,0624
Estella2A	subtentaculata	0,872	0,0069	0,0016	0,0132	0,1225
Estella3	subtentaculata	0,838	0,0077	0,0028	0,0230	0,1213
SantaFe1A	subtentaculata	0,894	0,0129	0,0008	0,0458	0,0164
SantaFe2A	subtentaculata	0,885	0,0117	0,0000	0,0453	0,0000
SantaFe3A	subtentaculata	0,929	0,0123	0,0000	0,0378	0,0000
Truchas1A	subtentaculata	0,901	0,0106	0,0030	0,0346	0,0871
Truchas2	subtentaculata	1	0,0146	0,0018	0,0496	0,0353
Truchas3	subtentaculata	0,824	0,0080	0,0015	0,0188	0,0799
Berga_2	etrusca-liguriensis	0,6820	0,0165	0,0018	0,0000	NA
Berga_3	etrusca-liguriensis	0,9640	0,0221	0,0053	0,0107	0,4907
Berga_5	etrusca-liguriensis	0,8900	0,0190	0,0015	0,0000	NA
Trilla_3	etrusca-liguriensis	0,6000	0,0173	0,0000	0,0000	NA
Trilla_5	etrusca-liguriensis	0,7270	0,0152	0,0000	0,0000	NA
Trilla_6	etrusca-liguriensis	0,8890	0,0189	0,0000	0,0000	NA

Supplementary Figures

Figure S1. Phylogenetic trees obtained in Chapter IV.1 using Maximum Likelihood (ML) and Multispecies Colaescent (MSC) model from two data sets of single copy orthologs. Dataset 1 contains all samples belonging to the Iberian-Apennines-Alps-plus clade. Dataset 2 exclude the samples from Berga locality. (A) ML tree from dataset 1. (B) MSC tree from data et 1. (C) ML tree from dataset 2. (D) MSC tree from dataset 2. The bootstrap values (ML-bv) and the branch support (MSC-bs) values are shown on the nodes in a color scale from minimum (white) to maximum (black). Scale bar: substitutions per site (ML) / Coalescent Units (MSC). FNQS: Final normalized quartet score. The arrows indicate the controversial points in the phylogeny.

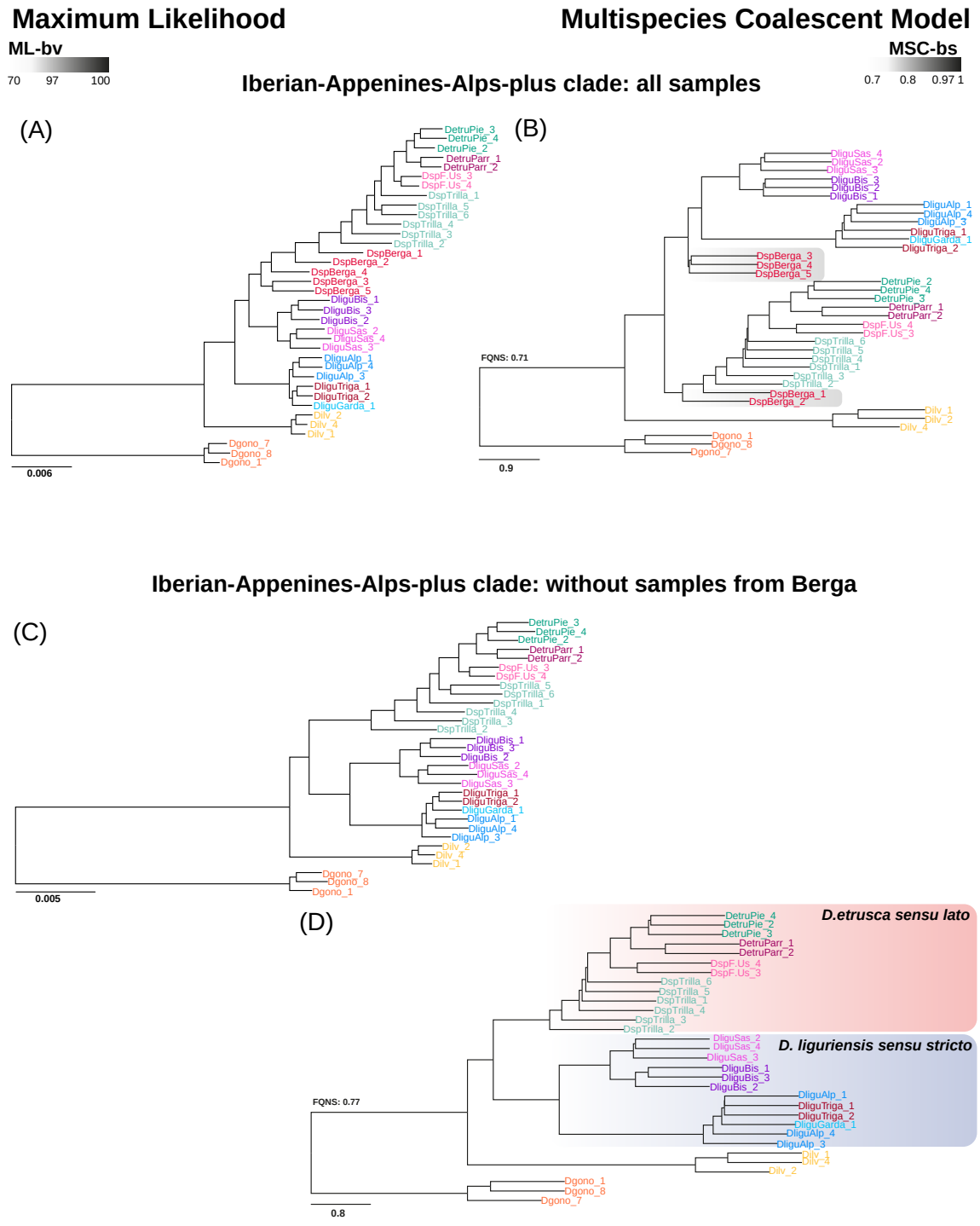


Figure S2. Average GC content in third codon positions (GC3%) and standard deviation calculated in 717 single copy orthologs of sexual (red) and asexual (green) individuals belonging to (A): Iberia-Apennines-Alps-plus clade and (B) *D. subtentaculata* and *D. vilafarrei* species.

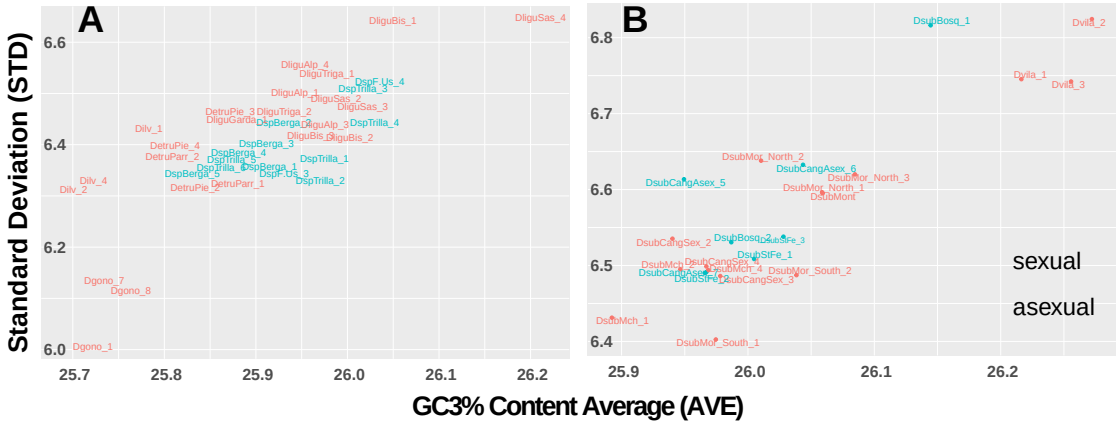


Figure S3. Haplotype Network of Dunuc12 marker. (A): Asexual individuals. (B): Sexual individuals. (C): Total haplotype network including all individuals. Each individual is indicated with a different color. Each circle represents a different haplotype and its size indicate the frequence of the haplotype in the individual. Branch lengths are proportional to the number of mutations.

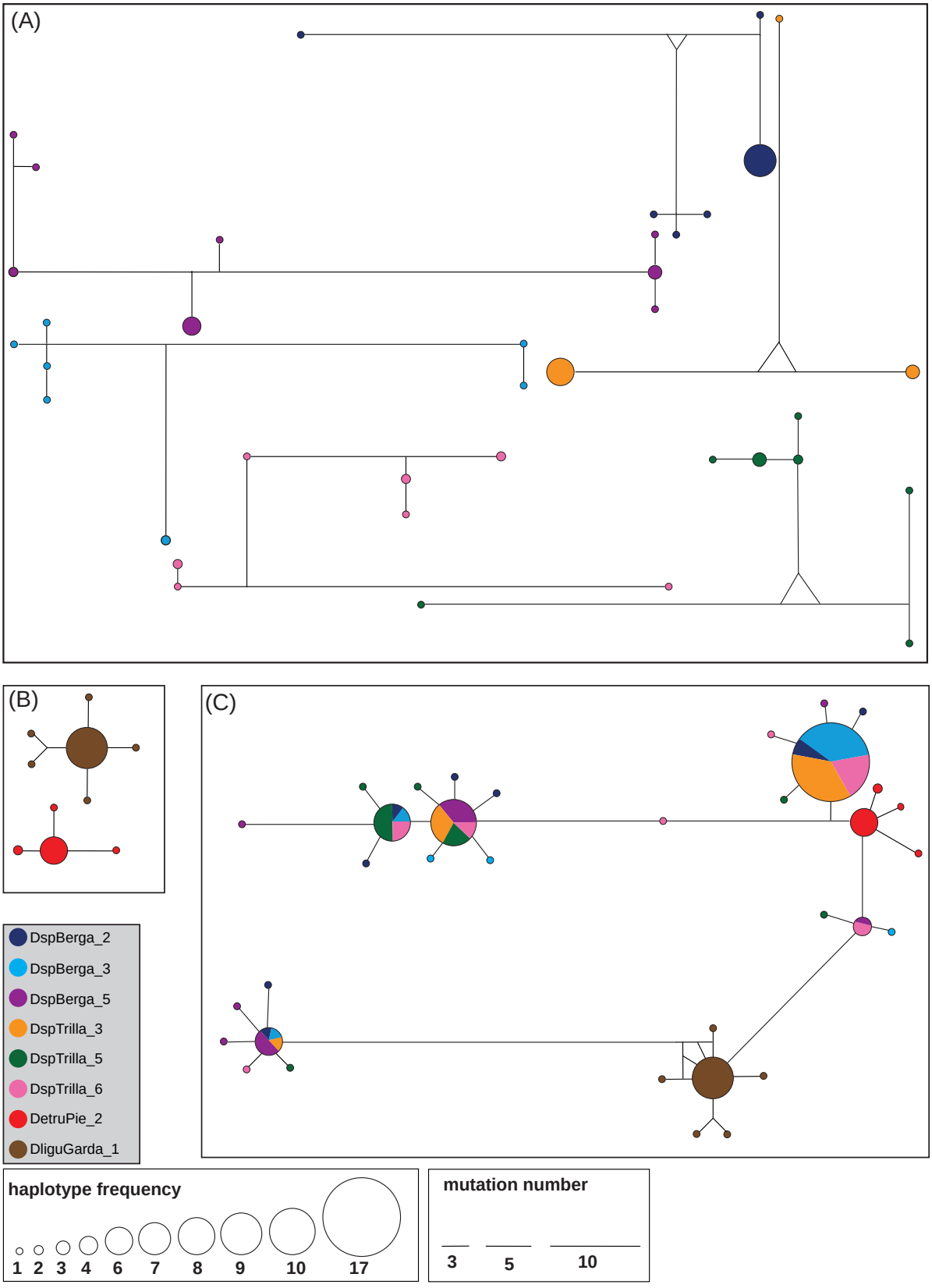


Figure S4. Pairwise distance matrix obtained from dataset 1 (1984 SC and all etrusca-liguriensis samples group) using the TN93 model. The color represent the distance values from lower (white) to higher (red).

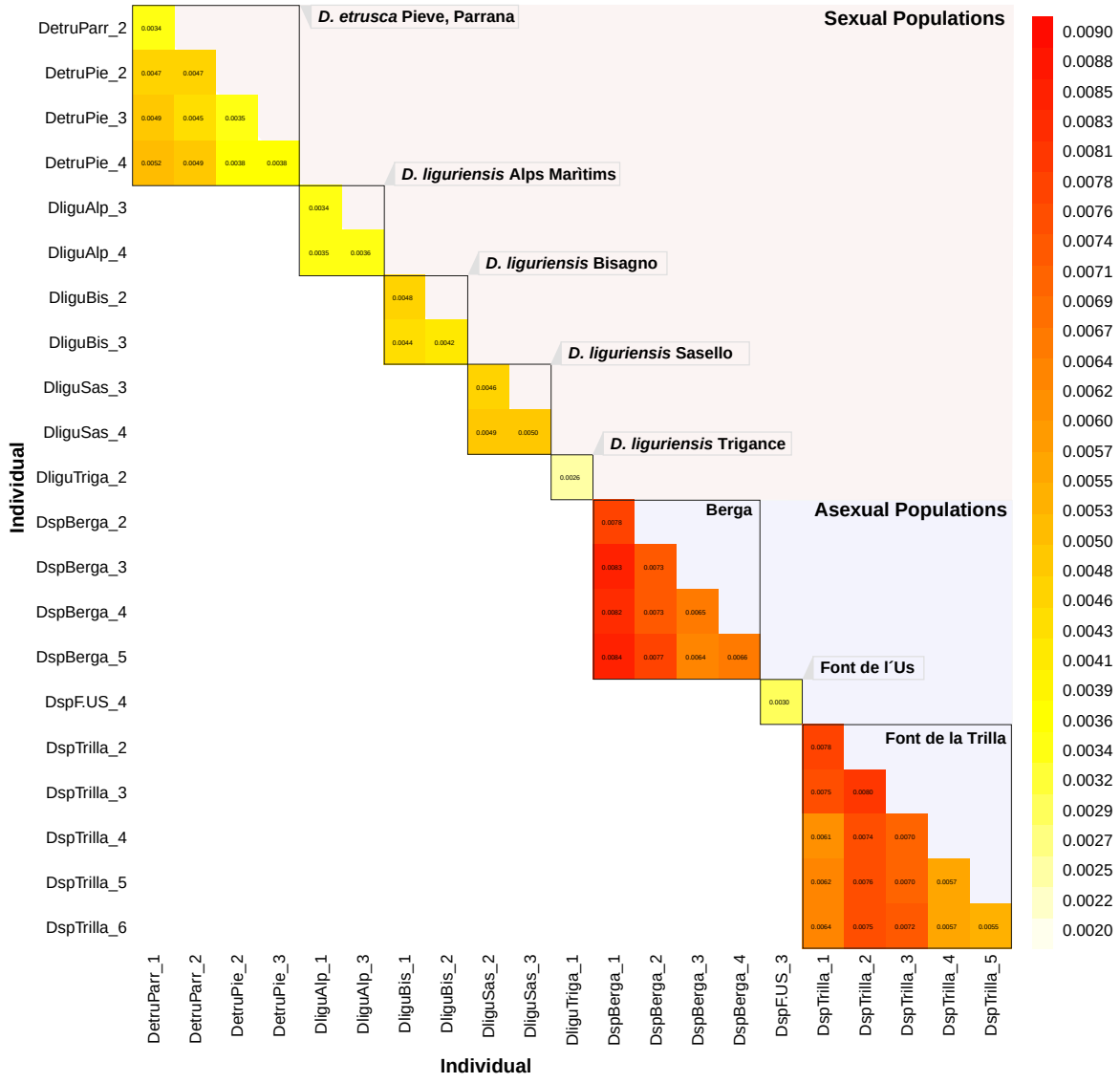
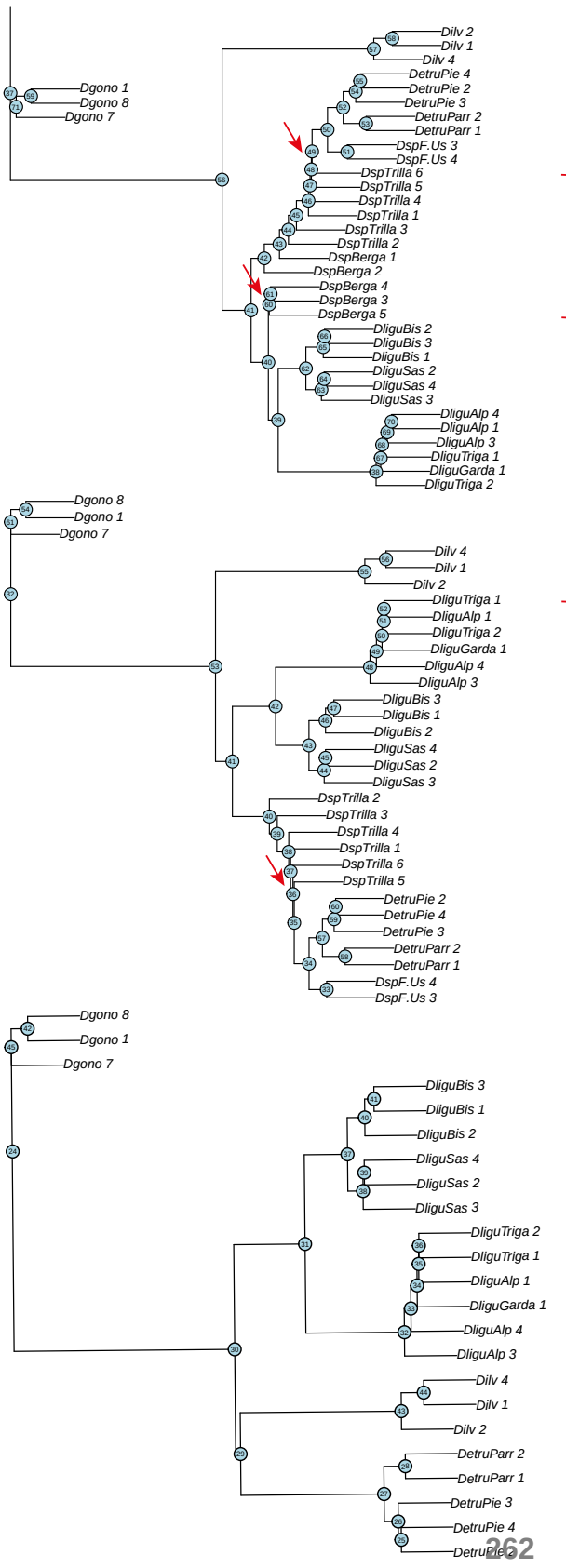


Figure S5

Per branch quartet support (q) and branch support (local posterior probability: pp) for the main (1) and the alternatives topologies (2, 3) for MSC trees obtained in ASTRAL-pro. The trees have been obtained in AstralPlane using ASTRAL-pro output and the blue circles indicate a numeric code for the nodes. The red arrows indicate discordant topologies and the correspondent values are highlated in grey on the tables. (A): Dataset 1: contains all samples belong to the Iberian-Apennines-Alps-plus clade. (B): Dataset 2: Samples from Berga locality have been removed. (C): Data set 3: Samples from Iberian Peninsula have been removed. See Supplementary Table S4 for more information.

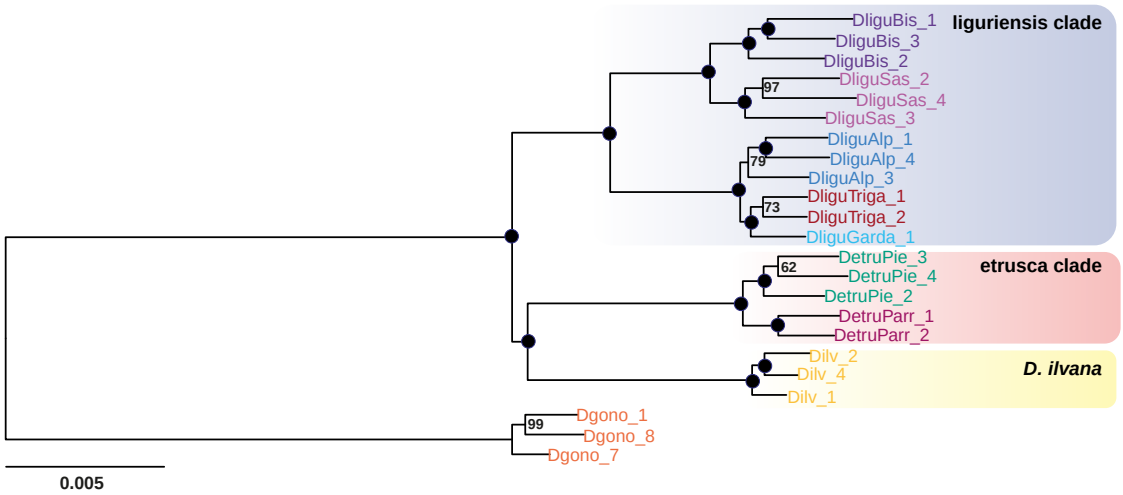


node	q1	q2	q3	1.00	pp2	pp3
37	NA	NA	NA			
38	NA	NA	NA	NA	NA	NA
39	0.91	0.05	0.04	1.00	0.00	0.00
40	0.46	0.40	0.15	1.00	0.00	0.00
41	0.53	0.11	0.36	1.00	0.00	0.00
42	0.49	0.39	0.12	1.00	0.00	0.00
43	0.51	0.29	0.20	1.00	0.00	0.00
44	0.45	0.39	0.17	1.00	0.00	0.00
45	0.43	0.33	0.24	1.00	0.00	0.00
46	0.48	0.27	0.25	1.00	0.00	0.00
47	0.36	0.36	0.29	0.51	0.49	0.01
48	0.35	0.34	0.31	0.78	0.19	0.03
49	0.34	0.34	0.32	0.51	0.42	0.07
50	0.51	0.19	0.30	1.00	0.00	0.00
51	0.56	0.24	0.20	1.00	0.00	0.00
52	0.52	0.26	0.22	1.00	0.00	0.00
53	0.58	0.24	0.18	1.00	0.00	0.00
54	0.49	0.27	0.24	1.00	0.00	0.00
55	0.39	0.32	0.29	1.00	0.00	0.00
56	0.63	0.16	0.21	1.00	0.00	0.00
57	0.97	0.02	0.01	1.00	0.00	0.00
58	0.54	0.24	0.21	1.00	0.00	0.00
59	0.51	0.24	0.25	1.00	0.00	0.00
60	0.35	0.324	0.330	0.72	0.11	0.17
61	0.35	0.32	0.33	0.89	0.04	0.07
62	0.62	0.17	0.20	1.00	0.00	0.00
63	0.52	0.24	0.24	1.00	0.00	0.00
64	0.37	0.33	0.30	1.00	0.00	0.00
65	0.53	0.23	0.24	1.00	0.00	0.00
66	0.35	0.32	0.33	0.87	0.05	0.08
67	0.39	0.31	0.30	1.00	0.00	0.00
68	0.35	0.31	0.33	0.89	0.03	0.08
69	0.40	0.32	0.28	1.00	0.00	0.00
70	0.39	0.34	0.27	1.00	0.00	0.00
71	0.99	0.01	0.00	1.00	0.00	0.00

32	NA	NA	NA	NA	NA	NA
33	NA	NA	NA	NA	NA	NA
34	0.54	0.20	0.27	1.00	0.00	0.00
35	0.51	0.32	0.17	1.00	0.00	0.00
36	0.35	0.34	0.31	0.78	0.18	0.04
37	0.37	0.35	0.28	0.94	0.06	0.00
38	0.36	0.35	0.29	0.86	0.13	0.01
39	0.47	0.29	0.24	1.00	0.00	0.00
40	0.43	0.33	0.24	1.00	0.00	0.00
41	0.69	0.07	0.24	1.00	0.00	0.00
42	0.73	0.17	0.10	1.00	0.00	0.00
43	0.67	0.19	0.15	1.00	0.00	0.00
44	0.51	0.23	0.25	1.00	0.00	0.00
45	0.35	0.31	0.34	0.81	0.03	0.17
46	0.53	0.26	0.21	1.00	0.00	0.00
47	0.44	0.29	0.28	1.00	0.00	0.00
48	0.91	0.05	0.05	1.00	0.00	0.00
49	0.42	0.30	0.29	1.00	0.00	0.00
50	0.41	0.32	0.28	1.00	0.00	0.00
51	0.36	0.31	0.33	0.95	0.01	0.03
52	0.34	0.33	0.33	0.40	0.27	0.33
53	0.53	0.20	0.27	1.00	0.00	0.00
54	0.51	0.25	0.24	1.00	0.00	0.00
55	0.97	0.02	0.02	1.00	0.00	0.00
56	0.57	0.25	0.19	1.00	0.00	0.00
57	0.49	0.24	0.27	1.00	0.00	0.00
58	0.58	0.21	0.21	1.00	0.00	0.00
59	0.48	0.27	0.26	1.00	0.00	0.00
60	0.35	0.32	0.33	0.92	0.03	0.05
61	0.99	0.01	0.00	1.00	0.00	0.00

24	NA	NA	NA	NA	NA	NA
25	NA	NA	NA	NA	NA	NA
26	0.37	0.32	0.31	1.00	0.00	0.00
27	049	0.26	0.26	1.00	0.00	0.00
28	0.56	0.26	0.18	1.00	0.00	0.00
29	0.96	0.02	0.02	1.00	0.00	0.00
30	0.39	0.33	0.28	1.00	0.00	0.00
31	0.83	0.09	0.08	1.00	0.00	0.00
32	0.90	0.05	0.05	1.00	0.00	0.00
33	0.41	0.29	0.30	1.00	0.00	0.00
34	0.41	0.31	0.28	1.00	0.00	0.00
35	0.36	0.32	0.33	0.94	0.02	0.04
36	0.34	0.33	0.33	0.47	0.21	0.31
37	0.71	0.17	0.12	1.00	0.00	0.00
38	0.50	0.24	0.26	1.00	0.00	0.00
39	0.35	0.31	0.34	0.77	0.03	0.20
40	0.53	0.27	0.21	1.00	0.00	0.00
41	0.44	0.29	0.27	1.00	0.00	0.00
42	0.52	0.25	0.24	1.00	0.00	0.00
43	0.97	0.02	0.02	1.00	0.00	0.00
44	0.57	0.24	0.20	1.00	0.00	0.00
45	0.99	0.00	0.01	1.00	0.00	0.00

Figure S6. ML tree obtained from dataset 3 in IQ-TREE for nucleotide data. Bootstrap values are shown on the nodes. The black dots represent the maximum value (100). Scale bar: substitutions per site.



Supporting Information Text S1

Methodology to search for Dunuc12 in the transcriptomic data

Dunuc12 (Dugesia nuclear 12) is a fragment of the nuclear gene Transmembrane p24 trafficking protein 9 (TMED9), which is a transport protein transmembrane domain. This nuclear marker was identified from low coverage genomic sequences of *Dugesia* sp blasted against the genome of *Schmidtea mediterranea* SXl v4.0 (Leria et al., 2020) and subsequently used for species delimitation and intraindividual genetic diversity analyses in different studies on the genus *Dugesia* (Dols-Serrate, Leria, Aguilar, Stocchino, & Riutort, 2020; Leria et al., 2020; Leria, Vila-Farré, Solà, & Riutort, 2019).

In this study, we cloned the Dunuc12 in 8 samples from the etrusca-liguriensis group to examine the intraindividual genetic diversity of sexual and fissiparous individuals. The obtained length of the Dunuc12 sequences was 680 positions, with an exonic region of 120 positions.

In this section, we describe the methodology used to find the exonic region of Dunuc12 in our transcriptomic data. We perform the steps as follow:

1. Extract the exonic region:

We aligned the entire sequences of the obtained clones of *D. etrusca* and *D. liguriensis* with sequences of Dunuc12 previously characterized and annotated (Leria et al., 2019) to find the exonic regions in the global alignment. We obtained 120 exonic positions of every cloned sequence. From the alignment of these 120 positions, we obtained a non resolved maximum likelihood tree (Fig. 1A in this text). This result is not rare, since the most variable regions in Dunuc12 have been found in the introns (Leria et al., 2019). A more resolutive tree was obtained when we included all the entire cloned sequences (Fig. 1B)

2. Search for the exonic region in the transcriptomic data:

We blasted every cloned sequence against the corresponding transcriptome using blastn with the script BLASTn_Dunuc.sh. We performed two blast searches, one against the transcriptome without filtering, just after the assembling step, and another one against the filtered transcripts after the Blobtools step (Chapter IV.1). The results

were identical in both cases, demonstrating that we retain Dunuc12, even after the filtering steps.

For every sample, the cloned sequences always matched against the same transcript, with very high identity percent, and bitscore, as well as low e-value (Table 1 in this text).

3. Transcript identification:

Taking advantage of the results obtained in the filtering steps with Blobtools, we used the Blastn of the transcriptome against the nucleotide database (Chapter IV.1) to validate the transcript identification. All the matched transcripts by the cloned sequences, yield also, hit against sequences from *Dugesia* identified as TMED9 in the nucleotide database of NCBI identified with the GeneBank accession codes MK713132-54 (Leria, et al., 2019).

4. Extract the coverage of the Dunuc12 region in the transcripts:

Working with the bam files previously obtained for Blobtools filtering, we obtained the per-base coverage of the 120 positions of Dunuc12 in the transcripts using samtools depth with the -a option, and the media coverage by region with mosdepth tool using the -n option (Fig. 2 in this text).

In addition, we visualized the transcripts in IGV, specifically the region where the cloned sequences matched. We explored these regions in IGV looking for polymorphic positions, but any position showed polymorphism; just as we expected taking into account the low variability of exonic regions of the Dunuc12.

5. Search for Dunuc12 in the orthogroups:

To check if Dunuc12 had been assembled in our transcriptome and to corroborate that this gene is a Single Copy, we looked for the matched transcripts in all orthogroups using the scripts look_Dunuc12_in_OG.sh and look_Dunuc12_in_OG_all_samples.sh. The exploration was done in the two orthologs search performed; the search with all samples (WM_717) and the search with samples from the etrusca-liguriensis group (etru-ligu_1984) (Chapter IV.1). We found that all transcripts formed only one orthogroup in the two searches: OG0010797 in etru-

ligu_1984, and OG0007336 in WM_717. Unfortunately, any of the two orthogroups was defined as SC orthogroup by Orthofinder.

In the case of etru-ligu_1984, the sample DliguSas_2 (not included in the cloning analysis) did not show any transcript into OG0010797, although one transcript matched with Temed9 in the blastn against nucleotide database of NCBI , indicatingthat this transcript was lost in the filtering steps. In addition, the sample Dgono_8 contributes with two transcripts to the orthogroup, being Dgono_8_DN38329_c0_g1_i1 a very rare sequence, that could be interpreted as a failure of the orthologs search.

In the case of the WM_717 search, in addition to the two problems explained above, we found more problematic sequences: DbenCors_North_3_DN37816_c0_g2_i1 was very rare, and Dvila_2_DN7972_c0_g1_i1.p2 was a partial isoform that was not eliminated with the filtering.

6. Maximum likelihood trees inference from Dunuc12 orthogroups:

We eliminated the problematic sequences in OG0010797 (etru-ligu_1984), aligned the block, and obtained trees from nucleotide and protein sequences using IQ-TREE (Fig. 3 in this text). The low resolution of these trees was congruent with the previous results, which describe the most variability in the intronic regions of Dunuc12 and not in the exons, more exposed to gene regulatory mechanisms. In addition, the trees show the topology of a region that has not suffered any duplication event in the evolutionary history of the group.

Main conclusion:

Although the OrthoFinder does not classify Dunuc12 as SC (because of artifacts in the assemblies, and failures in filtering or OrthoFinder runs), we have enough proofs to corroborate that Dunuc12 is not a paralog and it is a Single Copy gene in the *Dugesia* genome. In the first place, all sequences are grouped only in one orthogroup in every orthologue search. As a second point, the trees show the typical topology of Single Copy genes. These results validate the intronic region of Dunuc12 as a good marker for studies on intraindividual genetic variability in the genus *Dugesia*.

