Molecular Phylogenetics and Evolution 178 (2023) 107649

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

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

Lisandra Benítez-Álvarez [a], Laia Leria [a,b], Rosa Fernández [c], Eduardo Mateos [b,d], Younes El Ouanighi [e], Nard Bennas [e], Majida El Alami [e], Mohamed Yacoubi-Khebiza [f], Houssam Ayt Ougougdal [f], Marta Riutort [a,b,*]

[a] *Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Catalonia, Spain*
[b] *Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Catalonia, Spain*
[c] *Metazoa Phylogenomics Lab, Biodiversity Program, Institut de Biologia Evolutiva (CSIC- Universitat Pompeu Fabra), Passeig marítim de la Barceloneta 37-49, 08003 Barcelona, Catalonia, Spain*
[d] *Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Catalonia, Spain*
[e] *Laboratoire Ecologie, Systématique, Conservation de la Biodiversité (LESCB) URL-CNRST N° 18, FS, Abdelmalek Essaadi University, Avenue Sebta, Mhannech II, 93002 – Tetouan, Morocco*
[f] *Water, Biodiversity and Climate Change Laboratory, Semlalia Faculty of Sciences, Cadi Ayyad University, BP 2390, Avenue Le Prince Moulay Abdellah, 400 10 Marrakech, Morocco*

ARTICLE INFO

Keywords:
Asexuality
Biogeography
Fissiparity
Phylotranscriptomics
Phylogeny
Planaria

ABSTRACT

The Mediterranean is one of the most biodiverse areas of the Paleartic region. Here, basing on large data sets of single copy orthologs obtained from transcriptomic data, we investigated the evolutionary history of the genus *Dugesia* in the Western Mediterranean area. The results corroborated that the complex paleogeological history of the region was an important driver of diversification for the genus, speciating as microplates and islands were forming. These processes led to the differentiation of three main biogeographic clades: Iberia-Apennines-Alps, Corsica-Sardinia, and Iberia-Africa. The internal relationships of these major clades were analysed with several representative samples per species. The use of large data sets regarding the number of *loci* and samples, as well as state-of-the-art phylogenomic inference methods allowed us to answer different unresolved questions about the evolution of particular groups, such as the diversification path of *D. subtentaculata* in the Iberian Peninsula and its colonization of Africa. Additionally, our results support the differentiation of *D. benazzii* in two lineages which could represent two species. Finally, we analysed here for the first time a comprehensive number of samples from several asexual Iberian populations whose assignment at the species level has been an enigma through the years. The phylogenies obtained with different inference methods showed a branching topology of asexual individuals at the base of sexual clades. We hypothesize that this unexpected topology is related to long-term asexuality. This work represents the first phylotranscriptomic analysis of Tricladida, laying the first stone of the genomic era in phylogenetic studies on this taxonomic group.

1. Introduction

Freshwater planarians constitute one of the most diverse and broadly distributed groups of free-living flatworms. Due to their low dispersion capability and specific ecosystem requirements (Vila-Farré & Rink, 2018), the evolutionary history of these animals has been strongly shaped by geological changes as has been demonstrated for the genus *Dugesia* (Leria, et al., 2022; Solà et al., 2013; Solà et al., 2022). This

* Corresponding author.
E-mail addresses: lbenitezalvarez87@gmail.com (L. Benítez-Álvarez), l.leria@ceab.csic.es (L. Leria), rosa.fernandez@ibe.upf-csic.es (R. Fernández), emateos@ub.edu (E. Mateos), elouanighi.younes10@gmail.com (Y. El Ouanighi), nbennas@uae.ac.ma (N. Bennas), melalamielmoutaoukil@uae.ac.ma (M. El Alami), yacoubi@uca.ac.ma (M. Yacoubi-Khebiza), houssam.aytougougdal@ced.uca.ma (H. Ayt Ougougdal), mriutort@ub.edu (M. Riutort).

https://doi.org/10.1016/j.ympev.2022.107649
Received 15 July 2022; Received in revised form 13 October 2022; Accepted 18 October 2022
Available online 22 October 2022

genus has a broad distribution covering Eurasian, African and Australasian regions. Based on molecular data and biogeographic analyses it has been proposed that *Dugesia* arrived in Western Europe from Africa through terrestrial connections in the Eocene, splitting, around 30 mya, from the Eastern lineages that arrived in another wave from North Africa through the Arabian Peninsula and the ancient Aegean region (Solà et al., 2022).

Thirteen species of *Dugesia* are described as endemic from the Western Mediterranean and belonging to the molecularly defined Western Mediterranean clade (Leria et al., 2022; Solà et al, 2022); being *D. gonocephala* (Dugès, 1830), the only one that extends its distribution out of the region to the East and the North of Europe (Fig. 1). Moreover, two new lineages suspected to be new species have been recently found in Morocco (Leria et al., 2022). On the other hand, *D. sicula* Lepori, 1948 and *D. maghrebiana* Stocchino et al., 2009, also present in the Mediterranean region, belong to a different and very distant African lineage (Solà et al., 2022); although there are no molecular data for the latter species, its chromosome number and anatomy clearly point to its belonging to the African clade (Stocchino et al., 2009).

Despite its diversity, the evolutionary history of *Dugesia* in the Western Mediterranean is still not resolved. Morphological studies have been limited principally because of the presence of several asexual populations in the region (Fig. 1). Asexual reproduction prevents the assignment of populations to species due to their lack of reproductive structures, which are the principal source of evidence for taxonomic assignment. Thus, for many years most Mediterranean *Dugesia* populations were classified as *Dugesia gonocephala* sensu lato. This problem began to be solved when molecular data was applied to analyse the *Dugesia* populations present in the Mediterranean area (Baguñà et al., 1999, Lázaro et al., 2009), demonstrating that DNA sequences facilitated the assignment of individuals to its species, and also could yield information to envision their phylogenetic relationships. In that initial work, a basic scheme of major relationships within the group was found, but the use of only two molecular markers left many relationships poorly resolved. More recently, a study analysed representative samples of a large part of the species from the region using six molecular markers (Leria et al., 2022). They put the evolution of the genus in this area on a temporal frame, and with the help of niche modelling and ancestral areas reconstruction analyses, the authors proposed a complex and interesting biogeographic hypothesis (Leria et al., 2022). However, in their phylogeny non-supported nodes remained, and only one or two representatives per species were included, so the species' internal relationships and evolutionary history were not analysed.

In this respect, one of the most diverse and extended lineages in the Western Mediterranean *Dugesia* group is *D. subtentaculata* (Draparnaud, 1801). This nominal species was considered one single species in the Iberian Peninsula, South of France, Balearic Island, and Africa for a long time (De Vries, 1986, 1988). However, an integrative taxonomic study divided it into four species: *D. vilafarrei* Leria, 2020, *D. corbata* Leria, 2020 and *D. aurea* Leria, 2020, all sexual populations restricted to their type localities in the South of the Iberian Peninsula and two localities in the Mallorca Island respectively, and *D. subtentaculata* sensu stricto, with sexual, asexual, and facultative (sexual and asexual individuals in the same locality) populations distributed in all the Iberian Peninsula, South of France and the North of Africa (Leria et al., 2020). The broad distribution of *D. subtentaculata* makes the study of its population structure essential to understanding the evolutionary processes that drove its present distribution and evolution. However, Leria et al. (2020) found it was not possible to reconstruct a phylogeny of the populations, possibly due to the small number of markers used and the noise introduced by the Mosaic-Meselson effect, a genetic consequence of asexuality described just for this species (Leria et al., 2019).

Another interesting question regarding the species *D. etrusca* Benazzi, 1944 and *D. liguriensis* De Vries, 1988 has arisen recently. These species were considered endemic to the Tuscany and Liguria regions respectively (Benazzi, 1946; De Vries, 1988) and were described as strictly sexual. Even so, in the last decades new fissiparous reproducing populations from the Catalonian region in the Iberian Peninsula have

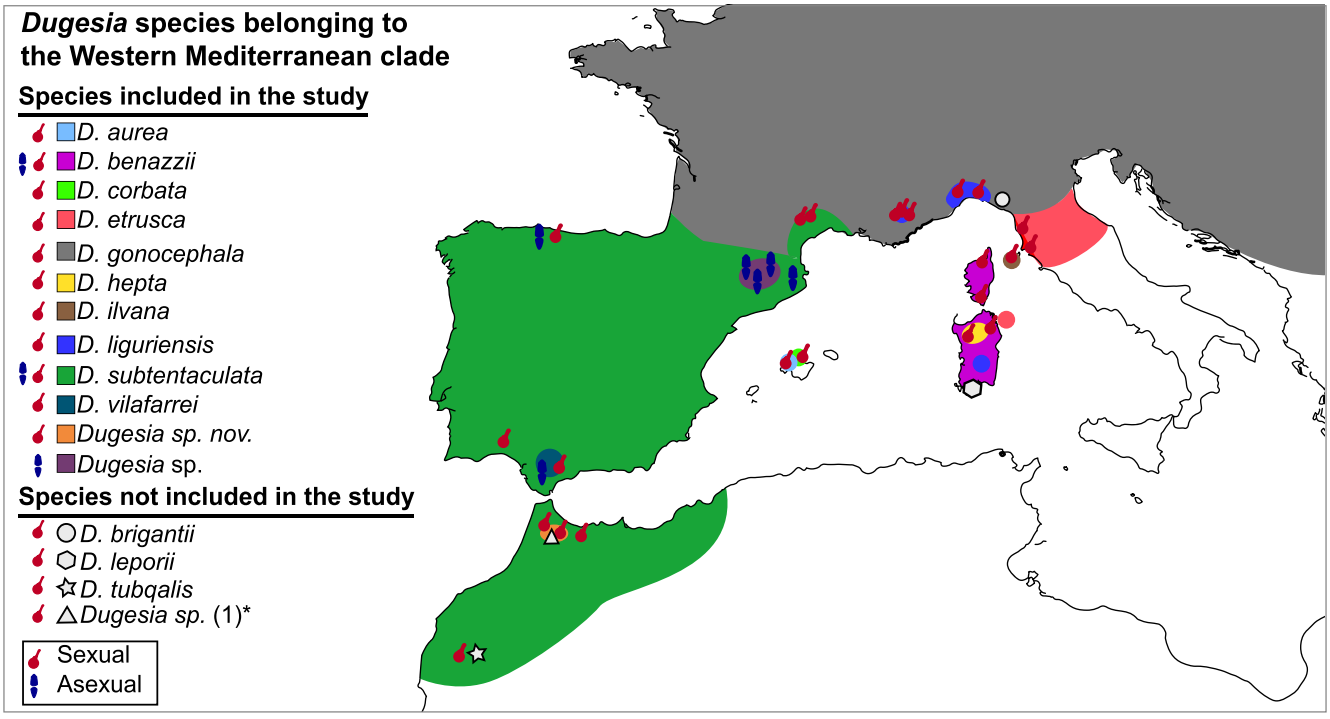


Fig. 1. Distribution map of all *Dugesia* species belonging to the Western Mediterranean clade and their reproductive strategies. The locations of the sampling points included in this study are shown with an icon that also indicates the reproductive strategy of the sampled population (red for sexual reproduction and blue for asexual reproduction). The species list and distribution are based on Leria et al., (2022). *: candidate new species in Leria et al., (2020, 2022). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

been assigned to this clade using molecular data, but without a precise assignment to species level (Baguñà et al., 1999; Lázaro et al., 2009). These populations are interesting since apparently, they represent a restricted asexual lineage geographically isolated from the sexual populations. In the Southern region of France situated between both groups of organisms no other population that could be assigned to one of these species has been found.

A new approach is now necessary to solve the remaining uncertainties in the evolutionary history of *Dugesia* in the Western Mediterranean. Nowadays, access to whole-genome information has opened the door to a new era in phylogenetic studies with a substantial increase in the informative regions to be analysed. Phylogenomics has been demonstrated to be a powerful approach to reconstruct molecular phylogenies and has aided to resolve old questions about the evolution of life (Fernández & Gabaldón, 2020; Guijarro-Clarke et al., 2020; Li et al., 2021). However, genomes are not always accessible, more if the studied group has never been sequenced or no references are available. In those cases, phylotranscriptomics arises as a good and cheaper alternative. Transcriptomic data has been used to resolve several phylogenetic questions in non-model organisms (Feng et al., 2021; Fernández et al., 2017; Foley et al., 2019; Laumer et al, 2015) and its correct performance compared to genomic data has been demonstrated (Cheon et al., 2020).

