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).

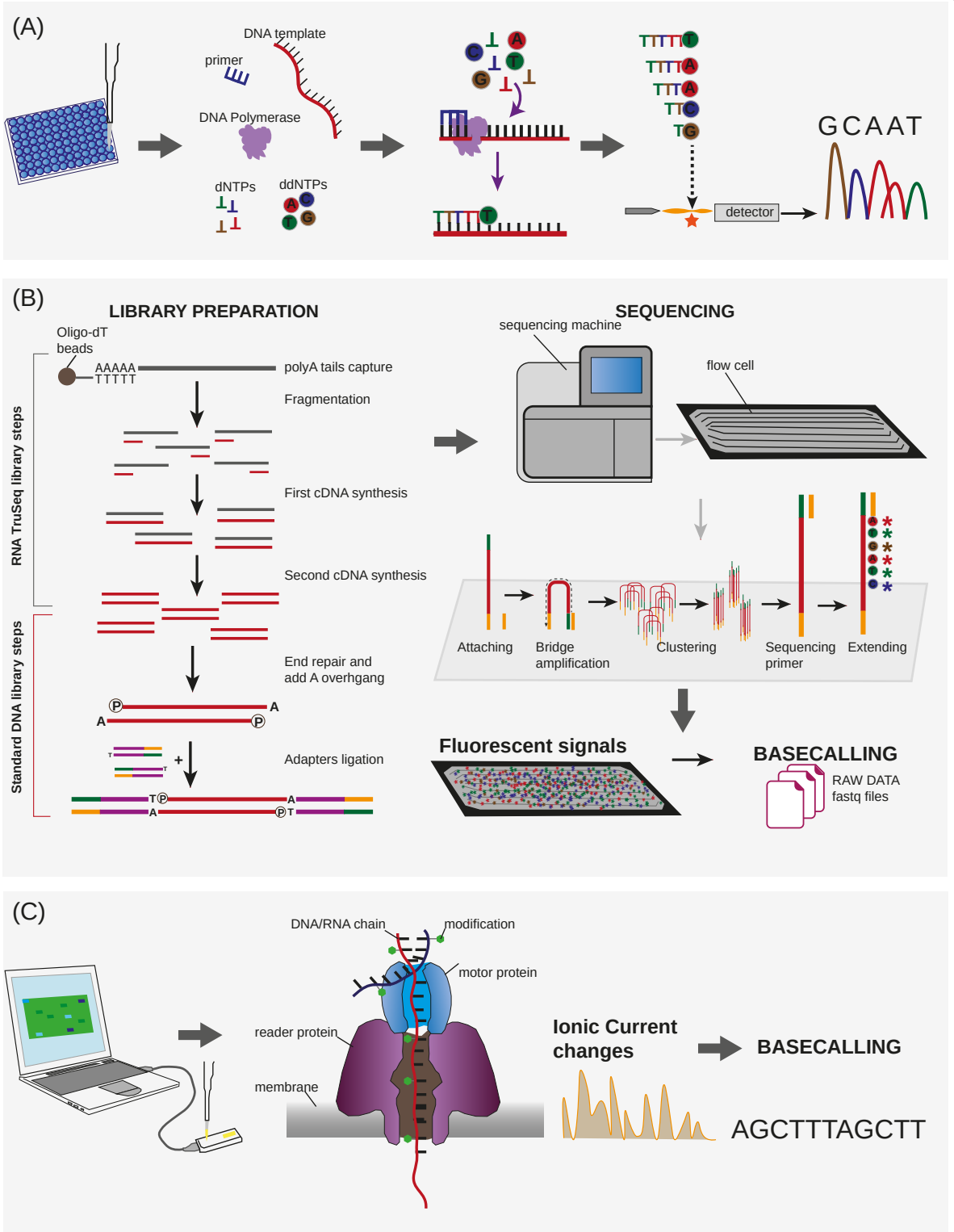


Figure 1. Sequencing systems. (A): Sanger method based on dideoxy technique. DNA template, primers, DNA polymerase, deoxyribonucleotides of DNA (dNTPs), and dideoxynucleotides (ddNTPs) are the components of the reaction. The ddNTPs lack the 3′ hydroxyl group that is required for the extension of DNA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

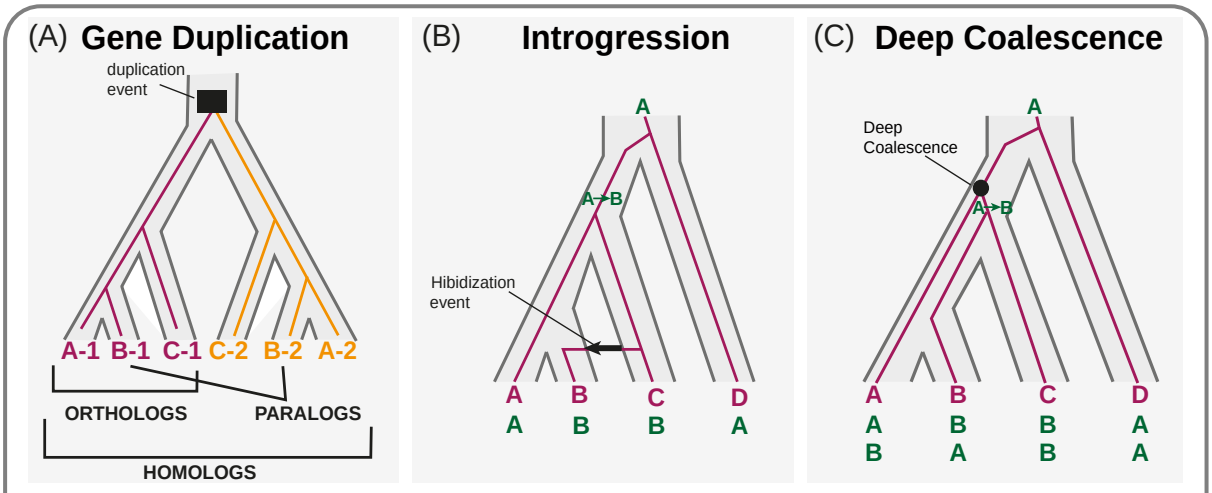


Figure 2. Schematic representation of principal sources of gene tree discordance. The species tree is highlighted in grey and the gene tree in yellow and/or garnet (A): Gene duplication. A duplication event (black square) lead to two gene copies (yellow and garnet gene trees). The concept of orthologs, 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).

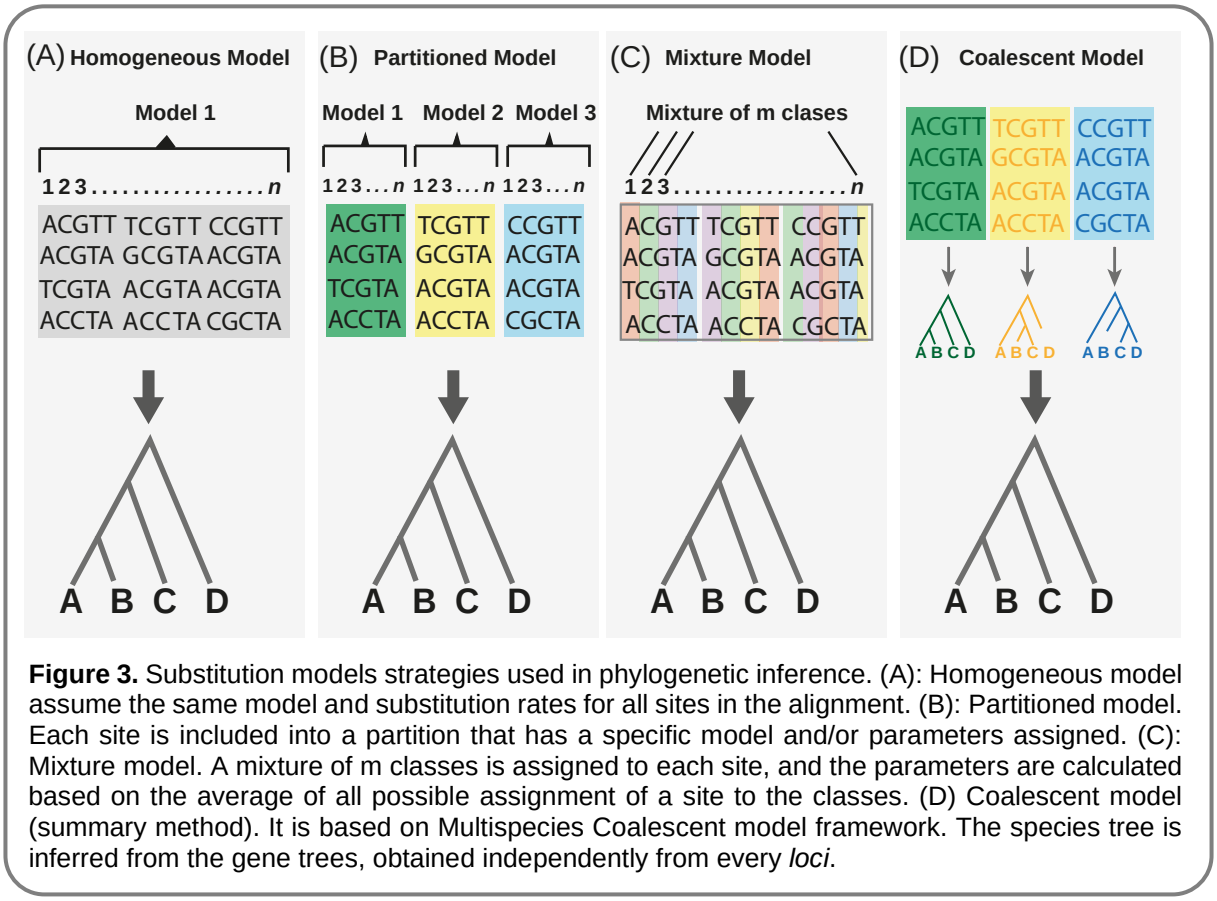


Figure 3. Substitution models strategies used in phylogenetic inference. (A): Homogeneous model assume the same model and substitution rates for all sites in the alignment. (B): Partitioned model. Each site is included into a partition that has a specific model and/or parameters assigned. (C): Mixture 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.