References in this text

Dols-Serrate et al., (2020). https://doi.org/10.1007/s13127-020-00438-z
Leria, L et al., (2020). https://doi.org/10.1016/j.ympev.2019.05.010
Leria, L et al., (2019). https://doi.org/10.1186/s12862-019-1440-1

Figures and Tables in this text

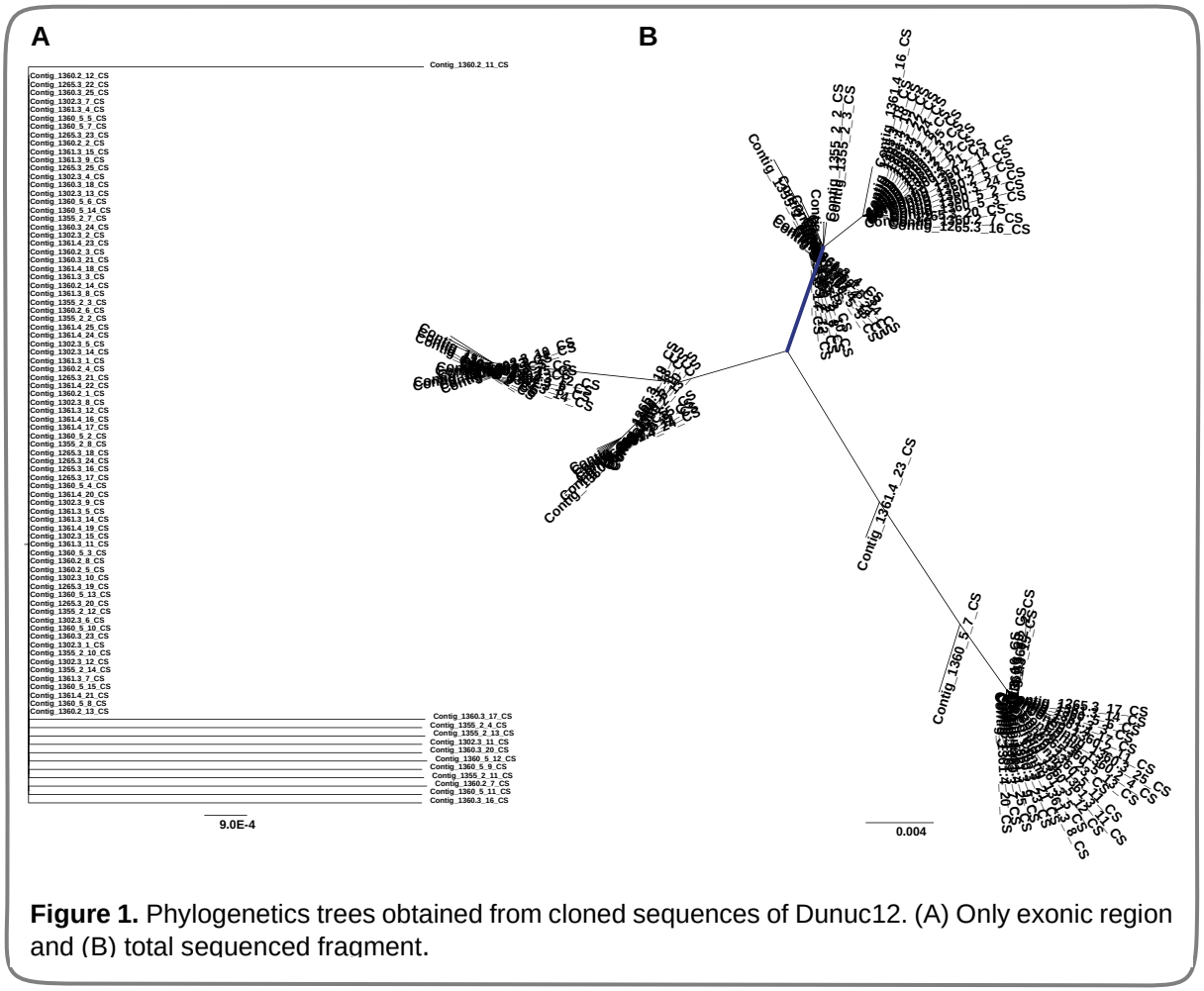


Figure 1. Phylogenetics trees obtained from cloned sequences of Dunuc12. (A) Only exonic region and (B) total sequenced fragment.

Table 1. Results of the Blastn of cloned sequences against their respective transcriptome.

Sample	MR ID	No. Clones	bitscore (lower/higher values)	Blastn results resume									
				Subject accession	% ident match	Alig length	mismatch	gapopen	query start	Query end	Subj. start	Subj. end	evalue
DspBerga_2	**1360.2**	**12**	217/222	TRINITY_DN1077_c0_g1_i3	99.167/100	120	1	0	1	120	404	285	2.87E-56
				TRINITY_DN1077_c0_g1_i2	99.17	120	1	0	1	120	404	285	2.87E-56
				TRINITY_DN1077_c0_g1_i4	99.17	120	1	0	1	120	404	285	2.87E-56
DspBerga_3	**1360.3**	**8**	217/222	TRINITY_DN3226_c0_g1_i2	99.167/100	120	1	0	1	120	468	349	4.06E-56
			217	TRINITY_DN3226_c0_g1_i1	99.17	120	1	0	1	120	468	349	4.06E-56
DetruBerg_5	**1360.5**	**14**	217/222	TRINITY_DN6190_c0_g1_i1	99.167/100	120	1	0	1	120	450	569	4.6E-56
DspTrilla_3	**1265.3**	**10**	222	TRINITY_DN1320_c0_g1_i2	100.00	120	0	0	1	120	473	592	9.36E-58
DspTrilla_5	**1361.3**	**11**	222	TRINITY_DN596_c0_g1_i1	100.00	120	0	0	1	120	590	471	8.57E-58
			222	TRINITY_DN596_c0_g1_i4	100.00	120	0	0	1	120	590	471	8.57E-58
DspTrilla_6	**1361.4**	**10**	222	TRINITY_DN784_c0_g1_i1	100.00	120	0	0	1	120	402	283	8.6E-58
			222	TRINITY_DN784_c0_g1_i2	100.00	120	0	0	1	120	402	283	8.6E-58
			222	TRINITY_DN784_c0_g1_i3	100.00	120	0	0	1	120	402	283	8.6E-58
DetruPie_2	**1355.2**	**10**	217/222	TRINITY_DN92587_c0_g1_i1	99.167/100	120	1	0	1	120	404	285	3.76E-56
DliguGarda_1	**1302.3**	**14**	222/217	TRINITY_DN77494_c0_g1_i1	100/99.17	120	0	0	1	120	450	569	7.2E-58

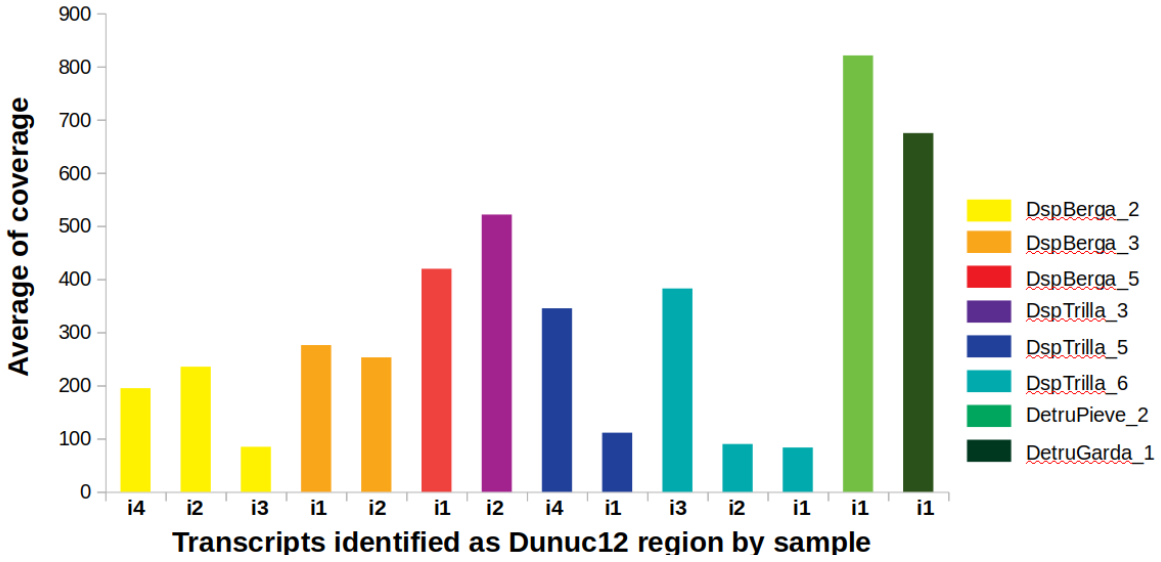


Figure 2. Average coverage of identified transcript as Dunuc12 region by sample. See Table 1 for complete transcripts names.

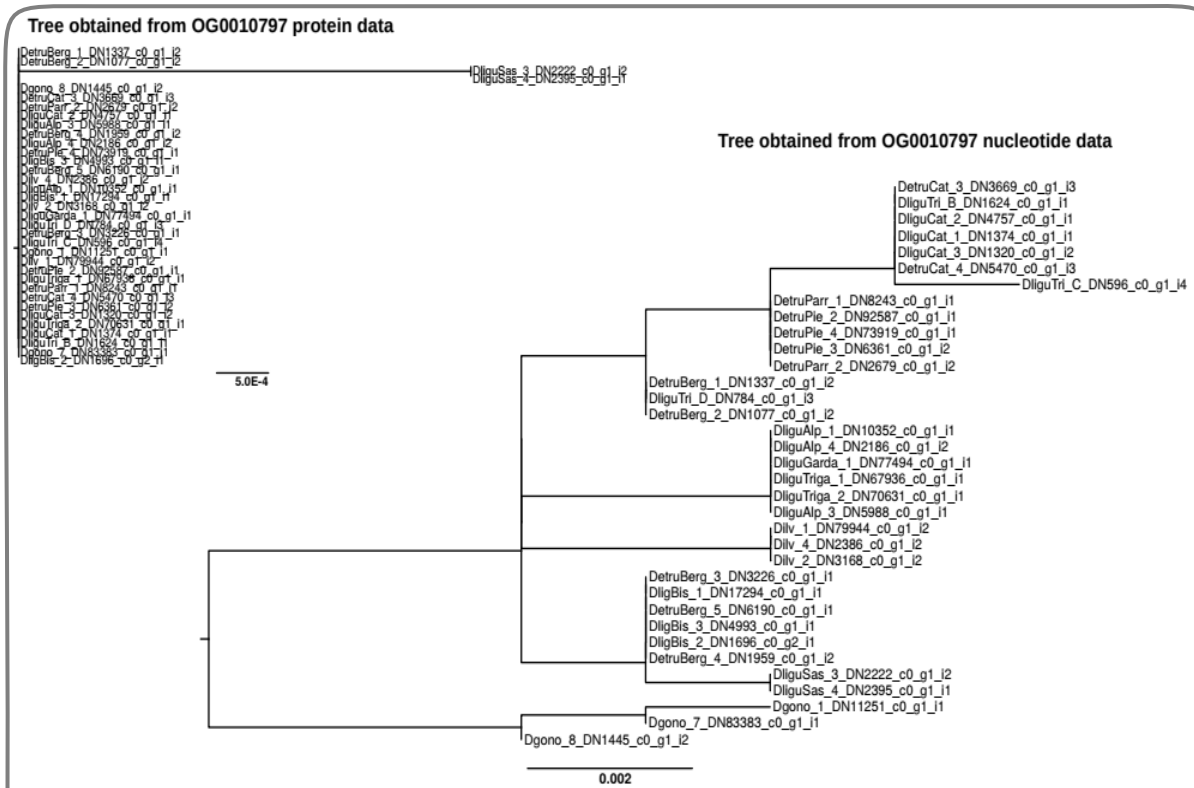


Figure 3. Phylogenetic trees obtained from the orthogroup OG0010797 identified as Dunuc12 in the ortholog search of etrusca-liguriensis group

References

Abril, J. F., Cebrià, F., Rodríguez-Esteban, G., Horn, T., Fraguas, S., Calvo, B., … Saló, E. (2010). Smed454 dataset: unravelling the transcriptome of *Schmidtea mediterranea*. *BMC Genomics2*, *11*(731). https://doi.org/10.1186/1471-2164-11-731

Altenhoff, A. M., Levy, J., Zarowiecki, M., Tomiczek, B., Vesztrocy, A. W., Dalquen, D. A., … Dessimoz, C. (2019). OMA standalone: Orthology inference among public and custom genomes and transcriptomes. *Genome Research*, *29*(7), 1152–1163. https://doi.org/10.1101/gr.243212.118

Altenhoff, A. M., Schneider, A., Gonnet, G. H., & Dessimoz, C. (2011). OMA 2011: Orthology inference among 1000 complete genomes. *Nucleic Acids Research*, *39*. https://doi.org/10.1093/nar/gkq1238

Baguñà, J., Carranza, S., Pala, M., Ribera, C., 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

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026036

Bast, J., Parker, D. J., Dumas, Z., Jalvingh, K. M., Van, P. T., Jaron, K. S., … Schwander, T. (2018). Consequences of asexuality in natural populations: Insights from stick insects. *Molecular Biology and Evolution*, *35*(7), 1668–1677. https://doi.org/10.1093/molbev/msy058

Benazzi, M. (1946). Sopra una nova planaria d'acqua dolce. *Estratto Dall' Archivo Zoologico Italiano*, *XXXI*.

Birky, C. W. (1996). Heterozygosity, Heteromorphy, and Phylogenetic Trees in Asexual Eukaryotes. *Genetics*, *144*, 427–437.

Boluda, C. G., Rico, V. J., Naciri, Y., Hawksworth, D. L., & Scheidegger, C. (2021). Phylogeographic reconstructions can be biased by ancestral shared alleles: The

case of the polymorphic lichen *Bryoria fuscescens* in Europe and North Africa. *Molecular Ecology*, *30*(19), 4845–4865. https://doi.org/10.1111/mec.16078

Bortolotti, V., Fazzuoli, M., Pandeli, E., Principi, G., Babbini, A., & Corti, S. (2001). Geology of central and eastern Elba island, Italy. *Ofioliti*, *26*(2 A), 97–105. https://doi.org/10.4454/ofioliti.v26i2a.137

Bortolotti, V., Pandeli, E., & Principi, G. (2001). The geology of the Elba Island: An historical introduction. *Ofioliti*, *26*(2 A), 79–96. https://doi.org/10.4454/ofioliti.v26i2a.136

Bouckaert, R. R. (2010). DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), 1372–1373. https://doi.org/10.1093/BIOINFORMATICS/BTQ110

Bouckaert, R. R., & Heled, J. (2014). DensiTree 2: Seeing Trees Through the Forest. *BioRxiv*, 012401. https://doi.org/10.1101/012401

Bowie, R. C. K., Fjeldså, J., Hackett, S. J., Bates, J. M., & Crowe, T. M. (2006). Coalescent models reveal the relative roles of ancestral polymorphism, vicariance, and dispersal in shaping phylogeographical structure of an African montane forest robin. *Molecular Phylogenetics and Evolution*, *38*(1), 171–188. https://doi.org/10.1016/j.ympev.2005.06.001

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(421). https://doi.org/10.1186/1471-2105-10-421

Capra, J. A., Hubisz, M. J., Kostka, D., Pollard, K. S., & Siepel, A. (2013). A Model-Based Analysis of GC-Biased Gene Conversion in the Human and Chimpanzee Genomes. *PLOS Genetics*, *9*(8), e1003684. https://doi.org/10.1371/JOURNAL.PGEN.1003684

Dapporto, L., & Cini, A. (2007). Faunal patterns in Tuscan archipelago butterflies: The dominant influence is recent geography not paleogeography. *European Journal of Entomology*, *104*(3), 497–503. https://doi.org/10.14411/eje.2007.070

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. (1988). Further contribution to the taxonomy and biogeography of the subgenus *Dugesia* (Platyhelminthes: Tricladida: Paludicola) in the Mediterranean region and the Middle East. *Isaarael Journal of Zoology*, *35*, 109–136.

Deutekom, E. S., Snel, B., & Van Dam, T. J. P. (2021). Benchmarking orthology methods using phylogenetic patterns defined at the base of Eukaryotes. *Briefings in Bioinformatics*, *22*(3), 1–9. https://doi.org/10.1093/bib/bbaa206

Di Nicola, M. R., & Vaccaro, A. (2020). New data on the presence of the Aesculapian snake *Zamenis longissimus* (Laurenti, 1768) (Serpentes Colubridae) on Elba Island (Tuscany, Italy). *Biodiversity Journal*, *11*(2), 611–614. https://doi.org/10.31396/biodiv.jour.2020.11.2.611.614

Du, Z., Hasegawa, H., Cooley, J. R., Simon, C., Yoshimura, J., Cai, W., … Li, H. (2019). Mitochondrial genomics reveals shared phylogeographic patterns and demographic history among three periodical cicada species groups. *Molecular Biology and Evolution*, *36*(6), 1187–1200. https://doi.org/10.1093/molbev/msz051

Duret, L., & Galtier, N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual Review of Genomics and Human Genetics*, *10*, 285–311. https://doi.org/10.1146/annurev-genom-082908-150001

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(157). https://doi.org/10.1186/s13059-015-0721-2

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

Fattorini, S. (2009). Both Recent and Pleistocene geography determine animal distributional patterns in the Tuscan Archipelago. *Journal of Zoology*, *277*(4), 291–301. https://doi.org/10.1111/j.1469-7998.2008.00540.x

Fehrer, J., Krak, K., & Chrtek, J. (2009). Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (Hieracium, Lactuceae, Asteraceae): Disentangling phylogenetic signal, reticulation, and noise. *BMC Evolutionary Biology*, *9*(1), 1–23. https://doi.org/10.1186/1471-2148-9-239

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, *78*(2), 737–756. https://doi.org/10.1093/genetics/78.2.737

Fernández, R., Gabaldon, T., Dessimoz, C., Fernández, R., Gabaldon, T., Dessimoz, C., … Phylo-, N. (2020). Orthology: Definitions, Prediction, and Impact on Species Phylogeny Inference. In C. Scornavacca, F. Delsuc, & N. Galtier (Eds.), *Phylogenetics in the Genomic Era* (pp. 2.4:1–2.4:14). https://doi.org/hal-02535414

Fontaneto, D., Herniou, E. A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., & Barraclough, T. G. (2007). Independently Evolving Species in Asexual Bdelloid Rotifers. *PLoS Biol*, *5*(4), 914–921. https://doi.org/10.1371/journal.pbio.0050087

Galtier, N. (2021). Fine-scale quantification of GC-biased gene conversion intensity in mammals. *Peer Community Journal*, *1*. https://doi.org/10.24072/pcjournal.22. hal-03435015v2

Gill, D. E., Chao Lin, Perkins, S. L., & Wolf, J. B. (1995). Genetic mosaicism in plants and clonal animals. *Annual Review of Ecology and Systematics*, *26*(40), 423–444. https://doi.org/10.1146/annurev.es.26.110195.002231

Hasegawa, M., Kishino, H., & Yano, T. aki. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution 1985 22:2*, *22*(2), 160–174. https://doi.org/10.1007/BF02101694

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

Hörandl, E. (2018). The classification of asexual organisms: Old myths, new facts, and a novel pluralistic approach. *Taxon*, *67*(6), 1066–1081. https://doi.org/10.12705/676.5

Hutter, C. (2020). *AstralPlane: R Package For Preparing, Running, Analyzing and Plotting from the Species Tree Program ASTRAL-III*. Retrieved from https://github.com/chutter/AstralPlane

Jukes, T. H., & Cantor, C. R. (1969). Evolution of Protein Molecules. *Mammalian Protein Metabolism*, 21–132. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7

Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, *21*(7), 428–444. https://doi.org/10.1038/s41576-020-0233-0

Kishino, H., & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution 1989 29:2*, *29*(2), 170–179. https://doi.org/10.1007/BF02100115

Kishino, H., Miyata, T., & Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution 1990 31:2*, *31*(2), 151–160. https://doi.org/10.1007/BF02109483

Lamolle, G., Fontenla, S., Rijo, G., Tort, J. F., & Smircich, P. (2019). Compositional Analysis of Flatworm Genomes Shows Strong Codon Usage Biases Across All Classes. *Frontiers in Genetics*, *10*(September), 1–12. https://doi.org/10.3389/fgene.2019.00771

Lázaro, E., Sluys, R., Pala, M., Stochino, 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

Le, S. Q., & Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, *25*(7), 1307–1320. https://doi.org/10.1093/MOLBEV/MSN067

Leria, L., Riutort, M., Romero, R., Ferrer, X., & Vila-Farré, M. Microplate tectonics and the environment as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography, In Press*.

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). https://doi.org/10.1186/s12862-019-1440-1

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

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

Lunt, D. H. (2008). Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. *BMC Evolutionary Biology*, *8*(1), 1–16. https://doi.org/10.1186/1471-2148-8-194

MacEachern, S., McEwan, J., & Goddard, M. (2009). Phylogenetic reconstruction and the identification of ancient polymorphism in the Bovini tribe (Bovidae, Bovinae). *BMC Genomics*, *10*, 1–17. https://doi.org/10.1186/1471-2164-10-177

Martens, K., Rossetti, G., & Baltanás, A. (1998). Reproductives modes and taxonomy. In K. Martens (Ed.), *Sex and parthenofenesis: evolutionary ecology of reproductive modes in non-marine ostracods* (pp. 197–214). Leiden: Backhuys Publishers.

Moran, P., & Kornfield, I. (1993). Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Molecular Biology and Evolution*, *10*(5), 1015–1029. https://doi.org/10.1093/oxfordjournals.molbev.a040063

Muller, H. (1932). Some genetic aspects of sex. *The American Naturalist*, *66*(703), 118–138.

Muller, H. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1*(1), 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Nelson Hall, R., Weill, U., Khariton, M., Leal-Ortiz, S., Drees, L., Chai, C., … Wang, B. (2021). Heterologous reporter expression in the planarian Schmidtea mediterranea through somatic mRNA transfection. *BioRxiv*, 1–43. Retrieved from https://doi.org/10.1101/2021.04.20.440701

Normark, B. B., Judson, O. P., & Moran, N. a. (2003). Genomic signatures of ancient asexual lineages. *Biological Journal of the Linnean Society*, *79*, 69–84. https://doi.org/10.1046/j.1095-8312.2003.00182.x

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*, 526–528.

Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. P. (2011). MACSE: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. *PLoS ONE*, *6*(9). https://doi.org/10.1371/journal.pone.0022594

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

Resch, A. M., Palakodeti, D., Lu, Y. C., Horowitz, M., & Graveley, B. R. (2012). Transcriptome analysis reveals strain-specific and conserved stemness genes in *Schmidtea mediterranea*. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0034447

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

Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., & Mesirov, J. P. (2017). Variant review with the integrative genomics viewer. *Cancer Research*, *77*(21), e31–e34. https://doi.org/10.1158/0008-5472.CAN-17-0337/SUPPLEMENTARY-VIDEO-S1

Romiguier, J., & Roux, C. (2017). Analytical biases associated with GC-content in molecular evolution. *Frontiers in Genetics*, *8*(FEB), 1–7. https://doi.org/10.3389/fgene.2017.00016

Saló, E., Abril, J. F., Adell, T., Cebrià, F., Eckelt, K., Fernández-Taboada, E., … Rodríguez-Esteban, G. (2009). Planarian regeneration: Achievements and future directions after 20 years of research. *International Journal of Developmental Biology*, *53*(8–10), 1317–1327. https://doi.org/10.1387/ijdb.072414es

Schwander, T., Henry, L., & Crespi, B. J. (2011). Molecular evidence for ancient asexuality in timema stick insects. *Current Biology*, *21*(13), 1129–1134. https://doi.org/10.1016/j.cub.2011.05.026

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE*, *11*(10), 1–10. https://doi.org/10.1371/journal.pone.0163962

Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, *51*(3), 492–508. https://doi.org/10.1080/10635150290069913

Shimodaira, H., & Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution*, *16*(8), 1114–1114. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026201

Solà, E., Leria, L., Stocchino, G. A., Bagherzadeh, R., Balke, M., Daniels, S. R., … Riutort, M. Three dispersal routes out of Africa: The puzzling biogeographical history of Dugesia freshwater planarians. *Journal of Biogeography, In Press*.

Stocchino, G. A., & Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology*, *80*(3), 319–328. https://doi.org/10.1080/11250003.2013.822025

Strimmer, K., & Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1487), 137–142. https://doi.org/10.1098/RSPB.2001.1862

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512–526. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040023

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, *17*, 57–86.

Team, R. C. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Retrieved from https://www.r-project.org/

Wang, B., Mojica, J. P., Perera, N., Lee, C. R., Lovell, J. T., Sharma, A., … Mitchell-Olds, T. (2019). Correction to: Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in *Boechera stricta*. *Genome Biology*, *20*(1), 1–15. https://doi.org/10.1186/s13059-019-1781-5

Wang, H. C., Minh, B. Q., Susko, S., & Roger, A. J. (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

Webster, M. T., & Hurst, L. D. (2012). Direct and indirect consequences of meiotic recombination: Implications for genome evolution. *Trends in Genetics*, *28*(3), 101–109. https://doi.org/10.1016/j.tig.2011.11.002

Welch, D. M., & Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, *288*(5469), 1211–1215. https://doi.org/10.1126/science.288.5469.1211

Yan, Z., Smith, M. L., Du, P., Hahn, M. W., & Nakhleh, L. (2022). Species Tree Inference Methods Intended to Deal with Incomplete Lineage Sorting Are Robust to the Presence of Paralogs. *Systematic Biology*, *71*(2), 367–381. https://doi.org/10.1093/sysbio/syab056

Zhang, C., Scornavacca, C., Molloy, E. K., & Mirarab, S. (2020). ASTRAL-pro: Quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution*, *37*(11), 3292–3307. https://doi.org/10.1093/molbev/msaa139

Chapter V

Differential expression analysis reveals different metabolic background in sexual and asexual freshwater planarians under wild conditions

LISANDRA BENÍTEZ-ÁLVAREZ, LAIA LERIA, AND MARTA RIUTORT

Differential expression analysis reveals different metabolic background in sexual and asexual freshwater planarians under wild conditions

Lisandra Benítez-Álvarez[1*], Laia Leria[1], and Marta Riutort[1]

[1]Departament de Genètica, Facultat de Biologia and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Catalonia, Spain.

*corresponding author: lbenitezalvarez87@gmail.com

Summary

Freshwater planarians are hermaphroditic organisms with a high regeneration capability which, in some species, is used to reproduce asexually by fission (Reddien & Alvarado, 2004; Saló et al., 2009). Individuals that reproduce by fission do not develop reproductive systems but produce new individuals by binary fission and regeneration of the lost structures in each part (Rink, 2013) The *Dugesia* genus presents both types of reproductive strategies in diverse combinations (Stochino & Manconi, 2013). Here, we performed a differential expression analysis between sexual and asexual individuals of several *Dugesia* species. We found differentially expressed genes between the two strategies indicating changes in the metabolic activity related to every reproductive strategy.

Main Text

Dugesia species show two types of reproduction, sexual with cross fertilization and asexual by fission (Fig. 1A), which can be combined in different ways. In the species from the Western Mediterranean it is possible to find the two strategies in the same population, populations of the same species presenting only one or the other strategy, or closely related species with different strategies (Fig. 1B). We used transcriptomic data extracted from wild individuals belonging to these types of populations to analyse the expression profiles of orthologous genes shared between sexual and asexual planarians.

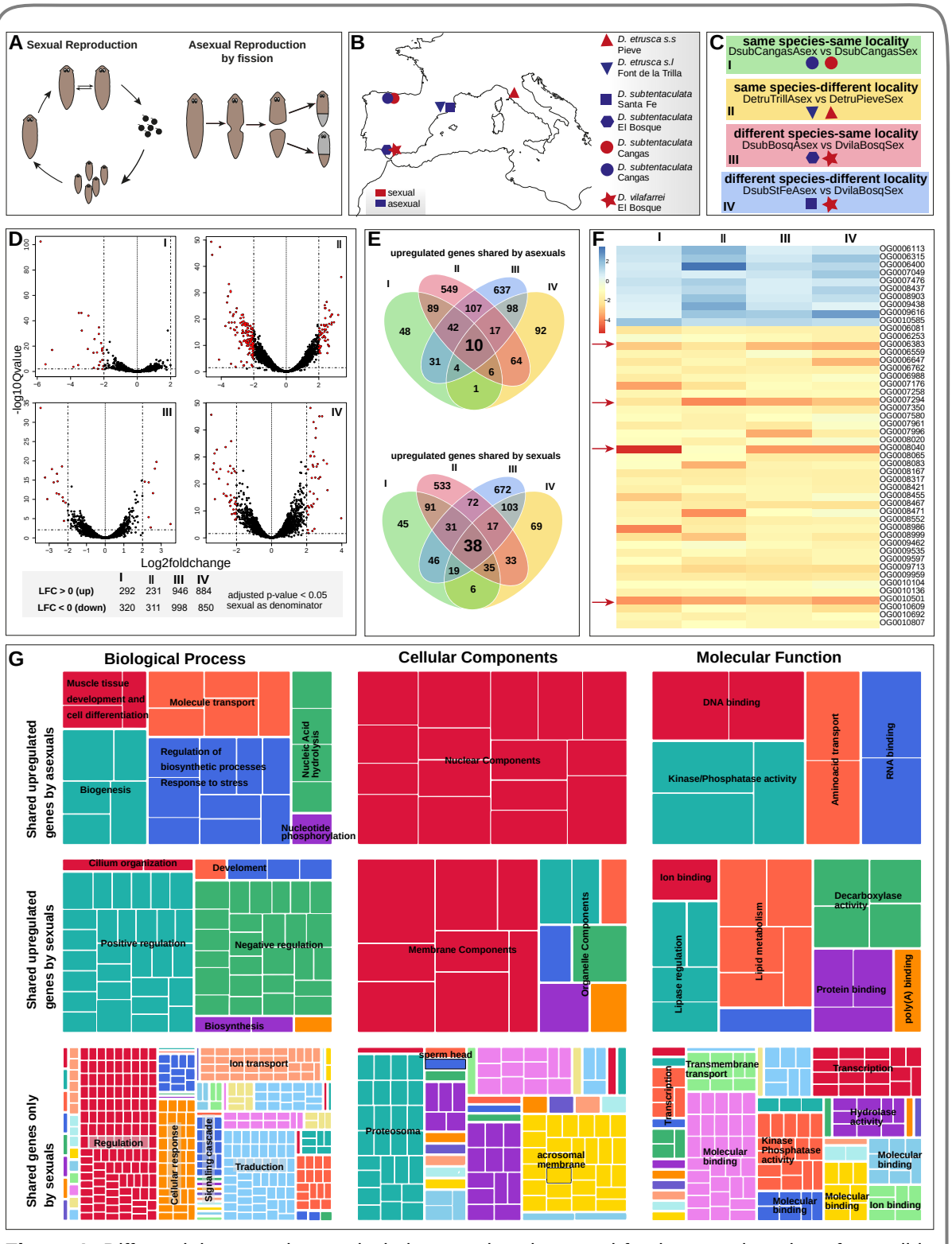


Figure 1. Differential expression analysis in sexual and asexual freshwater planarians from wild populations. (A) Scheme of sexual and asexual reproduction in planarians. (B) Localizaion of sampling points. (C) Comparison levels stablished for the analyses. (D) Volcano plots showing the differentially expressed genes found in every comparison. The sexual condition has been used as the denominator for the Fold Change (Log2FC) calculation. The red dots label the most differentially expresed genes. (E) Venn diagram showing the upregulated genes shared by asexuals (top) and the upregulated genes shared by sexuals (bottom).

The figure caption continues in the next page

Figure 1 cont.. (F) Heat map showing the Log2FC values of the upregulated genes shared by asexuals (blue) and the upregulated genes shared by sexuals (red). Red arrows indicate the most upregulated genes shared by sexuals. (G) Treemaps summarizing the Gene Ontology enrichment analysis of the set of upregulated genes shared by asexuals (top) and sexuals (middle). The panel at the bottom shows the treemaps summarizing the GO categories annotated for the genes shared only by sexual individuals.

Four levels of comparisons were established (Fig. 1C): I) Sexual and asexual individuals from the same species and the same population (DsubCangasAsex vs DsubCangasSex), II) Sexual and asexual individuals from the same species and different populations (DetruTrillAsex vs DetruPieveSex), III) Different sexual and asexual species from the same locality (DsubBosqAsex vs DvilaBosqSex), and finally IV) Different sexual and asexual species from different localities (DsubStFeAsex vs DvilaBosqSex).

Transcriptomic data was filtered, assembled and translated to protein sequence. The orthologs search was performed using all samples, obtaining an annotated reference of 3,373 single copy genes for every sample, following the strategy described by (Parker, Bast, Jalvingh, & Robinson-Rechavi, 2018). Each sample was mapped against their reference to perform the differential expression analysis in each comparison level, using sexuals as the denominator for the fold change calculation. The differentially expressed genes shared among all comparisons were selected and subjected to the Gene Ontology enrichment analysis. In addition, we obtained the set of genes shared only by sexual individuals, and that shared only by asexual individuals (Supplemental Information).