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

2. Material and methods

2.1. Taxon sampling

A thorough sampling effort was carried out throughout the Western Mediterranean to cover all the area. We visited known localities of *Dugesia* species as well as some new localities from April 2018 to March 2020 (Fig. 1). Two species, *Dugesia malickyi* De Vries, 1984 from Mexiates and a candidate new species from Eleonas (Sluys et al., 2013), were collected in Greece to be used as outgroup (Solà et al., 2013, 2022). In addition, different localities from Morocco were sampled looking for *D. tubqalis* and the new species reported in Leria et al. (2020). Unfortunately, no representatives of these taxa were found, but other samples collected in the region were included in the analyses (Table 1, Fig. 1, Table S1).

Animals were fixed *in situ* or in the laboratory. All the material used for sampling and handling the animals was cleaned with RNAase away, and the autoclavable material was sterilized twice (120 °C 20 min). For the transport of live animals, we used a portable refrigerator and the tubes with water from the river and the planarians were opened twice a day to aerate them. The fixations to preserve the RNA, *in situ* or in the laboratory, were done with RNAlater (SIGMA) following the recommendations of the manufacturer. When the size of the animal allowed it, a small portion of the posterior end was cut and fixed in absolute ethanol. In the case of very small animals, some individuals from the same sampling point were conserved in absolute alcohol and the rest in RNAlater. DNA for species identification was extracted from absolute ethanol fixed samples.

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

The remaining samples and nucleic acid extractions are stored in the freezers at the Department of Genetics, Microbiology and Statistics (Universitat de Barcelona) (Table 1). The methodology is summarized in a diagram highlighting the principal steps of all the procedure (Fig. 2).

2.2. Nucleic acids extraction and RNA library preparation

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

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

2.3. DNA-Barcoding identification

Samples were preliminary assigned to species level according to their locality based on information from previous studies (Lázaro et al., 2009; Leria et al., 2022). Once in the laboratory, a fragment of approximately 800 bp of mitochondrial *Cytochrome Oxidase I* (COI) was used as a marker to corroborate the species assignment done in the field. COI was amplified by Polymerase Chain Reaction (PCR), using 0.4 μM of the BarT (Álvarez-Presas et al, 2011) and COIR (Lázaro et al., 2009) primers in 25 μl of final reaction volume with $MgCl_2$ (2.5 mM), dNTPs (30 μM), and 0.75 U of Go Taq® DNA polymerase enzyme (Promega Madison, Wisconsin, USA) with its buffer (1X). The amplification conditions were : 1) 2″ at 95 °C, 2) 50′ at 94 °C, 3) 45′ at 43 °C, 4) 50′ at 72 °C, 5) 4″ at 72 °C, with 35 cycles of steps 2, 3, and 4.

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

The sequences were aligned with ClustalW Multiple Alignment on the BioEdit Sequence Alignment Editor (Hall, 1999) to a set of sequences downloaded from GenBank database (Table S2), used as references for all *Dugesia* species reported in the area of study. A Bayesian Inference tree was obtained using MrBayes v3.2.2 (Ronquist et al., 2012) with 10 million generations, sampling every 1000 generations, 25 % burn-in, and using three partitions by codon position. Individuals were assigned to the species with which they formed a monophyletic group.

2.4. Bioinformatic workflow for transcriptomic analysis

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

2.4.1. Quality control and trimming

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

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

Sample Code	Species	Reproductive Strategy	Region, Country	Locality
Daur_1	*D. aurea*	Sexual	Mallorca, Spain	Soller
Daur_2	*D. aurea*	Sexual	Mallorca, Spain	Soller
Daur_3	*D. aurea*	Sexual	Mallorca, Spain	Soller
DbenSard_6	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo
DbenSard_7	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo
DbenSard_9	*D. benazzii*	Sexual	Sardinia, Italy	Monte Albo
DbenCors_North_1	*D. benazzii*	Sexual	Corsica, Italy	Campile
DbenCors_North_2	*D. benazzii*	Sexual	Corsica, Italy	Campile
DbenCors_North_3	*D. benazzii*	Sexual	Corsica, Italy	Campile
DbenCors_South_5	*D. benazzii*	Sexual	Corsica, Italy	Monacia-d'Aullène
DbenCors_South_6	*D. benazzii*	Sexual	Corsica, Italy	Monacia-d'Aullène
Dcorb_1	*D. corbata*	Sexual	Mallorca, Spain	Sa Calobra
Dcorb_2	*D. corbata*	Sexual	Mallorca, Spain	Sa Calobra
Dcorb_3	*D. corbata*	Fissiparous	Mallorca, Spain	Sa Calobra
DetruParr_1	*D. etrusca*	Sexual	Italy	Parrana
DetruParr_2	*D. etrusca*	Sexual	Italy	Parrana
DetruPie_2	*D. etrusca*	Sexual	Italy	Pieve
DetruPie_3	*D. etrusca*	Sexual	Italy	Pieve
DetruPie_4	*D. etrusca*	Sexual	Italy	Pieve
Dgono_1	*D. gonocephala*	Sexual	France	Montpellier
Dgono_7	*D. gonocephala*	Sexual	France	Montpellier
Dgono_8	*D. gonocephala*	Sexual	France	Montpellier
Dhept_1	*D. hepta*	Sexual	Sardinia, Italy	Logulento
Dhept_2	*D. hepta*	Sexual	Sardinia, Italy	Logulento
Dhept_5	*D. hepta*	Sexual	Sardinia, Italy	Logulento
Dilv_1	*D. ilvana*	Sexual	Italy	Elba
Dilv_2	*D. ilvana*	Sexual	Italy	Elba
Dilv_4	*D. ilvana*	Sexual	Italy	Elba
DliguBis_1	*D. liguriensis*	Sexual	Italy	Bisagno
DliguBis_2	*D. liguriensis*	Sexual	Italy	Bisagno
DliguBis_3	*D. liguriensis*	Sexual	Italy	Bisagno
DliguAlp_1	*D. liguriensis*	Sexual	France	Alps Marìtims
DliguAlp_3	*D. liguriensis*	Sexual	France	Alps Marìtims
DliguAlp_4	*D. liguriensis*	Sexual	France	Alps Marìtims
DliguGarda_1	*D. liguriensis*	Sexual	France	La Garda
DliguSas_2	*D. liguriensis*	Sexual	Italy	Sassello
DliguSas_3	*D. liguriensis*	Sexual	Italy	Sassello
DliguSas_4	*D. liguriensis*	Sexual	Italy	Sassello
DliguTriga_1	*D. liguriensis*	Sexual	France	Trigance
DliguTriga_2	*D. liguriensis*	Sexual	France	Trigance
DsubMont	*D. subtentaculata*	Sexual	France	Montpellier
DsubBosq_1	*D. subtentaculata*	Fissiparous	Andalusia, Spain	El Bosque
DsubBosq_2	*D. subtentaculata*	Fissiparous	Andalusia, Spain	El Bosque
DsubCangAsex_5	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas
DsubCangAsex_6	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas
DsubCangAsex_7	*D. subtentaculata*	Facultative. Fiss.	Asturias, Spain	Cangas
DsubCangSex_2	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas
DsubCangSex_3	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas
DsubCangSex_4	*D. subtentaculata*	Facultative. Sex.	Asturias, Spain	Cangas
DsubMor_North_1	*D. subtentaculata*	Sexual	Morocco	Magoo Timriouen
DsubMor_North_2	*D. subtentaculata*	Sexual	Morocco	Magoo Timriouen
DsubMor_North_3	*D. subtentaculata*	Sexual	Morocco	Beni M'Hamed
DsubMor_South_1	*D. subtentaculata*	Sexual	Morocco	Imlil
DsubMor_South_2	*D. subtentaculata*	Sexual	Morocco	Imlil
DsubMch_1	*D. subtentaculata*	Sexual	Portugal	Monchique
DsubMch_2	*D. subtentaculata*	Sexual	Portugal	Monchique
DsubMch_4	*D. subtentaculata*	Sexual	Portugal	Monchique
DsubStFe_1	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe
DsubStFe_2	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe
DsubStFe_3	*D. subtentaculata*	Fissiparous	Catalonia, Spain	Santa Fe
Dvila_1	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque
Dvila_2	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque
Dvila_3	*D. vilafarrei*	Sexual	Andalusia, Spain	El Bosque
Dsp_nov_MorNorth_1	*Dugesia* sp. *nov*	Sexual	Morocco	Beni M'Hamed
Dsp_nov_MorNorth_6	*Dugesia* sp. *nov*	Sexual	Morocco	Beni M'Hamed
Dsp_nov_MorNorth_7	*Dugesia* sp. *nov*	Sexual	Morocco	Beni M'Hamed
DspF.US_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de ÍÚs
DspF.US_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de ÍÚs
DspTrilla_1	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspTrilla_2	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspTrilla_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspTrilla_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspTrilla_5	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla

(*continued on next page*)

Table 1 (*continued*)

Sample Code	Species	Reproductive Strategy	Region, Country	Locality
DspTrilla_6	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Font de la Trilla
DspBerga_1	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
DspBerga_2	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
DspBerga_3	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
DspBerga_4	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
DspBerga_5	*Dugesia* sp.	Fissiparous	Catalonia, Spain	Berga
Outgroup				
Dma_1	*D. malickyi*	Sexual	Greece	Mexiates
Dma_2	*D. malickyi*	Sexual	Greece	Mexiates
DspEast_1	*Dugesia* sp.	Sexual	Greece	Eleonas
DspEast_2	*Dugesia* sp.	Sexual	Greece	Eleonas

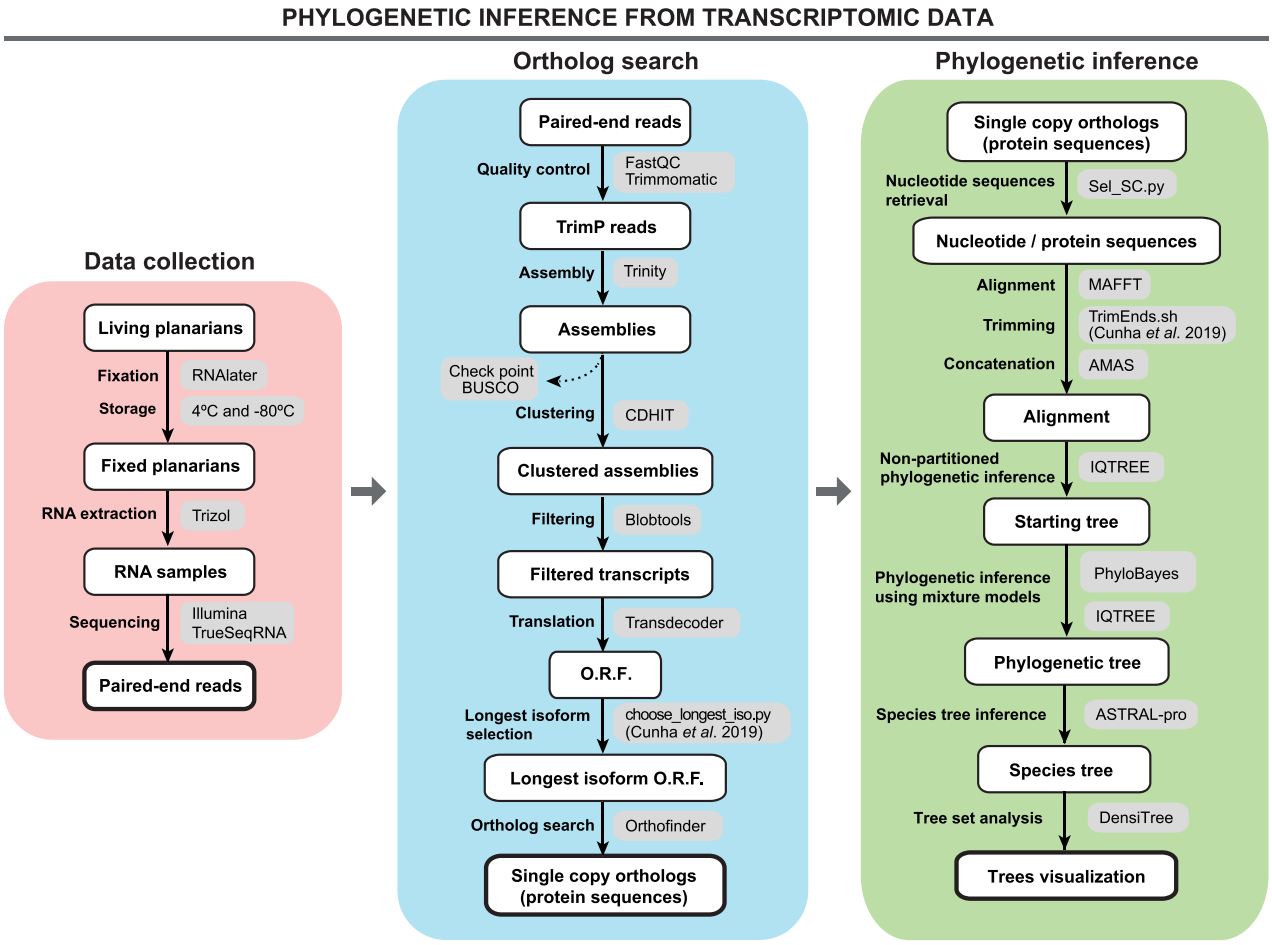


Fig. 2. Diagram summarizing the methodology followed for data collection, ortholog search, and phylogenetic inference.