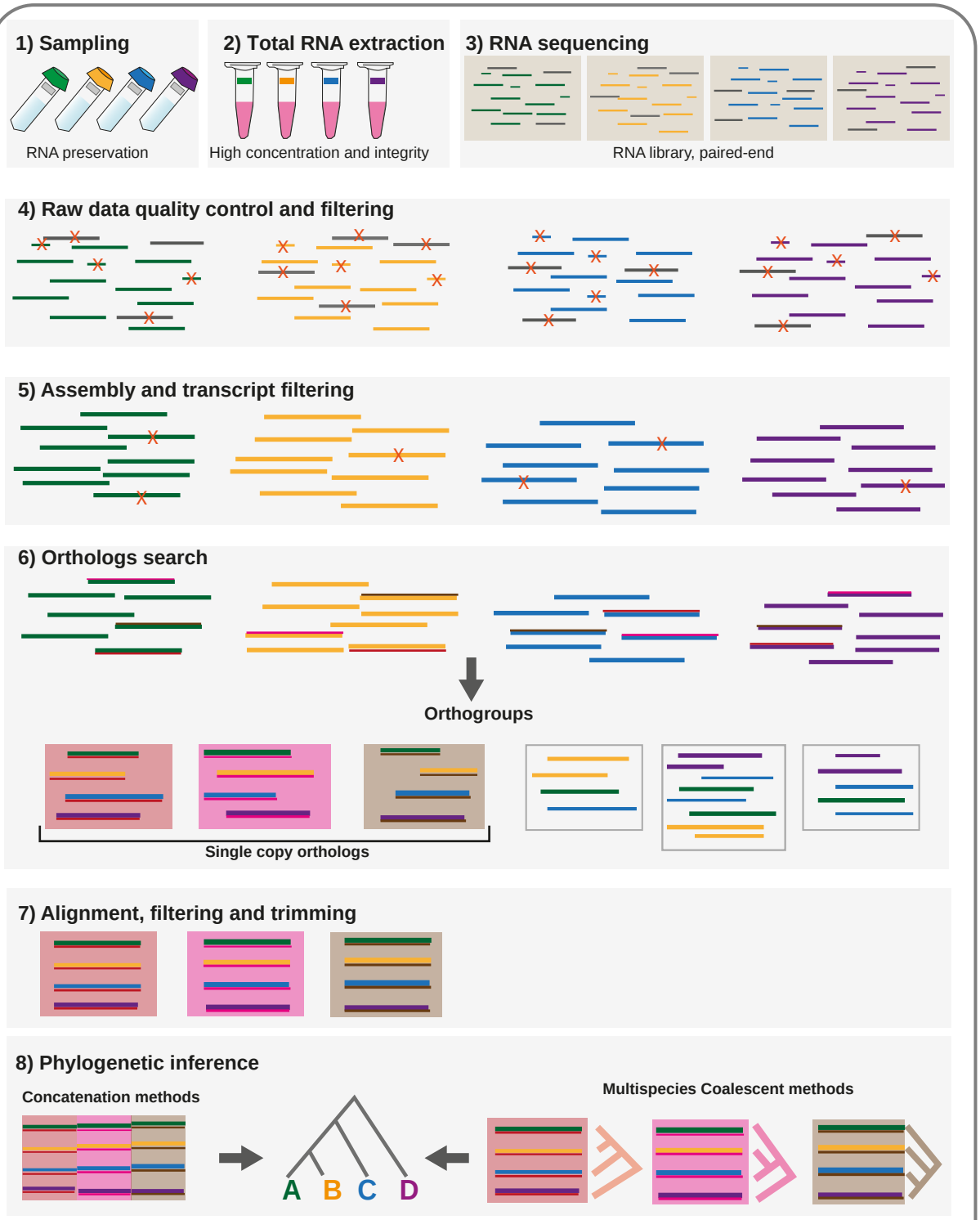


Figure 4. Phylotranscriptomic workflow. 1) The samples must be conserved in conditions that preserve the RNA against degradation. 2) High quality RNA extraction 3) RNA sequencing, paired-end reads are recommended. 4) Raw data quality control and filtering by quality, length and contaminants. 5) Assembly 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).

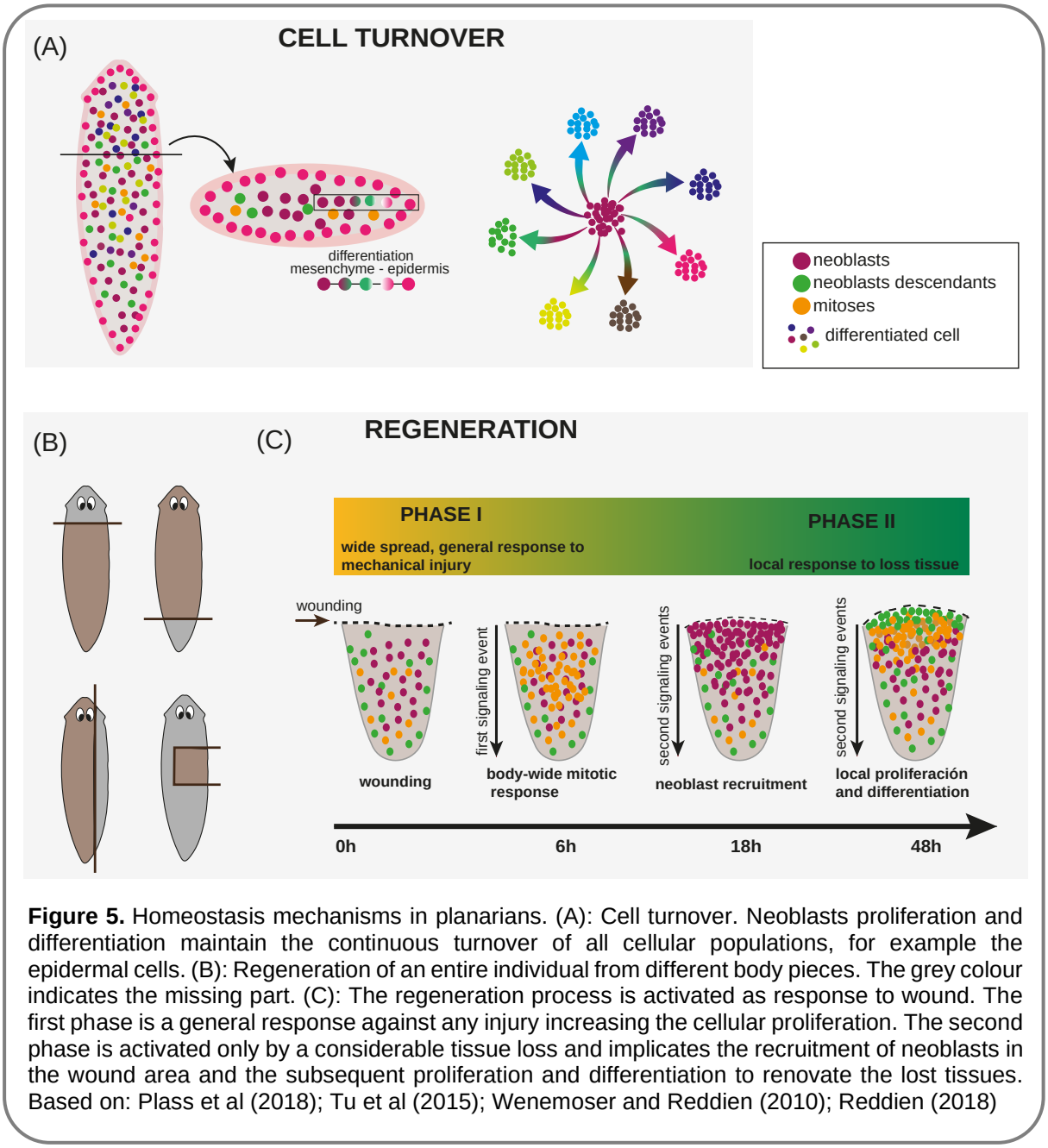


Figure 5. Homeostasis mechanisms in planarians. (A): Cell turnover. Neoblasts proliferation and differentiation maintain the continuous turnover of all cellular populations, for example the epidermal cells. (B): Regeneration of an entire individual from different body pieces. The grey colour indicates 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.