Of the several genes identified to be involved in regeneration processes in freshwater planarians (Reddien, 2018; Saló et al., 2009), only two were identified and annotated in our set of single copy orthologs. The Spk-1 gene, a protein tyrosine kinase continuously expressed during the regeneration process, and the Dth1 gene, a specific marker of gastrodermal cells. However, none of these genes was differentially expressed. Besides, none of the described genes related to sexual reproduction was retained in our reference. Because several functional genes related to the gametogenesis process are paralogs, such as the Cytoplasmic Polyadenylation Element Binding protein (CPEB) (Rouhana, Tasaki, Saberi, & Newmark, 2017), selecting only the single copy orthologs to build our reference we are aware that we lose a part of the information (Supplemental Information).

The differential expression profiles varied among comparisons (Fig. 1D). The number of differentially expressed genes was greater in the comparisons between individuals from different localities, possibly reflecting an expression profile adapted to the environment. To eliminate the noise introduced by environmental adaptation, we retained the upregulated genes shared among all comparisons, resulting in 10 upregulated genes shared among asexual individuals and 38 upregulated genes shared among sexual individuals (Fig. 1E). Several upregulated genes in sexuals show modular LogFC values higher than in asexual upregulated genes (Fig. 1F). The upregulated genes with high LogFC shared by sexual individuals were: OG0006383, OG0007294, OG0008040, and OG0010501. According to our annotation, these genes are related with the cilia and axoneme assembly, the spermatogenesis, the regulation of transcription, and the dephosphorylation functions respectively. These genes could be related to the gametogenesis process since they are upregulated in sexual individuals.

The gene ontology analysis of differentially expressed genes shared by asexuals are enriched in biological processes related to cellular replication (chromatin organization, DNA repair, nucleotide phosphorylation, protein deacetylation, cellular component biogenesis, and amino acid transport), response to stress (DNA damage detection and response, and stress activated protein kinase cascade), regulation of biosynthetic processes, muscle tissue development, and cell differentiation. In addition, this group of genes shows an enrichment of nuclear components (Fig. 1G top panel). These results show an enrichment of functions and components that could be related to the regeneration process.

On the other hand, the enriched biological processes in sexual individuals are related to anabolic and catabolic processes, metabolic regulation, and response to stimulus (osmotic response, thermotaxis, response to neurotrophic factors, and response to wounding), all related to the homeostasis maintenance. Components of the cell periphery and organelles are enriched in sexual individuals (Fig. 1G middle panel).

Regarding the orthogroups shared only by samples with the same reproductive strategy, we did not find any orthogroup strictly shared by the fissiparous samples. On the contrary, we found many orthogroups shared by sexuals (842) of which 119 were

fully annotated. The GO analysis of these 119 genes showed the presence of sperm components, possibly related to the gametogenesis process (Fig. 1G bottom panel).

It must be considered that we used data coming from animals sampled directly from the field. Thus, the expression profiles obtained here reflect the stage of the animals at the moment of sampling. Under homeostasis conditions, planarian stem cells are constantly replacing tissues through a cell proliferation process that drives the cellular turnover. A general increase in stem cell division, high transcriptional changes, and metabolic rates must be induced by injuries (Bohr, Shiroor, & Adler, 2021; Osuma, Riggs, Gibb, & Hill, 2018). Thus, it could be expected that the transcription profiles of fissiparity related genes, will be similar in sexual and asexual planarians if regeneration processes had not been induced by injuries or by fission in the second ones. This could explain the small number of upregulated genes shared by asexual individuals and the null number of genes shared only by asexual individuals.

Although we have used a limited reference of only 3,373 annotated single copy genes, it was possible to discover a differential expression profile shared by each of the reproductive strategies. Our results suggest a differentiated metabolic activity between sexual and asexual planarians shared across different environments and evolutionary lineages. Differences in the regulation of amino acid metabolism have been observed between asexual and experimentally sexualized strains of *Dugesia ryukyuensis* under laboratory conditions (Sekii et al 2019). Here, we found that the metabolic activity of asexual planarians in the wild is focused on the cellular division and the response to the stress of the regenerative process. The expression profile of sexual planarians, on the other hand, is aimed to respond to the environmental stimulus and maintenance of homeostasis. In addition, because sexual planarians maintain a high cellular turnover and the regenerative capacity in response to injuries (Reddien, 2018), they need to keep the expression of regeneration genes, while the genes related to gametogenesis and sexual reproduction process most probably are not expressed in fissiparous individuals.

Data accessibility

RNAseq data is accessible under the BioProject accession code PRJNA797284 at the NCBI SRA database and the scripts used are accessible at https://github.com/lisy87/dugesia-transcriptome

Acknowledgements

We are grateful to Silvia Hinojosa, who guided the first steps in the transcriptome assembling process, and Julio Rozas for his valuable suggestions. Additionally, we thank Darren Parker very much for answering our questions regarding the reference construction.

Authors contribution

LBA did the initial study design. LBA and MR contributed to data collection. LBA analysed the data and wrote the initial manuscript. LL helped to analyze the Gene Ontology results. MR, LBA, and LL wrote the final manuscript. Authors declare no conflicts of interest.

References

Bohr, T. E., Shiroor, D. A., & Adler, C. E. (2021). Planarian stem cells sense the identity of the missing pharynx to launch its targeted regeneration. *ELife*, *10*, 1–29. https://doi.org/10.7554/eLife.68830

Osuma, E. A., Riggs, D. W., Gibb, A. A., & Hill, B. G. (2018). High throughput measurement of metabolism in planarians reveals activation of glycolysis during regeneration. *Regeneration*, *5*(1), 78–86. https://doi.org/10.1002/reg2.95

Parker, D. J., Bast, J., Jalvingh, K., & Robinson-rechavi, M. (2018). *Repeated Evolution of Asexuality Involves Convergent Gene Expression Changes*. *36*(2), 350–364. https://doi.org/10.1093/molbev/msy217

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

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

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

Rouhana, L., Tasaki, J., Saberi, A., & Newmark, P. A. (2017). Genetic dissection of the planarian reproductive system through characterization of Schmidtea mediterranea CPEB homologs. *Developmental Biology*, *426*(1), 43–55. https://doi.org/10.1016/j.ydbio.2017.04.008

Saló, E., Abril, J. F., Adell, T., Cebrià, F., Eckelt, K., Fernández-Taboada, E., … Rodríguez-Esteban, G. (2009). Planarian regeneration: Achievements and future directions after 20 years of research. *International Journal of Developmental Biology*, *53*(8–10), 1317–1327. https://doi.org/10.1387/ijdb.072414es

Sekii, K., Yorimoto, S., Okamoto, H., Nagao, N., Maezawa, T., ...Kobayashi, K. (2019). Transcriptomic analysis reveals differences in the regulation of amino acid metabolism in asexual and sexual planarians. Scientific Reports, 9(1), 1-20. https://doi.org/10.1038/s41598-019-42025-z

Stochino, G & Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology, 80*(3), 319-328. https://doi.org/10.1080/11250003.2013.822025

SUPPLEMENTAL INFORMATION

1. Taxon sampling

The samples analysed are part of an evolutionary study of *Dugesia* species from the Western Mediterranean region (Chapter IV). The sampling localities were selected based on previous phylogenetic studies performed in *Dugesia* genus (Lázaro et al., 2009; Leria, Vila-Farré, Solà, & Riutort, 2019). For this study were selected 20 samples (Table S1) coming from 5 localities (Fig. 1B).

Table S1 Sample information including the accession codes in the NCBI-SRA repository

Sample Code	Species	Reproductive Strategy	Region, Country	Locality	Voucher ID	Accession code NCBI-SRA
DsubCangAsex_5	*D. subtentaculata*	Asexual	Asturias, Spain	Cangas	MR1297.5	SRR17642696
DsubCangAsex_6	*D. subtentaculata*	Asexual	Asturias, Spain	Cangas	MR1297.6	SRR17642695
DsubCangAsex_7	*D. subtentaculata*	Asexual	Asturias, Spain	Cangas	MR1297.7	SRR17642694
DsubCangSex_2	*D. subtentaculata*	Sexual	Asturias, Spain	Cangas	MR1297.2	SRR17642693
DsubCangSex_3	*D. subtentaculata*	Sexual	Asturias, Spain	Cangas	MR1297.3	SRR17642692
DsubCangSex_4	*D. subtentaculata*	Sexual	Asturias, Spain	Cangas	MR1297.4	SRR17642691
DetruTrilla_1	*D. etrusca s.l*	Asexual	Catalonia, Spain	Font de la Trilla	MR1265.1	SRR17642705
DetruTrilla_2	*D. etrusca s.l*	Asexual	Catalonia, Spain	Font de la Trilla	MR1265.2	SRR17642704
DetruTrilla_3	*D. etrusca s.l*	Asexual	Catalonia, Spain	Font de la Trilla	MR1265.3	SRR17642703
DetruPie_2	*D. etrusca*	Sexual	Toscana, Italy	Pieve	MR1355.2	SRR17642742
DetruPie_3	*D. etrusca*	Sexual	Toscana, Italy	Pieve	MR1355.3	SRR17642741
DetruPie_4	*D. etrusca*	Sexual	Toscana, Italy	Pieve	MR1355.4	SRR17642740
DsubBosq_1	*D. subtentaculata*	Asexual	Andalusia, Spain	El Bosque	MR1267.1	SRR17642698
DsubBosq_2	*D. subtentaculata*	Asexual	Andalusia, Spain	El Bosque	MR1267.2	SRR17642697
Dvila_1	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque	MR1266.1	SRR17642676
Dvila_2	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque	MR1266.2	SRR17642675
Dvila_3	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque	MR1266.3	SRR17642674
DsubStFe_1	*D. subtentaculata*	Asexual	Catalonia, Spain	Santa Fe	MR1082.1	SRR17642680
DsubStFe_2	*D. subtentaculata*	Asexual	Catalonia, Spain	Santa Fe	MR1082.2	SRR17642679
DsubStFe_3	*D. subtentaculata*	Asexual	Catalonia, Spain	Santa Fe	MR1082.3	SRR17642678

The animals were preserved *in situ* or in the laboratory with RNAlater (SIGMA) following the recommendations of the manufacturer. The sexuality or asexuality of the individuals was assessed observing them under the stereomicroscope and taking into account previous information about known populations. Sexual individuals were recognized by the presence of the gonopore, the external aperture of the copulatory apparatus. The asexual individuals were classified by the presence of the blastema, the regenerating bud formed were the fission of the individual has taken place or by the absence of gonopore. Each population was assigned to one or both reproductive strategies, depending on the anatomic characteristics of the sampled animals.

Total RNA was extracted using Trizol (Thermo Fisher Scientific, USA) following the manufacturer's instructions. The quantification and the integrity were assessed with Qubit and Bioanalyzer in the Centres Científics i Tecnològics, Universitat de Barcelona (CciTUB). Truseq stranded and ribo-zero libraries were constructed in Macrogen Inc., (Macrogen Europe, Madrid) to obtain Illumina paired-end reads (Chapter IV.1).

2. Testing different references

Since *Dugesia* genus do not have reference genome, we needed to find the best strategy to perform the differential expression analyses with our samples, that belong to different populations and different species. We tested different candidate genomes to select the reference with higher mapping percent after mapping with STAR (Dobin et al., 2013). For these tests we used the sexual and asexual samples from Cangas population and two candidate references: 1) the genome of *Schmidtea mediterranea* V4 downloaded from PlanMine database (Rozanski et al., 2018), and 2) the genome of *D. japonica* V1 downloaded from the genome browser PLANARIAN.JP (http://www.planarian.jp). The mapping percent of samples from Cangas was very low (15%) discarding the genome of *S. mediterranea* as reference. On the other hand, the percent of mapping against the *D. japonica* genome was around the 50%. However, the annotation of this genome is limited and hampering its use as reference.

In addition, we tried to assemble a transcriptome using all samples, but we only obtained a super transcriptome with the sum of all transcripts from all the samples used. This was probably a consequence of the diversity of samples analysed (Benítez-Álvarez et al., in preparation). Even using only sexual and asexual samples from Cangas (the same population and the same species) we obtained a supertranscriptome with an exaggerated number of transcripts.

We finally used the strategy described by (Parker, Bast, Jalvingh, & Robinson-Rechavi, 2018) to built a reference for every sample, including a set of single copy orthologs shared between all of them. The transcriptome of every sample was *de novo* assembled and used in a orthologs search. The set of single copy orthologs was selected, extracting the sequences belonging to every sample independently to built a reference for every sample. With this reference, we lost valuable information regarding paralogous genes with important functions in the fissiparity or sexuality, but using different populations and species, we could recover the differential expression patterns maintained across different environments and lineages.

3. Bioinformatic Workflow

We used data previously processed in Chapter IV.1. Briefly, the raw reads were filtered by quality and length with Trimmomatic (Bolger, Lohse, & Usadel, 2014). Paired reads were *de novo* assembled using Trinity v2.9.1 (Grabherr et al., 2011; B. J. Haas et al., 2013) following the default options. Transcripts were clustered using CD-HIT EST (Li & Godzik, 2006), applying a sequence identity threshold of 0.99 and filtered using a strategy based on Blobtools v 3.6 (Challis & Paulini, 2021) results. The transcripts matching against other groups different to Platyhelminthes were dropped out. Filtered transcripts were translated to proteins using TransDecoder v 5.5.0 (B. Haas & Papanicolaou, 2019) and the longest isoforms were selected using the script “choose_longest_iso.py” (Cunha & Giribet, 2019) (See (Benítez-Álvarez et al., in preparation) for a broad description of these steps).

4,533 single-copy orthogroups (SC) were identified by Orthofinder v 3.6 (Emms & Kelly, 2019) from the analysis of longer isoforms sets belonging to every sample. The nucleotide sequences of the SC were extracted from Orthofinder and Transdecoder outputs as is described in (Benítez-Álvarez et al., in preparation). The set of SC were each aligned independently with MAFFT v 7.487 (Katoh & Standley, 2013) using the --auto and --maxiterate options, with 1000 iterations. The alignments were trimmed and filtered by length (minimun 300 positions) using the scripts lignment_trimmer.py and fasta_select_by_len.py wrote by Darren Parker (Parker et al., 2018). The independent reference by species was obtained with simple bash commands. 20 files (one by sample) with 4,191 SC were retained at the end of the process.

To assess the diversity and completeness of the generated references, we predicted the ortholog genes found in the Metazoa database (ODB 10) using BUSCO V4 software (Manni, Berkeley, Seppey, Simão, & Zdobnov, 2021). Since the percent of BUSCO genes found in the references were very homogeneous (22.8-23.5%) and we were working with ortholog

groups, we selected one sample (DsubBosq_1) to perform the functional annotation of all references using the Trinotate pipeline (Bryant et al., 2017). 20 references with 3,373 annotated genes were built at the end. Orthologs shared only by asexuals and by sexuals were selected from OrthoFinder output with simple bash commands.

The abundances of the genes were calculated for every sample using Kallisto (Bray, Pimentel, Melsted, & Pachter, 2016) and imported in R (Team, 2021) to be analysed with the DESeq2 package (Love, Huber, & Anders, 2014). The enrichment analyses were done with the run_Goseq.pl script from Trinity package. The gene ontology results were summarized in REVIGO (Supek, Bošnjak, Škunca, & Šmuc, 2011) with the default values. The R script for run DGE analysis is available at (https://github.com/lisy87/dugesia-transcriptome)

Supplemental References

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*, 525–527. https://doi.org/10.1038/nbt.3519

Bryant, D., Johnson, K., DiTommaso, T., Tickle, T., Couger, M., Payzin-Dogru, D., … Whited, J. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep.*, *18*(3), 762–776. https://doi.org/10.1016/j.celrep.2016.12.063

Challis, R., & Paulini, M. (2021). *blobtoolkit/blobtools2: v2.6.1*. https://doi.org/10.5281/zenodo.5032179

Cunha, T. J., & Giribet, G. (2019). A congruent topology for deep gastropod relationships. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1898). https://doi.org/10.1098/rspb.2018.2776

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

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

Grabherr, M., Haas, B., Yassour, M., Levin, J., Thompson, D., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Haas, B., & Papanicolaou, A. (2019). *TransDecoder 5.5.0*. Retrieved from https://github.com/TransDecoder/TransDecoder/wiki

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

Lázaro, E. M., Sluys, R., Pala, M., Stochino, 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

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. https://doi.org/10.1186/s12862-019-1440-1

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Love, M., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*. https://doi.org/10.1186/s13059-014-0550-8

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Parker, D. J., Bast, J., Jalvingh, K., & Robinson-rechavi, M. (2018). Repeated Evolution of Asexuality Involves Convergent Gene Expression Changes. *Mol. Biol. Evol 36*(2), 350–364. https://doi.org/10.1093/molbev/msy217

Rozanski, A., Moon, H., Brandl, H., Martín-Durán, J. M., Grohme, M., Hüttner, K., … Rink, J. C. (2018). PlanMine 3.0—improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky1070

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). 'REVIGO summarizes and visualizes long lists of Gene Ontology terms'. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0021800

Team, R. C. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org/

GENERAL DISCUSSION

GENERAL DISCUSSION

In this thesis, I used genetic data to elucidate the mechanisms that shape the diversity and distribution of several freshwater planarian taxa. I used different methodologies to build phylogenetic trees and interpret the topologies obtained; from canonical relationships among species and groups to unexpected branching patterns, never seen before. The results showed complex evolutionary processes in Triclads, shaped by adaptation to new habitats, geological events, reproductive strategies, and even human intervention.

In this discussion section, I will develop the following issues: 1) the methodological frameworks used here and how every tool can contribute to obtain supported results, depending on the data accessibility and the available previous information about the in-question taxa; 2) the processes that drive the evolution of Triclads, such as the geological changes, the new habitat adaptation, and the reproduction mode; 3) the low GC content in the Dugesiidae family, and 4) the future perspectives and challenges of Evolutionary Biology in a fissiparity context .

1. From one to hundreds of genes. An improvement in resolution and tree inference

Phylogenetic studies have evolved from the use of holoenzymes to the analysis of entire genomic data sets. The use of genomic or transcriptomic data in phylogenetic inference constitutes an increasingly common practice. However, this data is not accessible for all taxa or for all research groups. Besides the cost of high-throughput sequencing, bioinformatic skills and computational power are necessary for data analyses. Also, a prior knowledge of the group in question should be taken into account. When there is no information regarding the evolutionary relationships of the group or those are only based on morphological characters, it may be convenient to use a first approach with few markers to generate a phylogenetic hypothesis on which to design the phylogenomic study. In this thesis, we used both approaches. Henceforth, I am going to comment on the advantages of each one, and when it is more convenient to use one or another. I will analyze these issues through the studies described in each of the previous chapters.

1.1. Few genes for a first phylogenetic approach

The use of molecular data constituted a great step in the evolutionary studies of flatworms since morphological synapomorphies to define the major clades are scarce (Baguñà & Riutort, 2004). Because of their homogeneous evolutionary rates across lineages, maintained by concerted evolution, ribosomal genes have a strong phylogenetic signal and have been broadly used in evolutionary studies (Hillis & Dixon, 1991; Mindell & Honeycutt, 1990; Patwardhan, Ray, & Roy, 2014). The 18S sequences allowed the reconstruction of the first molecular phylogeny of Platyhelminthes (Carranza, Baguñà, & Riutort, 1997) as well as the separation of acoels flatworm from this phylum, contributing to a better understanding of Bilaterians evolution (Ruiz-Trillo et al., 1999). These markers have been, and are currently, very useful in studies at high classification levels inside Platyhelminthes (Campos, Cummings, Reyes, & Laclette, 1998; Carranza et al., 1998; Laumer & Giribet, 2014; Olson & Caira, 1999; Oya & Kajihara, 2020; Pérez-Ponce de León & Hernández-Mena, 2019; Rohde et al., 1993; Waeschenbach, Webster, Bray, & Littlewood, 2007).

In Chapter I, fragments of 18S and 28S ribosomal genes were used to elucidate the internal relationships among Tricladida suborders. The scarce representatives of Cavernicola and their broad distribution around the world, make difficult the study of this group. However, here we were able to analyse, for the first time, the internal relationships of Tricladida with an almost complete taxon sampling of Cavernicola, except for genus *Balliania*. A supported phylogeny was obtained showing the monophyly of the three suborders and the sister relationship between Cavernicola and Maricola (Fig. 2 in Chapter I). Based on these results it was proposed an evolutionary hypothesis for the evolution of these major groups, which contrasts with the morphological proposal (Sluys et al., 2009; Sluys, 1990) and matches with previous molecular analysis performed with a very limited data set (Harrath et al., 2016). Without previous data in hand and poor accessibility to fresh samples of all representatives analysed here, the use of a concatenated data set of ribosomal markers was the best option to perform the first approach to a comprehensive molecular phylogeny of Tricladida at the suborder level and the internal relationship within Cavernicola.

Nuclear and mitochondrial genes also have been good markers for phylogenetic studies in triclads (see Álvarez-Presas & Riutort, 2014 for a broader

review). Since the ribosomal cluster is duplicated in Dugesiidae (Carranza et al., 1999; Carranza, Giribet, Ribera, Baguñà, & Riutort, 1996), their use in this family is more difficult because the orthologous sequences have to be correctly identified. In addition, at lower taxonomic levels mitochondrial and nuclear non-ribosomal genes are more useful to reconstruct phylogenetic relationships and population structure (Arif & Khan, 2009).

A region of the mitochondrial gene *COI* and a fragment of the *EF1a* nuclear gene were used to study, also for the first time, the diversification of the *Girardia* genus and its expansion out of its natural distribution area in Chapters II and III. Although the taxon sampling was limited, taking into account the great diversity of this genus, the Fig. 3 in Chapter II shows the largest *Girardia* phylogeny obtained until the moment and points out important issues regarding the evolution of this genus in the American continent, as well as its hidden diversity. With this study an important and old question, previously outlined by (Sluys et al., 2005) was resolved; the differentiation of the species *G. tigrina* from South and North American regions, erecting the South American lineage as a new species. Additionally, other important taxonomic issues were unveiled; a) the high differentiation of *G. schubarti* and related branches from the rest of *Girardia* species and their broad distribution, b) the etymologically incorrect name of *G. sinensis*, which does not describe the natural origin of the species.

Furthermore, the use of molecular data not only allows to analyse a great number of samples without the expertise and effort that is necessary for the morphological approach, but it is the only feasible method to include asexual populations. In Chapter III was used genetic information to describe the expansion of *Girardia* out of its native distribution, with a majority of asexual populations spread across non-American regions. Using *COI* and *EF1a* as DNA-Barcoding markers in a phylogenetic framework, the presence of *G. sinensis* was detected for the first time in Africa. Additionally, the expansion process of three North American species, *G. tigrina, G. dorotocephala, and G. sinensis,* through Europe and other regions in the world were also broadly discussed.

Along the three first chapters, specific molecular markers were used. Taking into account the scarce previous information for the taxa analysed in these sections, and the number of samples included in the studies, this strategy was the best option. In these chapters were obtained results that help to understand the evolution of the

three major groups of Triclads and one of the most diverse genera. In all cases, the best resolved and supported trees were obtained from concatenated data sets, applying the most appropriate evolutionary substitution models in the partition by gene and codon schemes. This strategy, combined with ML and BI inference methods is the most advisable and reliable methodology for small data sets analysis (Castoe, Doan, & Parkinson, 2004; Kainer & Lanfear, 2015; McGuire, Witt, Altshuler, & Remsen, 2007) , and it is frequently used (Baca, Toussaint, Miller, & Short, 2017; Guil et al., 2022; Hernández et al., 2022; Rota & Wahlberg, 2012). In the case of *Girardia,* only BI-based results were shown. Although ML analyses with intense bootstrapping were performed, obtaining the same topology, the bootstrap values were low. It was hypothesized that the high amount of missing data in the alignments due to differences in sequence length can affect the replacing process of the bootstrap calculation (Felsenstein, 1985b; Efron et al., 1996). The effect of the missing data in phylogenetic inference in concatenated data sets has been broadly discussed (Lemmon, Brown, Stanger-Hall, & Lemmon, 2009 and references therein). However, studies based on empirical and simulated data support the benefice of including missing data over the total elimination of genes or taxa (Fulton & Strobeck, 2006; Jiang, Chen, Wang, Li, & Wiens, 2014; Wiens, 2003, 2006) and the outperformance of the BI method when analysing alignments with high missing data content (Wiens & Morrill, 2011 and references therein). Therefore, we decided to include the majority of taxa and genetic information available and to use only the BI method.

1.2. Hundreds of genes: the phylogenomic framework

With the emergence of high through-put sequencing, phylogenomics has gained space in evolutionary studies. Moving from a few markers to hundreds of *loci* implicates a major effort and more time in data processing steps but allows exploring other frameworks and methodologies. The analysis of genomic data sets permits elucidating not only phylogenetic relationships, but also important milestones in the evolution of the Tree of Life, just as gene gain/loss, signatures of genomic adaptation, and chromosome evolution (Buffalo & Coop, 2019; Fernández & Gabaldón, 2020; Guijarro-Clarke et al., 2020; Iranzo, Wolf, Koonin, & Sela, 2019; Kapheim et al., 2015; Rodríguez et al., 2017; Roesti, Gavrilets, Hendry, Salzburger, & Berner, 2014; Sacerdot, Louis, Bon, Berthelot, & Roest Crollius, 2018; Salmón et al., 2021; E. G. Smith et al., 2022; Wu et al., 2022).

In Chapter IV transcriptomic data was used to study the evolution of the *Dugesia* genus in the Western Mediterranean. In this case, previous information was available (Lázaro et al., 2009; Leria et al., 2019; 2020; *in press*; Solà et al., 2013; *in press*) to design the study for the sake of answering the remaining questions about the evolutionary history of *Dugesia* in this region.

The new strategy is based on RNAseq data and it is quite different, from the sample conservation to the data analyses (Fig. 2 in Chapter IV.1), from previous strategies. For *de novo* assembly was used Trinity, which outperformance over other softwares has been broadly demonstrated (Henschel et al., 2012; Bankar, Todur, Shukla, & Vasudevan, 2015; Wagner, Fulton, & Henschel, 2016). Actually, Bridger (Chang et al., 2015), another assembler with high performance (Rana, Zadlock, Zhang, Murphy, & Bentivegna, 2016), was tested but in general better assembly statistics were obtained with Trinity (data not shown). Additionally, Trinity offers valuable tools for downstream analysis with RNAseq data (Haas et al., 2013).

One of the most sensitive steps in the workflow is ortholog inference due to its great connotation in phylogenetic inference. Although major errors can be achieved when distantly related taxa are analysed, the selection of the orthologs inference method is a critical step in any phylogenomic study (see Natsidis et al., 2021 for a broader review). OrthoFider was used to perform different ortholog searches in several data sets. It has been shown that this program represents an improvement in accuracy with respect to other *de novo* orthologs inference methods (Emms & Kelly, 2015). In addition, OMA; another method based on pre-established orthologs reference sets, was used to discard the idea that the ladder pattern seen for some taxa could be an artefact of orthologs inference (Chapter IV.2). Although OMA discovered a number of SC greater than OrthoFinder, both sets of SC yield the same tree topology. For transcriptome data, OMA could have advantages over OrthoFinder since it is computationally cheaper, and it is less susceptible to the incompleteness of data (Kapli, Yang, & Telford, 2020; Roth et al., 2008). However, this assumption must be taken with caution, since not all taxonomic groups have the same representation in databases. In the case of Platyhelminthes, it has been reported an important amount of hidden orthologs in transcriptomic data (Martín-Durán, Ryan, Vellutini, Pang, & Hejnol, 2017). In addition, OMA is known for having higher precision (proportion of predicted orthologs that are correct) but lower recall (proportion of true orthologs that are correctly

predicted) than most other methods (Adrian M Altenhoff et al., 2019). In this case, both OMA and OrthoFinder were previously tested in a small data set (data not shown), and certainly, OMA recovered more orthogroups than OrthoFinder, but 79% of the OrthoFinder orthogroups were included in the OMA output. Additionally, in this specific case, both programs took a similar time to run the analysis. Nevertheless, a high number of SC were discovered by OrthoFinder, which I think, has a very well-implemented algorithm that largely guarantees solid and credible results, without forgetting all the valuable information provided by its output.

Regarding the tree inference, mixture models were used to apply sequence substitution models, avoiding the problems associated with the variation of substitution rates across sites (Philippe et al., 2005). With both BI and ML frameworks, highly supported trees were obtained. However, most analyses were performed using the ML approach, since the analyses of large data sets with BI and mixture models are highly time-consuming and no differences in topology were found in the cases where both methods were used (Fig. 3 Chapter IV. 1). Thus, IQ-TREE, which has a very efficient implementation of an ML framework (Minh et al., 2020), was used to perform a major part of the analyses. For protein data 20 categories (or profiles) were used since three sets of profiles (C10, C20, and C60) were firstly tested, and no differences in topology, branch lengths, or support values were observed. Moreover, it has been demonstrated that C20 is enough to obtain accurate results (Quang et al., 2008). Since the CAT model is not implemented in IQ-TRE for nucleotide data, three components (JC, HKY, and GTR) with the MIX option were used to perform the analyses with DNA sequences (Minh et al., 2018).

With so many *loci* it is expected an increase in gene tree discordance. To reconcile this issue, in addition to the traditional BI and ML frameworks that work with concatenated matrix, an MSC summary method implemented in ASTRAL-pro was used. Matching with studies in other groups (Gonçalves, Simpson, Ortiz, Shimizu, & Jansen, 2019; Sann et al., 2021), inconsistencies between the topologies obtained by ML and MSC methods were found (Fig. S1 in Chapter IV.2). This contrast can rest in the differences behind concatenation methods and MSC (Fig. 3 in Introduction). The principal advantage of the MSC framework is that it allows the assessment of gene discordance sources, as ILS and introgression, while concatenation-based methods are unable to infer the correct species tree under high gene tree discrepancy (Degnan

& Rosenberg, 2009; Shen, Steenwyk, & Rokas, 2021). The application of the MSC method has been criticised based on the supposed violation of phylogenetics principles. However, these principles, applied mainly to concatenation methods, are invalid or misunderstood in a phylogenomic framework. Moreover, the efficiency and accuracy of this method for the analysis of large datasets are undeniable (Edwards et al., 2016).

In addition, ASTRAL-pro provides a statistically consistent estimation of the true species tree that agrees with the largest number of quartet trees induced by the set of the input gene trees (Mirarab et al., 2014). Based on the number of quartet trees in gene trees that are present in the species tree, the software calculates the quartet score of the tree topology, normalised by the total number of quartet trees in the input gene tree (Mirarab, 2019). Under this method, the support of the species tree is obtained for the branches. The local posterior probabilities are computed based on gene tree quartet frequencies around the branch implementing Bayes's rules. Thus, the algorithm yields the local posterior probability of the branch being correct, given the input gene trees (Sayyari & Mirarab, 2016).

These support measures are more informative than the traditional posterior probability and bootstrap support of BI and ML respectively and allow comparison of the effect of specific branches in the tree topology. In Chapter IV. 2, the final normalized quartets score (FNQS) and the quartet supports were used as measures of the tree discordance introduced in the species tree by asexual lineages (Fig. 5 Chapter IV.2). Therefore, under a complex evolutionary scenario such as the one proposed for the diversification of the Iberian-Apennines-Alps-plus clade, I think that MSC methods offer a more accurate framework for species tree inference and evolutionary process detecting.

Although several troubles are associated with phylogenetic inference (Sanderson & Shaffer, 2002), these new methods and strategies, applied to large informative data sets, overcome many of these issues (Philippe et al., 2005; Kapli, Yang, & Telford, 2020). Therefore, better resolution and support than the previous studies in *Dugesia* from the Western Mediterranean (Lázaro et al., 2009; Leria et al., *in press*) were obtained using these methodologies. The causes of this improvement lie in both the increase of the data set and the methodology used. In the first place, working with hundreds of coding regions under different evolutionary constraints, prevents the effect of intraindividual haplotype diversity described in this taxon. In

addition, the use of more sophisticated evolutionary model schemes and species tree inference methods allows for better retrieval of the genealogical information over the intraindividual genetic diversity, which can mask the phylogenetic relationships.

Another important result, with high methodological connotation, was demonstrating that the Dunuc-12 marker is not a paralog, validating its utility in intraindividual diversity analyses, and supporting the Mosaic-Meselson effect, described based on data provided by this marker (Leria et al., 2019). Then, Dunuc-12 was used to calculate intraindividual genetic diversity parameters as proof of long-term asexuality in *D. etrusca s.l* and *D. liguriensis s.l.* These analyses gave support to the hypothesis that the ladder-like pattern is an effect of ancient fissiparity.