2.4.2. *Assembly and clustering*

Paired reads were de novo assembled using Trinity v2.9.1 (Grabherr et al., 2011; Haas et al., 2013) following the default options. Some samples were selected for a completeness assessment with BUSCO v5.2.2 (Manni et al., 2021), averaging a completeness value close to 90 % of complete + fragmented BUSCOs using the metazoan database (OB10). Transcripts were clustered using CD-HIT EST (Fu et al., 2012; Li & Godzik, 2006), applying a sequence identity threshold of 0.99, and retaining an average of 97 % of the transcripts (Table S3).

2.4.3. *Transcript filtering*

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

2.4.4. *Ortholog search*

Filtered transcripts were translated to proteins using TransDecoder v

5.5.0 (Haas & Papanicolaou, 2019) and the longest isoforms were selected using the script "choose_longest_iso.py" (Cunha & Giribet, 2019; Fernández et al., 2014). An average of 21,215 longest isoforms by sample was obtained (Table S3).

To accomplish our three aims, we used three main groups of samples to run three ortholog searches. The first one included representatives of all taxa to obtain the phylogeny for the whole Western clade (Data set 1). A second group of samples was defined to study the phylogenetic relationships inside *D. subtentaculata*. This group included all samples assigned to this species and samples classified as *D. vilafarrei*, used as outgroup. We refer to this group as subtentaculata group (Data set 4). The last group, used to assign the unclassified samples from Font de l'Ús, Font de la Trilla, and Berga to species level, includes these samples and the representatives of *D. etrusca, D. liguriensis, D. ilvana* Lepori, 1948, *and D. gonocephala* (the last used as outgroup). We refer to this group as etrusca-liguriensis group (Data set 5).

OrthoFinder v 3.6 (Emms & Kelly, 2019) was used to perform the ortholog search using the longest isoforms. Three searches were performed, one for every group of samples (Tables S4-5). The protein sequences of SC were extracted from the OrthoFinder output and the corresponding nucleotide sequences were extracted from OrthoFinder and Transdecoder outputs, using two custom python scripts.

2.4.5. Alignment and concatenation

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

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

To perform the analyses to respond the three main questions in our study, six data sets were built using the output from the three ortholog searches with different compositions of samples and 100 % of gene occupancy (Table 2, Table S5, Fig. 3): 1) data sets 1–3 constructed from ortholog search 1, 2) data set 4 constructed from ortholog search 2, and 3) data sets 5 and 6 constructed from ortholog search 3. Data sets 2 and 3 were obtained after extracting samples from data set 1, and data set 6 was obtained after extracting samples from data set 5 (Table S5). These reduced data sets were re-aligned and re-processed after removing samples. For all data sets the protein and nucleotide information were analysed (Table 2, Table S6).

2.4.6. Phylogenetic inference

Maximum Likelihood (ML), Bayesian Inference (BI), and Multispecies Coalescence Model (MSC) were performed in IQ-TREE (Minh et al., 2020), PhyloBayes (Lartillot & Philippe, 2004), and ASTRAL-pro (Zhang et al., 2020) respectively. The parameters to carry out every analysis are detailed in Table S6, and a summary of performed analyses by data set is shown in Table 2.

2.4.6.1. Maximum Likelihood Trees: One partition. To explore the data and obtain starting trees for posterior analyses, we used the ML approximation implemented in IQ-TREE without defining partitions. These analyses were run using the ModelFinder Plus (MFP) option for the -m parameter, thus looking for the best-fit model (Kalyaanamoorthy et al., 2017) and 1,000 to 10,000 replicates of ultrafast bootstrap (Hoang et al., 2018) depending on the analysis (Table S6).

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

2.4.6.3. Bayesian inference Trees: Mixture models. Only the data set 2,

Table 2
Data sets used in the study and analysis performed with everyone of them. Samples included in data sets in bold font were used for ortholog search. See Supplementary Table S5 for further information on data set composition.

Datasets		Number of samples	Number of SC	Number of Positions		Analyses					
Name	Composition			Protein	Nucleotide	ML Exp. Prot/ Nuc	ML-Mixt Prot/ Nuc	BI-Mixt Prot/ Nuc	ASTRAL Prot/Nuc	DensiTree Prot/Nuc	ARC Nuc
Data set 1	All samples retained after filtering	82	717	274,960	839,555	X	X				
Data set 2	Reduced Dataset 1 to infer tree 1	29	717	276,745	841,425	X	X	X	X	X	
Data set 3	Reduced Dataset 1 to infer tree 2	12	717	–	831,918						X
Data set 4	Subtentaculata group	23	4175	1,764,320	5,376,224	X	X				
Data set 5	Etrusca-liguriensis group	36	1984	742,966	2,283,579	X	X		X		
Data set 6	Dataset 5 without samples from Berga	31	1984	744,252	2,255,464	X	X		X		

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

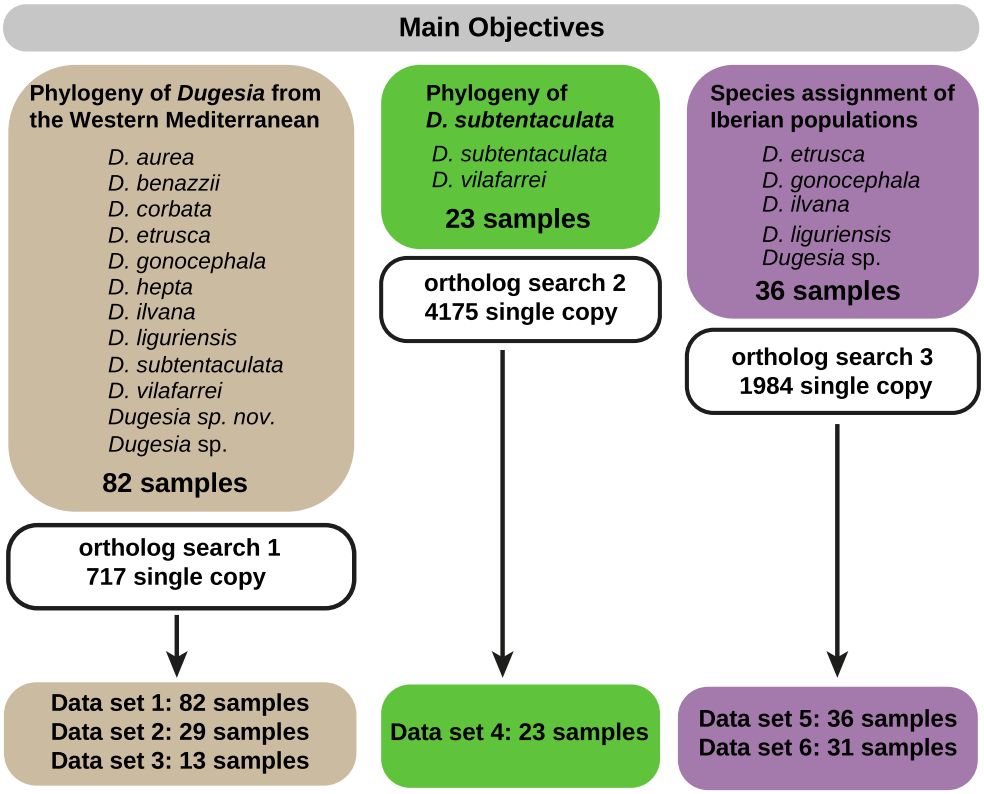


Fig. 3. Scheme summarizing the composition of the data sets. Considering the three main objectives of the present work were created three sample groups to perform the orthologs searches. The final data sets from each search are constituted by all the data or pruned sample compositions. For more details see Table 2, Table S4, and Table S5.

used to infer the phylogeny of all *Dugesia* species included in the study, was analysed with BI methods (Table 2); using 20 categories and the LG model for protein sequences, as well as the CAT GTR for nucleotide data. Two chains were launched with protein data, running until 10,368 and 20,807 iterations respectively and applying a burnin of 20 %. The effective sampling size (ESS) was over 1,000 and the discrepancy values were below 0.1 for all parameters. Additionally, the discrepancy observed across all bipartitions was equal to zero. With nucleotide data we launched two chains that ran until 18,926 and 18,771 iterations. After applying a burnin of 10 % taking into account the visualisation of tracer files in the Tracer program (Rambaut & Drummond, 2007), some values were slightly low for a few parameters (ESS between 100 and 300, discrepancy between 0.1 and 0.2) (Table S6). Obtaining the optimal values of ESS and discrepancy is very difficult when large data sets are analysed. If we take into account that the largest and mean discrepancy observed across all bipartitions are zero (maxdiff = 0, meandiff = 0) we can consider this run pretty acceptable (Lanfear et al., 2016; Schrempf et al., 2020).

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

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

2.4.7. *Individual gene trees visualisation*

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

2.4.8. *Ancestral character reconstruction (ACR)*

We inferred the probability of ancestral character states for the reproduction mode on the internal nodes of the ML tree obtained from the data set 3. In this case 717 SC and only 12 terminals were used, every-one representing one species and selecting only a sample of *D. malickyi* as outgroup (Table 2, Table S5). We set three states for the reproduction mode: Sexual (species with strictly sexual populations), Sexual + Asexual (species with sexual and asexual populations), and Asexual (species with strictly asexual populations). The current states were assigned to the terminals taking into account the reproductive strategies present in the whole species. Thereby, strictly asexual species have not been included in the analysis, since species with only asexual populations described are very rare and not present in the Mediterranean region. We estimated the ancestral states using the phytools package v.0.7.9 in R. The posterior probability for each state at nodes was determined from stochastic character-state mapping analysis, using the fitpolyMk function with the transient option as model, integrated in the make.simmap function and 10,000 simulations on Markov Chain Monte Carlo (MCMC).

3. Results

3.1. Sampling

We were able to obtain specimens from most of the species known from the Western Mediterranean clade (Fig. 1, Table1, Table S1). *D. brigantii* De Vries & Benazzi, 1983 and *D. leporii* Pala, Stocchino, Corso & Casu, 2000 have not been found since their initial description. In the case of *D. tubqalis* Harrath & Sluys, 2012, from Morocco, its type locality was in a very bad condition and no animals were found. Inspection of close localities did not reveal new locations for this species. In addition, due to weather conditions it was not possible to sample in Afaska, Morocco, the locality of the candidate new species *Dugesia* sp. (1) reported in Leria et al., (2020; 2022). However, in a close locality, Beni M'Hamed, *Dugesia* specimens were found. The DNA-Barcode analysis revealed that the new specimens show a great genetic distance with *Dugesia* sp (1) *sensu* Leria et al (2020; 2022), and probably represent a sister species, that here we name as *Dugesia* sp. nov., pending a thorough species delimitation study and the description of the new species.

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

3.2. Ortholog searches and data sets

We obtained transcriptome data from 83 specimens (Tables S1 and S3) representatives of 13 species (including the two outgroups from Greece) and undescribed new candidate species. After the transcript filtering step, 82 samples were retained, and three sample groups were analysed with OrthoFinder. The results of the three orthologs' searches performed are shown in Table S4. A total of 717, 4175 and 1984 Single Copy orthologs (SC) were obtained with all samples (search 1), subtentaculata samples group (search 2), and etrusca-liguriensis samples group (search 3) respectively.

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

3.3. Phylogeny of Dugesia in the Western Mediterranean

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

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

ML, BI and MSC analyses of data set 2 yielded the same topology for nucleotide and protein data. However, regarding the information content, some support values were lower when using protein data in ML and MSC methods, but not for BI, for which the support values were maximum (pp = 1) for nucleotide and protein data (Fig. 4, Fig. S2, Appendix A).