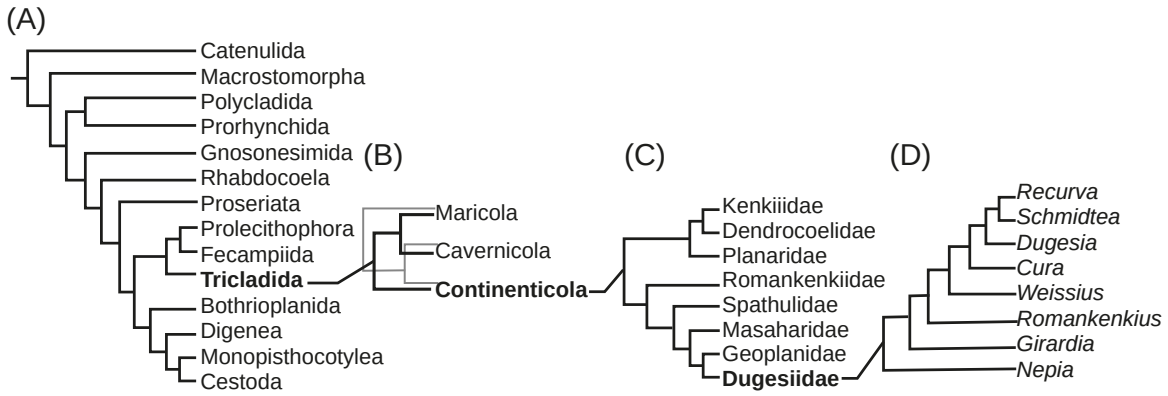


Figure 6. Schematic relationships of Tricladida at different taxonomic levels. A): Relationships for Platyhelminthes orders (Laumer et al., 2015). B): Tricladida relationships, the grey lines represent a different phylogenetic hypothesis based on morphological data (Harrath et al., 2016; Sluys et al 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).

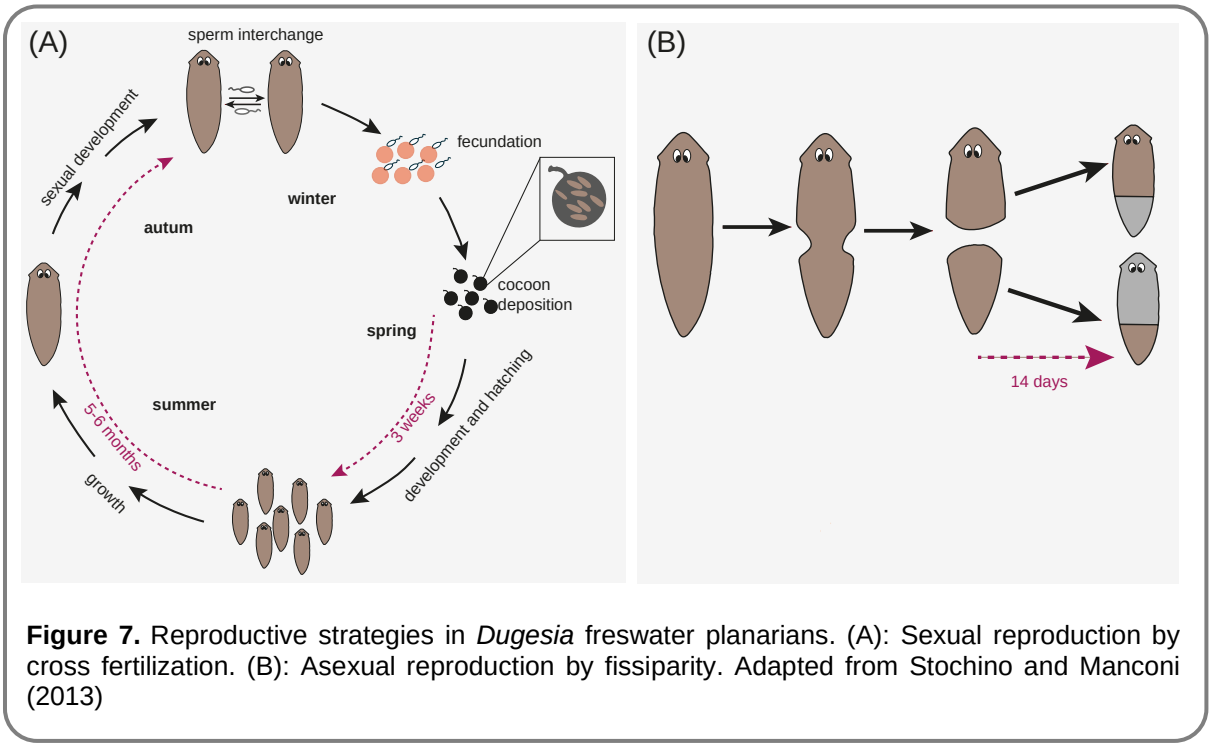


Figure 7. Reproductive strategies in *Dugesia* freshwater planarians. (A): Sexual reproduction by cross fertilization. (B): Asexual reproduction by fissiparity. Adapted from Stochino and Manconi (2013)

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).

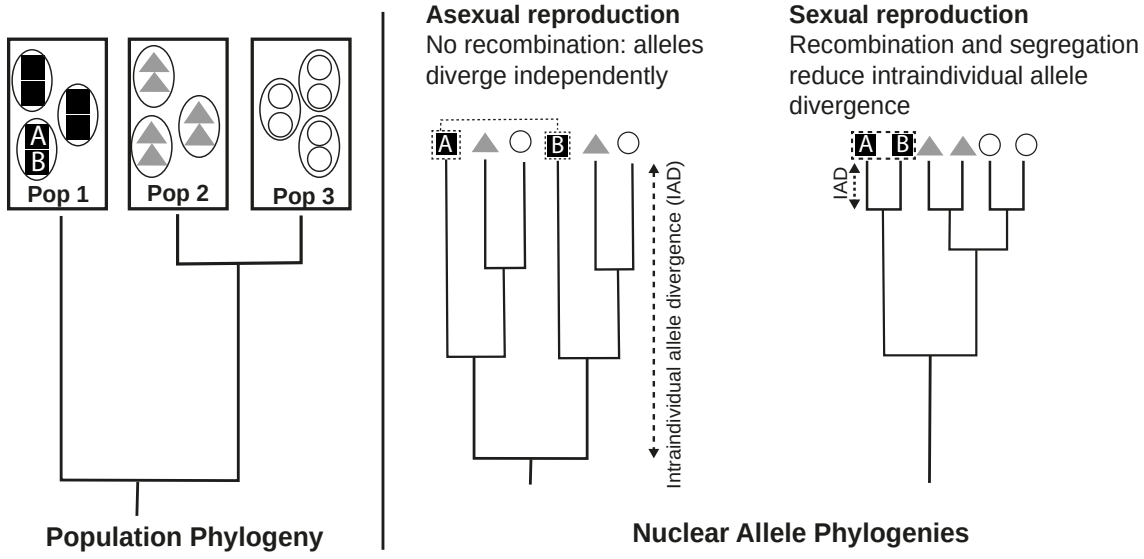


Figure 8. Meselson effect. The intraindividual allele divergence is higher in asexual individuals due to the lack of recombination. Scheme reproduced from Schawander et al., (2011).

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/

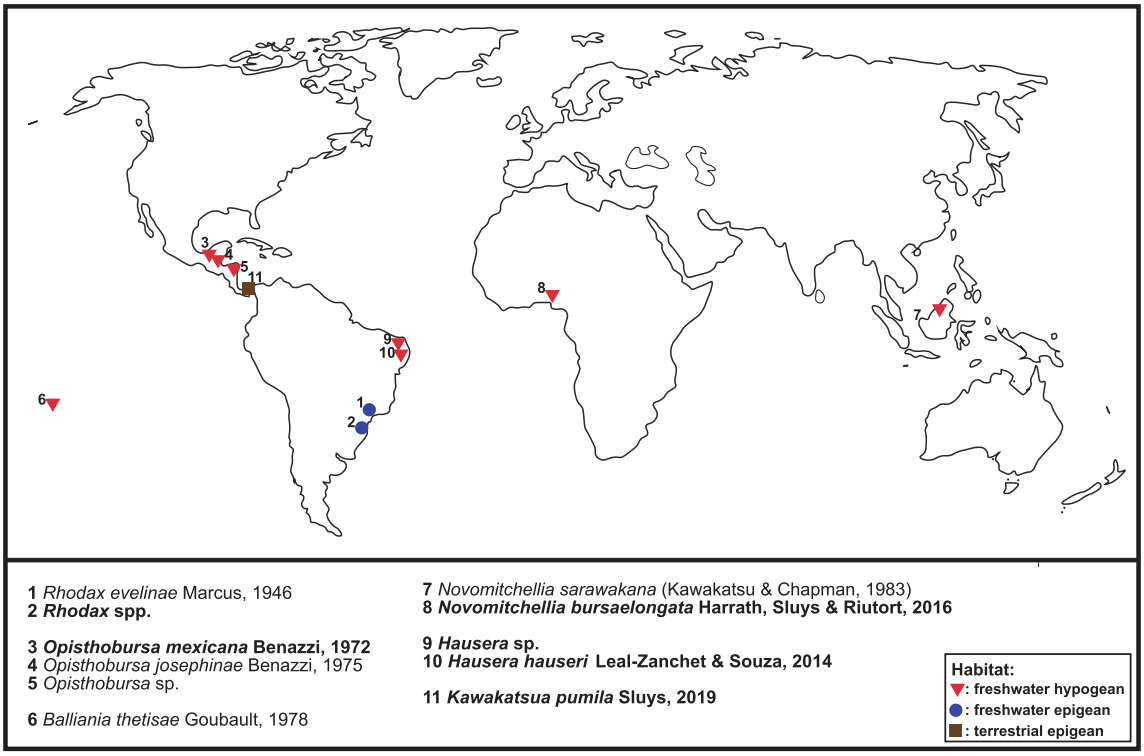


Fig. 1. Distribution of cavernicolan taxa; all known sites are shown. The taxa included in this study are highlighted in boldface.