Finally, with the great amount of transcriptome data available, and the concern around the sexual/asexual reproduction modes in freshwater planarians, it was very interesting to do a differential expression analysis between sexual and asexual individuals from different species and habitats combinations (Fig. 1 in Chapter V). This constitutes the first study that analyses the differential expression between sexual and asexual dugesids in natural habitats. The results suggested a differentiated metabolic activity between sexual and asexual planarians shared across different environments and evolutionary lineages. Because there is not a full annotated genome of *Dugesia*, a limited set of orthologs was used as the reference. Therefore, these results are very preliminary.

Although all results of the Chapters IV and V were based on transcriptomics, it should not be forgotten that DNA-Barcoding techniques were necessary for the first identification of the samples, proving again their value as the most effective and fast method for that in Dugesiidae, as has been demonstrated also in other taxonomic groups (DeSalle & Goldstein, 2019; Jannah, Hariri, Kasiamdari, & Handayani, 2021; Vu et al., 2019) .Besides, the Sanger method was used to obtain *COI* and *Dunuc-12* sequences. This shows that the methodologies are not isolated, and every study can combine them according to the objectives and the characteristics of the analysed taxa.

In addition, despite the demonstrated outperformance of phylogenomic analysis in evolutionary studies, it is not always possible to have access to high throughput sequencing methods, and the jump to these methodologies can be easier with previous information about the evolutionary scenery of the group in the study. All

methodologies are valuable, depending on the questions. The value of the work does not depend only on the methodology used, depends on major measures of the researchers' expertise and how they are able to use the means at their disposal to test hypotheses using the scientific method.

2. Processes that drive the diversification and distribution in freshwater planarians

Understanding the processes that govern the distribution of biological diversity is key to explaining the evolution of life. Throughout the chapters of this thesis, it can be seen how the diversification and distribution of freshwater planarians have been affected by different geological, ecological, genetic, and even anthropogenic processes, which have led to the diversification of the group in important biodiversity hotspots of the planet.

2.1. Geological events

Two biogeographic processes have been proposed to explain the distribution of biodiversity. 1) Vicariance: the distribution area of an ancestor became fragmented by geographic barriers (mountains, oceans). Then, the geographic division is followed by allopatric speciation, leading to geographically distant and isolated populations that share a common ancestor. 2) Dispersal: the ancestor inhabits an area that constitutes the "centre of origin," from which it disperses to other areas by crossing a geographic barrier. In the vicariance the barrier causes the separation, thus the geographical event sets a maximum for the resulting allopatric speciation. On the other hand, in dispersal, the barrier exists before the spread events (Sanmartín, 2012). Since freshwater planarians are not resistant to salinity, it was proposed that the major diversification events in the group are driven by vicariance (Ball, 1974; 1975). However, recent analyses based on molecular data of representatives of *Dugesia* all along its area of distribution have shown that also dispersal processes are needed to explain the diversification of the genus (Solà et al., *in press*). Therefore, the two types of processes have led to the diversification of this group.

The breakup of Pangea, the supercontinent that included a great part of the emerged land during the Paleozoic and Mesozoic eras (Frizon de Lamotte, Fourdan, Leleu, Leparmentier, & De Clarens, 2015; Guan, Geoffroy, & Xu, 2021), followed by the breakup of the Gondwana supercontinent in the Cretaceous-Paleogene periods

(Eagles & Vaughan, 2009; Jokat, Boebel, König, & Meyer, 2003; Veevers, 2004) have been the most studied geological processes producing vicariant events, because resulted in the diversification of most taxonomic groups. Tricladida are among them. In Chapter I, it was hypothesised that the Cavernicola suborder could have differentiated on Gondwana and perhaps already on Pangea. This hypothesis seems plausible given the broad distribution of the scarce representatives described in this group and their phylogenetic relationships (Figs. 1 and 3 in Chapter I). Recently, a new Cavernicola family, Amphibioplanidae, has been erected to house the *Amphibioplana onnisi* species, from caves on Sardinia island (Stocchino et al., 2021), adding a new point in the scattered distribution of this group. This distribution suggests that the common ancestor of Cavernicola lineages may have lived on Gondwana before this supercontinent started to break up and following vicariance and dispersion events shaped the disjunct distribution observed currently. Therefore, with the scarce information in hand, only bare hypotheses about the biogeography of this group can be outlined. Aditionally, in Tricladida Ball (1974) biogeographical hypothesis proposes the event of separation of America from Africa within the Gondwana breakup as the cause of *Girardia*'s separation from *Schmidtea* and *Dugesia genus.* Although recently the first split in *Dugesia* diversification has been dated approximately in 150 Mya, in the superior Jurassic based on genetic data (Solà et al., *in press*), predating the breakage of America from Africa. It must be taken into account that both molecular and geological datation are very imprecise. These important geological events explain also the diversification and the distribution patterns of flora and fauna in the Austral region (Chávez Hoffmeister, 2020; McLoughlin, 2001; Reguero et al., 2014). Moreover, a similar hypothesis to that of the Cavernicola suborder has been proposed to explain the diversification of freshwater amphipods (Cannizzaro & Berg, 2022).

Furthermore, dispersal events related to geological changes also have played a fundamental role in planarians diversification. The closure of the Panama Isthmus, for instance, has shaped the biodiversity in The Americas because it allowed the biotic dispersal between the two American landmasses, leading to the Great American Biotic Interchange (GABI) (Cody, Richardson, Rull, Ellis, & Pennington, 2010; Webb, 2006; Weir, Bermingham, & Schluter, 2009) . In Chapter II it was proposed that *Girardia* diversified on the South American portion of Gondwanaland, and the North American region was colonised by independent dispersion waves from the Neotropics through

freshwater tracks, during the emergence of the Isthmus of Panama. This closure has long been considered to have occurred in the Pliocene, around 3.5-3 Mya (Coates, 1992; Coates, Collins, Aubury, & Berggren, 2004; Coates & Stallard, 2013; Haug & Tiedemann, 1998; Haug, Tiedemann, Zahn, & Ravelo, 2001). However, new studies bring to light another hypothesis about this geological event. Based on molecular and fossil data Bacon et al., (2015) reported two significant waves of dispersal of terrestrial organisms approximately 20 and 6 Mya and parallel events splitting marine organisms between the Atlantic and Pacific oceans around 23 and 7 Mya. On the other hand, a new study based on stratigraphic analysis set the closure of the Central American Seaway in the Middle Miocene, approximately 11 Mya (Montes et al., 2015). However, (O'Dea et al., 2016) dated this geological event around 2.8 Mya or 3.5 Mya. Although some authors have been very critical of this work (Jaramillo et al., 2017; Molnar, 2017), others have set the closing of the Panama Isthmus at 3 Mya (Stange, Sánchez-Villagra, Salzburger, & Matschiner, 2018) and 0.6 Mya (Brikiatis, 2021). Recently, a new theory based on geodynamic mechanisms could explain the discrepancy in these hypotheses, reconciling the different lines of evidence used to explain the geological process underlying the Isthmus of Panama closure, characterised by intermittent connections (McGirr, Seton, & Williams, 2021). Several biogeographical studies show that different migratory waves occurred during the GABI, reflecting intermittent connections due to sea-level changes and climatic dynamics (Domingo, Tomassini, Montalvo, Sanz-Pérez, & Alberdi, 2020 and references therein). Therefore, with the present data it is difficult to set a date for the dispersion of *Girardia* from Southern land masses to Northern, but the phylogenetic tree (Fig. 3 in Chapter II) suggests at least two dispersal events, that could have occurred at different moments.

Meanwhile, *Dugesia* was diversifying in the Eastern hemisphere. Currently, the genus occupies freshwater bodies in Africa, Europe, the Middle East, South Asia, Far East, and Australasia, being the Mediterranean one of the zones with the most suitable habitats for *Dugesia* species (Solá et al., *in press*). The Mediterranean region is distinguished by its high diversity and endemism (Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000), product of diverse factors such as its geological history, high topographical and ecological heterogeneity, paleoclimatic and paleogeologic events, Messinian salinity crisis effects, sea-level oscillations during the glaciations of the Pleistocene (Allen, 2001). Different plate tectonic events such as continental

convergence, collisions, isolation, and microplate movements have moulded the current Mediterranean landscape. These geological processes have delimited two biogeographical zones separated by a geographic North-South line from the South of Sicily to the Cap Bon, in the northeast extreme of Tunisia, with high levels of diversity and endemism, and a differentiated biota in the two regions (Blondel, Aronson, Bodiou, Boeuf, & Fontaine, 2015).

In Chapter IV the biogeographical hypothesis proposed by Leria et al., (*in pres*) to explain the diversification of *Dugesia* in the Western Mediterranean was discussed. According to Leria et al., (*in press*), the paleogeographic events that occurred in the Western Mediterranean during the Oligocene-Miocene (33.9 - 5.3 Mya) played an important role driving the diversification of the *Dugesia* species in this region (Fig. 7 in Chapter IV.1). However, the inclusion of representatives of *D. ilvana* and of the asexual populations from the Iberian Peninsula belonging to the Iberian-Apennines-Alps-plus clade complicates the interpretation of the phylogeographical history of this special lineage (Chapter IV). The new data indicates that the diversification of this clade was quite complex, with the first split of *D. gonocephala* followed by the diversification of the group of *D. etrusca* and *D. liguriensis*. The ancestor of this clade could had passed to the Iberian Peninsula through fauna corridors present in the Eastern Pyrenees and remained there as an asexual lineage. The high altitude of the Alps and Apennines, on the other hand, could have acted as isolation barriers preventing the expansion of *D. liguriensis s.s* and *D. estrusca s.s* to lower elevations. Regarding *D. ilvana*, I think that the most reasonable scenario is a sister relationship with *D. etrusca s.s,* which is biogeographically most plausible. According to the geological history of Elba island and its connection with the Tuscany coast (Bortolotti, Fazzuoli, et al., 2001; Bortolotti, Pandeli, & Principi, 2001), *D. ilvana* could have split from *D. estrusca s.s* around 5 Mya.

These geological processes also have driven the phylogeographic patterns of other groups in this region. With a more ancient diversification process, but lower species number and minor distribution range than *Dugesia*, the current diversity of the *Schmidtea* genus in the Mediterranean basins has been shaped by paleogeological and paleoclimatic events during the Oligocene, the Last Glacial Maximum, and the fragmentation and migration of the microplates (Lázaro et al., 2011; Leria, Sluys, & Riutort, 2018) . Moreover, these events have affected the biogeographic history of other non Platyhelminthes lineages in Mediterranean regions, just as insects, spiders,

earthworms, land snails, and plants (Bidegaray-Batista & Arnedo, 2011; Cetlová, Zozomová-Lihová, Melichárková, Mártonfiová, & Španiel, 2021; Fochetti, Sezzi, Tierno de Figueroa, Modica, & Oliverio, 2009; Opatova, Bond, & Arnedo, 2016; Pérez-Losada, Breinholt, Porto, Aira, & Domínguez, 2011; Pfenninger et al., 2010) .

2.2. Habitat shifting; the caves as a refuge

Shift to newly available habitats is one of the first responses of organisms to a change in the conditions of their current habitat. Lately, this issue has gained strength in ecological studies because of its implications for the response of the organism to the current climate change (Devictor et al., 2012; Scheffers & Pecl, 2019; Wallingford et al., 2020). However, this is not a novel strategy. Important habitat shifts have occurred during the Quaternary climate changes (Davis & Shaw, 2001; Abellán, Benetti, Angus, & Ribera, 2011; Patnaik, Singh, Paul, & Sukumar, 2019). Additionally, habitats shift as an adaptive response to competence or depredation has shaped the evolutionary history of several taxa (Hata et al., 2017; Keren-Rotem, Bouskila, & Geffen, 2006; Latorre et al., 2021; Shy, 1984; Stoks & McPeek, 2006).

Shifts from epigean to hypogean habitats have been part of the evolutionary history of Triclads. In Chapter I, was analysed the evolution of the Cavernicola suborder pointing out a hypothesis to explain its complex diversification process. A scenario where an epigean freshwater diversifies by adapting to different continental habitats was sketched. Nevertheless, the Ancestral States Reconstruction analysis that supports this hypothesis, should be revised in the light of the new branch with an amphibious lifestyle described from Sardinia island recently (Stocchino et al., 2021). In any case, the diversity of habitats in this group and their internal phylogenetic relationships constitute irrefutable evidence that the hypogean habitat does not constitute an ancestral synapomorphic character of the clade. Therefore, the habitat shift could be a strategy to survive climate changes or competition, and the current scattered distribution shows the effect of important diversity loss throughout the evolutionary history of this old group.

Further, in America, the *Girardia* genus shows high diversity of cave-dwelling species. *G. mckenziei* was described in a cave from Chiapas, Mexico (Mitchell & Kawakatsu, 1973), while specimens classified as *G. arimana* were collected in Tamana cave in Trinidad Island (Trinidad y Tobago, Caribbean region), together with numerous

non identified flatworms (Kawakatsu & Mitchel, 1984). Additionally, in the last years, nine cave-dwelling species belonging to *Girardia* genus have been described from caves in Brazil, with high diversity in their troglobitic characters (De Souza, Nunes Morais, Cordeiro Medeiros, & Leal-Zanchet, 2015; De Souza, Morais, Bichuette, & Leal-Zanchet, 2016; Morais, Bichuette, Chagas-Júnior, & Leal-Zanchet, 2021; Hellmann, Ferreira, Rabelo, & Leal-Zanchet, 2020; Hellmann, Leal-Zanchet, & Lopes Ferreira, 2018). Moreover, two possible candidate species from distant caves were included in the analyses of Chapter II (samples 401 and 1178, Table S1, Fig. 3). Although epigean species outnumber hypogean ones, the high number of cave-dwelling species indicates that *Girardia* is highly successful in adapting to life in caves, and just as in Cavernicola, it could constitute a response to habitat conditions changes. Unfortunately, no representatives of most hypogean *Girardia* species were included in the molecular analyses, whereby no information regarding the evolutionary history of these species into the genus is currently achievable.

Transitions to hypogean lifestyle, with high variability in troglobitic characters, exist also in *Dugesia* genus. *D. brigantii*, supposedly belonging to the Iberian-Apennines-Alps-plus clade but not analysed here, was described from a cave in the Ligurian region (De Vries & Benazzi, 1983). *D. myopa* from Grotte d'Andranoboka in Madagascar, and from Indonesia *D. uenorum* and *D. leclerci* (De Vries, 1988; Kawakatsu & Mitchell, 1995). *D. batuensis* and *D. deharvengi* are two hypogean species described from different caves in the Southeast of Asia, and without sister relationship despite being from the same region (Khang, Tan, Panha, & Mohamed, 2017).

In addition, numerous subterranean species belonging to Dendrocoelidae and Planariidae families have been described around the globe (Gourbault, 1972; Gourbault, Benazzi, & Hellouet, 1976; Harrath, Sluys, Ghlala, & Alwasel, 2012; Sluys, 2012; Stocchino, Sluys, Marcia, & Manconi, 2013; Stocchino, Sluys, Montanari, & Manconi, 2017). Furthermore, it is remarkable the high number of species belonging to these families described in Romania. *D. obstinatum* highlighted because it was the first fully described triclad from the chemoautotrophic ecosystem of Movile Cave. This cave lacks any input of photoautotrophically based food; however, it is inhabited by a rich invertebrate community with high endemism value (Stocchino, Sluys, Kawakatsu, Sarbu & Manconi, 2017).

Therefore, the pass to hypogean habitats could constitute an adaptive strategy in Triclads, repeated across the group without any phylogenetic trace, and where the troglobite characters are reached by convergence. Even though the name of a monophyletic suborder makes reference to an underground lifestyle, these taxa could constitute relics of an ancient group sheltered in caves. The caves as fauna refuge during climate changing periods, and their role as isolated systems in speciation events, leading to numerous subterranean biodiversity hot spots have been standing out since Darwin times (Barr, 1968; Culver, Deharveng, Pipan, & Bedos, 2021 and references therein). It is also interesting the variability of troglobite characters in hypogean triclads; from the typical eyeless and unpigmented body to organisms without any apparent troglobiont feature. Since the occurrence of troglomorphic organisms has been positively correlated with the distance to the cave entrance (Manenti, 2014), this variability can indicate two possibilities not mutually exclusive; 1) different cave types or cave areas, or 2) different colonization times, so that in recent colonization events not enough time has passed for selection acting against epigeal characteristics.

Finally, it should be highlighted that the high diversity of cave-dwelling triclads found in Romania is not a coincidence, since this country is the cradle of speleology (Stocchino et al., 2017). Likewise, the great number of hypogean *Girardia* species described in recent years coincides with an intense sampling effort, with the specific objective of finding freshwater planarians, in Brazilian caves (Leal-Zanchet *pers. com.*). Both examples show that the number of hypogean planarians taxa could be underestimated due to a sampling bias. Nonetheless, the presence of planarians in groundwater ecosystems has been reviewed in various publications in order to draw attention to these taxa as important components of freshwater subterranean communities (Manenti et al., 2018; Manenti, Barzaghi, Tonni, Ficetola, & Melotto, 2019; Barzaghi, De Giorgi, Pennati, & Manenti, 2021).

2.3. Reproductive strategies

Planarians can present three types of reproduction, sexual by sperm interchange and asexual by fissiparity (Stocchino & Manconi, 2013) or by parthenogenesis (Pongratz, Storhas, Carranza, & Michiels, 2003). The shift among these alternative reproductive strategies in planarians, has been recognized as an adaptation to external conditions, especially temperature and food availability (Benazzi

& Gremigni, 1982). However, most *Girardia* populations out of America are fissiparous (Chapter III), and several strictly asexual populations belonging to different species occur in *Dugesia* (Chapters IV and V). Therefore, the role of asexual reproduction by fission in Dugesiidae family diversification may be much more important than previously thought. More than a punctual adaptive response, fissiparity constitutes an important trait in the life history of asexual populations, with important evolutionary implications.

In Chapter III were discussed the advantages of fissiparity for the establishment of alien *Girardia* populations in new habitats, and the potential invasive character that it represents. Asexual reproduction has been enumerated as one important advantage for alien species establishment (Arias, Giangrande, Gambi, & Anadon, 2013; Capel et al., 2017; Huang, Guo, & Chen, 2007; Pereira, Miller, & Kitahara, 2021), and the invasive freshwater crayfish has been used as a case study to expose the advantages of parthenogenetic reproduction for a world-wide spread (Chucholl, Morawetz, & Groß, 2012; Faulkes, 2010; Jones et al., 2009).

However, the success of invasive species represents a genetic paradox. The introductions, especially the accidental ones, implicate bottlenecked populations, with typically low genetic diversity, low evolutionary potential, and perhaps low reproductive fitness. Then, how is it possible that these stocks become invasive. Several strategies have been proposed to overcome this paradox, depending on the organism and how the introduction was carried out. Precisely, asexuality is one of them, helping to escape the inbreeding depression and providing high reproductive rates (Frankham, 2005). Additionally, although the asexual reproduction is generally associated with low genetic variability, and consequently reduced ability for environmental adaptation, it has been evidenced the adaptation, and not acclimation, to cooler temperatures in the obligate asexual insect *Adelgis tsugae* (Lombardo & Elkinton, 2017).

This genetic paradox, although has been conceptualized in the framework of invasive species, is also applicable to broadly distributed and well-established asexual populations without an invasive label. Therefore, asexuality has been defined as the most important factor in the spreading of *D. sicula* throughout Mediterranean basins. This species shows low genetic diversity in fissiparous populations, agreeing with a recent colonization process (Lázaro & Riutort, 2013). However, taking into account the high density of its populations and the habitat heterogeneity in that region, adaptive

processes, or at least acclimation, could have taken place along the broad distribution area of this species.

Little is known about the genetic footprint of fissiparity, beyond the evolutionary disadvantages that have always been attributed to asexuality (Hartfield, 2016). However, these drawbacks seems to be overcome in long-term asexual lineages (Barraclough, Birky, & Burt, 2003; Fontaneto, Tang, Obertegger, Leasi, & Barraclough, 2012). Since freshwater planarians, together with hydra, show the most regenerative capability in the animal kingdom, this group constitutes an efficient model to study the evolution of fissiparity as a reproductive strategy in animals. In this thesis, I account for the role of asexuality in the evolutionary history of *Dugesia* native species from the Western Mediterranean; analysing the phylogenetic relationships among sexual and asexual populations, some of its evolutionary consequences on the genetic background of species and the effects of fissiparity on tree inference.

In Chapter IV.1 was analysed the internal relationships inside *D. subtentaculata*, a species with a high density of strict or facultative asexual populations. This species has been used as a model to investigate the genetic consequences of asexuality (Leria et al., 2019). Mosaicism was found in both sexual and asexual individuals, presumably due to the inheritance of somatic mutations during fission and, additionally, the cellular turn-over and growth-degrowth dynamics that also affect sexual individuals. Furthermore, the Meselson effect but not Muller's ratchet were found in asexual populations. Regarding purifying selection signals found in all populations analysed, independently of the reproductive strategy, the authors proposed that shifts to sexuality during the evolution of current asexual populations could have favoured selection events, outlining the importance of shifting between fissiparity and sexuality in the diversification of this species across its distribution range. Despite the evident effects of asexuality at the intra-individual level demonstrated in *D. subtentaculata*, in Chapter IV.1 was possible to infer the phylogenetic relationships between sexual and asexual populations belonging to this species (Fig. 4 in Chapter IV.1), obtaining monophyletic and well-differentiated groups, regardless of their reproductive strategy.

However, a different situation was found in the more ancient Iberian-Apennines-Alps-plus clade. In this case, long-term asexuality has played a different role in its diversification. While sexual lineages evolved becoming differentiated

species in the Liguria and Tuscany regions, the asexual ones remained restricted to the Iberian Peninsula. Under longer asexuality, different fissiparous lineages accumulated high amounts of divergence between them inside the same population. Therefore, individuals belonging to these ancient divergent fissiparous lineages coalesce very back in time. This process is reflected by an anomalous branching topology (Fig. 6C in Chapter IV.2).

Contrary to *D. subtentaculata*, in this case, the evolutionary history of the group is more ancient and probably does not comprise shifting to sexual reproduction. This scenario is supported by the more pronounced Meselson effect and higher inter-individual genetic distances of the asexual populations belonging to the Iberian-Apennines-Alps-plus clade (Fig. 4A and C in Chapter IV.2). A hypothesis to explain how those fissiparous lineages are maintained through time without suffering the fatal increase of deleterious mutations could rest in hypothesis sketched by Leria et al., (2019). It is based on the possible existence of intraindividual selection processes thanks to the continuous tissues renewal present in these animals, which could prevent the appearance of non-synonymous and deleterious mutations under fissiparous reproduction, but this hypothesis has to be tested.

Nishimura et al (2015) found a high number of amino acid substitutions in a 20 years' old clonal line derived from a single individual in *D. japonica*. Non-synonymous changes were detected in 74% of coding genes, and genes with high-mutation levels were related to response to external environment changes and other functions involved in homeostasis maintenance. However, the planarians maintained normal body shape, behaviour, and physiological functions. Based on previous studies, these authors proposed that planarians might possess a mechanism of stem cell control that prevents the production of cancerous cells and death even bearing a large number of accumulated mutations. Additionally, based on their results they proposed that some non-fatal mutations could propagate in the planarian's body via the proliferation of mutated stem cells. In this way, genetic diversity could be acquired at the level of a single individual (Nishimura et al., 2015). Therefore, if these mechanisms are also effective in natural asexual populations, it could not only explain the maintenance of asexual populations through an intraindividual selection of non-deleterious non-synonymous mutations but also the high differentiation of fissiparity lineages, everyone with their intrinsic genetic variability.

Additionally, in the Iberian-Apennines-Alps-plus clade, the asexual populations from the Iberian Peninsula act as reservoirs of ancestral polymorphisms, attracting the related sexual branches, since asexual lineages coalesce with those sexual ones far back in time, when the sexual lineages were not differentiated yet. This information is misleading for the phylogenetic inference methods that fail to infer the real topology with respect to *D. ilvana*. Thus, only by not taking into account the information of asexual populations in the analyses, it is possible to recover the real relationship between *D. etrusca* and *D. ilvana*, which also is supported by a biogeographical hypothesis.

In summary, in the case of *G. tigrina*, *G. dorotocephala* and *G. sinensis,* asexual reproduction has favored the human-mediated colonization of non-American regions, and in a similar way in the case of *D. sicula* across the Mediterranean basins. The low genetic differentiation of asexual populations belonging to these species suggested that the fissiparity has had an important role in the expansion of their distribution range during the recent history of these species. On the other hand, in the case of *D. subtentaculata* the shift between sexual and asexual periods has favoured the diversification and expansion of this species across the Iberian Peninsula. On the contrary, in the Iberia-Apennines-Alps-plus clade, ancient asexual populations have remained confined in the Iberian Peninsula, while sexual lineages have diversified in other regions. In this case, asexuality may has favoured the maintenance of Iberian lineages, adding an uncommon source of diversity to the clade. Presumably, these populations have acted as reservoirs of ancestral polymorphisms preserved by long-term fissiparity. Additionally, the maintenance of different fissiparous lineages through time, without out-crossing and recombination events among them, is reflected in an unusual ladder-like pattern in phylogenetic trees; demonstrating the complex effects of long-term fissiparity in the diversification of this group. These examples demonstrate that fissiparity has played a different role in the diversification and distribution patterns inside the Dugesiidae family, and the importance of taking into account the diversity of reproductive strategies and their genetic footprint in evolutionary studies.

2.4. Human-mediated introductions

In Chapter III was analysed the expansion of *Girardia*, which has arrived on all continents, even Africa (Chapter III. 2). However, the dispersal ability of freshwater

planarians is supposed to be low due to their lack of protection against desiccation and non-salinity tolerance (Ball 1974; Sola et al., *in press*). Therefore, in the evolutionary context of *Girardia*, this disjunct distribution out of America can only be possible through human-mediated introductions and not by natural dispersion mechanisms, just as has been proposed in the case of the recent colonisation of *D. sicula* from Africa to all around the Mediterranean basin (Lázaro & Riutort, 2013).

Human activity has shaped the global species distributions since the first stages of human dispersion in the late Pleistocene. The subsequent periods, the Neolithic spread of agriculture, the era of island colonisation, and the emergence of early urbanised societies and commercial networks have led to changes in the world biodiversity, including extinctions, extirpation, and shifts in species composition, diversity, and community structure (Boivin et al., 2016 for a broader review). Currently, species introductions have become one of the principal causes of ecological changes in the world (Lowry et al., 2013). The intentional introductions of one species respond to economic interests, either for direct productive use or as biological control. On the other hand, non-intentional introductions generally need a vector to transport the species to the receptor region (Gallardo et al., 2016).

There are numerous examples of organisms with worldwide distribution due to human actions. Unfortunately, on many occasions, these introduced alien species become invasive with great damage to receptor ecosystems (Turbelin, Malamud, & Francis, 2017). Ecological disasters due to the introductions of exotic species in natural ecosystems are embarrassing examples, from which we do not seem to learn. Many of them have damaged the largest freshwater ecosystems in the world (Fierro et al., 2019; Kolar & Lodge, 2001; Lodge, Taylor, Holdich, & Skurdal, 2000; Mills, Leach, Carlton, & Secor, 1994).

Five phases have been described for the biological invasions: transport to a new region, release or escapement to the wild, establishment, dispersal or spread, and integration or impact (García-Berthou, 2007 and references therein). The introduction effort describes the number of introductions and individuals required for a species to become invasive and it constitutes one of the most influential characteristics of the invasion process (Cassey, Delean, Lockwood, Sadowski, & Blackburn, 2018). In addition, it is a significant predictor of invasive capability (Colautti, Grigorovich, & MacIsaac, 2006). In the case of intentional introductions, the two first phases run very

fast, but in accidental introductions, generally, the number of introduced individuals is low and the spreading phase depends on their biological characteristics. High offspring number by generation, large life span, and high physiological tolerances are the most essential characteristics of a species that becomes invasive (Devin & Beisel, 2007; Kolar & Lodge, 2001; Marchetti, Moyle, & Levine, 2004; Sol et al., 2012). In the case of freshwater planarians, their generative capacity makes them “immortals” (Sahu, Dattani, & Aboobaker, 2017), and theoretically, only one animal can reproduce by fission generating a clonal line in the new habitat in a short time. Therefore, fissiparity allows accomplishing three of the invasive species characteristics, low introduction effort, maximum life span, and fast population growth confering a great invasive potential to *Girardia* species (Chapter III).

It must be taken into account that human activities not only affect the distribution of the alien species but may displace native ones from their natural distribution range. Furthermore, human activities can affect the evolutionary history of native species by introducing new ecological interactions, which can change the selective pressures in the ecosystem. Our capacity to modify the world makes us responsible for all living beings that inhabit it. Being aware of this is primordial to conserving biodiversity and the evolutionary processes that lead to it.

3. Genomic signatures: the low GC content in freshwater planarians

Flatworms (Platyhelminthes) integrate one of the largest invertebrate phyla. The phylum groups a high diversity of life forms; from free-living to internal and external parasites with distinctive morphological, reproductive, and developmental traits (Caira & Littlewood, 2013; Collins, 2017). This diversity also results in high genomic variations regarding size, gene gain/loss, and repetitive elements (Coghlan et al., 2019). DNA base composition is one of the most variable characteristics of genomes. Processes such as gene conversion and high recombination rates lead to high GC content in specific genomic regions (Galtier, 2003; Duret & Arndt, 2008). Although GC content in genomic and transcriptomic data is not directly correlated, several studies use data from coding regions to estimate the level of GC richness and its correlation with codon usage bias across all domains of life (Barahimipour et al., 2015; Chen, Lee, Hottes, Shapiro, & McAdams, 2004; Duret, 2002; Figuet, Ballenghien, Romiguier, & Galtier,

2014; Guan et al., 2018; Guo, Bao, & Fan, 2007; Lightfield, Fram, & Ely, 2011; Parvathy, Udayasuriyan, & Bhadana, 2022; Zhou, Ning, Zhang, & Guo, 2014).

Flatworms exhibit high variability of GC content, without correlation with the evolutionary history of clades. Based on the GC content in the third codon position (GC3), three groups have been identified in Platyhelminthes; 1) low GC3 (free-living species and schistosomatids), 2) intermediate GC3 (with representatives of all major groups), and 3) high GC3 (other free-living); with broad variation in the global synonymous codon usage across the phylum too (Lamolle, Fontenla, Rijo, Tort, & Smircich, 2019). In Chapter IV.2, GC3 content was used as a measure of lack of recombination, but no differences between sexual and asexual individuals were found. As it has been reported before for freshwater planarians (Abril et al., 2010; Grohme et al., 2018; Solà et al., 2015), a low GC content in coding genes was found in the present analyses. In Chapter IV.2 it was hypothesised that the low GC content in this group could be an adaptive trait. It will favour the strands separation in the DNA replication process during the active mitosis in the regeneration phases. *Hydra*, the other organism recognized for its high regenerative capability, also shows a low GC content in its genome (Chapman et al., 2010). Additionally, *Plasmodium falciparum,* the most aggressive cause of malaria disease, is the organism with the lowest genomic GC content (Zhou, Bizzaro, & Marx, 2004) with a high difference against other *Plasmodium* species with less infective capability (Videvall, 2018). The infective phase of this parasite occurs through asexual reproduction by multiple divisions (schizogony), and it is characterized by a special mitosis mechanism during which the chromosomes do not appear to condense (Gerald, Mahajan, & Kumar, 2011). However, no studies have focused on linking rich AT content genomes with the high mitotic activity in these organisms.

On the other hand, the codon usage bias occurs because synonymous codons are not used with equal frequencies in genomic coding regions. It is the result of the equilibrium between mutational bias and natural selection. The last, presumably acting over highly expressed genes (Lamolle et al., 2019 and references therein; Novoa, Jungreis, Jaillon, Kellis, & Leitner, 2019). Therefore, the GC content can be related with the levels of gene expression. Taking advantage of the data generated in Chapter V, I will here briefly analyse this issue in freshwater planarians with low GC content. The GC content of the orthologs included in the references by samples was calculated.

Then, the normalised counts' values (a measure of expression levels) were plotted against the GC content of every gene, independently for the 20 samples. All graphs showed the same pattern (Fig. 1) without a correlation between GC content and expression levels. However, high GC content has been correlated with highly expressed genes in Taenidae species (Chen et al., 2013; Huang et al., 2017), chicken (Rao, Chai, Wang, Nie, & Zhang, 2013), and human (Lercher, Urrutia, Pavlícek, & Hurst, 2003; Vinogradov, 2005). Even, high GC content has been related to high expressed genes in human pluripotent embryonic stem cells (Bornelöv, Selmi, Flad, Dietmann, & Frye, 2019). Nevertheless, these results suggest that in freshwater planarians could exist a codon usage bias toward AT-rich codons, as has been proposed for GC-poor species (Iriarte, Lamolle, & Musto, 2021). Additionally, an A-T nucleotide frequency bias has been found at different regions in the mitogenomes of freshwater and terrestial planarian species (Solà et al., 2015). This low GC content is maintained in both sexual and asexual lineages by some mechanism, which is not related to the level of gene expression. Studies focused on the adaptive value of GC content across species are necessary to elucidate the implications of this genomic characteristic in the evolutionary processes.