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

The results from the three phylogenetic inference methods based on data set 2 have been summarized on the ML tree (Fig. 4). Three main clades have been differentiated. The first divergent clade including *D. gonocephala*, *D. ilvana*, *D. etrusca*, *D. liguriensis,* and Iberian populations from Catalonia region is mainly continental, except for *D. ilvana*, which is endemic from Elba Island in the Tuscan archipelago. Taking into account the unknown assignment of Iberian populations to *D. etrusca* or *D. liguriensis*, we decided to denote these branches as *D. etrusca* sensu lato and *D. liguriensis* sensu lato (s.l.) when they include fissiparous individuals, and use sensu stricto (s.s.) when only sexual populations are referred. We have named this lineage the Iberia-Apennines-Alps clade since it mostly includes populations from these geographic regions as well as *D. ilvana* from Elba island and *D. gonocephala,* distributed also in almost all continental Europe. Next, two sister clades are defined, one endemic from the Corsica and Sardinia islands, Corsica-Sardinia clade, groups *D. hepta* Pala, Casu & Vacca, 1981 and *D. benazzi* Lepori, 1951 species. The second clade, Iberia-Africa clade, includes a complex species group formed by: a) *D. aurea* and *D. corbata*, endemic from the Balearic Islands, b) *D. subtentaculata* broadly distributed in all the Iberian Peninsula and the North of Africa, c) *D. vilafarrei* restricted to one locality in the South of the Iberian Peninsula, and c) a new lineage from the Rift, in Africa (Fig. 4).

All methods showed high support values for all nodes, except for the node grouping *D. subtentaculata* and *D. vilafarrei*, where the branch support of MSC is relatively low (0.88), but the final normalised quartet score (a measure of support for the entire topology) is high (Fig. 4, Appendix A).

3.4. Phylogeny of D. subtentaculata in the Iberian Peninsula

The ML tree obtained from the analysis of data set 4 (Table 2) shows a clear structure in this species (Fig. 5). The population of Cangas, in the North of the Iberian Peninsula is the first to diverge. Although this population is represented by sexual and asexual individuals, no differentiation by reproductive strategy was shown. The next group to diverge includes samples from the North of Catalonia (Santa Fe) and South of France (Montpellier) to constitute the Northeast clade, followed by a Southwestern clade, including a population from South of Portugal (Monchique), that is the sister group of the Southern clade formed by

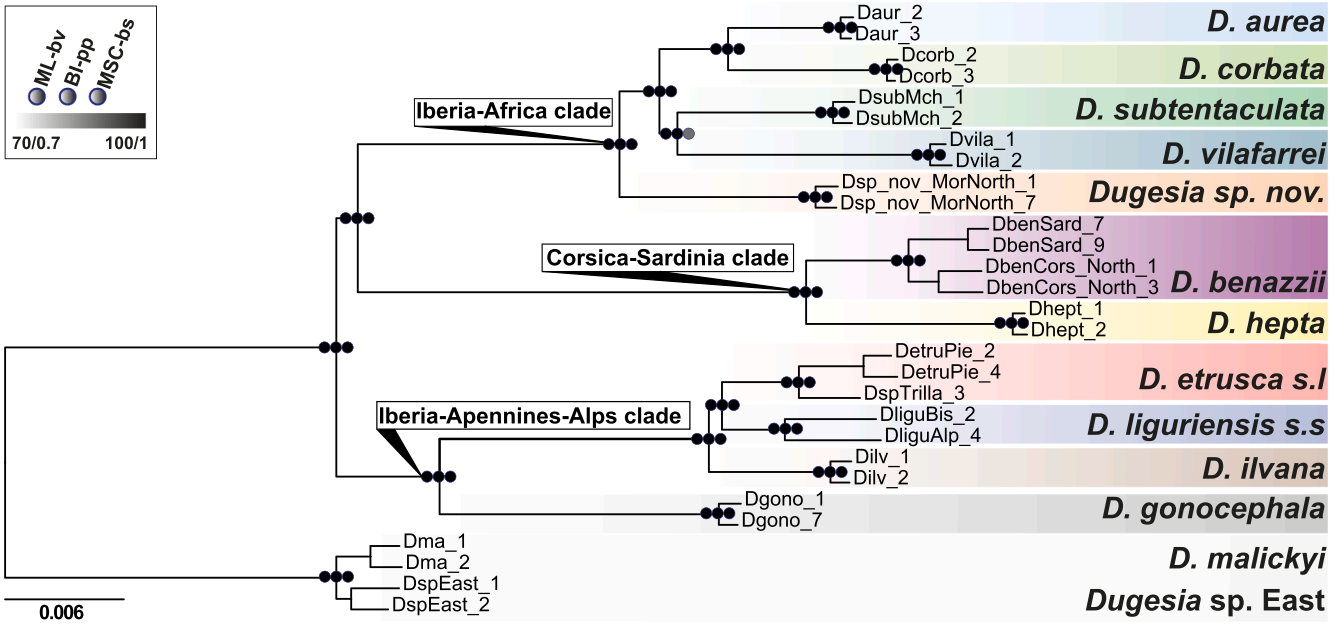


Fig. 4. Phylogenetic tree of *Dugesia* species from the Western Mediterranean obtained with transcriptomic data. The tree summarizes the results obtained with Maximum Likelihood (ML, IQ-TREE), Bayesian Inference (BI, PhyloBayes), and Multispecies Coalescence Model (MSC, ASTRAL-pro) analysing the nucleotide information of data set 2 (13 species, 29 samples, and 717 single copy orthologs). All approaches yield the same topology, shown here with the ML tree. Circles at nodes show the bootstrap value (ML-bv), the posterior probability (BI-pp), and the branch support (MSC-bs) in a grey scale as indicated in the legend. Scale bar: substitutions per site, s.s.: sensu stricto (the group does not include asexual representatives), s.l.: sensu lato (the group includes asexual representatives).

samples coming from the South of Spain (El Bosque) and two differentiated populations from Africa (Magoo in the North and Imlil in the South of Morocco).

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

The ML tree obtained from data set 5 (Table 2) shows an unexpected topology (Fig. 6A). *D. etrusca* s.s. is monophyletic (red in Fig. 6). However, *D. liguriensis* s.s. is divided in two clades and results in a paraphyletic group, since samples from Sasello and Bisagno group with *D. etrusca* s.l.. The unclassified samples from Font de l'Ús, Font de La Trilla, and Berga occupy an intermediate position within this group, showing a very strange branching pattern, where several individuals constitute each an independent lineage resulting in a ladder-like pattern (Fig. 6), different to the monophyletic grouping expected for a population that has evolved independently under panmixia. In addition, the bootstrap supports are low for many nodes (bs $<$ 97, grey circles in Fig. 6A).

The species trees obtained with Astral-pro using the MSC method show a different topology (Fig. 7A). In this case, *D. liguriensis* s.s. and *D. etrusca* s.s. form both monophyletic groups highly supported. The samples from Font de l'Ús group with high support with *D. etrusca* s.s. Although some internal nodes have low support, the samples from Font the la Trilla and two samples from Berga also group with high support in this clade, showing the ladder-like pattern. The rest of samples from Berga group with *D. liguriensis* s.s., also with a high support.

The samples from Font de l'ÚS and Font de La Trilla grouped with *D. etrusca* s.s. in all analyses (Fig. S1. Fig. 6, and Fig. 7), while the samples from Berga in some cases split into the two species (Fig. S1, Fig. 7A). Thus, we considered the samples from Berga were problematic and may be causing some artifactual groupings. To explore their effect in the phylogenetic inference, we extracted the representatives from Berga to obtain the data set 6 (Table 2, Fig. 3). From this data set, resolved trees were obtained, showing two monophyletic groups (Fig. 6B, Fig. 7B). Nonetheless, although the samples from Font de La Trilla grouped in *D. etrusca* s.l., they continued showing the ladder-like pattern.

3.6. Ancestral character reconstruction

We assigned a sexual + asexual reproductive strategy to *D. etrusca* s.l. and *D. liguriensis* s.l., because in some of the analyses both lineages include fissiparous individuals (Fig. S1, Fig. 7A). This assumption, nonetheless, must be taken with caution, since the evolutionary history of the Iberia-Apennines-Alps clade is complex and future analyses may show a different situation.

The ancestral character reconstructed for the reproductive mode of ancestors along the evolutionary tree of the Western Mediterranean *Dugesia* and their probability are shown in Fig. 8 and Fig. S3. The hypothesis of a strictly sexual ancestor was strongly supported in almost all nodes (pp $>$ 0.9), except for the node joining *D. etrusca* s.l. and *D. liguriensis* s.l., where the ancestor shows a high probability to be sexual and/or asexual (pp = 0.76) suggesting, in this case, the possibility of a facultative ancestor for this lineage.

4. Discussion

4.1. Phylotranscriptomics applied to the phylogenetic study of Dugesia

Two previous studies based on molecular data have tried to ascertain the phylogenetic relationships of *Dugesia* from the Western Mediterranean. The first (Lázaro et al., 2009) was based on only two markers while the second increased the markers to six (Leria et al., 2022) resulting in a substantial change in topology and resolution, validating not only the importance of the number of markers but also their nature. Lázaro and collaborators used fragments of the nuclear gene *Internal Transcribed Spacer* (ITS-1) and the mitochondrial region *Cytochrome Oxidase I* (COI). Although these markers have been widely used in taxonomic and phylogenetic studies in diverse taxa (Chen et al., 2015; DeSalle & Goldstein, 2019; Phillips et al., 2018; Vu et al., 2019) and particularly in planarians (for review articles see Álvarez-Presas & Riutort, 2014; Baguñà et al., 1999; Riutort et al., 2012), for Western Mediterranean *Dugesia* the use of 1,039 positions in a concatenated alignment did not yield enough information to obtain a completely resolved phylogeny. The addition of four markers (Leria et al., 2022)

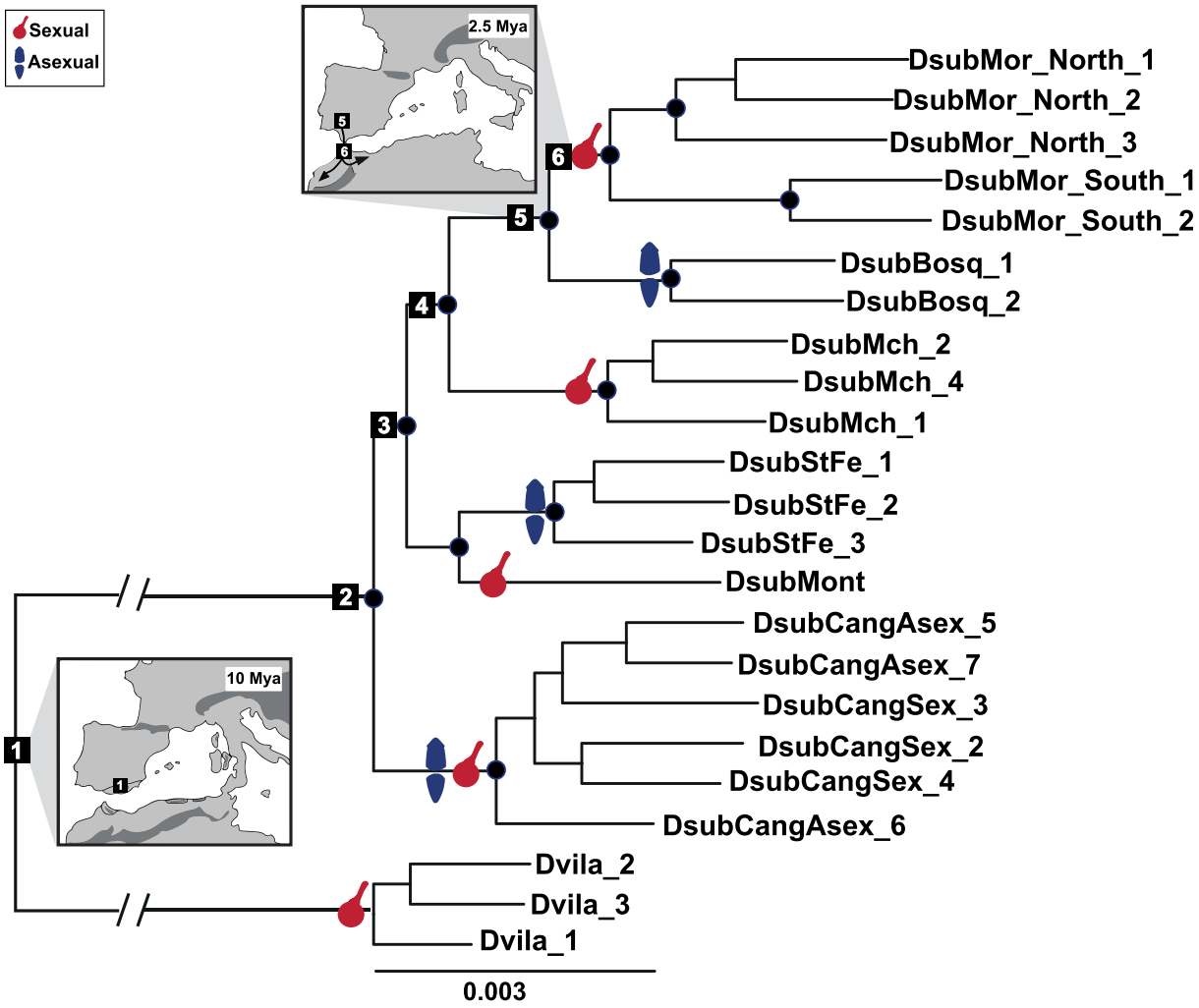


Fig. 5. Diversification of *D. subtentaculata* in the Iberian Peninsula and Africa. ML tree of *D. subtentaculata* obtained with IQ-TREE from Data set 4 (subtentaculata samples group: 2 species, 23 samples, 4175 SC). The black circles on the nodes indicate bs = 100 and the icons on the branches the reproductive strategy of populations. Scale bar: substitutions per site. The schematic maps represent the biogeographical events that gave rise to the split of *D. subtentacula* from its sister species (node 1), and the pass of the species to Africa with its diversification in the Southern and Northern regions (nodes 5 and 6).

increased the compared positions to 5,439 combining ribosomal and protein-coding genes from the nuclear and mitochondrial regions under different selective pressure. This broader data yielded a new topology that nonetheless still showed a few unsupported nodes, indicating that the molecular evolutionary rates of the regions sequenced were still insufficient to reflect the diversification process of the group in specific cases.