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]

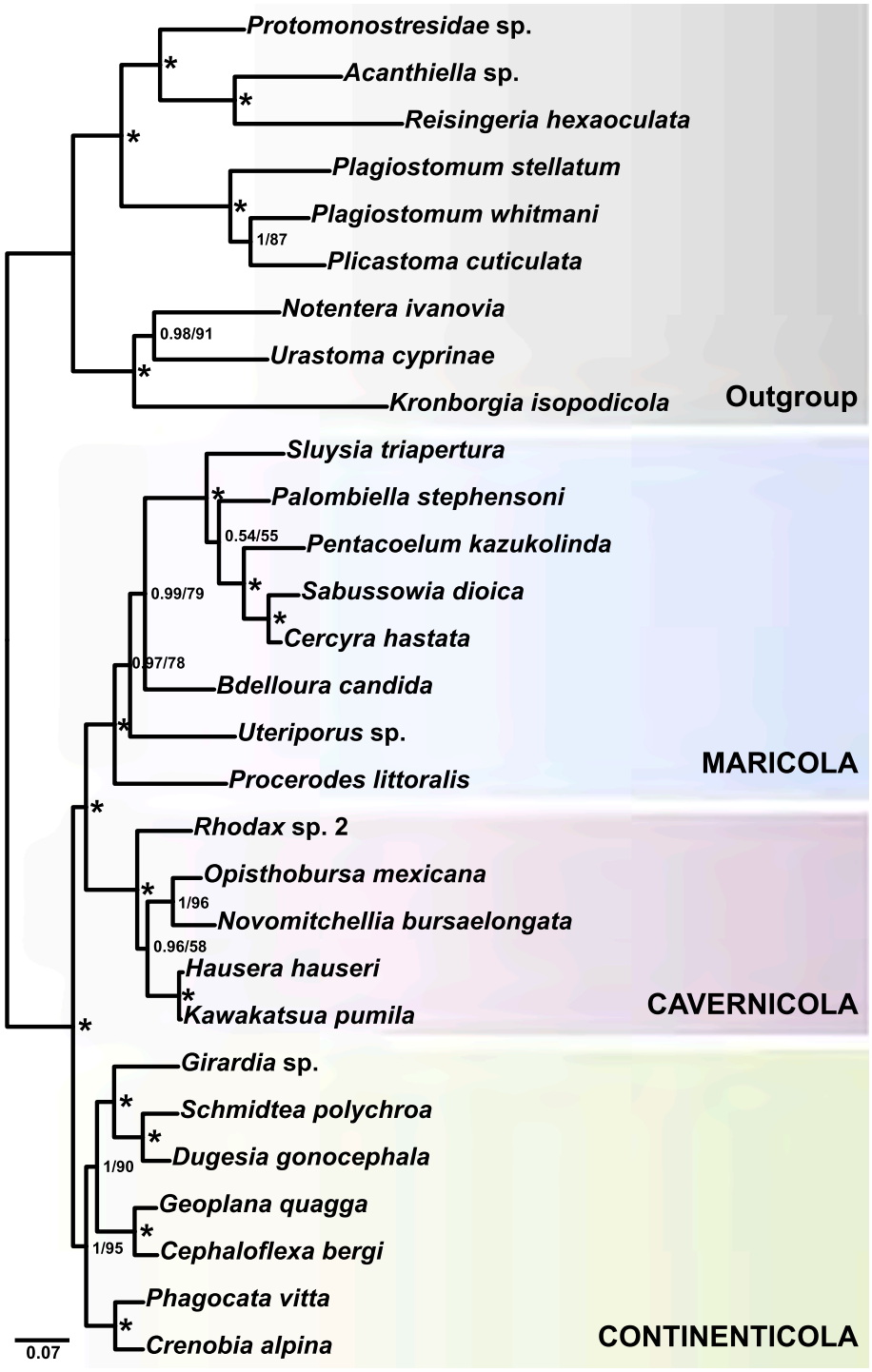


Fig. 2. Bayesian Inference tree inferred from the dataset including 18S and 28S sequences of representatives of the various Suborders of the Tricladida (dataset I-[18S + 28S]). Values at nodes correspond to posterior probability/bootstrap support. *: 1.00 and 100% values for BI and ML, respectively.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

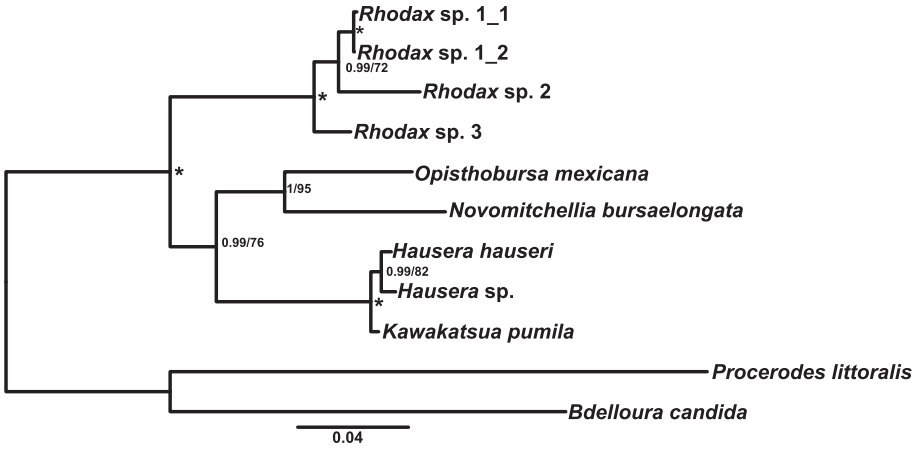


Fig. 3. Relationships within the Cavernicola inferred by Bayesian Inference from the dataset II-[18S + 28S]. Numbers at nodes indicate posterior probability/bootstrap supports for BI and ML, respectively. *: 1.00 and 100% values for BI and ML, respectively. Scale bar indicates number of substitutions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