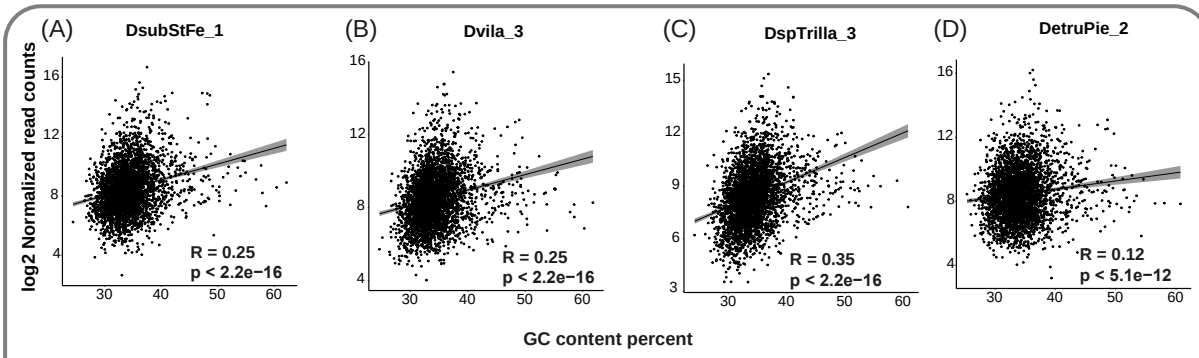


Figure 1. Relation between expression (log2 Normalized read counts) and GC content percent of orthologs genes from *Dugesia* representatives. (A): *D. subtentaculata* asexual. (B): *D. vilafarrei* sexual. (C): *Dugesia etrusca s.l. asexual*. (C): D. etrusca sexual

4. Evolutionary Biology: New challenges and perspectives in a fissiparity context

The life history of organisms shapes their evolution. Traits such as lifestyle, social behaviour, or reproduction mode determine the complex dynamics and the rearrangements of the genetic information inside populations, the basic unit of

evolution. Most of the theories that explain the evolutionary processes have been described for free-living, non-colonial, and sexual populations. However, a broad range of processes arise when we get out of this frame; demonstrating that the diversity of life is not only in the number of species but also in all the processes that lead to them.

Unfortunately, the role of asexuality has been underestimated due to the traditional assumptions regarding its negative effects. In this thesis, I analysed the effect of asexual reproduction by fission in the distribution and diversification of freshwater planarians. In this final section, I will discuss the principal challenges and future perspectives of evolutionary biology research in a fissiparity context.

In several invertebrates the regeneration process is linked to forms of propagation such as fragmentation, fission, and budding. Contrary to fragmentation, which generally depends on external forces, fission is usually an endogenous process strictly regulated. Not all organisms with regenerative capabilities are able to reproduce by fission, for which anterior-posterior regeneration, and other mechanisms are required. Hence, regeneration is the basis of asexual reproduction by fission (Brockes & Kumar, 2008). The important role of regeneration and stem cells for clonal asexual reproduction has been outlined in marine phyla (Sköld, 2009). In annelids, the regeneration capability is an ancestral characteristic in the group, but the asexual reproduction by fission has evolved secondarily and independently several times across the phylum, in the lineages that retain high regenerative capabilities (Zattara & Bely, 2016). Therefore, the shift to asexual reproduction is always possible if the machinery of stem cell proliferation is activated. The preliminary results obtained from differential expression analysis in Chapter V match this affirmation. Since the cell turnover machinery is maintained active in both sexual and asexual individuals, no expressed genes were found to be shared only by asexual individuals, while genes related to gametogenesis and sexual reproduction processes were shared only by sexual individuals in orthologs searches. However, to account for a better understanding of the genetic mechanism behind both reproduction modes using *Dugesia* as a model, it is necessary to obtain, in the first place, a fully annotated reference genome of this genus, and design investigations that involve both field and experimental data.

Additionally, this reference genome can be used with transcriptomic data to access intraindividual genetic variation, just as was done in the oribatid mite *Oppiella*

nova (Brandt et al., 2021), as long as the problem of ploidy differences, found in planarians, will be overcome in the bioinformatic pipelines. In addition, unveiling the underlying mechanisms that regulate the low content of GC in planarians and its implications in codon usage bias also constitute interesting questions that can be analysed with genomic and transcriptomic data.

Another important path is to investigate how intraindividual variation is distributed across different cell populations and if intraindividual selection processes occur in both neoblasts and differentiated cells. To account for these issues, are necessary studies analysing information from individual cells, belonging to different cell types. Additionally, it is necessary to look at a large number of genes under different regulatory and selection pressures. Single Cell sequencing techniques, already used in planarians (Plass et al., 2018), could allow to achieve this goal by helping to understand how long-term fissiparity lineages can be maintained over time.

But the future perspectives can not just hover around the collection of more genomic data. Other, conceptual and theoretical, challenges arise. Although the value of molecular data in systematic and taxonomic studies in Triclads was broadly exposed, the current frameworks of these disciplines fail under long-term fissiparity scenarios. With the genetic data at hand, and without morphological characters to evaluate, was impossible the assignation of asexual populations from the Iberian Peninsula to *D. etrusca* or *D. liguriensis* species. From my point of view, the most conservative solution, for now, is to classify as *sensu stricto* the sexual populations of these species, and *sensu lato* when asexual populations are referred. Although the populations from Font d l'Us and Font de La Trilla could be assigned to *D. etrusca* species, the evident effect of fissiparous individuals on phylogenetic inference casts doubt on this assignment. Evidently sexual and asexual lineages have evolved independently, and this has been demonstrated from a genetic and biogeographical point of view. In fact, having into account nowadays more accepted species concept, lineages evolving independently (De Queiroz 2007), each fissiparous lineage in this clade could be seen as a different species, something that, of course, does not make sense from a biological point of view. On the contrary, it clearly points to the fact that fissiparous organisms challenge even the more elaborated species concept.

The hypothesis sketched in Chapter IV. 2 to explain the diversification of this group and the ladder-like pattern is risky and needs to be tested in a broader theoretical

and methodological framework, but for the moment is not a crazy option. It is based on the idea that asexual populations retain ancestral polymorphisms, and fissiparity lineages evolve independently within the population, under the effects of non-recombination. These processes lead to what appears to be a deep coalescence phenomenon at the population level, since asexual populations coalesce back in time with sexual populations, even before the species differentiation. The shortcoming lies in the inefficiency of the current phylogenetic inference methods to interpret this information correctly and estimate the real species tree. Nevertheless, due to the advantages discussed, MSC methods are the best option for species tree inference under these scenarios.

Moreover, evolutionary models that take into account intraindividual genetic diversity and ploidy variation are necessary for a better inference of diversification processes, when asexual populations are present. In addition, species concepts and systematic biology frameworks have to be reviewed to consider the asexuality as a recurrent process in the evolution of the Tree of Life.

Conclusions

1. The molecular based phylogeny of Tricladida supports the sister-group relationship of the Cavernicola and the Maricola in contradiction with previous hypothesis. This supports a colonisation of marine habitats from freshwater and not the other way round.
2. The Cavernicola suborder possibly originated before the Gondwana breakup, in a freshwater habitat, and subsequently radiated and colonised both epigean and hypogean environments.
3. The first phylogeny of *Girardia* unveils hidden diversity of this genus in the American continent. Presumably, the genus diversified in South America and posteriorly colonised North landmasses through different dispersion waves.
4. The species *G. sinensis* described from China is native from North America
5. The North American lineage of *G. tigrina* is genetically differentiated from the South American one, which was erected as a new species: *G. clandistina sp. nov.*
6. Three species of North American origin; *G. tigrina, G. dorotocephala,* and *G. sinensis,* have been introduced multiple times across the world through human activity. All of them with a broad range of potential distribution in the Northern Hemisphere.
7. Fissiparity and environmental requirements (generalists with high suitability for human-modified habitats) favour the colonisation process of these three species.
8. In the Iberian Peninsula, the extensive areas of high suitability for *G. tigrina* and *G. sinensis*, overlapping with the more limited highly suitable areas of autochthonous freshwater planarians, point to a potential competition.
9. Future range suitability modelling, regardless of the climatic scenario, indicates that the three invasive species will extend their distributional range towards northern Europe, without diminishing the high suitability in the South. The expansion of the genus has also reached the African continent.

10. Transcriptomics yielded a phylogeny of *Dugesia* from the Western Mediterranean that corroborates a previous biogeographical hypothesis, increasing support values and adding new diversification events due to the inclusion of taxa not analysed before.
11. *D. subtentaculata* diversified across all the Iberian Peninsula from the Bethics, extending to the northern and southern regions, and crossing to Africa quite recently
12. In the *D. etrusca – D. liguriensis* species group, the unexpected topology of asexual individuals in the phylogenies could be reflecting the effects of long-term fissiparity in the most ancient clade of *Dugesia* from the Western Mediterranean.
13. This long-term fissiparity and the unexpected topology derived from it, hampers the assignation of the asexual Iberian populations to either *D. etrusca* and/or *D. liguriensis* species with certainty. Therefore, it is advisable to use the suffix *sensu lato* when these species are referred to including asexual individuals.
14. Processes such as new habitats adaptation, geological events, reproductive strategies, and even human intervention have influenced the diversification and present distribution of triclads.
15. Differentially expressed genes between sexual and asexual individuals indicate changes in the metabolic activity related to every reproductive strategy shared across different habitats and lineages.
16. A complete workflow for phylogenetic studies in freshwater planarians using transcriptomic data has been designed. It will be available in publications, a book chapter, and online platforms.
17. Although, this methodology has demonstrated to be a successful strategy to resolve Triclads phylogeny at genus level, new methods and frameworks are necessary to handle the genetic effects of asexuality, especially the fissiparity in planarians.

REFERENCES

Abellán, P., Benetti, C. J., Angus, R. B., & Ribera, I. (2011). A review of Quaternary range shifts in European aquatic Coleoptera. *Global Ecology and Biogeography*, *20*(1), 87–100. https://doi.org/https://doi.org/10.1111/j.1466-8238.2010.00572.x

Aberer, A. J., Kobert, K., & Stamatakis, A. (2014). Exabayes: Massively parallel bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, *31*(10), 2553–2556. https://doi.org/10.1093/molbev/msu236

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

Akashi, H. D., Cádiz Díaz, A., Shigenobu, S., Makino, T., & Kawata, M. (2016). Differentially expressed genes associated with adaptation to different thermal environments in three sympatric Cuban *Anolis* lizards. *Molecular Ecology*, *25*(10), 2273–2285. https://doi.org/10.1111/mec.13625

Alda, F., Tagliacollo, V. A., Bernt, M. J., Waltz, B. T., Ludt, W. B., Faircloth, B. C., … Chakrabarty, P. (2019). Resolving Deep Nodes in an Ancient Radiation of Neotropical Fishes in the Presence of Conflicting Signals from Incomplete Lineage Sorting. *Systematic Biology*, *68*(4), 573–593. https://doi.org/10.1093/sysbio/syy085

Allen, H. (2001). *Mediterranean Ecogeography* (1st ed.). https://doi.org/10.4324/9781315838526

Allman, E. S., Mitchell, J. D., & Rhodes, J. A. (2021). Gene Tree Discord, Simplex Plots, and Statistical Tests under the Coalescent. *Systematic Biology*, 1–14. https://doi.org/10.1093/sysbio/syab008

Altenhoff, A. M, Glover, N. M., & Dessimoz, C. (2019). Inferring Orthology and Paralogy. In Maria Anisimova (Ed.), *Evolutionary Genomics: Statistical and Computational Methods* (pp. 149–175). https://doi.org/10.1007/978-1-4939-9074-0_5

Altenhoff, A. M., & Dessimoz, C. (2009). Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS Computational Biology*, *5*(1). https://doi.org/10.1371/journal.pcbi.1000262

Altenhoff, A. M., Levy, J., Zarowiecki, M., Tomiczek, B., Vesztrocy, A. W., Dalquen, D. A., … Dessimoz, C. (2019). OMA standalone: Orthology inference among public and custom genomes and transcriptomes. *Genome Research*, *29*(7), 1152–1163. https://doi.org/10.1101/gr.243212.118

Alvarez, Y., Juste, J., Tabares, E., Garrido-Pertierra, A., Ibáñez, C., & Bautista, J. M. (1999). Molecular phylogeny and morphological homoplasy in fruitbats. *Molecular Biology and Evolution*, *16*(8), 1061–1067. https://doi.org/10.1093/oxfordjournals.molbev.a026195

Álvarez-Presas, M., Carbayo, F., Rozas, J., & Riutort, M. (2011). Land planarians (Platyhelminthes) as a model organism for fine-scale phylogeographic studies: understanding patterns of biodiversity in the Brazilian Atlantic Forest hotspot.

Journal of Evolutionary Biology, *24*, 887–896. https://doi.org/10.1111/j.1420-9101.2010.02220.x

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

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*, 555–568. https://doi.org/10.1016/j.ympev.2008.01.032

Álvarez-Presas, M., Mateos, E., Tudó, À., Jones, H., & Riutort, M. (2014). Diversity of introduced terrestrial flatworms in the Iberian Peninsula: a cautionary tale. *PeerJ*, *2:e430*. https://doi.org/10.7717/peerj.430

Anderson, E. (1949). *Introgressive Hybridization*. New York: John Wiley and Sons, Inc.

Arcila, D., Hughes, L. C., Meléndez-Vazquez, B., Baldwin, C. C., White, W. T., Carpenter, K. E., … Betancur-R, R. (2021). Testing the Utility of Alternative Metrics of Branch Support to Address the Ancient Evolutionary Radiation of Tunas, Stromateoids, and Allies (Teleostei: Pelagiaria). *Systematic Biology*, *70*(6), 1123–1144. https://doi.org/10.1093/sysbio/syab018

Ardui, S., Ameur, A., Vermeesch, J. R., & Hestand, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Research*, *46*(5), 2159–2168. https://doi.org/10.1093/nar/gky066

Ardura, A., Linde, A. R., Moreira, J. C., & Garcia-Vazquez, E. (2010). DNA barcoding for conservation and management of Amazonian commercial fish. *Biological Conservation*, *143*(6), 1438–1443. https://doi.org/10.1016/j.biocon.2010.03.019

Arenas, M. (2015). Trends in substitution models of molecular evolution. *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2015.00319

Arias, A., Giangrande, A., Gambi, M. C., & Anadon, N. (2013). Biology and new records of the invasive species *Branchiomma bairdi* (Annelida: Sabellidae) in the Mediterranean Sea. *Mediterranean Marine Science*, *14*(1), 162–171. https://doi.org/10.12681/mms.363

Arif, I. A., & Khan, H. A. (2009). Molecular markers for biodiversity analysis of wildlife animals: a brief review. *Animal Biodiversity and Conservation*, *32*(1), 9–17. https://doi.org/10.32800/abc.2009.32.0009

Armstrong, K. F., & Ball, S. L. (2005). DNA barcodes for biosecurity: invasive species identification. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1813–1823. https://doi.org/10.1098/RSTB.2005.1713

Avise, J. C. (1994). *Molecular Markers, Natural History and Evolution*. https://doi.org/10.1007/978-1-4615-2381-9

Baca, S. M., Toussaint, E. F. A., Miller, K. B., & Short, A. E. Z. (2017). Molecular phylogeny of the aquatic beetle family Noteridae (Coleoptera: Adephaga) with an emphasis on data partitioning strategies. *Molecular Phylogenetics and Evolution*, *107*, 282–292. https://doi.org/https://doi.org/10.1016/j.ympev.2016.10.016

Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P., & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24). https://doi.org/10.1073/pnas.1509107112

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., & Riutort, M. (2004). Molecular phylogeny of the Platyhelminthes. *Canadian Journal of Zoology*, *82*(168–193). https://doi.org/10.1139/z03-214

Baguñà, J., Carranza, S., Pala, M., Ribera, C., 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

Balart-García, P., Cieslak, A., Escuer, P., Rozas, J., Ribera, I., & Fernández, R. (2021). Smelling in the dark: phylogenomic insights into the chemosensory system of a subterranean beetle. *Molecular Ecology*, (November 2020), 1–18. https://doi.org/10.1111/mec.15921

Ball, I. R. (1971). Systematic and Biogeographical Relationships of Some Dugesis Species (Tricladida, Paludicola) from Central and South America. *Ameriacan Museum Novitates*, *2472*, 1–25.

Ball, I. R. (1974a). A contribution to the phylogeny and biogeography of the freshwater triclads (Platyhelminthes, Turbellaria). In *Biology of the Turbellaria* (pp. 339–401). New York: McGraw-Hill.

Ball, I. R. (1974b). A new genus of freshwater triclad from Tasmania with reviews of the related genera Cura and Neppia (Turbellaria, Tricladida). *Life Sci. Contr. R. Ont. Mus*, *99*, 1–48.

Ball, I. R. (1975). Nature and Formulation of biogeographical hypotheses. *Systematic Zoology*, *24*(4), 407–430.

Bankar, K. G., Todur, V. N., Shukla, R. N., & Vasudevan, M. (2015). Ameliorated de novo transcriptome assembly using Illumina paired end sequence data with Trinity Assembler. *Genomics Data*, *5*, 352–359. https://doi.org/https://doi.org/10.1016/j.gdata.2015.07.012

Barahimipour, R., Strenkert, D., Neupert, J., Schroda, M., Merchant, S. S., & Bock, R. (2015). Dissecting the contributions of GC content and codon usage to gene expression in the model alga Chlamydomonas reinhardtii. *Plant Journal*, *84*(4), 704–717. https://doi.org/10.1111/tpj.13033

Barido-Sottani, J., Vaughan, T. G., & Stadler, T. (2020). A Multitype Birth Death Model for Bayesian Inference of Lineage-Specific Birth and Death Rates. *Systematic Biology*, *69*(5), 973–986. https://doi.org/10.1093/sysbio/syaa016

Barr, T. C. (1968). Cave Ecology and the Evolution of Troglobites. In *Evolutionary Biology* (pp. 35–102). https://doi.org/doi:10.1007/978-1-4684-8094-8_2

Barraclough, T. G., Birky, C. W., & Burt, A. (2003). Diversification in sexual and asexual organisms. *Evolution*, *57*(9), 2166–2172. https://doi.org/10.1111/J.0014-3820.2003.TB00394.X

Barzaghi, B., De Giorgi, D., Pennati, R., & Manenti, R. (2021). Planarians, a Neglected Component of Biodiversity in Groundwaters. *Diversity*, Vol. 13. https://doi.org/10.3390/d13050178

Benazzi, L. G. (1970). Gametogenesis and Egg Fertilization in Planarians. *International Review of Cytology*, *27*, 101–179. https://doi.org/10.1016/S0074-7696(08)61247-4

Benazzi, M., & Gremigni, V. (1982). Developmental biology of triclad turbellarians (Planaria). In *Developmental biology of freshwater invertebrates* (pp. 151–211). New York: Liss.

Berry, V., & Gascuel, O. (1996). On the interpretation of bootstrap trees: Appropriate threshold of clade selection and induced gain. *Molecular Biology and Evolution*, *13*(7), 999–1011. https://doi.org/10.1093/molbev/13.7.999

Bezeng, B. S., & Van Der Bank, H. F. (2019). DNA barcoding of southern African crustaceans reveals a mix of invasive species and potential cryptic diversity. *PLOS ONE*, *14*(9), e0222047. https://doi.org/10.1371/JOURNAL.PONE.0222047

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), 1–15. https://doi.org/10.1186/1471-2148-11-317/FIGURES/5

Birky, C. W. (1996). Heterozygosity, Heteromorphy, and Phylogenetic Trees in Asexual Eukaryotes. *Genetics*, *144*, 427–437.

Blondel, J., Aronson, J., Bodiou, J., Boeuf, G., & Fontaine, C. (2015). The Mediterranean Region: Biological Diversity in Space and Time. https://doi.org/10.1086/656852

Bohmann, K., Mirarab, S., Bafna, V., & Gilbert, M. T. P. (2020). Beyond DNA barcoding: The unrealized potential of genome skim data in sample identification. *Molecular Ecology*, *29*(14), 2521–2534. https://doi.org/https://doi.org/10.1111/mec.15507

Boivin, N. L., Zeder, M. A., Fuller, D. Q., Crowther, A., Larson, G., Erlandson, J. M., … Petraglia, M. D. (2016). Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6388–6396. https://doi.org/10.1073/pnas.1525200113

Bornelöv, S., Selmi, T., Flad, S., Dietmann, S., & Frye, M. (2019). Codon usage optimization in pluripotent embryonic stem cells. *Genome Biology*, *20*(1), 119. https://doi.org/10.1186/s13059-019-1726-z

Bortolotti, V., Fazzuoli, M., Pandeli, E., Principi, G., Babbini, A., & Corti, S. (2001). Geology of central and eastern Elba island, Italy. *Ofioliti*, *26*(2 A), 97–105. https://doi.org/10.4454/ofioliti.v26i2a.137

Bortolotti, V., Pandeli, E., & Principi, G. (2001). The geology of the Elba Island: An historical introduction. *Ofioliti*, *26*(2 A), 79–96. https://doi.org/10.4454/ofioliti.v26i2a.136

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, *15*(4), e1006650. Retrieved from https://doi.org/10.1371/journal.pcbi.1006650

Brandt, A., Van, P. T., Bluhm, C., Anselmetti, Y., Dumas, Z., Figuet, E., … Bast, J. (2021). Haplotype divergence supports long-term asexuality in the oribatid mite Oppiella nova. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(38). https://doi.org/10.1073/pnas.2101485118

Brikiatis, L. (2021). Evidence for a Recent Formation of the Isthmus of the Isthmus of Panama 0.6 Mya. *Research Square*. https://doi.org/10.21203/rs.3.rs-604885/v1

Briski, E., Cristescu, M. E., Bailey, S. A., & MacIsaac, H. J. (2011). Use of DNA barcoding to detect invertebrate invasive species from diapausing eggs. *Biological Invasions*, *13*(6), 1325–1340. https://doi.org/10.1007/S10530-010-9892-7/TABLES/4

Brockes, J. P., & Kumar, A. (2008). Comparative aspects of animal regeneration. *Annual Review of Cell and Developmental Biology*, *24*, 525–549. https://doi.org/10.1146/annurev.cellbio.24.110707.175336

Bruno, W. J., Socci, N. D., & Halpern, A. L. (2000). Weighted neighbor joining: A likelihood-based approach to distance-based phylogeny reconstruction. *Molecular Biology and Evolution*, *17*(1), 189–197. https://doi.org/10.1093/oxfordjournals.molbev.a026231

Buffalo, V., & Coop, G. (2019). The Linked Selection Signature of Rapid Adaptation in Temporal Genomic Data. *Genetics*, *213*(3), 1007–1045. https://doi.org/10.1534/genetics.119.302581

Burke, T, & Bruford, M. W. (1987). DNA fingerprinting in birds. *Nature*, *327*(6118), 149–152. https://doi.org/10.1038/327149a0

Caira, J. N., & Littlewood, D. T. J. (2013). Worms, Platyhelminthes. In *Encyclopedia of Biodiversity: Second Edition* (Vol. 7). https://doi.org/10.1016/B978-0-12-384719-5.00166-0

Campos, A., Cummings, M. P., Reyes, J. L., & Laclette, J. P. (1998). Phylogenetic Relationships of Platyhelminthes Based on 18S Ribosomal Gene Sequences. *Molecular Phylogenetics and Evolution*, *10*(1), 1–10. https://doi.org/https://doi.org/10.1006/mpev.1997.0483

Cannizzaro, A. G., & Berg, D. J. (2022). Gone with Gondwana: Amphipod diversification in freshwaters followed the breakup of the supercontinent. *Molecular Phylogenetics and Evolution*, *171*(March), 107464. https://doi.org/10.1016/j.ympev.2022.107464

Capel, K., Toonen, R. J., Rachid, C. T. C. C., Creed, J. C., Kitahara, M. V., Forsman, Z., & Zilberberg, C. (2017). Clone wars: asexual reproduction dominates in the invasive range of *Tubastraea* spp. (Anthozoa: Scleractinia) in the South-Atlantic Ocean. *PeerJ*, *5*, e3873. https://doi.org/10.7717/peerj.3873

Carbayo, F., Alvarez-Presas, M., Olivares, C. T., Marques, F. P. L., Froehlich, E., & Riutort, M. (2013). Molecular phylogeny of Geoplaninae (Platyhelminthes) challenges current classification: proposal of taxonomic actions. *Zoologica Scripta*, *42*(5), 508–528. https://doi.org/10.1111/zsc.12019

Carranza, S., Baguña, J., & Riutort, M. (1997). Are the Platyhelminthes a Monophyletic Primitive Group? An Assessment Using 18s rDNA Sequences. *Mol. Biol. Evol*, *14*(5), 485–497.

Carranza, S., Baguña, J., & Riutort, M. (1999). Origin and Evolution of Paralogous rRNA Gene Clusters Within the Flatworm Family Dugesiidae (Platyhelminthes, Tricladida). *Journal of Molecular Evolution*, *49*, 250–259.

Carranza, S., Giribet, G., Ribera, C., Baguña, J., & Riutort, M. (1996). Evidence that Two Types of 18s rDNA Coexist in the Genome of Dugesia. *Mol. Biol. Evol*, *13*(6), 824–832.

Carranza, S., Littlewood, D. T. J., Clough, K. A., Ruiz-Trillo, I., Baguña, J., & Riutort, M. (1998). A robust molecular phylogeny of the Tricladida (Platyhelminthes: Seriata) with a discussion on morphological synapomorphies. *Proc. R. Soc. Lond*, *265*, 631–640.

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

Cassey, P., Delean, S., Lockwood, J. L., Sadowski, J. S., & Blackburn, T. M. (2018). Dissecting the null model for biological invasions: A meta-analysis of the propagule pressure effect. *PLoS Biology*, *16*(4), 1–15. https://doi.org/10.1371/journal.pbio.2005987

Castoe, T. A., Doan, T. M., & Parkinson, C. L. (2004). Data partitions and complex models in Bayesian analysis: The phylogeny of gymnophthalmid lizards. *Systematic Biology*, *53*(3), 448–469. https://doi.org/10.1080/10635150490445797

Castresana, J. (2001). Cytochrome b Phylogeny and the Taxonomy of Great Apes and Mammals. *Molecular Biology and Evolution*, *18*(4), 465–471. https://doi.org/10.1093/oxfordjournals.molbev.a003825

Cavalli-Sforza, L. L., & Edwards, A. W. (1967). Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, *19*(3), 233–257. Retrieved from https://pubmed.ncbi.nlm.nih.gov/6026583

Cetlová, V., Zozomová-Lihová, J., Melichárková, A., Mártonfiová, L., & Španiel, S. (2021). Multiple Drivers of High Species Diversity and Endemism Among Alyssum Annuals in the Mediterranean: The Evolutionary Significance of the Aegean Hotspot. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.627909

Chang, Z., Li, G., Liu, J., Zhang, Y., Ashby, C., Liu, D., … Huang, X. (2015). Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biology*, *16*(1), 30. https://doi.org/10.1186/s13059-015-0596-2

Chapman, J. A., Kirkness, E. F., Simakov, O., Hampson, S. E., Mitros, T., Weinmaier, T., … Steele, R. E. (2010). The dynamic genome of Hydra. *Nature*, *464*(7288), 592–596. https://doi.org/10.1038/nature08830

Charbagi-Barbirou, K., Álvarez-Presas, M., Vila, M., Gammoudi, M., Tekaya, S., & Riutort, M. (2011). Marine triclads (Platyhelminthes, Tricladida, Maricola), a preliminary molecular approach to their phylogeny. *Cahiers de Biologie Marine*, *52*(3), 303–311.

Chávez Hoffmeister, M. F. (2020). From Gondwana to the Great American Biotic Interchange: The Birth of South American Fauna. In M. Pino & G. Astorga (Eds.), *Pilauco: A Late Pleistocene Archaeo-paleontological Site. The Latin American Studies Book Series* (pp. 13–32). https://doi.org/10.1007/978-3-030-23918-3_2

Chen, L., Liu, T., Yang, D., Nong, X., Xie, Y., Fu, Y., … Yang, G. (2013). Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. *Biochemical and Biophysical Research Communications*, *430*(4), 1344–1348. https://doi.org/10.1016/j.bbrc.2012.12.078

Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L., & McAdams, H. H. (2004). Codon usage between genomes is constrained genome-wide mutational processes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3480–3485. https://doi.org/10.1073/pnas.0307827100

Chen, X., Chen, Y., Wu, C., & Wang, A. (2015). A new species of the genus *Girardia* (Tricladida: Dugesiidae) from China. *Zoological Systematic*, *40*(2), 166–178. https://doi.org/10.11865/zs.20150202

Cheon, S., Zhang, J., & Park, C. (2020). Is Phylotranscriptomics as Reliable as Phylogenomics? *Molecular Biology and Evolution*, *37*(12), 3672–3683. https://doi.org/10.1093/molbev/msaa181

Chifman, J., & Kubatko, L. (2014). Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics*, *30*(23), 3317–3324. https://doi.org/10.1093/BIOINFORMATICS/BTU530

Chinone, A., Nodono, H., & Matsumoto, M. (2014). Production of diploid and triploid offspring by inbreeding of the triploid planarian *Dugesia ryukyuensis*. *Chromosoma*, *123*, 265–272. https://doi.org/10.1007/s00412-008-0148-6

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., & Warnow, T. (2015). A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, *16*(Suppl 10), 1–11. https://doi.org/10.1186/1471-2164-16-S10-S2

Chucholl, C., Morawetz, K., & Groß, H. (2012). The clones are coming - Strong increase in Marmorkrebs [Procambarus fallax (Hagen, 1870) f. virginalis] records from eEurope. *Aquatic Invasions*, *7*(4), 511–519. https://doi.org/10.3391/ai.2012.7.4.008

Coates, A. G. (1992). Closure of the Isthmus of Panama: the near-shore marine record of Costa Rica and western Panama. *Geological Society of America Bulletin*, *104*(7), 814–828. https://doi.org/10.1130/0016-7606(1992)104<0814:COTIOP>2.3.CO;2

Coates, A. G., & Stallard, R. F. (2013). How old is the Isthmus of Panama. *Bulletin of Marine Science*, *89*(4), 801–813. https://doi.org/10.5343/bms.2012.1076

Coates, A. G., Collins, L. S., Aubury, M. P., & Berggren, W. A. (2004). The geology of the Darien, Panama, and the late Miocene-Pliocene collision of the Panama arc with northwestern South America. *Bulletin of the Geological Society of America*, *116*(11–12), 1327–1344. https://doi.org/10.1130/B25275.1

Cody, S., Richardson, J. E., Rull, V., Ellis, C., & Pennington, R. T. (2010). The Great American Biotic Interchange revisited. *Ecography*, *33*(2), 326–332. https://doi.org/https://doi.org/10.1111/j.1600-0587.2010.06327.x

Coghlan, A., Tyagi, R., Cotton, J. A., Holroyd, N., Rosa, B. A., Tsai, I. J., … Consortium, I. H. G. (2019). Comparative genomics of the major parasitic worms. *Nature Genetics*, *51*(1), 163–174. https://doi.org/10.1038/s41588-018-0262-1

Colautti, R. I., & MacIsaac, H. I. (2004). A neutral terminology to define 'invasive' species. *Diversity and Distributions*, *10*(2), 135–141. https://doi.org/10.1111/j.1366-9516.2004.00061.x

Colautti, R. I., Grigorovich, I. A., & MacIsaac, H. J. (2006). Propagule pressure: A null model for biological invasions. *Biological Invasions*, *8*(5), 1023–1037. https://doi.org/10.1007/s10530-005-3735-y

Collins, J. J. (2017). Platyhelminthes. *Current Biology*, *27*(7), R252–R256. https://doi.org/10.1016/j.cub.2017.02.016

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE*, *12*(12), 1–18. https://doi.org/10.1371/journal.pone.0190152

Coyne, J. A., Felton, A. A., & Lewontin, R. C. (1978). Extent of genetic variation at a highly polymorphic esterase locus in *Drosophila pseudoobscura*. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(10), 5090–5093. https://doi.org/10.1073/pnas.75.10.5090

Cristescu, M. E. (2014). From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends in Ecology & Evolution*, *29*(10), 566–571.