In the present work, we moved to a strategy based on phylotranscriptomics and developed the pipeline of programs and scripts needed to perform all the analyses for the first time in freshwater planarians. With this new strategy, we obtained a strongly supported topology for all nodes using 717 single copy orthologs (>800,000 bp, Table S6) identified from coding regions. The transcriptomic strategy has the advantage of increasing to hundreds of thousands the positions analysed, while the price and the time needed to obtain them are equal or even lower than the PCR amplification of only a few markers. On the other hand, the bioinformatic processing of the data increases the time spent in the analyses, especially the initial quality controls, cleaning, and the search of orthologs. But with the advantage that once the pipeline has been established, it can be repeated including any new data. The development of these large data sets has moreover stimulated the improvement of the ways to implement the evolutionary models needed in the probabilistic based methods (Holder & Lewis, 2003) which may also result in the improvement observed in our case (Table S6) with respect to previous studies. From the application of the same model for the entire alignment to the use of different models for partitions by gene or codon, diverse strategies have been developed. Since the upraise of genomic data sets, the use of Mixture Models has expanded in the field of phylogenomic analyses, replacing the traditional partition schemes. In this new strategy, the phylogenetic inference algorithm evaluates for each site the more adequate model to be applied and groups similar sites in categories (Quang et al., 2008) making it unnecessary to give beforehand any partition scheme. Overall, the good performance of this method when applied to phylogenetic inference has been demonstrated (Schrempf et al., 2020).

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

Regarding the use of different data types, we also have been able to

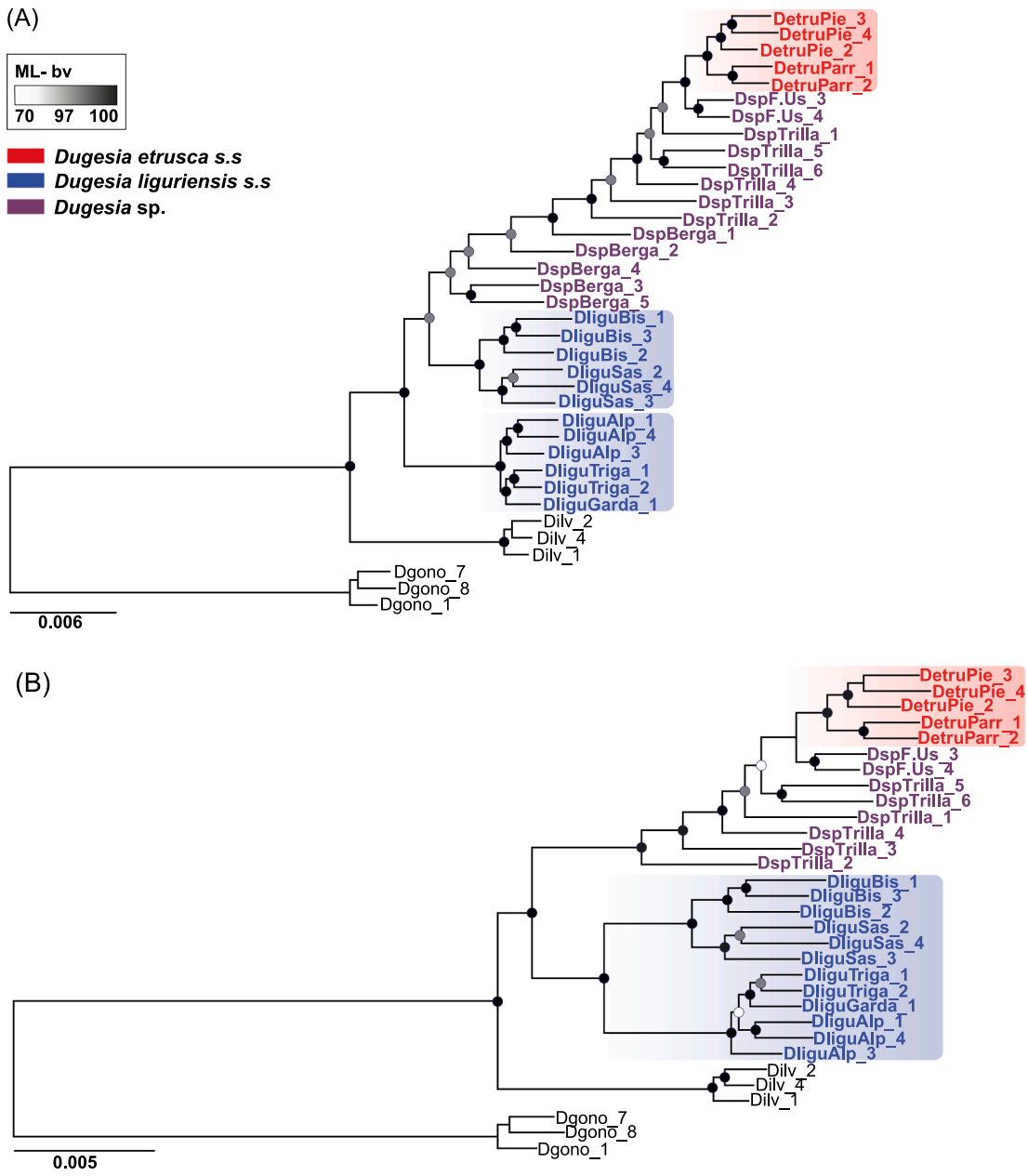


Fig. 6. ML trees of Iberian-Apennines-Alps clade obtained with IQ-TREE using nucleotide data from (A): Data set 5 (etrusca-liguriensis samples group: 4 species, 36 samples, 1984 SC) and (B): Data set 6 (Data set 5 without Berga representatives: 4 species, 31 samples, 1984 SC). Circles at nodes show bootstrap values (ML-bv) in a grey scale as indicated in the legend. Scale bar: substitutions per site. The representatives of *D. etrusca* s.s, *D. liguriensis* s.s and asexual populations from the Iberian Peninsula are highlighted in different font colours; red, blue, and purple respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

analyse aminoacidic data and compare its performance against nucleic data to answer our questions. The DensiTree graph showed a most blurred pattern in protein trees indicating higher gene tree discordance in protein data (Appendix A bottom), which explains that nucleotide sequences rendered more maximum supported nodes with all methods. These results indicated that the amino acidic sequences do not contain enough information to resolve the phylogeny. The protein data is more conserved than nucleotide sequences and its information is more useful to study the evolution on a broad scale of time (Nei & Kumar, 2000), while the diversification of *Dugesia* in the Western Mediterranean is relatively recent (Sola et al., 2022). For these reasons, we based our main results and discussion on the analyses of nucleotide data.

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

4.2. Evolutionary history of Dugesia in the Western Mediterranean

In the first and very preliminar molecular approximation to study the evolutionary history of *Dugesia,* Lázaro et al. (2009) included a few individuals of several species belonging to the Western Mediterranean area. Those included four populations of *D. subtentaculata* sensu stricto from Iberian Peninsula, as well as the populations currently considered different species (*D. aurea* and *D. corbata*), also *D. gonocephala*,

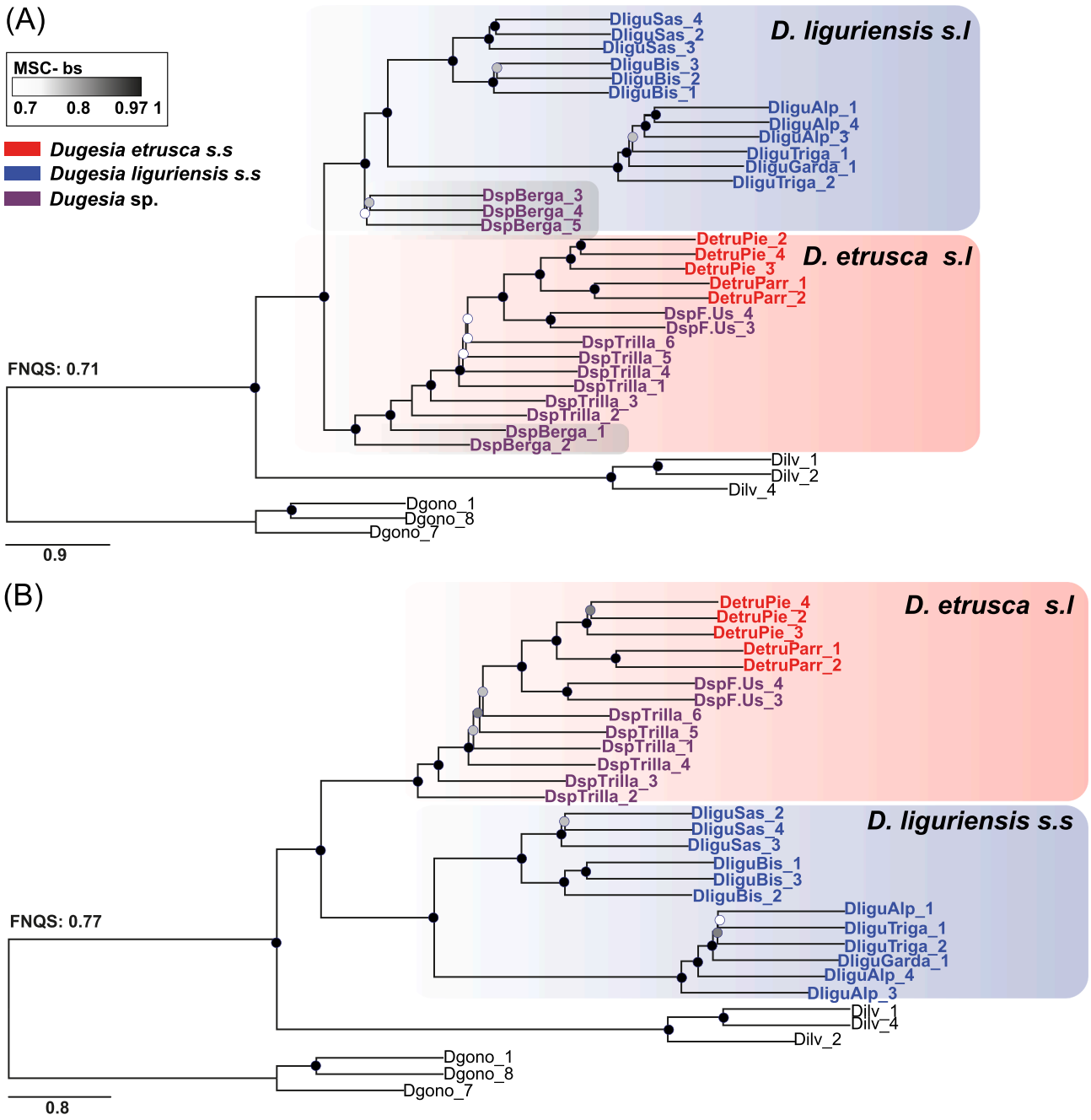


Fig. 7. MSC tree of Iberian-Apennines-Alps clade obtained with ASTRAL-pro using nucleotide data and analysing (A): Data set 5 (etrusca-liguriensis samples group: 4 species, 36 samples, 1984 SC) and (B): Data set 6 (Data set 5 without Berga representatives: 4 species, 31 samples, 1984 SC). Circles at nodes show branch supports (MSC-bs) in a grey scale as indicated in the legend. Scale bar: Coalescent Units. FNQS: Final normalized quartet score. The representatives of *D. etrusca* s.s, D. *liguriensis* s.s and asexual populations from Iberian Peninsula are highlighted in different font colours; red, blue, and purple respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

D. etrusca, D. liguriensis, D. ilvana, D. hepta, and *D. benazzii*. The results showed for the first time the differentiation in the three main clades that are clearly defined in our phylogeny but did not resolve the relationships among them nor even the relationships within some of them (for some of which the sampling was very poor).