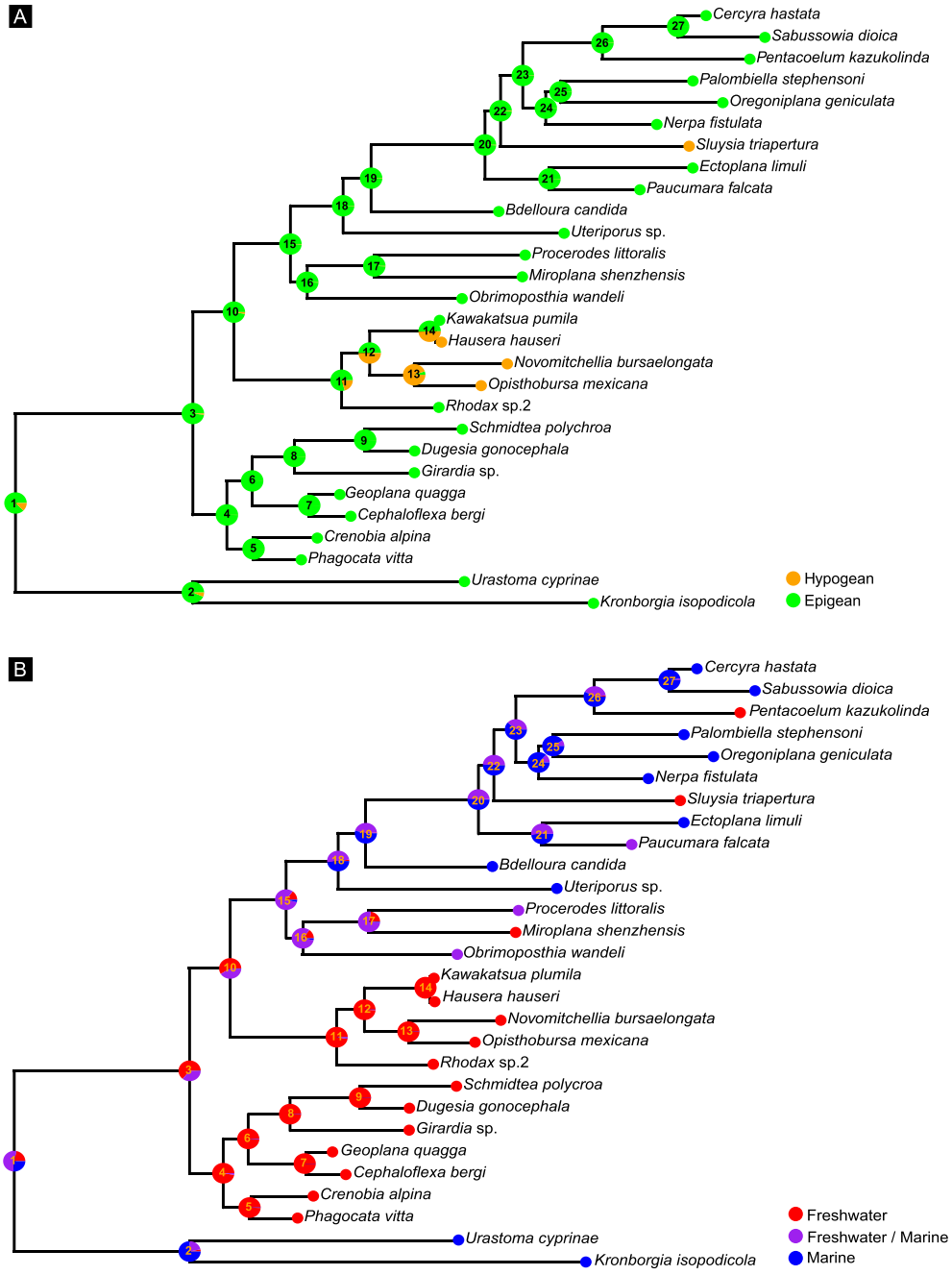


Fig. 4. Results of the Ancestral States Reconstruction analysis based on the Bayesian Inference tree obtained from the dataset III-[18S]. Pie charts at nodes represent the posterior probabilities of ASR analysis for A: epigean (green) and hypogean (yellow) habitat, and B: freshwater (red), freshwater-marine (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.)

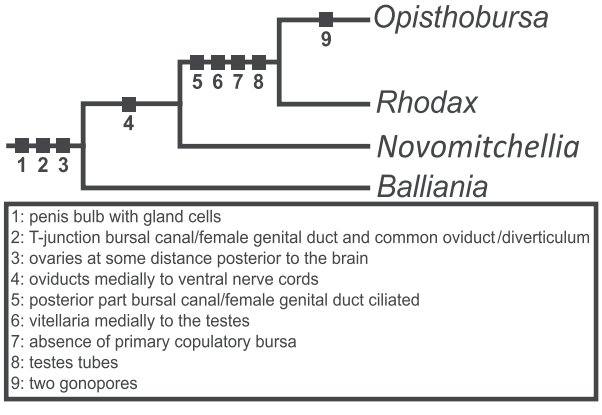


Fig. 5. Morphological phylogeny of Cavernicola based on Sluys (1990). Black rectangles represent morphological apomorphies.

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

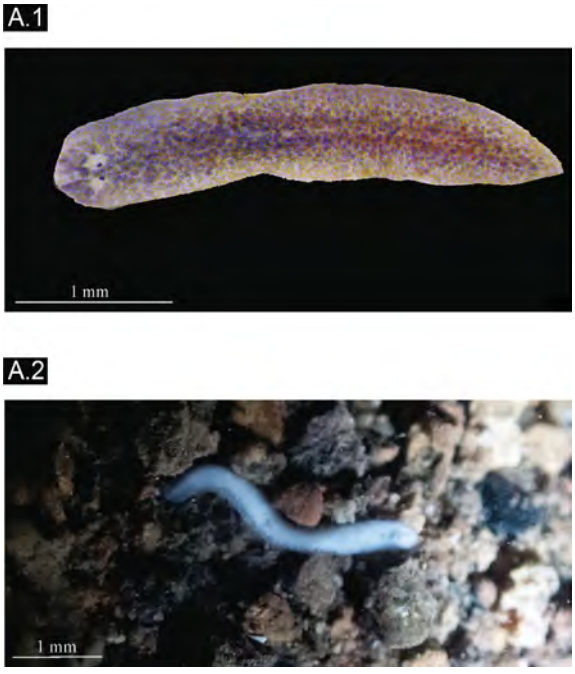
We thank Christopher Laumer for kindly providing information on the locality and some external characteristics of *Kawakatsua pumila* before the species was formally described. We also thank Halim Harrath for making available to us a photograph of the type locality of *Novomitchellia bursaelongata.*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

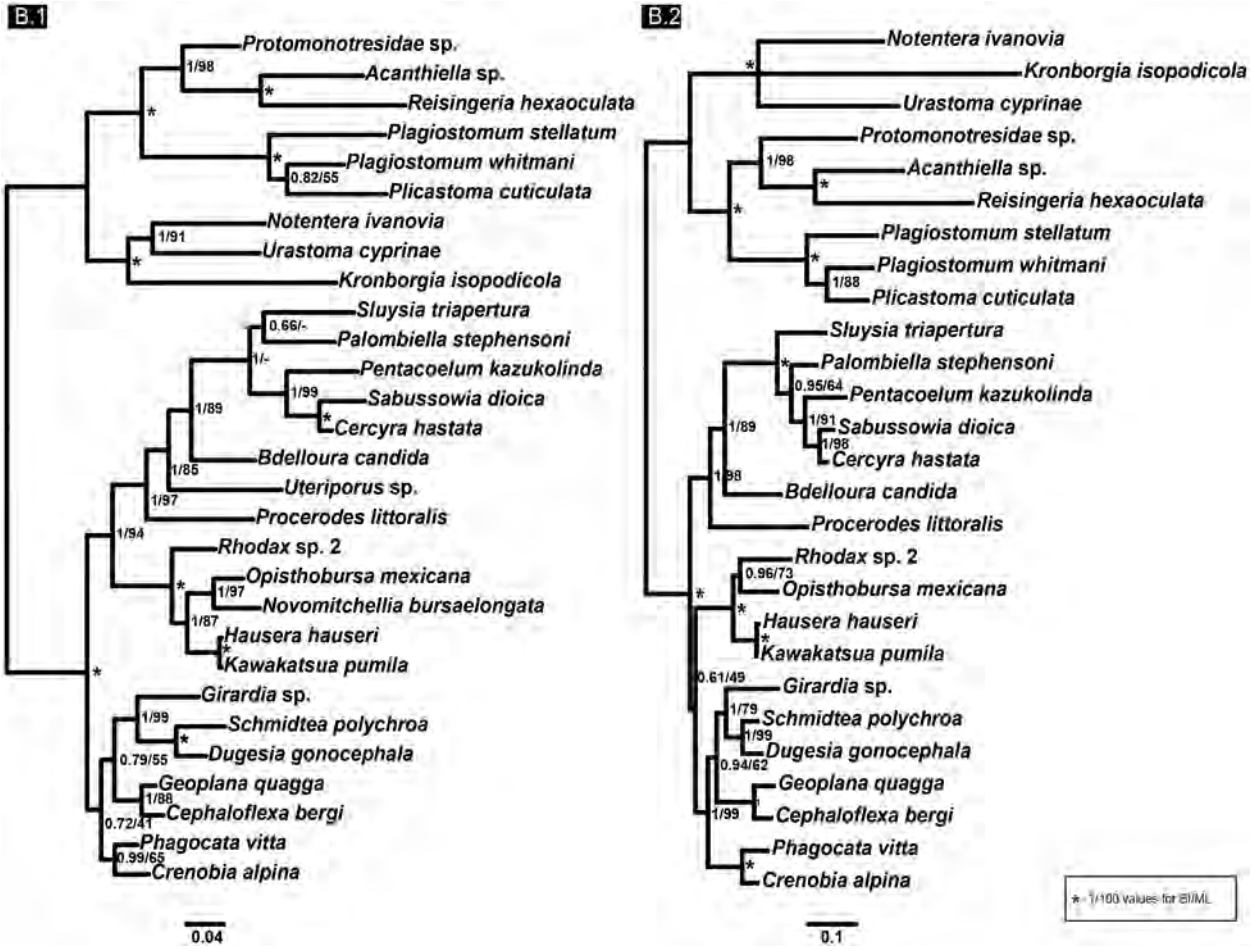
Appendix A

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.