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), 1–14. https://doi.org/10.1038/s41598-019-44952-3

Culver, D. C., Deharveng, L., Pipan, T., & Bedos, A. (2021). An Overview of Subterranean Biodiversity Hotspots. *Diversity*, Vol. 13. https://doi.org/10.3390/d13100487

Cunha, T. J., & Giribet, G. (2019). A congruent topology for deep gastropod relationships. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1898). https://doi.org/10.1098/rspb.2018.2776

D´Souza, T. G., Storhas, M., Schulenburg, H., Beukeboom, L. W., & Michiels, N. K. (2004). Occasional sex in an 'asexual' polyploid hermaphrodite. *Proc. R. Soc. Lond. B*, *271*, 1001–1007. https://doi.org/10.1098/rspb.2004.2675

D'Souza, T. G., Schulte, R. D., Schulenburg, H., & Michiels, N. K. (2006). Paternal inheritance in parthenogenetic forms of the planarian *Schmidtea polychroa*. *Heredity*, *97*(2), 97–101. https://doi.org/10.1038/sj.hdy.6800841

Dang, T., & Kishino, H. (2019). Stochastic variational inference for Bayesian phylogenetics: A case of CAT model. *Molecular Biology and Evolution*, *36*(4), 825–833. https://doi.org/10.1093/molbev/msz020

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). JModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

Davis, M., & Shaw, R. (2001). Range Shifts and Adaptive Responses to Quaternary Climate Change. *Science*, *292*(5517), 673–679. https://doi.org/10.1126/science.292.5517.673

Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978). A model of evolutionary change in protein. In M. Dayhoff (Ed.), *Atlas of Protein Sequence and Structure Vol. 5,*

suppl. 3 (pp. 345–352). Washington, D.C: National Biomedical Research Foundation.

De Meeûs, T., Prugnolle, F., & Agnew, P. (2007). Asexual reproduction: Genetics and evolutionary aspects. *Cellular and Molecular Life Sciences*, *64*(11), 1355–1372. https://doi.org/10.1007/s00018-007-6515-2

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

De Souza, S. T., Nunes Morais, A. L., Cordeiro Medeiros, L. M., & Leal-Zanchet, A. M. (2015). The first troglobitic species of freshwater flatworm of the suborder Continenticola (Platyhelminthes) from South America. *ZooKeys*, *16*, 1–16. https://doi.org/10.3897/zookeys.470.8728

De Souza, S., Morais, A. L., Bichuette, M. E., & Leal-Zanchet, A. (2016). Two new species of freshwater flatworms (Platyhelminthes: Tricladida: Continenticola) from South American caves. *Zootaxa*, *4092*(1), 107–121.

De Vries, E. J., Baguñà, J., & Ball, I. R. (1984). Chromosomal polymorphism in planarians (Turbellaria, Tricladida) and the plate tectonics of the western Mediterranean. *Genetica*, *62*(3), 187–191. https://doi.org/10.1007/BF00056435

De Vries, Elizabeth J, & Sluys, R. (1991). Phylogenetic relationships of the genus Dugesia (Platyhelminthes, Tricladida, Paludicola). *J. Zool., Lond.*, *223*, 103–116.

De Vries, Elizabeth J. (1988). Further contribution to the taxonomy and biogeography of the subgenus *Dugesia* (Platyhelminthes: Tricladida: Paludicola) in the Mediterranean region and the Middle East. *Isarael Journal of Zoology*, *35*, 109–136.

De Vries, Elizabeth J., & Benazzi, M. (1983). *Dugesia brigantii* n.sp., A freshwater planarian found in an Italian cave. *Bolletino Di Zoologia*, *50*(3–4), 263–268. https://doi.org/10.1080/11250008309439451

Degnan, J. H. (2018). Modeling hybridization under the network multispecies coalescent. *Systematic Biology*, *67*(5), 786–799. https://doi.org/10.1093/sysbio/syy040

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

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, *24*(6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009

Degnan, J. H., Rosenberg, N. A., & Stadler, T. (2012). The probability distribution of ranked gene trees on a species tree. *Mathematical Biosciences*, *235*(1), 45–55. https://doi.org/10.1016/j.mbs.2011.10.006

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus*. *Journal of Applied Ecology*, *49*(4), 953–959. https://doi.org/10.1111/J.1365-2664.2012.02171.X

DeSalle, R., & Goldstein, P. (2019). Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution*, *7*(September), 1–11. https://doi.org/10.3389/fevo.2019.00302

Devictor, V., van Swaay, C., Brereton, T., Brotons, L., Chamberlain, D., Heliölä, J., … Jiguet, F. (2012). Differences in the climatic debts of birds and butterflies at a continental scale. *Nature Climate Change*, *2*(2), 121–124. https://doi.org/10.1038/nclimate1347

Devin, S., & Beisel, J. N. (2007). Biological and ecological characteristics of invasive species: A gammarid study. *Biological Invasions*, *9*(1), 13–24. https://doi.org/10.1007/s10530-006-9001-0

Dolman, G., & Phillips, B. (2004). Single copy nuclear DNA markers characterized for comparative phylogeography in Australian wet tropics rainforest skinks. *Molecular Ecology Notes*, *4*(2), 185–187. https://doi.org/https://doi.org/10.1111/j.1471-8286.2004.00609.x

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*(3), 369–386. https://doi.org/10.1007/s13127-020-00438-z

Domingo, L., Tomassini, R. L., Montalvo, C. I., Sanz-Pérez, D., & Alberdi, M. T. (2020). The Great American Biotic Interchange revisited: a new perspective from the stable isotope record of Argentine Pampas fossil mammals. *Scientific Reports*, *10*(1), 1–10. https://doi.org/10.1038/s41598-020-58575-6

Dong, W., Liu, J., Yu, J., Wang, L., & Zhou, S. (2012). Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. *PLOS ONE*, *7*(4). Retrieved from https://doi.org/10.1371/journal.pone.0035071

Dubé, C. E., Planes, S., Zhou, Y., Berteaux-Lecellier, V., & Boissin, E. (2017). On the occurrence of intracolonial genotypic variability in highly clonal populations of the hydrocoral *Millepora platyphylla* at Moorea (French Polynesia). *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-14684-3

Duret, L. (2002). Evolution of synonymous codon usage in metazoans. *Current Opinion in Genetics & Development*, *12*(6), 640–649. https://doi.org/https://doi.org/10.1016/S0959-437X(02)00353-2

Duret, L., & Arndt, P. F. (2008). The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genetics*, *4*(5). https://doi.org/10.1371/journal.pgen.1000071

Duret, L., & Galtier, N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual Review of Genomics and Human Genetics*, *10*, 285–311. https://doi.org/10.1146/annurev-genom-082908-150001

Dutheil, J. Y., Galtier, N., Romiguier, J., Douzery, E. J. P., Ranwez, V., & Boussau, B. (2012). Efficient selection of branch-specific models of sequence evolution. *Molecular Biology and Evolution*, *29*(7), 1861–1874. https://doi.org/10.1093/molbev/mss059

Eagles, G., & Vaughan, A. P. M. (2009). Gondwana breakup and plate kinematics: Business as usual. *Geophysical Research Letters*, *36*(10). https://doi.org/https://doi.org/10.1029/2009GL037552

Edwards, S. V., Liu, L., & Pearl, D. K. (2007). High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(14), 5936–5941. https://doi.org/10.1073/pnas.0607004104

Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., … Davis, C. C. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, *94*, 447–462. https://doi.org/10.1016/j.ympev.2015.10.027

Efron, B., Halloran, E., & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(23), 13429–13434. https://doi.org/10.1073/pnas.93.23.13429

Egger, B., Lapraz, F., Tomiczek, B., Müller, S., Dessimoz, C., Girstmair, J., … Telford, M. J. (2015). A Transcriptomic-Phylogenomic Analysis of the Evolutionary Relationships of Flatworms. *Current Biology*, *25*(10), 1347–1353. https://doi.org/10.1016/J.CUB.2015.03.034

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., … Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*(5910), 133–138. https://doi.org/10.1126/SCIENCE.1162986/SUPPL_FILE/EID.SOM.PDF

Eisen, J. A. (2000). Horizontal gene transfer among microbial genomes: New insights from complete genome analysis. *Current Opinion in Genetics and Development*, *10*(6), 606–611. https://doi.org/10.1016/S0959-437X(00)00143-X

Eisenhoffer, G. T., Kang, H., & Alvarado, A. S. (2008). Molecular Analysis of Stem Cells and Their Descendants during Cell Turnover and Regeneration in the Planarian *Schmidtea mediterranea*. *Cell Stem Cell*, *3*(3), 327–339. https://doi.org/10.1016/j.stem.2008.07.002

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(157). https://doi.org/10.1186/s13059-015-0721-2

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

Engelstädter, J. (2008). Constraints on the evolution of asexual reproduction. *BioEssays*, *30*(11–12), 1138–1150. https://doi.org/10.1002/bies.20833

Epp, L. S., Boessenkool, S., Bellemain, E. V. A. P., Haile, J., Esposito, A., Riaz, T., … Brochmann, C. (2012). New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Molecular Ecology*, *21*(8), 1821–1833. https://doi.org/https://doi.org/10.1111/j.1365-294X.2012.05537.x

Faulkes, Z. (2010). The spread of the parthenogenetic marbled crayfish, Marmorkrebs (Procambarus sp.), in the North American pet trade. *Aquatic Invasions*, *5*(4), 447–450. https://doi.org/10.3391/ai.2010.5.4.16

Felix, D. A., Gutiérrez-Gutiérrez, Ó., Espada, L., Thems, A., & González-Estévez, C. (2019). It is not all about regeneration: Planarians striking power to stand starvation. *Seminars in Cell and Developmental Biology*, *87*, 169–181. https://doi.org/10.1016/j.semcdb.2018.04.010

Felsenstein, J. (1983). Parsimony in Systematics: Biological and Statistical Issues. *Annual Review of Ecology and Systematics*, *14*(1), 313–333. https://doi.org/10.1146/annurev.es.14.110183.001525

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, *78*(2), 737–756. https://doi.org/10.1093/genetics/78.2.737

Felsenstein, J. (1978). Cases in which Parsimony or Compatibility Methods Will be Positively Misleading. *Systematic Zoology*, *27*(4), 401–410.

Felsenstein, J. (1981). Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach. *Journal of Molecular Evolution*, *17*, 368–376.

Felsenstein, J. (1985a). Phylogenies and the comparative method. *The America Naturalist*, *125*(1), 1–15. https://doi.org/10.4135/9781412974615.n115

Felsenstein, J. (1985b). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*, *39*(4), 783. https://doi.org/10.2307/2408678

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

Fernández, R., Edgecombe, G. D., & Giribet, G. (2016). Exploring Phylogenetic Relationships within Myriapoda and the Effects of Matrix Composition and Occupancy on Phylogenomic Reconstruction. *Systematic Biology*, *65*(5), 871–889. https://doi.org/10.1093/SYSBIO/SYW041

Fernández, R., Gabaldon, T., Dessimoz, C., Fernández, R., Gabaldon, T., Dessimoz, C., … Phylo-, N. (2020). Orthology: Definitions, Prediction, and Impact on Species Phylogeny Inference. In C. Scornavacca, F. Delsuc, & N. Galtier (Eds.), *Phylogenetics in the Genomic Era* (pp. 2.4:1–2.4:14). https://doi.org/hal-02535414

Fierro, P., Valdovinos, C., Arismendi, I., Díaz, G., Ruiz De Gamboa, M., & Arriagada, L. (2019). Assessment of anthropogenic threats to Chilean Mediterranean freshwater ecosystems: Literature review and expert opinions. *Environmental Impact Assessment Review*, *77*, 114–121. https://doi.org/https://doi.org/10.1016/j.eiar.2019.02.010

Figuet, E., Ballenghien, M., Romiguier, J., & Galtier, N. (2014). Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome Biology and Evolution*, *7*(1), 240–250. https://doi.org/10.1093/gbe/evu277

Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M., & Reddien, P. W. (2018). Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science*, *360*(6391). https://doi.org/10.1126/science.aaq1736

Fitch, W. M. (1971). Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Systematic Zoology*, *20*(4), 406–416.

Fitch, W. M. (2000). Homology a personal view. *Trends in Genetics*, *16*(5). https://doi.org/10.1016/S0168-9525(00)02005-9

Fitch, W. M., & Markowitz, E. (1970). An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochemical Genetics*, *4*(5), 579–593. https://doi.org/10.1007/BF00486096

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2020). A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. *Molecular Biology and Evolution*, *37*(4), 1211–1223. https://doi.org/10.1093/molbev/msz296

Fochetti, R., Sezzi, E., Tierno de Figueroa, J. M., Modica, M. V, & Oliverio, M. (2009). Molecular systematics and biogeography of the western Mediterranean stonefly genus *Tyrrhenoleuctra* (Insecta, Plecoptera). *Journal of Zoological Systematics and Evolutionary Research*, *47*(4), 328–336. https://doi.org/https://doi.org/10.1111/j.1439-0469.2008.00508.x

Fontaneto, D., Tang, C. Q., Obertegger, U., Leasi, F., & Barraclough, T. G. (2012). Different Diversification Rates Between Sexual and Asexual Organisms. *Evolutionary Biology*, *39*(2), 262–270. https://doi.org/10.1007/S11692-012-9161-Z/TABLES/2

Forster, P., Forster, L., Renfrew, C., & Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9241–9243. https://doi.org/10.1073/pnas.2004999117

Fradera-Sola, A., Thomas, A., Gasior, D., Harper, J., Hegarty, M., Armstead, I., & Fernandez-Fuentes, N. (2019). Differential gene expression and gene ontologies associated with increasing water-stress in leaf and root transcriptomes of perennial ryegrass (*Lolium perenne*). *PLoS ONE*, *14*(7), 1–29. https://doi.org/10.1371/journal.pone.0220518

Frankham, R. (2005). Resolving the genetic paradox in invasive species. *Heredity*, *94*(4), 385. https://doi.org/10.1038/SJ.HDY.6800634

Frézal, L., & Leblois, R. (2008). Four years of DNA barcoding: Current advances and prospects. *Infection, Genetics and Evolution*, *8*(5), 727–736. https://doi.org/10.1016/j.meegid.2008.05.005

Frizon de Lamotte, D., Fourdan, B., Leleu, S., Leparmentier, F., & De Clarens, P. (2015). Style of rifting and the stages of Pangea breakup. *Tectonics*, *34*(5), 1009–1029. https://doi.org/10.1002/2014TC003760

Fu, Y. X., & Li, W. H. (1999). Coalescing into the 21st century: An overview and prospects of coalescent theory. *Theoretical Population Biology*, *56*(1), 1–10. https://doi.org/10.1006/tpbi.1999.1421

Fulton, T. L., & Strobeck, C. (2006). Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. *Molecular Phylogenetics and Evolution*, *41*(1), 165–181. https://doi.org/https://doi.org/10.1016/j.ympev.2006.05.025

Gadagkar, S. R., Rosenberg, M. S., & Kumar, S. (2005). Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *304*(1), 64–74. https://doi.org/10.1002/jez.b.21026

Gallardo, B., Clavero, M., Sánchez, M. I., Vilà, M., Havel, J. E., Kovalenko, K. E., … Kats, L. B. (2016). Global ecological impacts of invasive species in aquatic ecosystems. *Global Change Biology*, *22*(1), 151–163. https://doi.org/10.1111/gcb.13004

Galtier, N. (2003). Gene conversion drives GC content evolution in mammalian histones. *Trends in Genetics*, *19*(2), 65–68. https://doi.org/10.1016/S0168-9525(02)00002-1

García-Berthou, E. (2007). The characteristics of invasive fishes: What has been learned so far? *Journal of Fish Biology*, *71*(SUPPL. D), 33–55. https://doi.org/10.1111/j.1095-8649.2007.01668.x

Gee, H., & Young, J. O. (1993). The food niches of the invasive *Dugesia tigrina* (Girard) and indigenous *Polycelis tenuis* Ijima and *P. nigra* (Muller) (Turbellaria; Tricladida) in a Welsh lake. *Hydrobiologia*, *154*(2), 99–106. https://doi.org/https://doi.org/10.1007/BF00014313

Gee, H., Pickavance, J. R., & Young, J. O. (1998). A comparative study of the population biology of the American immigrant triclad *Dugesia tigrina* (Girard) in two British lakes. *Hydrobiologia*, *361*, 135–143.

Gerald, N., Mahajan, B., & Kumar, S. (2011). Mitosis in the human malaria parasite plasmodium falciparum. *Eukaryotic Cell*, *10*(4), 474–482. https://doi.org/10.1128/EC.00314-10

Gill, D. E., Chao Lin, Perkins, S. L., & Wolf, J. B. (1995). Genetic mosaicism in plants and clonal animals. *Annual Review of Ecology and Systematics*, *26*(40), 423–444. https://doi.org/10.1146/annurev.es.26.110195.002231

Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, *11*(5), 725–736. https://doi.org/10.1093/oxfordjournals.molbev.a040153

Goloboff, P. A., Torres Galvis, A., & Arias, J. S. (2018). Parsimony and model-based phylogenetic methods for morphological data: comments on O'Reilly et al. *Palaeontology*, *61*(4), 625–630. https://doi.org/10.1111/pala.12353

Gonçalves, D. J. P., Simpson, B. B., Ortiz, E. M., Shimizu, G. H., & Jansen, R. K. (2019). Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular Phylogenetics and Evolution*, *138*, 219–232. https://doi.org/10.1016/j.ympev.2019.05.022

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

González-Estévez, C., Felix, D. A., Rodríguez-Esteban, G., & Aziz Aboobaker, A. (2012). Decreased neoblast progeny and increased cell death during starvation-induced planarian degrowth. *International Journal of Developmental Biology*, *56*(1–3), 83–91. https://doi.org/10.1387/ijdb.113452cg

Goss, E. M., Cardenas, M. E., Myers, K., Forbes, G. A., Fry, W. E., Restrepo, S., & Grunwald, N. J. (2011). The Plant Pathogen *Phytophthora andina* Emerged via Hybridization of an Unknown Phytophthora Species and the Irish Potato Famine Pathogen, *P.infestans* . *PLoS ONE*, *6*(9).

Gourbault, N. (1972). Recherches sur les Triclades Paludicoles hypogés. *Mém. Mus. Nat. Hist. Nat. Paris, Série A*, *73*, 249p.

Gourbault, N., Benazzi, M., & Hellouet, M. (1976). Triclades obscuricoles des Pyrénées V. Études morphologique et cytotaxonomique de *Dendrocoelopsis bessoni* n. sp. *Bulletin Mus. Nat. Hist. Nat. Paris, Série 3, No 406, Zoologie*, *283*, 1–12.

Grant, L. J. (2017). *Biodiversity, Biogeography and Phylogeny of Australian Freshwater Triclads (Platyhelminthes: Tricladida: Continenticola)* (James Cook University). Retrieved from https://researchonline.jcu.edu.au/51512/1/51512-grant-2016-thesis.pdf

Graur, D., & Li, W.-H. (2000). *Fundamentals of Molecular Evolution*. Sunderland, Massachusetts: Sinauer Asociates, Inc.

Grohme, M. A., Schloissnig, S., Rozanski, A., Pippel, M., Young, G. R., Winkler, S., … Rink, J. C. (2018). The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature 2018*. https://doi.org/10.1038/nature25473

Grover, A., & Sharma, P. C. (2016). Development and use of molecular markers: Past and present. *Critical Reviews in Biotechnology*, *36*(2), 290–302. https://doi.org/10.3109/07388551.2014.959891

Guan, D. L., Ma, L. Bin, Khan, M. S., Zhang, X. X., Xu, S. Q., & Xie, J. Y. (2018). Analysis of codon usage patterns in *Hirudinaria manillensis* reveals a preference for GC-ending codons caused by dominant selection constraints. *BMC Genomics*, *19*(1), 1–14. https://doi.org/10.1186/s12864-018-4937-x

Guan, H., Geoffroy, L., & Xu, M. (2021). Magma-assisted fragmentation of Pangea: Continental breakup initiation and propagation. *Gondwana Research*, *96*, 56–75. https://doi.org/https://doi.org/10.1016/j.gr.2021.04.003

Guardiola, M., Uriz, M. J., Taberlet, P., Coissac, E., Wangensteen, S., & Turon, X. (2015). Deep-Sea, Deep-Sequencing: Metabarcoding Extracellular DNA from Sediments of Marine Canyons. *PLOS ONE*, *10*(10), 1–26. https://doi.org/10.5061/dryad.520gq

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

Guil, N., Guidetti, R., Cesari, M., Marchioro, T., Rebecchi, L., & Machordom, A. (2022). Molecular phylogenetics, speciation, and long distance dispersal in tardigrade evolution: A case study of the genus Milnesium. *Molecular Phylogenetics and Evolution*, *169*, 107401. https://doi.org/https://doi.org/10.1016/j.ympev.2022.107401

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*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Guo, X., Bao, J., & Fan, L. (2007). Evidence of selectively driven codon usage in rice: Implications for GC content evolution of Gramineae genes. *FEBS Letters*, *581*(5), 1015–1021. https://doi.org/10.1016/j.febslet.2007.01.088

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., … Regev, A. (2013). De novo transcript sequence recostruction from RNA-Seq: reference generation and analysis with Trinity. In *Nature protocols* (Vol. 8). https://doi.org/10.1038/nprot.2013.084.De

Harrath, A. H., Mansour, L., Lagnika, M., Sluys, R., Boutin, C., Alwasel, S., … 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*, *in press*.

Harrath, A. H., Sluys, R., Ghlala, A., & Alwasel, S. (2012). The first subterranean freshwater planarians from North Africa, with an analysis of adenodactyl structure in the genus *Dendrocoelum* (platyhelminthes, tricladida, dendrocoelidae). *Journal of Cave and Karst Studies*, *74*(1), 48–57. https://doi.org/10.4311/2011LSC0215

Harrison, R. G. (1990). Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology*, *7*, 69–128.

Harrison, R. G. (1993). *Hybrid Zones and the Evolutionary Process*. New York: Oxford University Press, Inc.

Harrison, Richard G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(S1), 795–809. https://doi.org/10.1093/jhered/esu033

Hartfield, M. (2016). Evolutionary genetic consequences of facultative sex and outcrossing. *Journal of Evolutionary Biology*, *29*, 5–22. https://doi.org/10.1111/jeb.12770

Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution 1985 22:2*, *22*(2), 160–174. https://doi.org/10.1007/BF02101694

Hata, H., Sogabe, A., Tada, S., Nishimoto, R., Nakano, R., Kohya, N., … Kawanishi, R. (2017). Molecular phylogeny of obligate fish parasites of the family Cymothoidae (Isopoda, Crustacea): evolution of the attachment mode to host fish and the habitat shift from saline water to freshwater. *Marine Biology*, *164*(5), 105. https://doi.org/10.1007/s00227-017-3138-5

Haug, G. H., & Tiedemann, R. (1998). Effect of the formation of the Isthmus of Panama on Atlantic Ocean thermohaline circulation Gerald. *Nature*, *393*, 673–676. https://doi.org/10.1038/31447

Haug, G. H., Tiedemann, R., Zahn, R., & Ravelo, A. C. (2001). Role of Panama uplift on oceanic freshwater balance. *Geology*, *29*(3), 207–210. https://doi.org/10.1130/0091-7613(2001)029<0207:ROPUOO>2.0.CO;2

Havel, J. E., Kovalenko, K. E., Thomaz, S. M., Amalfitano, S., & Kats, L. B. (2015). Aquatic invasive species: challenges for the future. *Hydrobiologia*, *750*(1), 147–170. https://doi.org/10.1007/s10750-014-2166-0

He, L., Schneider, H., Hovenkamp, P., Marquardt, J., Wei, R., Wei, X., … Xiang, Q. (2018). A molecular phylogeny of selligueoid ferns (Polypodiaceae): Implications for a natural delimitation despite homoplasy and rapid radiation. *TAXON*, *67*(2), 237–249. https://doi.org/https://doi.org/10.12705/672.1

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hebert, P. D. N., Cywinska, A., Ball, S. L., & R, J. (2003). Biological identifications through DNA barcodes. *Proc. R. Soc. Lond*, *270*, 313–321. https://doi.org/10.1098/rspb.2002.2218

Hellmann, L., Ferreira, R. L., Rabelo, L., & Leal-Zanchet, A. M. (2020). Enhancing the still scattered knowledge on the taxonomic diversity of freshwater triclads (Platyhelminthes: Dugesiidae) in caves from two Brazilian Biomes. *Studies on Neotropical Fauna and Environment*. https://doi.org/10.1080/01650521.2020.1829901

Hellmann, L., Leal-Zanchet, A. M., & Lopes Ferreira, R. (2018). Two new sympatric troglobitic freshwater flatworms (Platyhelminthes: Dugesiidae) from a hotspot of subterranean biodiversity in the Neotropics. *Zootaxa*, *4438*(3), 561–574–561–574. https://doi.org/10.11646/ZOOTAXA.4438.3.8

Henschel, R., Lieber, M., Wu, L.-S., Nista, P. M., Haas, B. J., & LeDuc, R. D. (2012). Trinity RNA-Seq Assembler Performance Optimization. *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the EXtreme to the Campus and Beyond*. https://doi.org/10.1145/2335755.2335842

Hernández, C., Alvarado, M., Salgado-Roa, F. C., Ballesteros, N., Rueda-M, N., Oliveira, J., … Ramírez, J. D. (2022). Phylogenetic relationships and evolutionary patterns of the genus *Psammolestes* Bergroth, 1911 (Hemiptera: Reduviidae: Triatominae). *BMC Ecology and Evolution*, *22*(1), 1–15. https://doi.org/10.1186/s12862-022-01987-x

Hibbins, M. S., & Hahn, M. W. (2022). Corrigendum to: Phylogenomic approaches to detecting and characterizing introgression. *Genetics*, *220*(2). https://doi.org/10.1093/genetics/iyab220

Hillis, D. M., & Dixon, M. T. (1991). Ribosomal DNA: Molecular Evolution and Phylogenetic Inference. *The Quarterly Review of Biology*, *66*(4). https://doi.org/https://doi.org/10.1086/417338

Hoang, D. T., Chernomor, O., Haeseler, A. von, Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, *35*, 518–522. https://doi.org/10.1093/molbev/msx281

Höhna, S., Coghill, L. M., Mount, G. G., Thomson, R. C., & Brown, J. M. (2018). P3: Phylogenetic posterior prediction in RevBayes. *Molecular Biology and Evolution*, *35*(4), 1028–1034. https://doi.org/10.1093/molbev/msx286

Holder, M., & Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*, *4*(4), 275–284. https://doi.org/10.1038/nrg1044

Holmes, S. (2003). Bootstrapping Phylogenetic Trees: Theory and Methods. *Statistical Science*, *18*(2), 241–255. https://doi.org/10.1214/ss/1063994979

Huang, H., Guo, S., & Chen, G. (2007). Reproductive biology in an invasive plant *Solidago canadensis*. *Frontiers of Biology in China 2007 2:2*, *2*(2), 196–204. https://doi.org/10.1007/S11515-007-0030-6

Huang, X., Xu, J., Chen, L., Wang, Y., Gu, X., Peng, X., & Yang, G. (2017). Analysis of transcriptome data reveals multifactor constraint on codon usage in *Taenia*

multiceps. *BMC Genomics*, *18*(1), 308. https://doi.org/10.1186/s12864-017-3704-8

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, *294*(5550), 2310–2314. https://doi.org/10.1126/science.1065889

Huelsenbeck, John P., & Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*, *28*(7), 437–466. https://doi.org/10.1146/annurev.ecolsys.28.1.437

Huelsenbeck, John 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

Huey, R. B., Garland, T. J., & Turelli, M. (2019). Revisiting a Key Innovation in Evolutionary Biology: Felsenstein's 'Phylogenies and the Comparative Method'. *The American Naturalist*, *193*(6), 755–772. https://doi.org/10.1086/703055

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

Huttener, R., Thorrez, L., In'T Veld, T., Granvik, M., Snoeck, L., Van Lommel, L., & Schuit, F. (2019). GC content of vertebrate exome landscapes reveal areas of accelerated protein evolution. *BMC Evolutionary Biology*, *19*(1), 1–11. https://doi.org/10.1186/s12862-019-1469-1

Hyman, E. D. (1988). A new method of sequencing DNA. *Analytical Biochemistry*, *174*(2), 423–436. https://doi.org/https://doi.org/10.1016/0003-2697(88)90041-3

Illumina. (2022a). Illumina. Retrieved 14 March 2022, from Introduction to DNA Sequencing website: https://emea.illumina.com/techniques/sequencing/dna-sequencing.html

Illumina. (2022b). *RNA Sequencing Methods Collection. An overview of recent RNA-Seq publications featuring Illumina® technology*. Retrieved from https://www.illumina.com/content/dam/illumina-marketing/documents/products/research_reviews/rna-sequencing-methods-review-web.pdf

Iranzo, J., Wolf, Y. I., Koonin, E. V, & Sela, I. (2019). Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nature Communications*, *10*(1), 5376. https://doi.org/10.1038/s41467-019-13429-2

Iriarte, A., Lamolle, G., & Musto, H. (2021). Codon Usage Bias: An Endless Tale. *Journal of Molecular Evolution*, *89*(9), 589–593. https://doi.org/10.1007/s00239-021-10027-z

Issigonis, M., & Newmark, P. A. (2019). From worm to germ: Germ cell development and regeneration in planarians. In *Current Topics in Developmental Biology* (1st ed., Vol. 135). https://doi.org/10.1016/bs.ctdb.2019.04.001

IUCN. (2022). Invasive species. Retrieved 1 January 2022, from https://www.iucn.org/theme/species/our-work/invasive-species

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*. https://doi.org/10.1242/dev.167684

Jain, M., Olsen, H. E., Paten, B., Akeson, M., Branton, D., Daniel, B., … Koren, S. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*(1), 239. https://doi.org/10.1186/s13059-016-1103-0

Jannah, M., Hariri, M. R., Kasiamdari, R. S., & Handayani, N. S. N. (2021). The use of DNA barcoding and phylogenetic analysis to improve identification of usnea spp. based on ITS rDNA. *Journal of Tropical Biodiversity and Biotechnology*, *6*(1). https://doi.org/10.22146/JTBB.58635

Janssen, R., & Liu, P. (2021). Comparing the topology of phylogenetic network generators. *Journal of Bioinformatics and Computational Biology*, *19*(6). https://doi.org/10.1142/S0219720021400126

Jaramillo, C., Montes, C., Cardona, A., Silvestro, D., Antonelli, A., & Bacon, C. D. (2017). Comment (1) on " Formation of the Isthmus of Panama " by O ' Dea et al . *Science Advances*, *3*(6). https://doi.org/DOI: 10.1126/sciadv.1602321

Jayaswal, V., Wong, T. K. F., Robinson, J., Poladian, L., & Jermiin, L. S. (2014). Mixture models of nucleotide sequence evolution that account for heterogeneity in the substitution process across sites and across lineages. *Systematic Biology*, *63*(5), 726–742. https://doi.org/10.1093/sysbio/syu036

Jiang, W., Chen, S.-Y., Wang, H., Li, D.-Z., & Wiens, J. J. (2014). Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution*, *80*, 308–318. https://doi.org/10.1016/j.ympev.2014.08.006

Jokat, W., Boebel, T., König, M., & Meyer, U. (2003). Timing and geometry of early Gondwana breakup. *Journal of Geophysical Research: Solid Earth*, *108*(B9). https://doi.org/10.1029/2002JB001802

Joly, S., McLenachan, P. A., & Lockhart, P. J. (2009). A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist*, *174*(2). https://doi.org/10.1086/600082

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, *8*(3), 275–282. https://doi.org/10.1093/BIOINFORMATICS/8.3.275

Jones, J. P. G., Rasamy, J. R., Harvey, A., Toon, A., Oidtmann, B., Randrianarison, M. H., … Ravoahangimalala, O. R. (2009). The perfect invader: a parthenogenic crayfish poses a new threat to Madagascar's freshwater biodiversity. *Biological Invasions*, *11*(6), 1475–1482. https://doi.org/10.1007/s10530-008-9334-y

Judson, O. P., & Normark, B. B. (1996). Ancient asexual scandals. *Trends in Ecology and Evolution*, *11*(2 SPEC. ISS.), 41–46. https://doi.org/10.1016/0169-5347(96)81040-8

Jukes, T. H., & Cantor, C. R. (1969). Evolution of Protein Molecules. *Mammalian Protein Metabolism*, 21–132. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7

Justine, J. Lou, Winsor, L., Barrière, P., Fanai, C., Gey, D., Han, A. W. K., … Tsatsia, F. (2015). The invasive land planarian Platydemus manokwari (Platyhelminthes,

Geoplanidae): Records from six new localities, including the first in the USA. *PeerJ*, *2015*(6), e1037. https://doi.org/10.7717/PEERJ.1037/FIG-7

Justine, J. Lou, Winsor, L., Gey, D., Gros, P., & Thévenot, J. (2018). Giant worms chez moi! Hammerhead flatworms (Platyhelminthes, Geoplanidae, *Bipalium* spp., *Diversibipalium* spp.) in metropolitan France and overseas French territories. *PeerJ*, *2018*(5), e4672. https://doi.org/10.7717/PEERJ.4672/SUPP-3

Kainer, D., & Lanfear, R. (2015). The effects of partitioning on phylogenetic inference. *Molecular Biology and Evolution*, *32*(6), 1611–1627. https://doi.org/10.1093/molbev/msv026

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*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kanana, Y., & Riutort, M. (2019). First record of freshwater planarian *Girardia tigrina* (Platyhelminthes, Tricladida, Continenticola) in Eastern Europe. *Zootaxa*, *4624*(4), 597–600. https://doi.org/10.11646/zootaxa.4624.4.13

Kapheim, K. M., Pan, H., Li, C., Salzberg, S. L., Puiu, D., Magoc, T., … Venkat, A. (2015). Genomic signatures of evolutionary transitions from solitary to group living. *Science*, *348*(6239), 1139–1144.

Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, *21*(7), 428–444. https://doi.org/10.1038/s41576-020-0233-0

Kawakatsu, M., & Mitchel, R. W. (1984). Redescription of *Dugesia arimana* Hyman, 1957, based upon material from Trinidad St. Vicent and Venezuela (Turbellaria, Tricladida, Paludicola). *Bull. Fuji Women´s College*, *II*(22), 63–77.

Kawakatsu, M., & Mitchell, R. W. (1995). Two new freshwater cavernicolous planarians (Turbellaria, Tricladida, Paludicola) from Sulawesi (Celebes), Indonesia. *Special Bulletin of the Japanese Society of Coleopterology*, *4*, 81–103.

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

Kenk, R. (1974). Index to the genera and species of the freshwater triclads (Turbellaria) of the world. *Smithson Contrib Zool*, *183*, 1–90.

Keren-Rotem, T., Bouskila, A., & Geffen, E. (2006). Ontogenetic habitat shift and risk of cannibalism in the common chameleon (*Chamaeleo chamaeleon*). *Behavioral Ecology and Sociobiology*, *59*(6), 723–731. https://doi.org/10.1007/s00265-005-0102-z

Khang, T. F., Tan, S. H., Panha, S., & Mohamed, Z. (2017). Molecular phylogenetics and sequence analysis of two cave-dwelling dugesia species from Southeast Asia (Platyhelminthes: Tricladida: Dugesiidae). *Raffles Bulletin of Zoology*, *65*(September), 515–524.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution 1980 16:2*, *16*(2), 111–120. https://doi.org/10.1007/BF01731581

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. https://doi.org/DOI: 10.1017/CBO9780511623486

Kingman, J. F. C. (1982a). On the Genealogy of Large Populations. *Journal of Applied Probability*, *19*, 27–43. https://doi.org/10.2307/3213548

Kingman, J. F. C. (1982b). The coalescent. *Stochastic Processes and Their Applications*, *13*, 235–248. https://doi.org/10.1016/0304-4149(82)90011-4

Knakievicz, T., Lau, A. H., Prá, D., & Erdtmann, B. (2007). Biogeography and Karyotypes of Freshwater Planarians (Platyhelminthes, Tricladida, Paludicola) in Southern Brazil. *Zoological Science*, *24*, 123–129. https://doi.org/doi:10.2108/zsj.24.123

Knakievicz, T., Vieira, S. M., Erdtmann, B., & Bunselmeyer Ferreira, H. (2006). Reproductive modes and life cycles of freshwater planarians (Platyhelminthes, Tricladida, Paludicula) from southern Brazil. *Invertebrate Biology*, *125*(3), 212–221. https://doi.org/10.1111/j.1744-7410.2006.0054.x

Köhler, F., & Criscione, F. (2015). A molecular phylogeny of camaenid land snails from north-western Australia unravels widespread homoplasy in morphological characters (Gastropoda, Helicoidea). *Molecular Phylogenetics and Evolution*, *83*, 44–55. https://doi.org/https://doi.org/10.1016/j.ympev.2014.11.009

Kolar, C. S., & Lodge, D. M. (2001). Progress in invasion biology: predicting invaders. *Ecology & Evolution*, *16*(4), 199–-204. https://doi.org/10.1016/S0169-5347(01)02101-2

Kostka, D., Hubisz, M. J., Siepel, A., & Pollard, K. S. (2012). The role of GC-Biased gene conversion in shaping the fastest evolving regions of the human genome. *Molecular Biology and Evolution*, *29*(3), 1047–1057. https://doi.org/10.1093/molbev/msr279

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/BIOINFORMATICS/BTZ305

Kristensen, D. M., Wolf, Y. I., Mushegian, A. R., & Koonin, E. V. (2011). Computational methods for Gene Orthology inference. *Briefings in Bioinformatics*, *12*(5), 379–391. https://doi.org/10.1093/bib/bbr030

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, *47*(D1), D807–D811. https://doi.org/10.1093/NAR/GKY1053

Kubatko, L. S., Carstens, B. C., & Knowles, L. L. (2009). STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, *25*(7), 971–973. https://doi.org/10.1093/BIOINFORMATICS/BTP079

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, *56*(1), 17–24. https://doi.org/10.1080/10635150601146041

Lamolle, G., Fontenla, S., Rijo, G., Tort, J. F., & Smircich, P. (2019). Compositional Analysis of Flatworm Genomes Shows Strong Codon Usage Biases Across All

Classes. *Frontiers in Genetics*, *10*(September), 1–12. https://doi.org/10.3389/fgene.2019.00771

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). Partitionfinder 2: 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

Lartillot, N. & Philippe, H. (2004). A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Molecular Biology and Evolution*, *21*(6), 1095–1109. https://doi.org/10.1093/molbev/msh112

Lartillot, N. & Philippe, H. (2006). Computing Bayes factors using thermodynamic integration. *Systematic Biology*, *55*, 195–207. https://doi.org/10.1080/10635150500433722

Lartillot, N. & Philippe, H. (2008). Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1496), 1463–1472. https://doi.org/10.1098/rstb.2007.2236

Lartillot, N., Brinkmann, H., & Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Ecology and Evolution*, *7*(S4). https://doi.org/10.1186/1471-2148-7-S1-S4

Latorre, F., Deutschmann, I. M., Labarre, A., Obiol, A., Krabberød, A. K., Pelletier, E., … Logares, R. (2021). Niche adaptation promoted the evolutionary diversification of tiny ocean predators. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(25), 1–10. https://doi.org/10.1073/pnas.2020955118

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*, 570–588.

Laumer, C. E., Hejnol, A., & Giribet, G. (2015). Nuclear genomic signals of the “microturbellarian” roots of platyhelminth evolutionary innovation. *ELife*, *2015*(4). https://doi.org/10.7554/eLife.05503

Laver, T., Harrison, J., O’Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., & Studholme, D. J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification*, *3*, 1–8. https://doi.org/10.1016/j.bdq.2015.02.001

Lawrence, J. G. (1999). Gene transfer, speciation, and the evolution of bacterial genomes. *Current Opinion in Microbiology*, *2*(5), 519–523. https://doi.org/https://doi.org/10.1016/S1369-5274(99)00010-7

Lázaro, E. M., & 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. M., Harrath, A., 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*(1), 1–15. https://doi.org/10.1186/1471-2148-11-274/FIGURES/3

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

Le, S. Q., & Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, *25*(7), 1307–1320. https://doi.org/10.1093/MOLBEV/MSN067

Lee, J.-Y., Kim, H.-S., Ha, B. J., & Park, Y.-H. (2006). Phylogenetic analysis by RFLP and sequencing of mitochondrial DNA in a Korean population. *Archives of Pharmacal Research*, *29*(1), 88–95. https://doi.org/10.1007/BF02977474

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., … Wong, G. K. S. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, *574*(7780), 679–685. https://doi.org/10.1038/s41586-019-1693-2

Lefort, V., Longueville, J. E., & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, *34*(9), 2422–2424. https://doi.org/10.1093/molbev/msx149

Lemer, S., Bieler, R., & Giribet, G. (2019). Resolving the relationships of clams and cockles: Dense transcriptome sampling drastically improves the bivalve tree of life. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1896). https://doi.org/10.1098/rspb.2018.2684

Lemmon, A. R., Brown, J. M., Stanger-Hall, K., & Lemmon, E. M. (2009). The Effect of Ambiguous Data on Phylogenetic Estimates Obtained by Maximum Likelihood and Bayesian Inference. *Systematic Biology*, *58*(1), 130–145. https://doi.org/10.1093/sysbio/syp017

Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., & Gascuel, O. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of Big Data. *Nature*, *556*, 452–456. https://doi.org/10.1038/s41586-018-0043-0

Lercher, M. J., Urrutia, A. O., Pavlícek, A., & Hurst, L. D. (2003). A unification of mosaic structures in the human genome. *Human Molecular Genetics*, *12*(19), 2411–2415. https://doi.org/10.1093/hmg/ddg251

Leria, L., Riutort, M., Romero, R., Ferrer, X., & Vila-Farré, M. Microplate tectonics and the environment as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography, In Press*.

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). https://doi.org/10.1186/s12862-019-1440-1

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

Li, H., Zhang, H., Guan, R., & Miao, X. (2013). Identification of differential expression genes associated with host selection and adaptation between two sibling insect species by transcriptional profile analysis. *BMC Genomics*, *14*(1). https://doi.org/10.1186/1471-2164-14-582

Li, J., Cui, Y., Jiang, J., Yu, J., Niu, L., Deng, J., … Li, J. (2017). Applying DNA barcoding to conservation practice: a case study of endangered birds and large mammals in China. *Biodiversity and Conservation*, *26*(3), 653–668. https://doi.org/10.1007/s10531-016-1263-y

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*, *13*(9), 2178. https://doi.org/10.1101/GR.1224503

Li, Q., Yang, H., & Zhong, T. P. (2015). Regeneration across metazoan phylogeny: Lessons from model organisms. *Journal of Genetics and Genomics*, *42*(2), 57–70. https://doi.org/10.1016/j.jgg.2014.12.002

Li, W. X., Sluys, R., Vila-Farré, M., Chen, J. J., Yang, Y., Li, S. F., & Wang, A. T. (2019). A new continent in the geographic distribution of the genus *Oregoniplana* (Platyhelminthes: Tricladida: Maricola), its rediscovery in South Africa and its molecular phylogenetic position. *Zoological Journal of the Linnean Society*, *187*(1), 82–99. https://doi.org/10.1093/zoolinnean/zlz013

Lightfield, J., Fram, N. R., & Ely, B. (2011). Across bacterial phyla, distantly-related genomes with similar genomic GC content have similar patterns of amino acid usage. *PLoS ONE*, *6*(3). https://doi.org/10.1371/journal.pone.0017677

Liu, L. (2008). BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*, *24*(21), 2542–2543. https://doi.org/10.1093/BIOINFORMATICS/BTN484

Liu, L., & Yu, L. (2011). Estimating Species Trees from Unrooted Gene Trees. *Systematic Biology*, *60*(5), 661–667. https://doi.org/10.1093/SYSBIO/SYR027

Liu, L., Anderson, C., Pearl, D., & Edwards, S. V. (2019). Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model. In Maria Anisimova (Ed.), *Evolutionary Genomics: Statistical and Computational Methods, Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-9074-0_7

Liu, L., Wu, S., & Yu, L. (2015). Coalescent methods for estimating species trees from phylogenomic data. *Journal of Systematics and Evolution*, *53*(5), 380–390. https://doi.org/10.1111/JSE.12160

Liu, L., Yu, L., Kubatko, L., Pearl, D. K., & Edwards, S. V. (2009). Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution*, *53*(1), 320–328. https://doi.org/10.1016/j.ympev.2009.05.033

Liu, L., Yu, L., Pearl, D. K., & Edwards, S. V. (2009). Estimating Species Phylogenies Using Coalescence Times among Sequences. *Systematic Biology*, *58*(5), 468–477. https://doi.org/10.1093/SYSBIO/SYP031

Liu, Y., Edrisi, M., Ogilvie, H. A., & Nakhleh, L. (2022). *NestedBD: Bayesian Inference of Phylogenetic Trees From Single-Cell DNA Copy Number Profile Data Under a Birth-Death Model*. 1–9.

Lodge, D. M., Taylor, C. A., Holdich, D. M., & Skurdal, J. (2000). Nonindigenous Crayfishes Threaten North American Freshwater Biodiversity: Lessons from Europe. *Fisheries*, *25*(8), 7–20. https://doi.org/10.1577/1548-8446(2000)025<0007:NCTNAF>2.0.CO;2

Lombardo, J. A., & Elkinton, J. S. (2017). Environmental adaptation in an asexual invasive insect. *Ecology and Evolution*, *7*(14), 5123–5130. https://doi.org/10.1002/ECE3.2894

Lopez, P., Casane, D., & Philippe, H. (2002). Heterotachy, an important process of protein evolution. *Molecular Biology and Evolution*, *19*(1), 1–7. https://doi.org/10.1093/oxfordjournals.molbev.a003973

Love, M., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*. https://doi.org/10.1186/s13059-014-0550-8

Lowry, E., Rollinson, E. J., Laybourn, A. J., Scott, T. E., Aiello-Lammens, M. E., Gray, S. M., … Gurevitch, J. (2013). Biological invasions: A field synopsis, systematic review, and database of the literature. *Ecology and Evolution*, *3*(1), 182–196. https://doi.org/10.1002/ece3.431

Lozano-Fernandez, J., Tanner, A. R., Giacomelli, M., Carton, R., Vinther, J., Edgecombe, G. D., & Pisani, D. (2019). Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nature Comunications*, *10*(2295). https://doi.org/10.1038/s41467-019-10244-7

Lynch, M. (1991). Analysis of Population Genetic Structure by DNA Fingerprinting. In Terry Burke, G. Dolf, A. J. Jeffreys, & R. Wolff (Eds.), *DNA Fingerprinting: Approaches and Applications* (pp. 113–126). https://doi.org/10.1007/978-3-0348-7312-3_9

Maddison, W. P. (1997). Gene Trees in Species Trees. *Systematic Biology*, *46*(3), 523–536. https://doi.org/10.1093/SYSBIO/46.3.523

Maddison, W., & Knowles, L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, *55*(1), 21–30. https://doi.org/10.1080/10635150500354928

Mahmood, K., Orabi, J., Kristensen, P. S., Sarup, P., Jørgensen, L. N., & Jahoor, A. (2020). De novo transcriptome assembly, functional annotation, and expression profiling of rye (*Secale cereale* L.) hybrids inoculated with ergot (*Claviceps purpurea*). *Scientific Reports*, *10*(1), 1–16. https://doi.org/10.1038/s41598-020-70406-2

Mahmoud, M., Zywicki, M., Twardowski, T., & Karlowski, W. M. (2019). Efficiency of PacBio long read correction by 2nd generation Illumina sequencing. *Genomics*, *111*(1), 43–49. https://doi.org/10.1016/j.ygeno.2017.12.011

Mallet, J. (2005). Hybridization as an invasion of the genome. *Evolution*, *20*(5), 229–237.

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*, 279–283.

Manenti, R, Barzaghi, B., Lana, E., Stocchino, G. A., Manconi, R., & Lunghi, E. (2018). The stenoendemic cave-dwelling planarians (Platyhelminthes, Tricladida) of the Italian Alps and Apennines: conservation issues. *Journal for Nature Conservation*, *45*, 90–97.

Manenti, Raoul, Barzaghi, B., Tonni, G., Ficetola, G. F., & Melotto, A. (2019). Even worms matter: cave habitat restoration for a planarian species increased environmental suitability but not abundance. *Oryx*, *53*(2), 216–221. https://doi.org/DOI: 10.1017/S0030605318000741

Manenti, Raoul. (2014). Role of cave features for aquatic troglobiont fauna occurrence: Effects on 'accidentals' and troglomorphic organisms distribution. *Acta Zoologica Academiae Scientiarum Hungaricae*, *60*(3), 257–270.

Marchetti, M. P., Moyle, P. B., & Levine, R. (2004). Invasive species profiling? Exploring the characteristics of non-native fishes across invasion stages in California. *Freshwater Biology*, *49*(5), 646–661. https://doi.org/10.1111/j.1365-2427.2004.01202.x

Marmi, J., López-Giráldez, J. F., & Domingo-Roura, X. (2004). Phylogeny, evolutionary history and taxonomy of the Mustelidae based on sequences of the cytochrome b gene and a complex repetitive flanking region. *Zoologica Scripta*, *33*(6), 481–499. https://doi.org/10.1111/j.0300-3256.2004.00165.x

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Martín-Durán, J. M., Monjo, F., & Romero, R. (2012). Planarian embryology in the era of comparative developmental biology. *International Journal of Developmental Biology*, *56*(1–3), 39–48. https://doi.org/10.1387/ijdb.113442jm

Martín-Durán, J. M., Ryan, J. F., Vellutini, B. C., Pang, K., & Hejnol, A. (2017). Increased taxon sampling reveals thousands of hidden orthologs in flatworms. *Genome Research*, *27*, 1263–1272. https://doi.org/doi: 10.1101/gr.216226.116

Martins, E. P. (2000). Adaptation and the comparative method. *Trends in Ecology and Evolution*, *15*(7), 296–299. https://doi.org/10.1016/S0169-5347(00)01880-2

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

Mazza, G., Menchetti, M., Sluys, R., Solà, E., Riutort, M., Tricarico, E., … Mori, E. (2016). First report of the land planarian *Diversibipalium multilineatum* (Makino & Shirasawa, 1983) (Platyhelminthes, Tricladida, Continenticola) in Europe. *Zootaxa*, *4067*(5), 577–580.

McGirr, R., Seton, M., & Williams, S. (2021). Kinematic and geodynamic evolution of the Isthmus of Panama region: Implications for Central American Seaway closure. *GSA Bulletin*, *133*(3–4), 867–884. https://doi.org/10.1130/B35595.1

McGuire, J. A., Witt, C. C., Altshuler, D. L., & Remsen, J. V. (2007). Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy. *Systematic Biology*, *56*(5), 837–856. https://doi.org/10.1080/10635150701656360

McLoughlin, S. (2001). The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, *49*(3), 271–300. Retrieved from https://doi.org/10.1071/BT00023

Meier, J. I., Salazar, P. A., Kučka, M., Davies, R. W., Dréau, A., Aldás, I., … Chan, Y. F. (2021). Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(25). https://doi.org/10.1073/pnas.2015005118

Meinken, H. (1925). Ein verkannter Laihruher. *Wochesher, Aquarien Terrarienkunde*, *22*, 94–96.

Meleshko, O., Martin, M. D., Korneliussen, T. S., Schröck, C., Lamkowski, P., Schmutz, J., … Stenøien, H. K. (2021). Extensive Genome-Wide Phylogenetic Discordance Is Due to Incomplete Lineage Sorting and Not Ongoing Introgression in a Rapidly Radiated Bryophyte Genus. *Molecular Biology and Evolution*, *38*(7), 2750–2766. https://doi.org/10.1093/molbev/msab063

Men, A. E., Wilson, P., Siemering, K., & Forrest, S. (2008). Sanger DNA Sequencing. In M. Janitz (Ed.), *Next Generation Genome Sequencingg: Towards Personalized Medicine*. https://doi.org/10.1002/9783527625130.ch1

Meyer, X., Dib, L., Silvestro, D., & Salamin, N. (2019). Simultaneous Bayesian inference of phylogeny and molecular coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(11), 5027–5036. https://doi.org/10.1073/pnas.1813836116

Miller, J. C., & Tanksley, S. D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus Lycopersicon. *Theoretical and Applied Genetics*, *80*(4), 437–448. https://doi.org/10.1007/BF00226743

Mills, E. L., Leach, J. H., Carlton, J. T., & Secor, C. L. (1994). Exotic Species and the Integrity of the Great Lakes. *BioScience*, *44*(10), 666–676. https://doi.org/10.2307/1312510

Mindell, D. P., & Honeycutt, R. L. (1990). Ribosomal RNA in Vertebrates: Evolution and Phylogenetic Applications. *Annual Review of Ecology and Systematics*, *21*(1), 541–566. https://doi.org/10.1146/annurev.es.21.110190.002545

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Haeseler, A. von, & Lanfear, R. (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

Minh, B., Lanfear, R., Ly-Trong, N., Trifinopoulos, J., Schrempf, D., & Schmidt, H. A. (2018). *IQ-TREE version 2.2.0: Tutorials and Manual Phylogenomic software by maximum likelihood*. Retrieved from http://www.iqtree.org

Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., S. Swenson, M., & Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation.

Bioinformatics, *30*(17), i541–i548. https://doi.org/10.1093/BIOINFORMATICS/BTU462

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

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*(3), 366–380. https://doi.org/10.1093/sysbio/syu063

Mirarab, S., Nakhleh, L., & Warnow, T. (2021). Multispecies Coalescent: Theory and Applications in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, *52*, 247–268. https://doi.org/10.1146/annurev-ecolsys-012121-095340

Mirarab, S. (2019). Species Tree Estimation Using ASTRAL: Practical Considerations. *ArXiv*. https://doi.org/10.48550/arXiv.1904.03826

Mitchell, R. W., & Kawakatsu, M. (1973). A new cave-adapted planarian (Tricladida, Paludicola, Planariidae) from Chiapas, Mexico. In R. W. Mitchell & J. R. Reddell (Eds.), *Studies on the Cavernicole Fauna of Mexico and Adjacent Regions* (pp. 165–170). Ass. Mexican Cave Studies.

Molnar, P. (2017). Comment (2) on " Formation of the Isthmus of Panama " by O ' Dea et al . *Science Advances*, *3*(6). https://doi.org/DOI: 10.1126/sciadv.1602320

Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J. C., Valencia, V., … Niño, H. (2015). Middle Miocene closure of the Central American Seaway. *Science*, *348*(6231), 226–229. https://doi.org/DOI: 10.1126/science.aaa2815

Mooney, H. A., & Cleland, E. E. (2001). The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(10), 5446–5451. https://doi.org/10.1073/pnas.091093398

Morais, A. L., Bichuette, M. E., Chagas-Júnior, A., & Leal-Zanchet, A. (2021). Under threat: Two new troglobitic species of Girardia (Platyhelminthes: Tricladida) from sandstone and limestone caves in Brazil. *Zoologischer Anzeiger*, *293*, 292–302. https://doi.org/10.1016/j.jcz.2021.06.015

Morales-Briones, D. F., Kadereit, G., Tefarikis, D. T., Moore, M. J., Smith, S. A., Brockington, S. F., … Yang, Y. (2021). Disentangling Sources of Gene Tree Discordance in Phylogenomic Data Sets: Testing Ancient Hybridizations in Amaranthaceae s.l. *Systematic Biology*, *70*(2), 219–235. https://doi.org/10.1093/sysbio/syaa066

Morel, B., Schade, P., Lutteropp, S., Williams, T. A., Szöllosi, G. J., & Stamatakis, A. (2022). SpeciesRax: A Tool for Maximum Likelihood Species Tree Inference from Gene Family Trees under Duplication, Transfer, and Loss. *Molecular Biology and Evolution*, *39*(2). https://doi.org/10.1093/molbev/msab365

Mori, E., Magoga, G., Panella, M., Montagna, M., Winsor, L., Justine, J. Lou, … Mazza, G. (2021). Discovering the Pandora's box: the invasion of alien flatworms in Italy. *Biological Invasions 2021 24:1*, *24*(1), 205–216. https://doi.org/10.1007/S10530-021-02638-W

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

Mossel, E., & Roch, S. (2010). Incomplete lineage sorting: Consistent phylogeny estimation from multiple loci. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *7*(1), 166–171. https://doi.org/10.1109/TCBB.2008.66

Mount, G. G., & Brown, J. M. (2022). Comparing Likelihood Ratios to Understand Genome-Wide Variation in Phylogenetic Support. *Systematic Biology*. https://doi.org/10.1093/sysbio/syac014

Muller, H. (1932). Some genetic aspects of sex. *The American Naturalist*, *66*(703), 118–138.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1*(1), 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Müller, T., & Vingron, M. (2000). Modeling Amino Acid Replacement. *Journal of Computational Biology*, *7*(6), 761–776. https://doi.org/10.1089/10665270050514918

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Natsidis, P., Kapli, P., Schiffer, P. H., & Telford, M. J. (2021). Systematic errors in orthology inference and their effects on evolutionary analyses. *IScience*, *24*(2), 102110. https://doi.org/10.1016/j.isci.2021.102110

Naumenko, S. A., Logacheva, M. D., Popova, N. V, Klepikova, A. V, Penin, A. A., Bazykin, G. A., … Yampolsky, L. Y. (2017). Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accompanied by frequent episodes of positive selection. *Molecular Ecology*, *26*(2), 536–553. https://doi.org/10.1111/mec.13927

Neigel, J., Domingo, A., & Stake, J. (2007). DNA barcoding as a tool for coral reef conservation. *Coral Reefs*, *26*(3), 487. https://doi.org/10.1007/s00338-007-0248-4

Nekrutenko, A., & Li, W. H. (2000). Assessment of compositional heterogeneity within and between eukaryotic genomes. *Genome Research*, *10*(12), 1986–1995. https://doi.org/10.1101/gr.10.12.1986

Newmark, P. A., Wang, Y., & Chong, T. (2008). Germ cell specification and regeneration in planarians. *Cold Spring Harbor Symposia on Quantitative Biology*, *73*, 573–581. https://doi.org/10.1101/sqb.2008.73.022

Newmark, Phillip A, & Sánchez Alvarado, A. (2000). Bromodeoxyuridine Specifically Labels the Regenerative Stem Cells of Planarians. *Developmental Biology*, *220*(2), 142–153. https://doi.org/https://doi.org/10.1006/dbio.2000.9645

Nichio, B. T. L., Marchaukoski, J. N., & Raittz, R. T. (2017). New tools in orthology analysis: A brief review of promising perspectives. *Frontiers in Genetics*, *8*(OCT), 1–12. https://doi.org/10.3389/fgene.2017.00165

Nichols, R. (2001). Gene trees and species trees are not the same. *Trends in Ecology & Evolution*, *16*(7), 358–364. https://doi.org/https://doi.org/10.1016/S0169-5347(01)02203-0

Nishimura, O., Hosoda, K., Kawaguchi, E., Yazawa, S., Hayashi, T., Inoue, T., … Agata, K. (2015). Unusually Large Number of Mutations in Asexually Reproducing Clonal Planarian Dugesia japonica. *PLOS ONE*, *10*(11), e0143525. https://doi.org/10.1371/JOURNAL.PONE.0143525

Normark, B. B., Judson, O. P., & Moran, N. a. (2003). Genomic signatures of ancient asexual lineages. *Biological Journal of the Linnean Society*, *79*, 69–84. https://doi.org/10.1046/j.1095-8312.2003.00182.x

Novoa, E. M., Jungreis, I., Jaillon, O., Kellis, M., & Leitner, T. (2019). Elucidation of Codon Usage Signatures across the Domains of Life. *Molecular Biology and Evolution*, *36*(10), 2328–2339. https://doi.org/10.1093/molbev/msz124

O'Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L., … Jackson, J. B. C. (2016). Formation of the Isthmus of Panama. *Science Advances*, *2*(8). https://doi.org/10.1126/sciadv.1600883

O'Meara, B. C. (2010). New Heuristic Methods for Joint Species Delimitation and Species Tree Inference. *Systematic Biology*, *59*(1), 59–73. https://doi.org/10.1093/SYSBIO/SYP077

O'Reilly, J. E., Puttick, M. N., Pisani, D., & Donoghue, P. C. J. (2018). Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology*, *61*(1), 105–118. https://doi.org/10.1111/pala.12330

Olasagasti, F., Lieberman, K. R., Benner, S., Cherf, G. M., Dahl, J. M., Deamer, D. W., & Akeson, M. (2010). Replication of individual DNA molecules under electronic control using a protein nanopore. *Nature Nanotechnology*, *5*(11), 798–806. https://doi.org/10.1038/nnano.2010.177

Olson, P. D., & Caira, J. N. (1999). Evolution of the Major Lineages of Tapeworms (Platyhelminthes: Cestoidea) Inferred from 18S Ribosomal DNA and Elongation Factor-1α. *The Journal of Parasitology*, *85*(6), 1134–1159. https://doi.org/10.2307/3285679

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

Orii, H., Sakurai, T., & Watanabe, K. (2005). Distribution of the stem cells (neoblasts) in the planarian *Dugesia japonica*. *Development Genes and Evolution*, *215*(3), 143–157. https://doi.org/10.1007/s00427-004-0460-y

Ottenburghs, J. (2018). Exploring the hybrid speciation continuum in birds. *Ecology and Evolution*, *8*(24), 13027–13034. https://doi.org/10.1002/ece3.4558

Oya, Y., & Kajihara, H. (2020). Molecular Phylogenetic Analysis of Acotylea (Platyhelminthes: Polycladida). *Zoological Science*, *37*(3), 271–279. https://doi.org/10.2108/zs190136

Palmer, J. D., Jansen, R. K., Michaels, H. J., Chase, M. W., & Manhart, J. R. (1988). Chloroplast DNA Variation and Plant Phylogeny. *Annals of the Missouri Botanical Garden*, *75*(4), 1180–1206. https://doi.org/10.2307/2399279

Palumbo, F., Scariolo, F., Vannozzi, A., & Barcaccia, G. (2020). NGS-based barcoding with mini-COI gene target is useful for pet food market surveys aimed at mislabelling detection. *Scientific Reports 2020 10:1*, *10*(1), 1–8. https://doi.org/10.1038/s41598-020-74918-9

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

Paradis, E. (2018). Analysis of haplotype networks: The randomized minimum spanning tree method. *Methods in Ecology and Evolution*, *9*(5), 1308–1317. https://doi.org/10.1111/2041-210X.12969

Parvathy, S. T., Udayasuriyan, V., & Bhadana, V. (2022). Codon usage bias. *Molecular Biology Reports*, *49*(1), 539–565. https://doi.org/10.1007/s11033-021-06749-4

Patnaik, R., Singh, N. P., Paul, D., & Sukumar, R. (2019). Dietary and habitat shifts in relation to climate of Neogene-Quaternary proboscideans and associated mammals of the Indian subcontinent. *Quaternary Science Reviews*, *224*, 105968. https://doi.org/https://doi.org/10.1016/j.quascirev.2019.105968

Patwardhan, A., Ray, S., & Roy, A. (2014). Molecular Markers in Phylogenetic Studies-A Review. *Journal of Phylogenetics & Evolutionary Biology*, *2*(2). https://doi.org/10.4172/2329-9002.1000131

Pellettieri, J., & Alvarado, A. Ś. (2007). Cell turnover and adult tissue homeostasis: From humans to planarians. *Annual Review of Genetics*, *41*, 83–105. https://doi.org/10.1146/annurev.genet.41.110306.130244

Penny, D. (1992). The comparative method in evolutionary biology. In *Journal of Classification* (Vol. 9). https://doi.org/10.1007/bf02618482

Pereira, L. B. L., Miller, D. J., & Kitahara, M. V. (2021). High regenerative capacity is a general feature within colonial dendrophylliid corals (Anthozoa, Scleractinia). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *336*(3), 281–292. https://doi.org/10.1002/JEZ.B.23021

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

Pérez-Ponce de León, G., & Hernández-Mena, D. I. (2019). Testing the higher-level phylogenetic classification of Digenea (Platyhelminthes, Trematoda) based on nuclear rDNA sequences before entering the age of the 'next-generation' Tree of Life. *Journal of Helminthology*, *93*(3), 260–276. https://doi.org/DOI: 10.1017/S0022149X19000191

Pfenninger, M., Véla, E., Jesse, R., Elejalde, M. A., 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

Philippe, H., Delsuc, F., Brinkmann, H., & Lartillot, N. (2005). Phylogenomics. *Annual Review of Ecology, Evolution, and Systematics*, *36*, 541–562. https://doi.org/10.1146/annurev.ecolsys.35.112202.130205

Pickavance, J. (1971). The diet of the Immigrant planarian *Dugesia tigrina* (Girard). Food in the wild and comparison with some British species. *Journal of Animal Ecology*, *40*, 637–650.