In the study by Leria et al. (2022) the sampling was broadened to include representatives of *Dugesia* from Morocco and a new species from the Iberian Peninsula (*D. vilafarrei*). However, in that work only one or two representatives per species were included, not allowing the analysis of intraspecific genetic structure and evolutionary history. Their results showed the three main clades of the previous study (Lázaro et al., 2009), but with the same topology that we obtained here (Fig. 4), although with still low support for the node joining the Iberia-Africa clade with the Corsica-Sardinia clade. In the present work, the support values for all the nodes have been maximum giving full support to that topology (Fig. 8).

Leria et al., (2022) also dated their phylogeny and performed ancestral range reconstruction and niche modelling analyses that allowed them to put forward a biogeographical hypothesis. They showed that the evolution of *Dugesia* has been shaped by the paleo-geological processes that formed the Western Mediterranean as well as climatic changes. The good support we give to their topology allows us to put our tree in the time frame they proposed. We have summarized Leria et al., (2022) hypothesis in Fig. 8B-G over a scheme of our phylogenetic tree (Fig. 8A). The scheme shows the divergence times

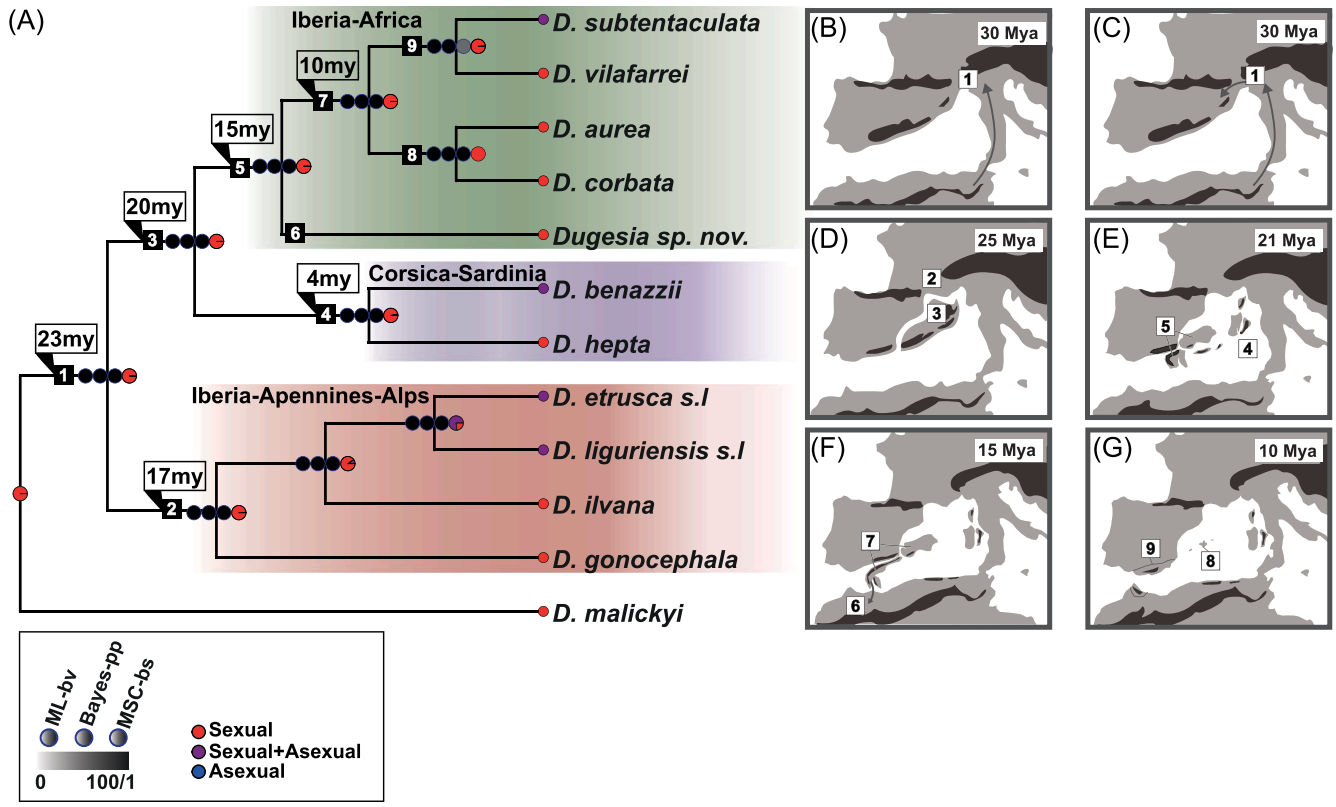


Fig. 8. Scheme summarizing the evolutionary history of *Dugesia* from the Western Mediterranean. (A): a forced ultrametric tree obtained from data set 3 (12 species, 12 samples, and 717 SC) highlighting the three main clades. The divergence times shown on the nodes are extracted from Leria et al., (2022) based on a calibrated phylogeny obtained using 6 molecular markers and different taxon composition. B-F: schemes showing the geological events that shaped the diversification of *Dugesia* in the Western Mediterranean modified from Leria et al., (2022). Circles at nodes show support values on a grey scale and, in color, the ancestral character reconstruction as indicated in the legend. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

obtained by Leria et al., (2022) using a different data set to infer the calibrated phylogeny in BEAST. Therefore, the branch length of our scheme has small discrepancies to their divergence times, but it shows the coincident topology found in both studies, supporting their biogeographical hypothesis. The hypothesis locates the ancestor of the Western clade arriving in Europe through the Italian Peninsula 30 Mya (Fig. 8B), matching the results of a recent biogeographic study of *Dugesia* genus (Sola et al., 2022). This ancestor, could start dispersing throughout the continental region, passing also to the Iberian Peninsula (Fig. 8C).

The first diversification event of *Dugesia* from the Western Mediterranean clade occurred around 23 Mya, coinciding with the breakage of Eastern Iberia and Southern France from the continent (Rosenbaum et al., 2002). This diversification event putatively isolated the ancestor of the Iberia-Apennines-Alps clade, which remained in the continent, from the ancestor of the Corsica-Sardinia and Iberia-Africa clades, which remained in the landmass that would become the Corsica and Sardinia islands, the Balearic Islands, the Betic region and part of the North of Africa (Fig. 8D).

4.3. The Corsica-Sardinia clade

Important facts about the evolution of this clade in the Corsica and Sardinia islands arise from our results. In the first place, as stated above, it is the first time that the sister relationship of this clade with the Iberia-Africa clade receives maximum support in phylogenetic analysis. Three endemic species have been described from the archipelago formed by Corsica and Sardinia; *D. benazzii, D. hepta,* and *D. leporii.* However, the latter has not been found since their initial description, so never been included in a molecular study. For *D. benazzii* and *D. hepta*, a recent study based on two molecular markers and including multiple populations of both species (Dols-Serrate et al., 2020) along their distribution has shown a very complex evolutionary history characterized by a putative process of hybridization, the combination of different reproductive strategies, and chromosome rearrangements. In that work it was put forward the hypothesis that *D. benazzi* could be separated into two species; one from each island (Dols-Serrate et al., 2020). Our work shows *D. benazzii* representatives from the islands of Sardinia and Corsica forming two monophyletic well-differentiated groups (Fig. 4 and Fig. S1). Our data hence renders further support to *D. benazzii* being in fact two species, one distributed in Corsica and the other one in Sardinia. Moreover, we show a clear differentiation between *D. benazzii* from the South and the North of Corsica, indicating a geographic genetic structure of this lineage in that island.

4.4. The D. subtentaculata journey: From the Betic-Riff plate to Iberia and back to Africa

According to the biogeographic hypothesis of Leria et al., (2022), the ancestor of the Iberian Peninsula and Balearic Island representatives belonging to the Iberia-Africa clade (Fig. 8, node 7) originated in the Betic-Riff plate. Breakage of this plate separated this ancestor from its sister group in the North of Africa (Fig. 8F), here represented by a non-described species. Later, after the Balearic Islands split, the ancestor of *D. vilafarrei* and *D. subtentaculata* remained in the Iberian Peninsula. Here we corroborate this biogeographic history and moreover obtain for the first time a resolved phylogeny of *D. subtentaculata* populations (Fig. 5).

Even though not all reported populations have been included, our phylogeny is enough to propose a first hypothesis on the colonization

process of Iberian Peninsula by *D. subtentaculata*. Based on our tree topology (Fig. 5), we hypothesize that the ancestor of this species dispersed from the ancient Betic plate to the North of the Iberian Peninsula. The population from Montpellier (Southeastern France) may be the product of one anthropogenic introduction, since by that time, the Pyrenees were fully formed (Dèzes et al., 2004, 2005) or could represent a natural dispersion through the well-recognised corridor between the Pyrenees and the Mediterranean Sea (Martinez Rica & Montserrat Recoder, 1990). The diversification of the species continued to the south of the Iberian Peninsula with the split of the Southwestern populations in Portugal, and the population from El Bosque in Andalusia. A similar biogeographic pattern showing an earlier diversification of the lineages from the Eastern basins of the Iberian Peninsula than the lineages from the Western basins has been found for the native freshwater fishes of this region (Filipe et al., 2009), which may be related to the hydrographical evolution of the Iberian Peninsula (De Vicente et al., 2011). The last dispersion of *D. subtentaculata* concerned its crossing to Africa. Leria et al., (2022) dated the divergence between one Iberian and one African population around 1.6 Mya, indicating that the pass putatively occurred during the Pleistocene, when the sea level in the Gibraltar strait was lower.

The lack of phylogenetic resolution reported in Leria et al., (2020) for this species was a consequence of the Mosaic-Meselson effect generating high intraindividual diversity due to its reproductive strategy (fissiparity combined with occasional events of sexual reproduction) affecting both genes analysed, COI and Dunuc12 (Leria et al., 2019). In the present study, the use of thousands of exonic regions from single-copy orthologs has resulted in a resolved and supported phylogeny within *D. subtentaculata*. The advantage of our methodology resides, in the first place, in the use of coding regions under different selective pressure that possibly restricts the emergence of mosaicism. Thus, many conserved sites probably presenting fixed substitutions among populations can counter the effect of intraindividual diversity in a few sites and result in a phylogenetically informative set of data. In this aspect, it is important to notice that most of the intraindividual variability found in the Dunuc12 marker by Leria et al., (2019) was situated in the intronic region. In addition, the methodology used in the present study including new strategies for the application of evolutionary models, possibly allowed us to better retrieve most of this information.

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

In relation to the evolutionary history of *D. subtentaculata* in the Northwest of Africa, we report two differentiated populations from the North and the South of the Atlas in Morocco. The presence of this species in Africa has been previously reported (Harrath et al., 2012; Stocchino et al., 2012), but no phylogenetic analyses had included representatives of *D. subtentaculata* from this geographical area before. Moreover, it is remarkable that the new species from North Morocco analysed here (Dsp_nov_Mor North in Fig. 4) is different from the candidate species included in previous analyses (Leria et al., 2020, 2022), indicating that at least three species belonging to the Western Mediterranean clade are present in the Atlas area (*D. tubqalis*, and the two new candidate species), apart from the representatives of *D. subtentaculata* mentioned above. This suggests there is a high species richness hidden in the Atlas and the Rif region.

4.5. The Iberia-Apennines-Alps clade: Geographically broader than thought

D. etrusca s.s and *D. liguriensis* s.s. were described from individuals coming from Tuscany (Benazzi, 1946) and Liguria (De Vries, 1988) respectively. In a first attempt to infer a molecular phylogeny for Tricladida (Baguñà et al., 1999) a single individual of *Dugesia* coming from Northeastern Spain surprisingly grouped with representatives of *D. etrusca* (*D. liguriensis* was not included in that study). Later, in Lázaro et al. (2009) a single individual coming from Sardinia and the previously cited locality from Northeastern Spain were ascribed to *D. liguriensis*. Finally, the presence of putative *D. etrusca* and *D. liguriensis* from Aragón and Catalonia in the Northeastern section of the Iberian Peninsula was detected by DNA-Barcoding in posterior samplings of Iberian regions (Riutort *pers. com.*). These findings raised the question of which of both species the Iberian Peninsula populations belong to and how biogeographically we could explain their existence.