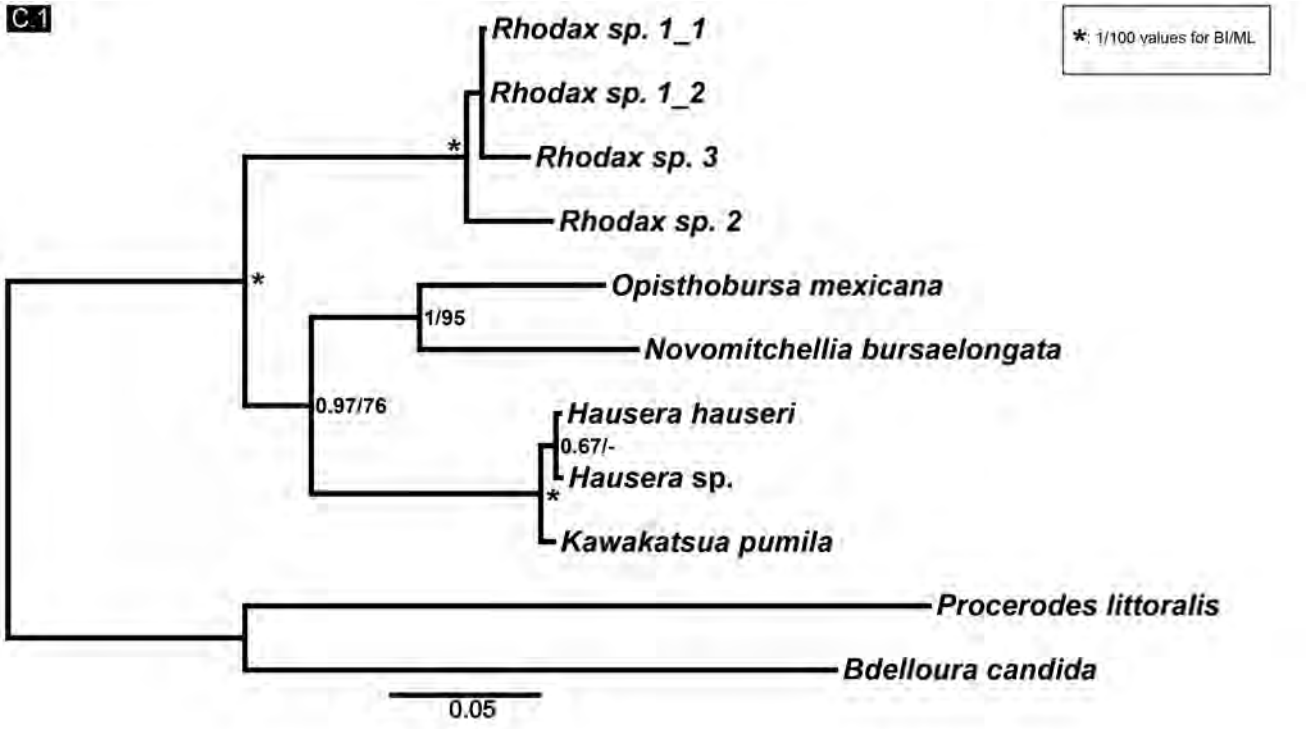
Appendix B

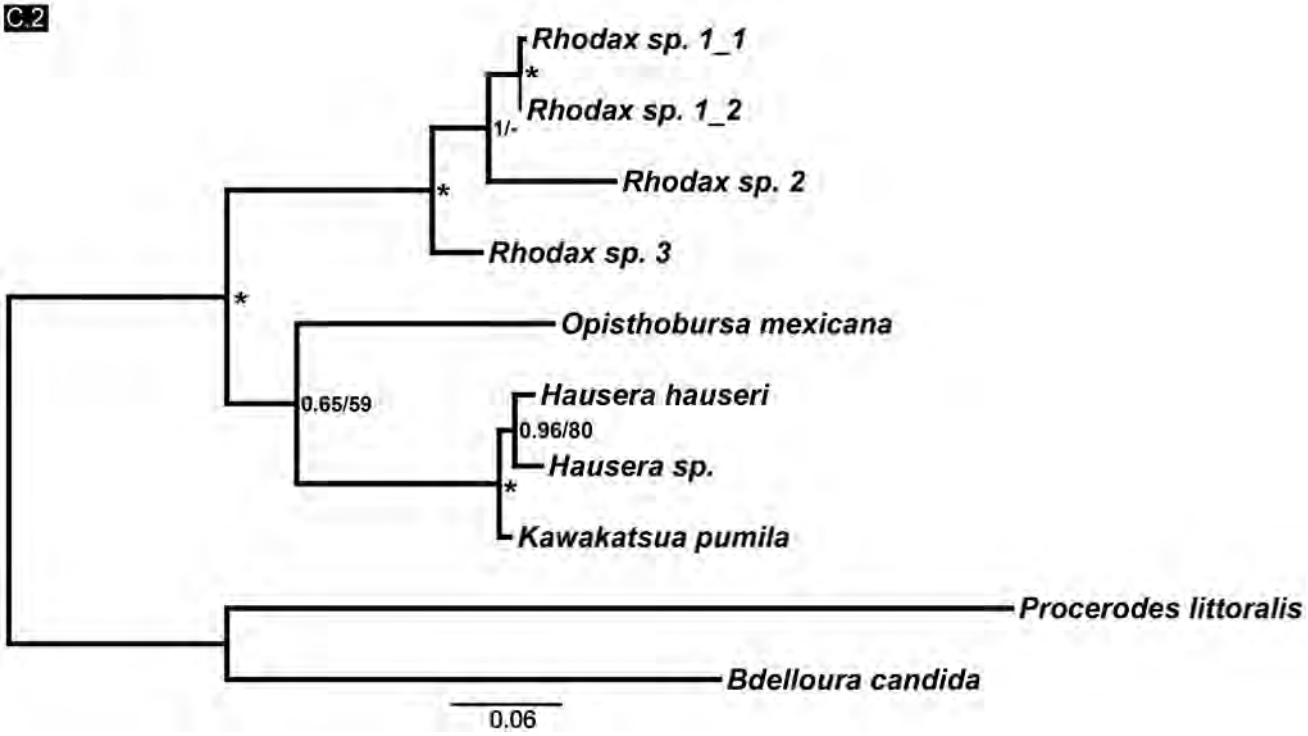
Bayesian Inference tree of Tricladida suborders based on the dataset I-[18S] (B1) and dataset I-[28S] (B2) datasets. Numbers at nodes indicate posterior probability/bootstrap supports for BI and ML, respectively. *: 1.00/100% values for BI/ML. Scale bar indicates number of substitutions per site.



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 "micro-turbellarian" 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).

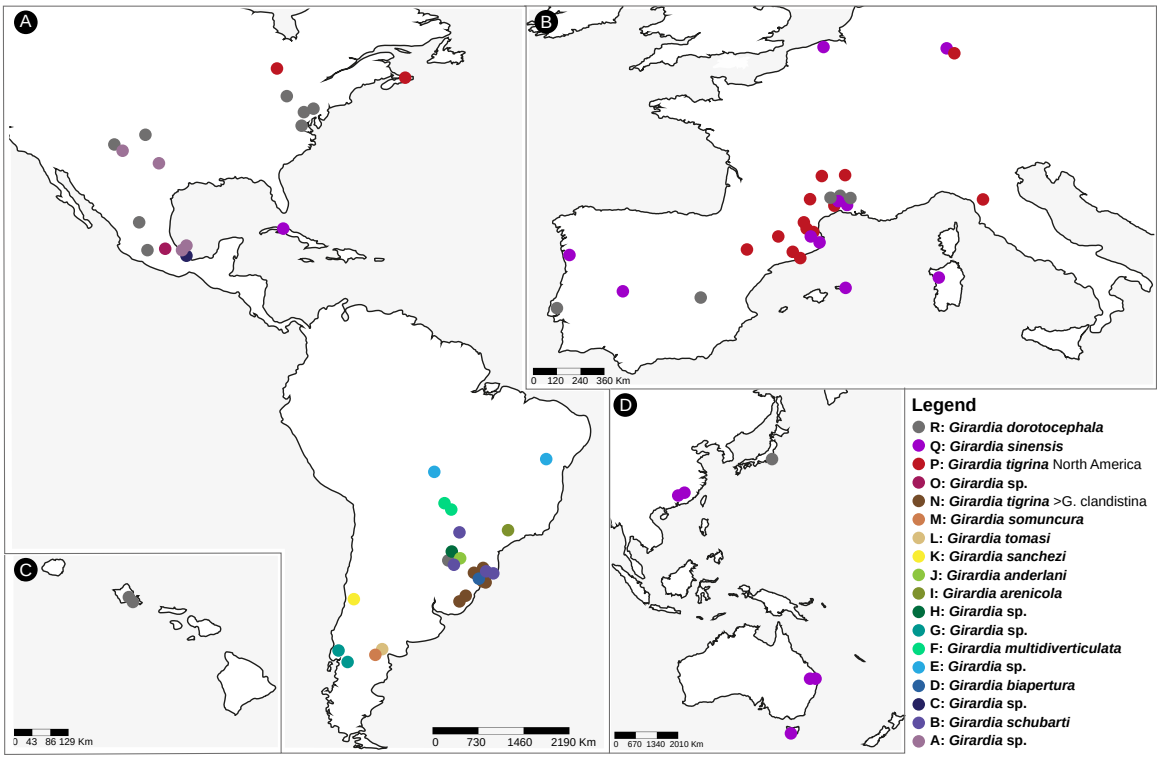


Figure 1. Maps showing the sampling localities for all individuals analyzed in the present study, including those corresponding to GenBank sequences. (A): The Americas; (B): Western Europe; (C): Hawaii; (D): Asia and Oceania. For a finer resolution, visit the interactive map at: https://www.ub.edu/planarian-maps/

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*.