Pineda-Krch, M., & Lehtilä, K. (2004). Costs and benefits of genetic heterogeneity within organisms. *Journal of Evolutionary Biology*, *17*(6), 1167–1177. https://doi.org/10.1111/j.1420-9101.2004.00808.x

Plass, M., Solana, J., Alexander Wolf, F., Ayoub, S., Misios, A., Glažar, P., … 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., 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*, 1–15. https://doi.org/10.1186/1471-2148-3-23

Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, *53*(5), 793–808. https://doi.org/10.1080/10635150490522304

Posada, D., & Crandall, K. A. (2001). Selecting the best-fit model of nucleotide substitution. *Systematic Biology*, *50*(4), 580–601. https://doi.org/10.1080/10635150118469

Poss, K. D. (2010). Advances in understanding tissue regenerative capacity and mechanisms in animals. *Nature Reviews Genetics*, *11*(10), 710–722. https://doi.org/10.1038/NRG2879

Puccinelli, I., & Deri, P. (1991). Comparative karyological analysis of some American planarians belonging to the genus *Dugesia* (subgenus *Girardia*) (Platyhelminthes, Tricladida). *Caryologia*, *44*(3–4), 225–232. https://doi.org/10.1080/00087114.1991.10797189

Puttick, M. N., O'Reilly, J. E., Pisani, D., & Donoghue, P. C. J. (2019). Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeontology*, *62*(1), 1–17. https://doi.org/10.1111/pala.12388

Quang, L. S., Gascuel, O., & Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*, *24*(20), 2317–2323. https://doi.org/10.1093/bioinformatics/btn445

Rana, S. B., Zadlock IV, F. J., Zhang, Z., Murphy, W. R., & Bentivegna, C. S. (2016). Comparison of De Novo Transcriptome Assemblers and k-mer Strategies Using the Killifish, Fundulus heteroclitus. *PLOS ONE*, *11*(4), e0153104. Retrieved from https://doi.org/10.1371/journal.pone.0153104

Rancilhac, L., Irisarri, I., Angelini, C., Arntzen, J. W., Babik, W., Bossuyt, F., … Vences, M. (2021). Phylotranscriptomic evidence for pervasive ancient hybridization among Old World salamanders. *Molecular Phylogenetics and Evolution*, *155*, 106967. https://doi.org/https://doi.org/10.1016/j.ympev.2020.106967

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). Using DNA Sequences From Multiple Loci. *Genetics*, *164*, 1645–1656.

Rannala, B., Edwards, S. V, Leaché, A., Yang, Z., Rannala, B., Edwards, S. V, … Leaché, A. (2020). The Multi-species Coalescent Model and Species Tree Inference. In C. Scornavacca, F. Delsuc, & N. Galtier (Eds.), *Phylogenetics in the Genomic Era*. Retrieved from https://hal.archives-ouvertes.fr/hal-02535622

Rao, Y. S., Chai, X. W., Wang, Z. F., Nie, Q. H., & Zhang, X. Q. (2013). Impact of GC content on gene expression pattern in chicken. *Genetics Selection Evolution*, *45*(1), 9. https://doi.org/10.1186/1297-9686-45-9

Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355. https://doi.org/10.1111/J.1471-8286.2007.01678.X

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

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

Redmond, A. K., & McLysaght, A. (2021). Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-22074-7

Reguero, M. A., Gelfo, J. N., López, G. M., Bond, M., Abello, A., Santillana, S. N., & Marenssi, S. A. (2014). Final Gondwana breakup: The Paleogene South American native ungulates and the demise of the South America–Antarctica land connection. *Global and Planetary Change*, *123*, 400–413. https://doi.org/https://doi.org/10.1016/j.gloplacha.2014.07.016

Ribas, M., Riutort, M., & Baguña, J. (1989). Morphological and biochemical variation in populations of *Dugesia (G.) tigrina* (Turbellaria, Tricladida, Paludicola) from the western Mediterranean: biogeographical and taxonomical implications. *J. Zool., Lond.*, *218*, 609–626.

Richardson, D. M., Pyšek, P., Rejmánek, M., Barbour, M. G., Dane Panetta, F., & West, C. J. (2000). Naturalization and invasion of alien plants: Concepts and definitions. *Diversity and Distributions*, *6*(2), 93–107. https://doi.org/10.1046/j.1472-4642.2000.00083.x

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. *Int. J. Dev. Biol.*, *56*, 5–17. https://doi.org/10.1387/ijdb.113441mr

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/BIOINFORMATICS/BTP616

Rodríguez, A., Rusciano, T., Hamilton, R., Holmes, L., Jordan, D., & Wollenberg Valero, K. C. (2017). Genomic and phenotypic signatures of climate adaptation in an Anolis lizard. *Ecology and Evolution*, *7*(16), 6390–6403. https://doi.org/https://doi.org/10.1002/ece3.2985

Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W., & Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, *23*(16), 3944–3956. https://doi.org/https://doi.org/10.1111/mec.12720

Rogers, J., & Gibbs, R. A. (2014). Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature Reviews Genetics*, *15*(5), 347–359. https://doi.org/10.1038/nrg3707

Rohde, K., Hefford, C., Ellis, J. T., Baverstock, P. R., Johnson, A. M., Watson, N. A., & Dittmann, S. (1993). Contributions to the phylogeny of Platyhelminthes based on partial sequencing of 18S ribosomal DNA. *International Journal for Parasitology*, *23*(6), 705–724. https://doi.org/https://doi.org/10.1016/0020-7519(93)90067-9

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (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

Rosenberg, M. S., & Kumar, S. (2003). Heterogeneity of nucleotide frequencies among evolutionary lineages and phylogenetic inference. *Molecular Biology and Evolution*, *20*(4), 610–621. https://doi.org/10.1093/molbev/msg067

Rosenberg, N. A. (2013). Discordance of Species Trees with Their Most Likely Gene Trees: A Unifying Principle. *Molecular Biology and Evolution*, *30*(12), 2709–2713. https://doi.org/10.1093/MOLBEV/MST160

Rota, J., & Wahlberg, N. (2012). Exploration of data partitioning in an eight-gene data set: phylogeny of metalmark moths (Lepidoptera, Choreutidae). *Zoologica Scripta*, *41*(5), 536–546. https://doi.org/https://doi.org/10.1111/j.1463-6409.2012.00551.x

Roth, A. C. J., Gonnet, G. H., & Dessimoz, C. (2008). Algorithm of OMA for large-scale orthology inference. *BMC Bioinformatics*, *9*, 1–10. https://doi.org/10.1186/1471-2105-9-518

Rothberg, J. M., & Leamon, J. H. (2008). The development and impact of 454 sequencing. *Nature Biotechnology*, *26*(10), 1117–1124. https://doi.org/10.1038/nbt1485

Ruiz-Trillo, I., Riutort, M., Littlewood, D. T. J., Herniou, E. A., & Baguña, J. (1999). Acoel flatworms: earliest extant bilaterian metazoans, not members of the Platyhelminthes. *Science*, *283*, 1919–1923.

Runemark, A., Vallejo-Marin, M., & Meier, J. I. (2019). Eukaryote hybrid genomes. *PLoS Genetics*, *15*(11), 1–22. https://doi.org/10.1371/journal.pgen.1008404

Ruppert, E. E., Fox, R. S., & Barnes, R. D. (2004). *Invertebrate Zoology: A Functional Evolutionary Approach*. Thomson-Brooks/Cole.

Sacerdot, C., Louis, A., Bon, C., Berthelot, C., & Roest Crollius, H. (2018). Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biology*, *19*(1), 166. https://doi.org/10.1186/s13059-018-1559-1

Sahu, S., Dattani, A., & Aboobaker, A. A. (2017). Secrets from immortal worms: What can we learn about biological ageing from the planarian model system? *Seminars in Cell and Developmental Biology*, *70*, 108–121. https://doi.org/10.1016/j.semcdb.2017.08.028

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., … Erlich, H. A. (1988). Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science*, *239*(4839), 487–491. https://doi.org/10.1126/SCIENCE.2448875

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985). Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science*, *230*(4732), 1350–1354. https://doi.org/10.1126/SCIENCE.2999980

Salmón, P., Jacobs, A., Ahrén, D., Biard, C., Dingemanse, N. J., Dominoni, D. M., … Isaksson, C. (2021). Continent-wide genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nature Communications*, *12*(1), 2983. https://doi.org/10.1038/s41467-021-23027-w

Saló, E., Abril, J. F., Adell, T., Cebrià, F., Eckelt, K., Fernández-Taboada, E., … Rodríguez-Esteban, G. (2009). Planarian regeneration: Achievements and future directions after 20 years of research. *International Journal of Developmental Biology*, *53*(8–10), 1317–1327. https://doi.org/10.1387/ijdb.072414es

Saló, E., Baguñà, J., & Romero, R. (1980). Les Planàries d 'aigües dolces a Catalunya i les illes Balears. Clau sistemàtica i distribució geogràfica. *Butlletí de La Institució Catalana d'Història Natural*, *45*, 15–30.

Sanderson, M. J., & Shaffer, H. B. (2002). Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics*, *33*, 49–72. https://doi.org/10.1146/annurev.ecolsys.33.010802.150509

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. https://doi.org/10.1073/PNAS.74.12.5463

Sanmartín, I. (2012). Historical Biogeography: Evolution in Time and Space. *Evolution: Education and Outreach*, *5*(4), 555–568. https://doi.org/10.1007/s12052-012-0421-2

Sann, M., Meusemann, K., Niehuis, O., Escalona, H. E., Mokrousov, M., Ohl, M., … Schmid-Egger, C. (2021). Reanalysis of the apoid wasp phylogeny with additional taxa and sequence data confirms the placement of Ammoplanidae as sister to bees. *Systematic Entomology*, *46*(3), 558–569. https://doi.org/https://doi.org/10.1111/syen.12475

Santos, C. A., Sonoda, G. G., Cortez, T., Coutinho, L. L., & Andrade, S. C. S. (2021). Transcriptome expression of biomineralization genes in *Littoraria flava* gastropod in Brazilian rocky shore reveals evidence of local adaptation. *Genome Biology and Evolution*, *13*(March), 1–20. https://doi.org/10.1093/gbe/evab050

Sayyari, E., & Mirarab, S. (2016). Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular Biology and Evolution*, *33*(7), 1654–1668. https://doi.org/10.1093/MOLBEV/MSW079

Scheffers, B. R., & Pecl, G. (2019). Persecuting, protecting or ignoring biodiversity under climate change. *Nature Climate Change*, *9*(8), 581–586. https://doi.org/10.1038/s41558-019-0526-5

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/sl0750-007-9002-8

Schrempf, D., Lartillot, N., & Szöllősi, G. (2020). Scalable Empirical Mixture Models That Account for Across-Site Compositional Heterogeneity. *Molecular Biology and Evolution*, *37*(12), 3616–3631. https://doi.org/10.1093/molbev/msaa145

Schwander, T., Henry, L., & Crespi, B. J. (2011). Molecular evidence for ancient asexuality in timema stick insects. *Current Biology*, *21*(13), 1129–1134. https://doi.org/10.1016/j.cub.2011.05.026

Scornavacca, C., & Galtier, N. (2017). Incomplete lineage sorting in mammalian phylogenomics. *Systematic Biology*, *66*(1), 112–120. https://doi.org/10.1093/sysbio/syw082

Shen, X. X., Steenwyk, J. L., & Rokas, A. (2021). Dissecting Incongruence between Concatenation- and Quartet-Based Approaches in Phylogenomic Data. *Systematic Biology*, *70*(5), 997–1014. https://doi.org/10.1093/sysbio/syab011

Shy, E. (1984). Habitat shift and geographical variation in North American tanagers (Thraupinae: Piranga). *Oecologia*, *63*(2), 281–285. https://doi.org/10.1007/BF00379890

Sibley, C. G., & Ahlquist, J. E. (1984). The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *Journal of Molecular Evolution*, *20*(1), 2–15. https://doi.org/10.1007/BF02101980

Simon, C. (2020). An Evolving View of Phylogenetic Support. *Systematic Biology*. https://doi.org/10.1093/sysbio/syaa068

Sköld, H. N. (2009). Stem Cells in Asexual reproduction of Marine Invertebrates. In B. Rinkevich & V. Matranga (Eds.), *Stem Cells in Marine Organisms* (pp. 105–137). https://doi.org/10.1007/978-90-481-2767-2_5

Slade, R. W., Moritz, C., Heideman, A., & Hale, P. T. (1993). Rapid assessment of single-copy nuclear DNA variation in diverse species. *Molecular Ecology*, *2*(6), 359–373. https://doi.org/https://doi.org/10.1111/j.1365-294X.1993.tb00029.x

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), 1–11. https://doi.org/10.1002/cpmb.59

Sluys, R, & Kawakatsu, M. (2001). Contribution to an inventory of the freshwater planarians of Australia and New Zealand (Platyhelminthes, Tricladida, Dugesiidae), with distribution maps of the species examined. *Beaufortia/Bull Zool Mus Univ Amsterdam*, *51*(10), 163–198.

Sluys, R, & Riutort, M. (2018). Planarian Diversity and Phylogeny. In J. C. Rink (Ed.), *Planarian Regeneration. Methods and protocols. Methods in Molecular Biology Series.* https://doi.org/10.1007/978-1-4939-7802-1_1

Sluys, R, Grant, L. J., & Blair, D. (2007). Freshwater planarians from artesian springs in Queensland, Australia (Platyhelminthes, Tricladida, Paludicola). *Contribution to Zoology*, *76*, 9–19.

Sluys, R. (2012). A new, sibling species of cave flatworm from Switzerland (Platyhelminthes, Tricladida, Dendrocoelidae). *Revue Suisse de Zoologie*, *119*, 181–188. https://doi.org/10.5962/bhl.part.150024

Sluys, R. & Kawakatsu, M. (2007). Biodiversity of marine planarians revisited (Platyhelminthes, Tricladida, Maricola). *Journal of Natural History*, *39*(6), 445–467. https://doi.org/10.1080/00222930410001671309

Sluys, R., Kawakatsu, M., & Ponce De León, R. (2005). Morphological stasis in an old and widespread group of species: Contribution to the taxonomy and biogeography of the genus *Girardia* (Platyhelminthes, Tricladida, Paludicola). *Studies on Neotropical Fauna and Environment*, *40*(2), 155–180. https://doi.org/10.1080/01650520500070220

Sluys, R., Kawakatsu, M., & Yamamoto, K. (2010). Exotic freshwater planarians currently known from Japan. *Belgian Journal of Zoology*, *140*, 103–109.

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*, 523–547. https://doi.org/10.1111/zoj.12077

Sluys, R., Vila-Farré, M., Álvarez-Presas, M., Riutort, M., Kawakatsu, M., & Tulp, A. S. (2015). The diet and distribution of *Pentacoelum kazukolinda* (Kawakatsu & Mitchell, 1984), a maricolan planarian with a freshwater ecology. *Zoologica Scripta*, *44*(1), 72–91. https://doi.org/10.1111/ZSC.12084

Sluys, R. (1989a). Phylogenetic relationships of the triclads (Platyhelminthes, Seriata, Tricladida). *Bijdragen Tot de Dierkunde*, *59*(1), 3–25.

Sluys, R. (1989b). *A Monograph of the Marine Triclads*. CRC Press.

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). Toward a phylogenetic classification and characterization of dugesiid genera (Platyhelminthes, Tricladida, Dugesiidae): A morphological perspective. In D. T. J. Littlewood & R. A. Bray (Eds.), *Interrelationship of the Platyhelminthes.* (pp. 57–73). London and New York: TAYLOR & FRANCIS.

Sluys, R., Kawakatsu, M., Riutort, M., & Baguña, J. (2009). A new higher classification of planarian flatworms (Platyhelminthes, Tricladida). *Journal of Natural History*, *43*(29–30), 1763–1777. https://doi.org/10.1080/00222930902741669

Smith, E. G., Hazzouri, K. M., Young, C. J., Delaney, P., Al-Kharafi, M., Howells, E. J., … Burt, J. A. (2022). Signatures of selection underpinning rapid coral adaptation to the world's warmest reefs. *Science Advances*, *8*(2). https://doi.org/10.1126/sciadv.abl7287

Smith, M. L., & Hahn, M. W. (2021). New Approaches for Inferring Phylogenies in the Presence of Paralogs. *Trends in Genetics*, *37*(2), 174–187. https://doi.org/10.1016/j.tig.2020.08.012

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

Sol, D., Maspons, J., Vall-llosera, M., Bartomeus, I., García-Peña, G. E., Piñol, J., & Freckleton, R. P. (2012). Unraveling the Life History of Successful Invaders. *Science*, *337*(6094), 580–583. https://doi.org/10.1126/science.1221523

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). https://doi.org/10.1371/journal.pone.0120081

Solà, E., Leria, L., Stocchino, G. A., Bagherzadeh, R., Balke, M., Daniels, S. R., … Riutort, M. Three dispersal routes out of Africa: The puzzling biogeographical history of *Dugesia* freshwater planarians. *Journal of Biogeography, In Press*.

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

Solà, E., Sluys, R., Segev, O., Blaustein, L., & Riutort, M. (2015). The taxonomic status of *Dugesia biblica* from Israel and Turkey (Platyhelminthes, Tricladida, Dugesiidae). *ZooKeys*, *506*, 1–12. https://doi.org/10.3897/zookeys.506.9663

Soltis, P. S., & Soltis, D. E. (2003). Applying the Bootstrap in Phylogeny Reconstruction. *Statistical Science*, *18*(2), 256–267. https://doi.org/10.1214/ss/1063994980

Soneson, C., & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, *14*. https://doi.org/10.1186/1471-2105-14-91

Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/0.1093/bioinformatics/btu033

Stamatakis, Alexandros, Hoover, P., & Rougemont, J. (2008). A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology*, *57*(5), 758–771. https://doi.org/10.1080/10635150802429642

Stange, M., Sánchez-Villagra, M. R., Salzburger, W., & Matschiner, M. (2018). Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports miocene closure of the Panamanian Isthmus. *Systematic Biology*, *67*(4), 681–699. https://doi.org/10.1093/sysbio/syy006

Stocchino, G. A, Sluys, R., Montanari, A., & Manconi, R. (2017). A new species of stygobiont freshwater planarian (Platyhelminthes, Tricladida) from a chemoautotrophic ecosystem: the Frasassi karst in Italy. *Zootaxa*, *4323*(4). https://doi.org/10.11646/zootaxa.4323.4.7

Stocchino, G. A., & Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology*, *80*(3), 319–328. https://doi.org/10.1080/11250003.2013.822025

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., 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*(2), 236–256. https://doi.org/10.1163/18759866-20191406

Stocchino, G. A., Sluys, R., Kawakatsu, M., Sarbu, S. M., & Manconi, R. (2017). A new species of freshwater flatworm (Platyhelminthes, Tricladida, Dendrocoelidae) inhabiting a chemoautotrophic groundwater ecosystem in Romania. *European Journal of Taxonomy*, *2017*(342), 1–21. https://doi.org/10.5852/ejt.2017.342

Stocchino, G. A., 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*(2), 93–112. https://doi.org/10.4311/2011LSC0252

Stoks, R., & McPeek, M. A. (2006). A Tale of Two Diversifications: Reciprocal Habitat Shifts to Fill Ecological Space along the Pond Permanence Gradient. *The American Naturalist*, *168*(S6), S50–S72. https://doi.org/10.1086/509045

Streelman, J. T., & Karl, S. A. (1997). Reconstructing labroid evolution with single–copy nuclear DNA. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*(1384), 1011–1020. https://doi.org/10.1098/rspb.1997.0140

Stringer, C. E. (1917). The Means of Locomotion in Planarians. *Proceedings of the National Academy of Sciences*, *3*(12), 691–692. https://doi.org/10.1073/pnas.3.12.691

Stringlis, I. A., Proietti, S., Hickman, R., Van Verk, M. C., Zamioudis, C., & Pieterse, C. M. J. (2018). Root transcriptional dynamics induced by beneficial rhizobacteria and microbial immune elicitors reveal signatures of adaptation to mutualists. *Plant Journal*, *93*(1), 166–180. https://doi.org/10.1111/tpj.13741

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

Taguchi, T., Tagami, E., Mezaki, T., Vacarizas, J. M., Canon, K. L., Avila, T. N., … Kubota, S. (2020). Karyotypic mosaicism and molecular cytogenetic markers in the scleractinian coral *Acropora pruinosa* Brook, 1982 (Hexacorallia, Anthozoa, Cnidaria). *Coral Reefs*, *39*(5), 1415–1425. https://doi.org/10.1007/s00338-020-01975-x

Talbot, J., & Schötz, E. M. (2011). Quantitative characterization of planarian wild-type behavior as a platform for screening locomotion phenotypes. *Journal of Experimental Biology*, *214*(7), 1063–1067. https://doi.org/10.1242/JEB.052290

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, *17*, 57–86.

Thomas, R. H., Schaffner, W., Wilson, A. C., & Pääbo, S. (1989). DNA phylogeny of the extinct marsupial wolf. *Nature*, *340*(6233), 465–467. https://doi.org/10.1038/340465a0

Thomson, R. C., & Brown, J. M. (2022). On the Need for New Measures of Phylogenomic Support. *Systematic Biology*. https://doi.org/10.1093/sysbio/syac002

Tripp, E. A., Fatimah, S., Darbyshire, I., & Mcdade, L. A. (2013). Origin of African Physacanthus (Acanthaceae) via Wide Hybridization. *PLoS ONE*, *8*(1).

Trivedi, S., Aloufi, A. A., Ansari, A. A., & Ghosh, S. K. (2016). Role of DNA barcoding in marine biodiversity assessment and conservation: An update. *Saudi Journal of Biological Sciences*, *23*(2), 161–171. https://doi.org/10.1016/j.sjbs.2015.01.001

Turbelin, A. J., Malamud, B. D., & Francis, R. A. (2017). Mapping the global state of invasive alien species: patterns of invasion and policy responses. *Global Ecology and Biogeography*, *26*(1), 78–92. https://doi.org/10.1111/geb.12517

Tyler, S., Schilling, S., Hooge, M., & Bush, L. F. (comp). (2022). Turbellarian taxonomic database. Version 2.0. Retrieved from http://turbellaria.umaine.edu

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 der Velde, G. (1975). The immigrant triclad flatworm *Dugesia tigrina* (Girard) (Plalyhelminthes, Turbellaria). Range extension and ecological position in the Netherlands. *Hydrobiological Bulletin*, *9*, 123–130. https://doi.org/https://doi.org/10.1007/BF02263331

Van Oppen, M. J. H., Souter, P., Howells, E. J., Heyward, A., & Berkelmans, R. (2011). Novel genetic diversity through somatic mutations: Fuel for adaptation of reef corals? *Diversity*, *3*(3), 405–423. https://doi.org/10.3390/d3030405

Veevers, J. J. (2004). Gondwanaland from 650-500 Ma assembly through 320 Ma merger in Pangea to 185-100 Ma breakup: Supercontinental tectonics via stratigraphy and radiometric dating. *Earth-Science Reviews*, *68*(1–2), 1–132. https://doi.org/10.1016/j.earscirev.2004.05.002

Venditti, C., Meade, A., & Pagel, M. (2008). Phylogenetic Mixture Models Can Reduce Node-Density Artifacts. *Systematic Biology*, *57*(2), 286–293. https://doi.org/10.1080/10635150802044045

Videvall, E. (2018). Plasmodium parasites of birds have the most AT-rich genes of eukaryotes. *Microbial Genomics*, *4*(2). https://doi.org/10.1099/mgen.0.000150

Vinogradov, A. E. (2005). Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth. *Trends in Genetics*, *21*(12), 639–643. https://doi.org/https://doi.org/10.1016/j.tig.2005.09.002

Vu, D., Groenewald, M., De Vries, M., Gehrmann, T., Stielow, B., Eberhardt, U., … Verkley, G. J. M. (2019). Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. *Studies in Mycology*, *92*, 135–154. https://doi.org/10.1016/j.simyco.2018.05.001

Waeschenbach, A., Webster, B. L., Bray, R. A., & Littlewood, D. T. J. (2007). Added resolution among ordinal level relationships of tapeworms (Platyhelminthes: Cestoda) with complete small and large subunit nuclear ribosomal RNA genes. *Molecular Phylogenetics and Evolution*, *45*(1), 311–325. https://doi.org/https://doi.org/10.1016/j.ympev.2007.03.019

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/SUPPL_FILE/1203983.WAGNER.SOM.PDF

Wagner, M., Fulton, B., & Henschel, R. (2016). Performance Optimization for the Trinity RNA-Seq Assembler. In A. Knüpfer, T. Hilbrich, C. Niethammer, J. Gracia, W. E. Nagel, & M. M. Resch (Eds.), *Tools for High Performance Computing 2015*. https://doi.org/10.1007/978-3-319-39589-0_3

Walker, R. S., Gurven, M., Burger, O., & Hamilton, M. J. (2008). The trade-off between number and size of offspring in humans and other primates. *Proceedings. Biological Sciences / The Royal Society*, *275*(1636), 827–833. https://doi.org/10.1098/rspb.2007.1511

Wallingford, P. D., Morelli, T. L., Allen, J. M., Beaury, E. M., Blumenthal, D. M., Bradley, B. A., … Sorte, C. J. B. (2020). Adjusting the lens of invasion biology to focus on the impacts of climate-driven range shifts. *Nature Climate Change*, *10*(5), 398–405. https://doi.org/10.1038/s41558-020-0768-2

Wang, H. C., Minh, B. Q., Susko, S., & Roger, A. J. (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

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology 2021 39:11*, *39*(11), 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. *PeerJ*. https://doi.org/10.7717/peerj.4705

Ward, R. D., Hanner, R., & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, *74*(2), 329–356. https://doi.org/10.1111/J.1095-8649.2008.02080.X

Ward, Robert D, Holmes, B. H., White, W. T., & Last, P. R. (2008). DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. *Marine and Freshwater Research*, *59*(1), 57–71. Retrieved from https://doi.org/10.1071/MF07148

Webb, S. D. (2006). The great American Biotic Interchange: patterns and processes. *Annals of the Missouri Botanical Garden*, *93*(2), 245–257. https://doi.org/10.3417/0026-6493(2006)93[245:TGABIP]2.0.CO;2

Weber, C. C., Boussau, B., Romiguier, J., Jarvis, E. D., & Ellegren, H. (2014). Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian

base composition. *Genome Biology*, *15*(12), 549. https://doi.org/10.1186/s13059-014-0549-1

Webster, M. T., & Hurst, L. D. (2012). Direct and indirect consequences of meiotic recombination: Implications for genome evolution. *Trends in Genetics*, *28*(3), 101–109. https://doi.org/10.1016/j.tig.2011.11.002

Weir, J. T., Bermingham, E., & Schluter, D. (2009). The Great American Biotic Interchange in birds. *Proceedings of the National Academy of Sciences*, *106*(51), 21737–21742. https://doi.org/10.1073/pnas.0903811106

Weir, W., Capewell, P., Foth, B., Clucas, C., Pountain, A., Steketee, P., … Macleod, A. (2016). Population genomics reveals the origin and asexual evolution of human infective trypanosomes. *ELife*, *5*(JANUARY2016). https://doi.org/10.7554/ELIFE.11473

Welch, D. M., & Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, *288*(5469), 1211–1215. https://doi.org/10.1126/science.288.5469.1211

Wenemoser, D., & Reddien, P. W. (2010). Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Developmental Biology*, *344*(2), 979–991. https://doi.org/10.1016/J.YDBIO.2010.06.017

Whelan, S., & Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution*, *18*(5), 691–699. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A003851

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., … Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), E4859–E4868. https://doi.org/10.1073/pnas.1323926111

Wiens, J. J. (2003). Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology*, *52*(4), 528–538. https://doi.org/10.1080/10635150390218330

Wiens, J. J. (2006). Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics*, *39*(1 SPEC. ISS.), 34–42. https://doi.org/10.1016/j.jbi.2005.04.001

Wiens, J. J., & Morrill, M. C. (2011). Missing data in phylogenetic analysis: Reconciling results from simulations and empirical data. *Systematic Biology*, *60*(5), 719–731. https://doi.org/10.1093/sysbio/syr025

Witt, J. D. S., Threloff, D., & Hebert, P. D. N. (2006). DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Molecular Ecology*, *15*(10), 3073–3082. https://doi.org/https://doi.org/10.1111/j.1365-294X.2006.02999.x

Wright, J. F. (1987). Colonization of rivers and canals in Great Britain by *Dugesia tigrina* (Girard) (Platyhelminthes: Tricladida). *Freshwater Biology*, *17*, 69–78. Retrieved from https://doi.org/10.1111/j.1365-2427.1987.tb01029.x

Wu, G., Miyauchi, S., Morin, E., Kuo, A., Drula, E., Varga, T., … Martin, F. M. (2022). Evolutionary innovations through gain and loss of genes in the ectomycorrhizal

Boletales. *New Phytologist*, *233*(3), 1383–1400. https://doi.org/https://doi.org/10.1111/nph.17858

Yang, C., Zheng, Y., Tan, S., Meng, G., Rao, W., Yang, C., … Liu, S. (2020). Efficient COI barcoding using high throughput single-end 400 bp sequencing. *BMC Genomics*, *21*(1), 1–10. https://doi.org/10.1186/S12864-020-07255-W/FIGURES/4

Yang, H.-M., Sluys, R., Kawakatsu, M., & Min, G.-S. (2018). New molecular sequences for two genera of marine planarians facilitate determination of their position in the phylogenetic tree, with new records for two species (Platyhelminthes, Tricladida, Maricola). *ZooKeys*, *781*, 1–17. https://doi.org/10.3897/zookeys.781.26324

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. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854–865. https://doi.org/10.1093/CZOOLO/61.5.854

Yang, Z., & Rannala, B. (1997). Bayesian Phylogenetic Inference Using DNA Sequences: A Markov Chain Monte Carlo Method. *Mol. Biol. Evol*, *14*(7), 717–724. Retrieved from http://mbe.oxfordjournals.org/cgi/content/abstract/14/7/717

Yang, Z., & Rannala, B. (2005). Branch-length prior influences Bayesian posterior probability of phylogeny. *Systematic Biology*, *54*(3), 455–470. https://doi.org/10.1080/10635150590945313

Yang, Z., & Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature*, *13*, 303–314. https://doi.org/10.1038/nrg3186

Yang, Z., & Zhu, T. (2018). Bayesian selection of misspecified models is overconfident and may cause spurious posterior probabilities for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(8), 1854–1859. https://doi.org/10.1073/pnas.1712673115

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, *11*(2). https://doi.org/10.1186/gb-2010-11-2-r14

Zattara, E. E., & Bely, A. E. (2016). Phylogenetic distribution of regeneration and asexual reproduction in Annelida: regeneration is ancestral and fission evolves in regenerative clades. *Invertebrate Biology*, *135*(4), 400–414. https://doi.org/10.1111/IVB.12151

Zhang, C., Scornavacca, C., Molloy, E. K., & Mirarab, S. (2020). ASTRAL-pro: Quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution*, *37*(11), 3292–3307. https://doi.org/10.1093/molbev/msaa139

Zhang, H., Jain, C., & Aluru, S. (2020). A comprehensive evaluation of long read error correction methods. *BMC Genomics*, *21*(Suppl 6), 1–15. https://doi.org/10.1186/s12864-020-07227-0

Zhang, Z. H., Jhaveri, D. J., Marshall, V. M., Bauer, D. C., Edson, J., Narayanan, R. K., … Zhao, Q. Y. (2014). A comparative study of techniques for differential expression analysis on RNA-seq data. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0103207

Zhao, A., Qin, H., & Fu, X. (2016). What Determines the Regenerative Capacity in Animals? *BioScience*, *66*(9), 735–746. https://doi.org/10.1093/BIOSCI/BIW079

Zhou, H. Q., Ning, L. W., Zhang, H. X., & Guo, F. B. (2014). Analysis of the relationship between genomic GC content and patterns of base usage, codon usage and amino acid usage in prokaryotes: Similar GC content adopts similar compositional frequencies regardless of the phylogenetic lineages. *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0107319

Zhou, X., Shen, X. X., Hittinger, C. T., & Rokas, A. (2017). Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets. *Molecular Biology and Evolution*, *35*(2), 486–503. https://doi.org/10.1093/molbev/msx302

Zhou, Y., Bizzaro, J. W., & Marx, K. A. (2004). Homopolymer tract length dependent enrichments in functional regions of 27 eukaryotes and their novel dependence on the organism DNA (G+C)% composition. *BMC Genomics*, *5*, 1–16. https://doi.org/10.1186/1471-2164-5-95

Zimin, A. V., & Salzberg, S. L. (2022). The SAMBA tool uses long reads to improve the contiguity of genome assemblies. *PLoS Computational Biology*, *18*(2), 1–11. https://doi.org/10.1371/journal.pcbi.1009860

Zimmermann, J., Voss, H., Schwager, C., Stegemann, J., & Ansorge, W. (1988). Automated Sanger dideoxy sequencing reaction protocol. *FEBS Letters*, *233*(2), 432–436. https://doi.org/https://doi.org/10.1016/0014-5793(88)80477-0

Zurawski, G., & Clegg, M. T. (1987). Evolution of Higher-Plant Chloroplast DNA-Encoded Genes: Implications for Structure-Function and Phylogenetic Studies. *Annual Review of Plant Physiology*, *38*(1), 391–418. https://doi.org/10.1146/annurev.pp.38.060187.002135

Barcelona
2022