In the present work, we include for the first-time multiple representatives of several Iberian Peninsula localities harbouring specimens belonging to this clade as well as *D. ilvana,* endemic from Elba Island, not analysed in Leria et al. (2022). With this broad representation of the species and geographic diversity within the group, we expected to set a supported hypothesis on the relationships, species assignations and biogeographical history for this group. Nonetheless, the asexuality of the Iberian representatives have hampered a part of our aims, opening new and interesting questions on the evolution of this group of *Dugesia* species and of asexual lineages in general, as we develop in the following.

Our phylogenetic tree shows *D. ilvana* as the first split within the clade. The geological history of Elba island (Bortolotti et al., 2001a, 2001b) sets a geological maximum for the divergence of *D. ilvana* of approximately between 8 and 5 Mya. However, Leria et al (2022) set the split of *D. etrusca* s.s. + *D. liguriensis* s.s. clade around 9 Mya. Therefore, the geological history of Elba Island might seem slightly too young to have preceded the *D. liguriensis-D. etrusca* split. Since the evolutionary history of Elba Island fauna has been linked to connections with Corsica regions and the Tuscany coast (Fattorini, 2009 and references therein; Dapporto & Cini, 2007; Di Nicola & Vaccaro, 2020) it is difficult to drag a biogeographic scenario for the split of *D. ilvana* from the ancestor of *D. etrusca* s.l. and *D. liguriensis* s.l. Therefore, the phylogenetic position of *D. ilvana* shown here must be reviewed under more detailed analyses, taking into account the complex diversification process of Iberia-Apennines-Alps clade.

According to the topology (Figs. 4 and 8) the clade including *D. etrusca* s.l. and *D. liguriensis* s.l. would have diversified occupying the area from the Apennines region to the Northeastern in the Iberian Peninsula. It has been hypothesized that the Alps prevented the expansion of the Apennines-Alps lineage to Central Europe, while the Pyrenees possibly limited its entry to the Iberian Peninsula (Leria et al., 2022). However, the new populations found in Catalonia contradict the latter hypothesis. We propose that this pass was possible because the Eastern Pyrenees orogenesis encompassed two stages; a first stage from Early Cretaceous to middle Lutetian time (99 to 47 Mya) and the second one from middle Lutetian to late Oligocene (47 to 23 Mya). The first stage was characterized by a low topography, because the increase in relief was partially compensated by downward flexion of the Iberian plate, as well as high levels of mountain erosion. While the last stage, with more orogenic activity, was not concluded until approximately 23 Mya (Vergés et al., 1995). This complex geological history led to the occurrence of several fauna corridors along the Pyrenees landscape (Ninot et al., 2017). Considering these orogenic processes, their geological dates, and the broad uncertainty around divergence times estimated by Leria et al (2022), the Eastern Pyrenees could have acted as the corridor for the ancestors of the populations found in the North of Iberian Peninsula (Fig. 8C). Why these species did not expand further in the Iberian Peninsula and remained restricted to the localities described here, is a question we cannot resolve with the present data.

In summary, we detect some incongruencies and a certain difficulty to explain the evolutionary history of the group and its biogeography. When we analysed in more detail the relationships within and among the *D. etrusca* s.l. and the *D. liguriensi* s.l. clades (Fig. 6, 7, and S1) to try to have a better understanding of the relationships among the Iberian populations and the rest, and to assign them to one or the other species, the scenario became much more complex, probably due to the

evolutionary consequences of asexuality on the distribution pattern of genetic diversity in populations and on its effect on phylogenetic inference, as we discuss in the following.

4.6. The Iberia-Apennines-Alps clade: A complex evolutionary history driven by asexuality?

In the species identification analysis with COI, individuals from Font de la Trilla and Berga already showed an anomalous ladder-like pattern like the one we later found in the transcriptome-based trees (Figs. 6, 7 and S1). For this reason, a broader number of samples from these populations were included in the transcriptomic analyses. Despite the strange branching pattern, all our trees show that the Iberian fissiparous populations belong to the Iberia-Apennines-Alps clade. However, regarding their species assignment to *D. etrusca* or *D. liguriensis* species, our results are more than disquieting, exciting. Based on our phylogenetic trees, populations from Font de l'Ús and Font de la Trilla can be assigned to *D. etrusca*. But, from these trees three questions arise: i) the species assignment of individuals from Berga, ii) what is causing the strange ladder pattern of individuals from Font de la Trilla and Berga, and iii) whether this anomalous branching pattern could affect the phylogenetic inference.

Two putative explanations for these issues could be related to a case of hybridization between both species or, alternatively, to the long term asexuality of the Iberian populations. We discard the hybrid origin hypothesis for the Iberian specimens since no hybrid populations have been found in the Apennine-Alps region, where the two sexual lineages inhabit. In fact, there is no knowledge of both species dwelling in the same water course or even the same area anywhere. There is neither knowledge of sexual populations of the species in the Iberian Peninsula, a region that has been thoroughly sampled (Leria et al., 2020), nonetheless the presence of both species in a still unknown locality, cannot be completely discarded. In any case, the hybridization between the two species would have had to take place a long time ago, so that the hybrid lineage could pass through the Pyrenees, with the subsequent extinction of its populations in France and Italy, leaving only the Iberian hybrid populations.

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

The ACR analysis in the present study results in a high probability of the existence of ancestral populations with both, sexual and asexual reproduction for the node joining *D. etrusca* s.l. and *D.liguriensis* s.l. (Fig. 8, Fig. S3), which support the idea of an ancestral fissiparous lineage arriving to the Iberian Peninsula.

The long-term asexuality of these ancestral populations could be a factor explaining the ladder topology showed by their descendant specimens in our trees (Figs. 1, 6, and 7). The pattern observed looks like that described for nuclear alleles in asexual populations under the Meselson effect, when alleles show high divergence by the accumulation of mutations independently of each other (Schwander et al., 2011). Under long-term fissiparous reproduction different clonal lines are established in the population, and present-day individuals can belong to independent lineages of fissiparity inside the populations, with their alleles being related far away in time. However, *D. subtentaculata* also presenting fissiparous reproduction, shows the typical monophyletic pattern expected for populations geographically separated (Fig. 5). This difference could be explained by a more recent diversification of *D. subtentacula* (Leria et al., 2022) and the alternation of fissiparity with sexual periods described in this species. Thus, we suspect that the Iberian populations of the Iberia-Apennines-Alps clade show haplotypes derived independently from the haplotypes present in their common ancestor with the ancestor of *D. etrusca* s.s. and of *D.liguriensis* s.s. before the divergence of the two species. When the two species diverged, the haplotypes present in the ancestor would have been lost differentially and those present within each species would have recombined and then accumulated new changes, resulting in the typical monophyletic groups with a basal stem. On the other hand, in the asexual Iberian lineages since no recombination events took place today haplotypes will have derived directly from the ancestral sequences, and hence in the present trees branch close to the base of the lineage. The grouping of Berga individuals at the base of both *D. etrusca* and *D.liguriensis* lineages, could indicate the presence in the Berga population of haplotypes derived from the haplotypes present in the ancestor of both sexual lineages, while the other Iberian populations would have only conserved haplotypes shared with the ancestor of *D. etrusca* s.s.

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

5. Final remarks

The study of the processes that shape biodiversity is key to understanding the evolution of life on our planet. This knowledge is useful not only to understand basic evolutionary processes but to apply this information to conserve the diversity and their functions in the ecosystems. The Mediterranean region constitutes a hotspot of diversity (Myers et al., 2000), and its palaeogeographical history is one of the most complex in the world (Allen, 2001). Large mountain systems, unstable climate periods, volcanic activity, and plate fragmentation events act as modelers of the biodiversity in this region. Here, we contribute to the study of one important component of the Mediterranean freshwater ecosystem, the free-living planarians. Focused on the Western Mediterranean region, we help to elucidate the evolutionary history of the *Dugesia* genus using transcriptomic data. Our work represents a step forward in the phylogenetic studies in this group, passing from the analysis of a few markers to hundreds of them, supporting previous information, and contributing with new valuable data to the knowledge on these species. We corroborated, with maximum support for the first time, a biogeographic hypothesis that explains the diversification of *Dugesia* in the Western Mediterranean; affected by the tectonic dynamics of the region during the Cenozoic. The new transcriptomic data has also allowed overcoming the problems generated by mosaicism to solve *D. subtentaculata* internal relationships and show their power to discover intraspecific genetic structuring and putative new species, as in the case of the species from the Corsica-Sardinia archipelago and those from Morocco. Finally, the new data, together with a wider sampling, has allowed the unveiling of a complex evolutionary history in the Iberian-Apennines-Alps clade, possibly driven in part by fissiparity.

Future work with transcriptome data, but if possible, also with good reference genomes for some of the species, would be key to understanding the evolution of the asexual populations of the different *Dugesia*

species, and the consequences of this type of reproduction in each species' genetic background. It is necessary to use species such as *D. subtentaculata, D. benazzii, D. etrusca* s.l., *D. liguriensis* s.l., and *D. sicula*, which integrate the asexuality in different evolutionary scenarios to understand the effect of asexuality in the natural process of resistance, resilience, and diversification of life. These studies will also allow analysing how the disparity in reproductive modes can affect the phylogenetic inference and the species delimitation methods, which are based on species concepts that become obsolete under asexuality. Species concepts and systematic biology frameworks must be reviewed, to take into account the asexuality as a recurrent process in the evolution of the tree of life, and to understand the effect of the reproduction modes in the diversification of the species. In this way, we will begin to contemplate asexual reproduction as a strategy, more than an evolutionary constraint.

Data accessibility

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

In Dryad under https://doi.org/10.5061/dryad.9kd51c5mw are deposited the intermediate results: 1) the protein sequences used in the OrthoFinder searches; 2) the protein and nucleotide sequences of the single copy (SC) orthologs obtained in the three OrthoFinder searches; 3) the 6 datasets used for the analyses.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

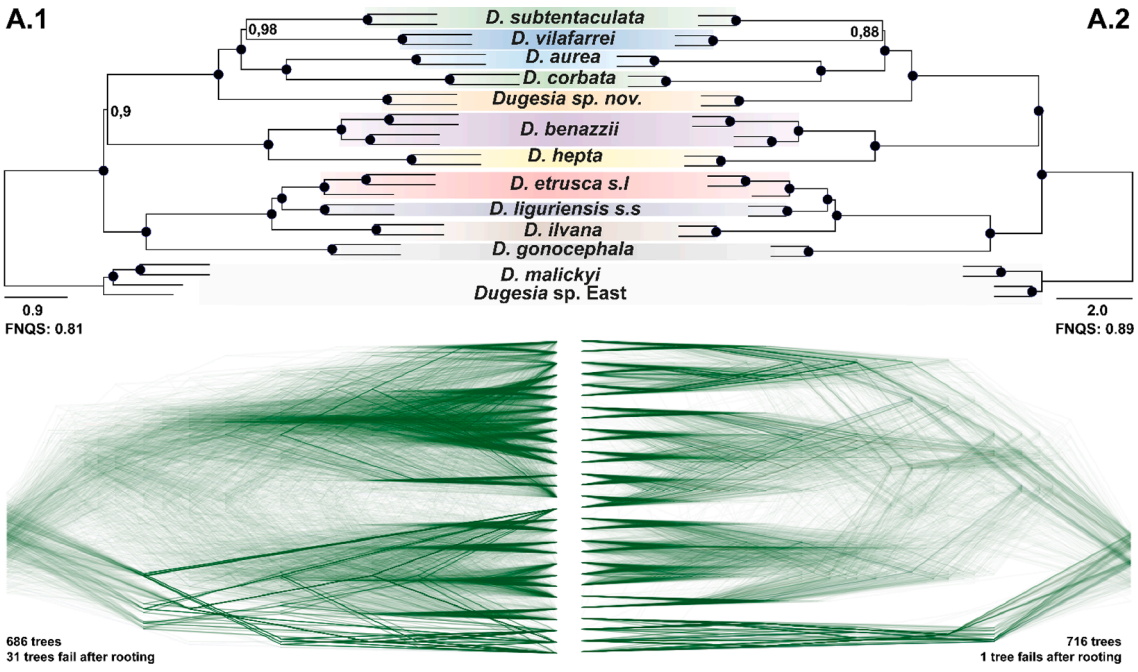
Acknowledgements

We are grateful to Julio Rozas for sharing with us the computational cluster of the Evolutionary Genomics & Bioinformatics research group. We also thank Silvia Hinojosa for guiding the first bioinformatics steps and Juan Manuel Lentijo Mondejar for his advice with the python scripts. In addition, we thank Ignacio Tenaguillo Arriola for the help launching some analyses. Also, thanks to Eduard Solà for the sampling campaign in Greece. Our thanks also to the Laboratoire Ecologie, Systématique, Conservation de la Biodiversité (LESCB), Unité de Recherche Labellisée -CNRST N° 18, Faculté des Sciences, Université Abdelmalek Essaadi and the Water, Biodiversity and Climate Change Laboratory, Cadi Ayyad University by the support with the sampling and transportation permissions in Morocco. Sampling Permission in Morocco: Decision N° 19/2019 HCEFLCD/DLCDPN/DPRN/CFF.