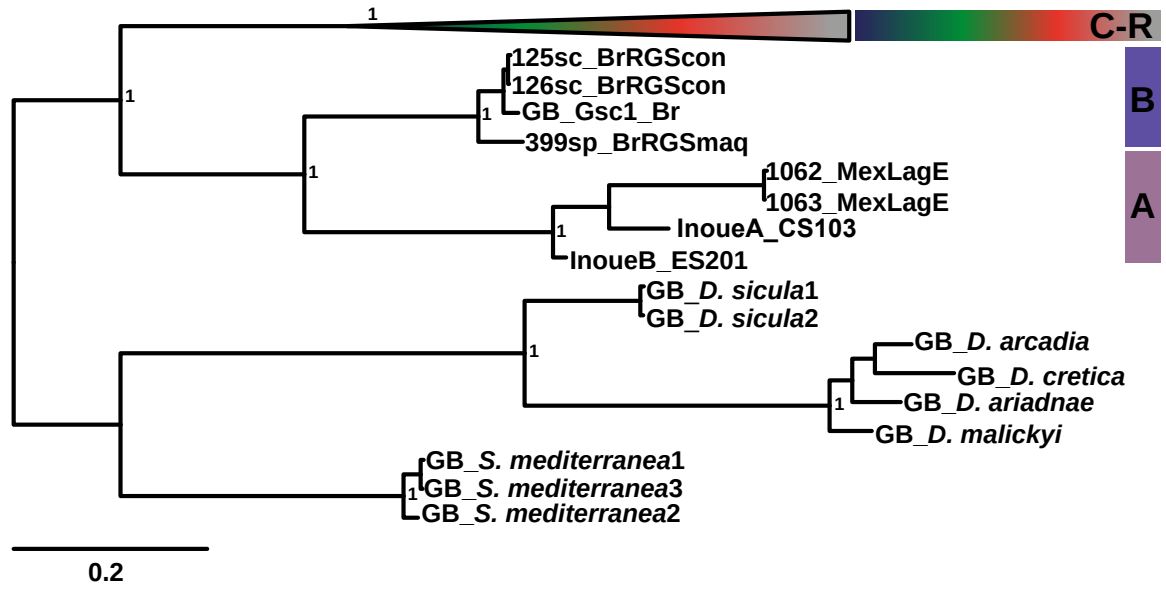


Figure 2. Bayesian Inference tree inferred from Dataset 2 (COI with outgroup). Clades C to R have been collapsed for the sake of clarity. Clade A is constituted by unclassified samples from Mexico and Texas (USA); Clade B includes identified individuals of Girardia schubarti from Brazil and other unidentified 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

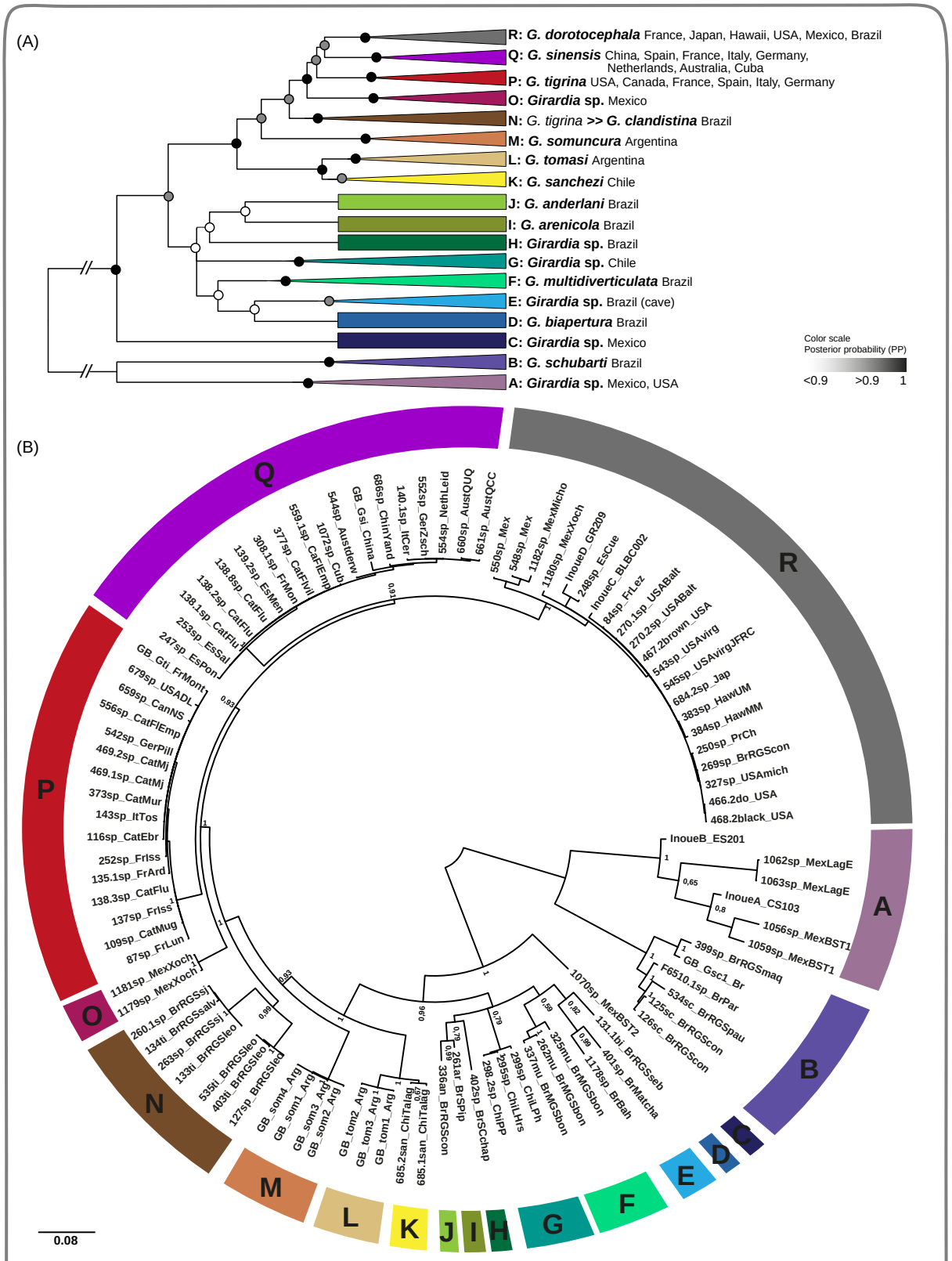


Figure 3. Bayesian Inference tree inferred from Dataset5 (concatenated no outgroup). Different clades indicated by letters and colours. (A): Schematic representation of the tree with collapsed clades, showing species identifications, when available, and countries of origin of the various terminals; shaded 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

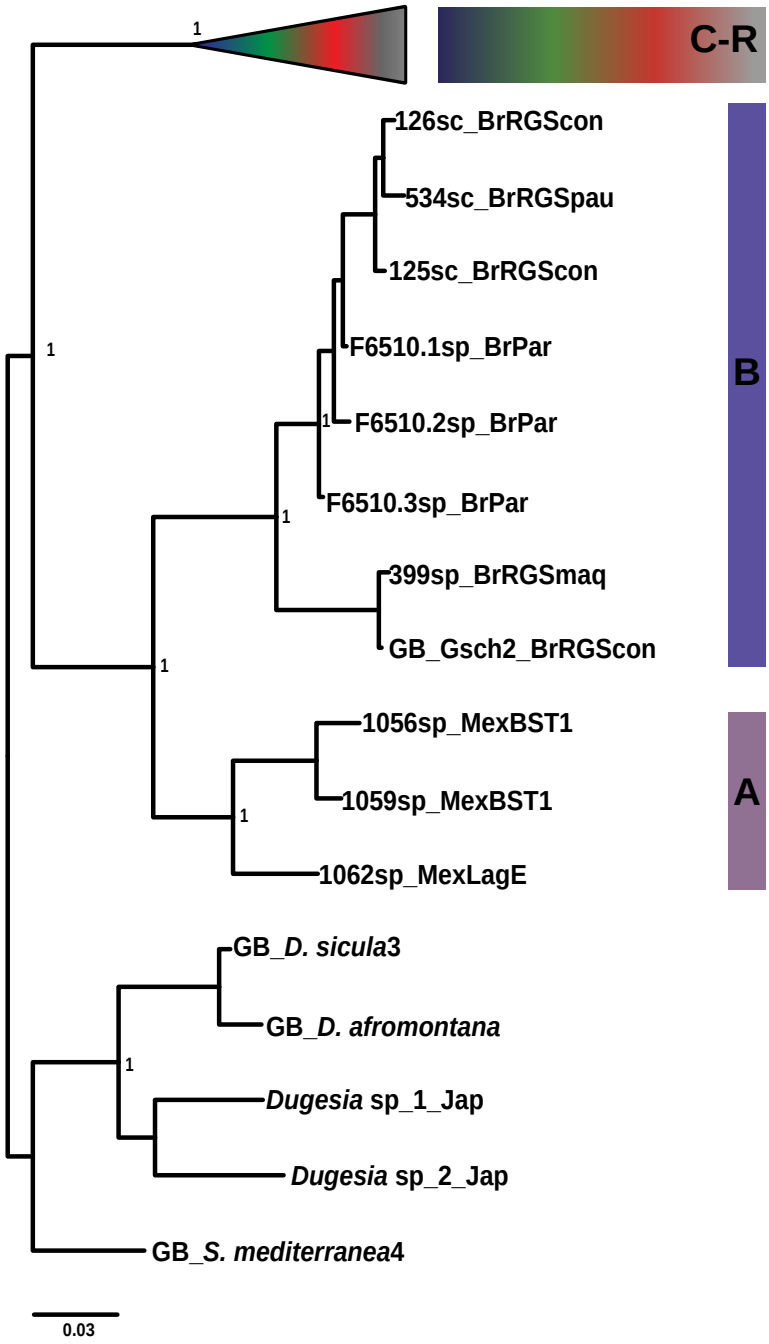
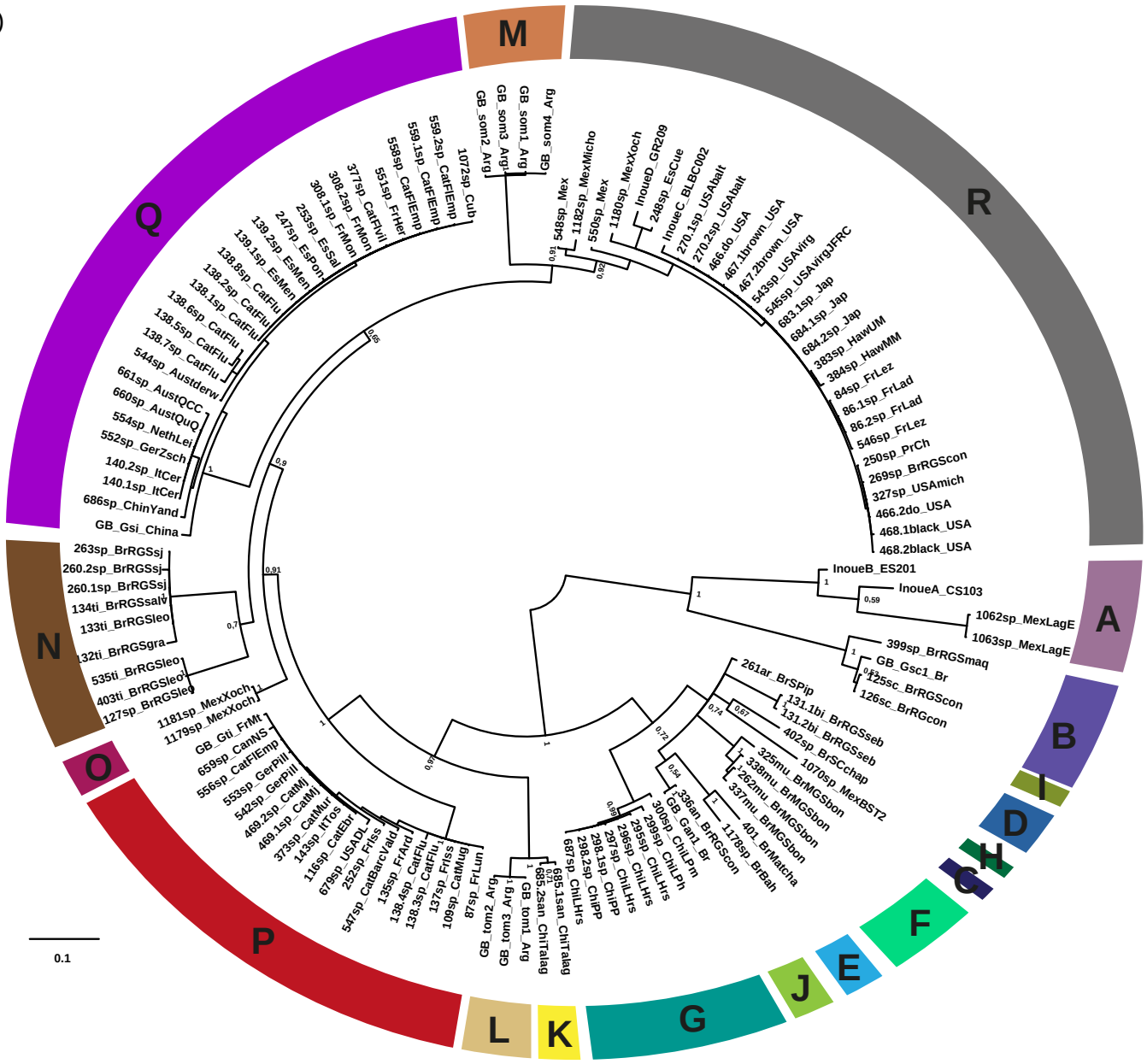


Figure S1. Bayesian Inference tree from Dataset4 (EF1a with outgroup). The sister group of *Girardia schubarti* and unclassified samples from Mexico have been collapsed. Values at nodes correspond to posterior probability support. Scale bar: number of substitutions per nucleotide position.

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.

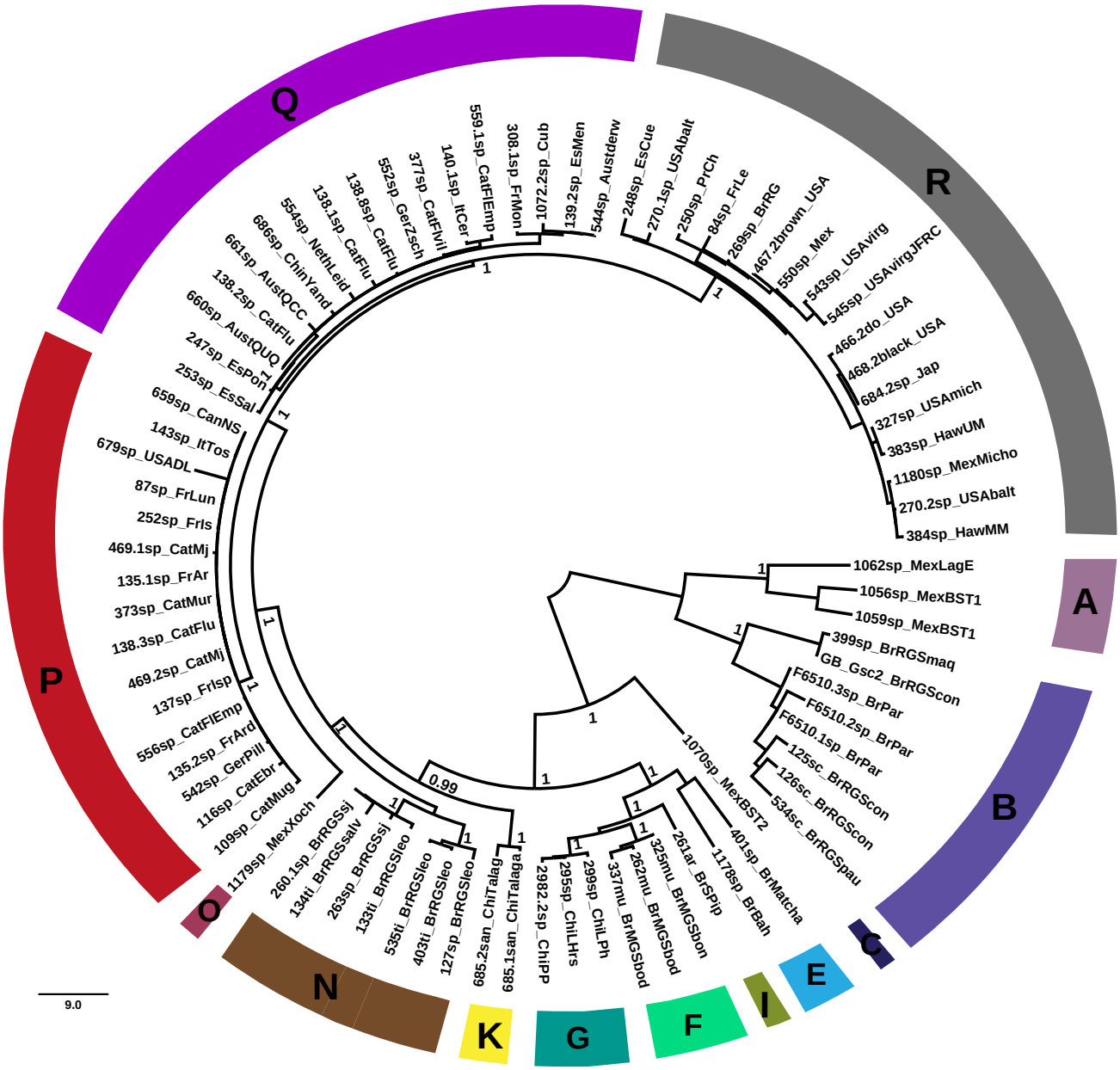
(A)



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