Funding

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

Appendix A



Appendix A. Species trees obtained with ASTRAL-pro analysing the data set 2 (13 species, 29 samples, and 717 SC), using A.1: protein data and A.2: nucleotide data. The black dots on the nodes represent branch support = 1. Scale bar: Coalescent Units. FNQS: Final normalized quartet score. Individual trees visualized in DensiTree are shown in green at the bottom.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107649.

References

Allen, H., 2001. In: Mediterranean Ecogeography. Routledge, p. 1st ed.. https://doi.org/10.4324/9781315838526.

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

Álvarez-Presas, M., Riutort, M., 2014. Planarian (Platyhelminthes, Tricladida) Diversity and Molecular Markers: A New View of an Old Group. Diversity 6, 323–338. https://doi.org/10.3390/d6020323.

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

Baguñà, J., Carranza, S., Pala, M., Ribera, C., Arnedo, M.A., Ribas, M., Riutort, M., 1999. From morphology and karyology to molecules. New methods for taxonomical identification of asexual populations of freshwater planarians. A tribute to Professor Mario Benazzi. Ital. J. Zool. 66, 207–214. https://doi.org/10.1080/11250009909356258.

Benazzi, M., 1946. Sopra una nova planaria d'acqua dolce. Archivo Zoologico Italiano, XXXI, pp. 93–102.

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

Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660.

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

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

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

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

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

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

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

Cheon, S., Zhang, J., Park, C., 2020. Is Phylotranscriptomics as Reliable as Phylogenomics? Mol. Biol. Evol. 37 (12), 3672–3683. https://doi.org/10.1093/molbev/msaa181.

Cunha, T.J., Giribet, G., 2019. A congruent topology for deep gastropod relationships. Proc. R. Soc. B: Biol. Sci. 286 (1898) https://doi.org/10.1098/rspb.2018.2776.

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

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

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

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

DeSalle, R., Goldstein, P., 2019. Review and Interpretation of Trends in DNA Barcoding. Front. Ecol. Evol. 7 (September), 1–11. https://doi.org/10.3389/fevo.2019.00302.

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

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

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

Dols-Serrate, D., Leria, L., Aguilar, J.P., Stocchino, G.A., Riutort, M., 2020. *Dugesia hepta* and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex? Org. Divers. Evol. 20 (3), 369–386. https://doi.org/10.1007/s13127-020-00438-z.

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

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20 (238) https://doi.org/10.1186/s13059-019-1832-y.

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

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

Fernández, R., Gabaldón, T., 2020. Gene gain and loss across the metazoan tree of life. Nat. Ecol. Evol. 4, 524–533. https://doi.org/10.1038/s41559-019-1069-x.

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

Fernández, R., Sharma, P.P., Tourinho, A.L., Giribet, G., 2017. The Opiliones tree of life: shedding light on harvestmen relationships through transcriptomics. Proc. R. Soc. B: Biol. Sci. 284 (1849) https://doi.org/10.1098/RSPB.2016.2340.

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

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

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

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

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

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

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

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

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

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

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

Holder, M., Lewis, P.O., 2003. Phylogeny estimation: traditional and Bayesian approaches. Nat. Rev. Genet. 4 (4), 275–284. https://doi.org/10.1038/NRG1044.

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

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

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14 (6), 587–589. https://doi.org/10.1038/nmeth.4285.

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

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

Kobayashi, K., Maezawa, T., Nakagawa, H., Hoshi, M., 2012. Existence of Two Sexual Races in the Planarian Species Switching between Asexual and Sexual Reproduction. Zool. Sci. 29 (4), 265. https://doi.org/10.2108/zsj.29.265.
Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. F1000Research 6 (1287). https://doi.org/10.12688/f1000research.12232.1.
Lanfear, R., Hua, X., Warren, D.L., 2016. Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. Genome Biol. Evol. 8 (8), 2319–2332. https://doi.org/10.1093/gbe/evw171.
Lartillot, N., Philippe, H., 2004. A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. Mol. Biol. Evol. 21 (6), 1095–1109. https://doi.org/10.1093/molbev/msh112.
Laumer, C.E., Hejnol, A., Giribet, G., 2015. Nuclear genomic signals of the "microturbellarian" roots of platyhelminth evolutionary innovation. ELife 2015 (4). https://doi.org/10.7554/eLife.05503.
Lázaro, E.M., Riutort, M., 2013. *Dugesia sicula* (Platyhelminthes, Tricladida): the colonizing success of an asexual Planarian. BMC Evol. Biol. 13 (1), 268. https://doi.org/10.1186/1471-2148-13-268.
Lázaro, E.M., Sluys, R., Pala, M., Stochino, G.A., Baguñà, J., Riutort, M., 2009. Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). Mol. Phylogenet. Evol. 52, 835–845. https://doi.org/10.1016/j.ympev.2009.04.022.
Le, S.Q., Gascuel, O., 2008. An Improved General Amino Acid Replacement Matrix. Mol. Biol. Evol. 25 (7), 1307–1320. https://doi.org/10.1093/MOLBEV/MSN067.
Leria, L., Vila-Farré, M., Solà, E., Riutort, M., 2019. Outstanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. BMC Evol. Biol. 19 (1) https://doi.org/10.1186/s12862-019-1440-1.
Leria, L., Vila-Farré, M., Álvarez-Presas, M., Sánchez-Gracia, A., Rozas, J., Sluys, R., Riutort, M., 2020. Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhelminthes, Tricladida): Extreme intraindividual genetic diversity, morphological stasis, and karyological variability. Mol. Phylogenet. Evol. 143, 106496 https://doi.org/10.1016/j.ympev.2019.05.010.
Leria, L., Riutort, M., Romero, R., Ferrer, X., Vila-Farré, M., 2022. Microplate tectonics and the environment as distribution drivers in Western Mediterranean freshwater planarians. J. Biogeogr. 49, 1124–1136. https://doi.org/10.1111/jbi.14373.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22 (13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
Li, Y., Shen, X.X., Evans, B., Dunn, C.W., Rokas, A., 2021. Rooting the Animal Tree of Life. Mol. Biol. Evol. 38 (10), 4322–4333. https://doi.org/10.1093/molbev/msab170.
Liu, L., Anderson, C., Pearl, D., Edwards, S.V., 2019. Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model. In: Anisimova, M. (Ed.), *Evolutionary Genomics. Methods in molecular biology* (Vol. 1910, pp. 211–239). https://doi.org/10.1007/978-1-4939-9074-0_7.
Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. 38 (10), 4647–4654. https://doi.org/10.1093/molbev/msab199.
Martinez Rica, J.P., Montserrat Recoder, P., 1990. Biogeographic features of the Pyrenean range. Mt. Res. Dev. 10 (3), 235–240.
Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.
Mirarab, S., Nakhleh, L., Warnow, T., 2021. Multispecies Coalescent: Theory and Applications in Phylogenetics. Annu. Rev. Ecol. Evol. Syst. 52, 247–268. https://doi.org/10.1146/annurev-ecolsys-012121-095340.
Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403 (6772), 853–858. https://doi.org/10.1038/35002501.
Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press Inc., New York.
Ninot, J. M., Carrillo, E., Ferré, A., 2017. The Pyrenees BT - *The Vegetation of the Iberian Peninsula: Volume 1* (J. Loidi, Ed.). https://doi.org/10.1007/978-3-319-54784-8_8.
Phillips, A.J.L., Hyde, K.D., Alves, A., Liu, J.K., 2018. Families in Botryosphaeriales: a phylogenetic, morphological and evolutionary perspective. Fungal Diversity 94 (1), 1–22. https://doi.org/10.1007/S13225-018-0416-6.
Quang, L.S., Gascuel, O., Lartillot, N., 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24 (20), 2317–2323. https://doi.org/10.1093/bioinformatics/btn445.
Rambaut, A., Drummond, A.J., 2007. *Tracer v1.4.* Available at: beast.bio.ed.ac.uk/Tracer.
Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3 (2), 217–223. https://doi.org/10.1111/J.2041-210X.2011.00169.X.
Riutort, M., Álvarez-Presas, M., Lázaro, E., Solà, E., Paps, J., 2012. Evolutionary history of the Tricladida and the Platyhelminthes: an up-to-date phylogenetic and systematic account. Int. J. Dev. Biol. 56, 5–17. https://doi.org/10.1387/ijdb.113441mr.
Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Huelsenbeck, J.P., 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542. https://doi.org/10.1093/sysbio/sys029.
Rosenbaum, G., Lister, G.S., Duboz, C., 2002. Reconstruction of the tectonic evolution of the Western Mediterranean since the Oligocene. J. Virtual Explorer 8, 107–130. https://doi.org/10.3809/jvirtex.2002.00053.
Schrempf, D., Lartillot, N., Szöllősi, G., 2020. Scalable Empirical Mixture Models That Account for Across-Site Compositional Heterogeneity. Mol. Biol. Evol. 37 (12), 3616–3631. https://doi.org/10.1093/molbev/msaa145.
Schwander, T., Henry, L., Crespi, B.J., 2011. Molecular evidence for ancient asexuality in timema stick insects. Curr. Biol. 21 (13), 1129–1134. https://doi.org/10.1016/j.cub.2011.05.026.
Sluys, R., Solà, E., Gritzalis, K., Vila-farré, M., Mateos, E., Riutort, M., 2013. Integrative delineation of species of Mediterranean freshwater planarians (Platyhelminthes: Tricladida: Dugesiidae). Zool. J. Linn. Soc. 169, 523–547. https://doi.org/10.1111/zoj.12077.
Solà, E., Sluys, R., Gritzalis, K., Riutort, M., 2013. Fluvial basin history in the northeastern Mediterranean region underlies dispersal and speciation patterns in the genus *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae). Mol. Phylogenet. Evol. 66 (3), 877–888. https://doi.org/10.1016/j.ympev.2012.11.010.
Solà, E., Leria, L., Stocchino, G.A., Bagherzadeh, R., Balke, M., Daniels, S.R., Riutort, M., 2022. Three dispersal routes out of Africa: The puzzling biogeographical history of Dugesia freshwater planarians. J. Biogeogr. 49, 1219–1233. https://doi.org/10.1111/jbi.14371.
Stocchino, G.A., Manconi, R., Corso, G., Sluys, R., Casu, S., Pala, M., 2009. African planarians: Morphology and karyology of *Dugesia maghrebiana* sp. n. (Platyhelminthes, Tricladida) from Tunisia. Italian J. Zool. 76 (1), 83–91. https://doi.org/10.1080/11250000802141683.
Stocchino, G.A., Manconi, R., 2013. Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). Italian J. Zool. 80 (3), 319–328. https://doi.org/10.1080/11250003.2013.822025.
Stocchino, G.A., Sluys, R., Manconi, R., 2012. A new species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from the Afromontane forest in South Africa, with an overview of freshwater planarians from the African continent. Zootaxa 58 (3551), 43–58. https://doi.org/10.11646/zootaxa.3551.1.3.
Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. 17, 57–86.
Team, R.C., 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.r-project.org/.
Vergés, J., Millán, H., Roca, E., Muñoz, J.A., Marzo, M., Cirés, J., Cloetingh, S., 1995. Eastern Pyrenees and related foreland basins: pre-, syn- and post-collisional crustal-scale cross-sections. Mar. Pet. Geol. 12 (8), 903–915. https://doi.org/10.1016/0264-8172(95)98854-X.
Vila-Farré, M., Rink, J.C., 2018. The Ecology of Freshwater Planarians. In J. C. Rink (Ed.), *Planarian Regeneration. Methods in Molecular Biology, vol 1774* (pp. 173–205). https://doi.org/10.1007/978-1-4939-7802-1_3.
Vu, D., Groenewald, M., de Vries, M., Gehrmann, T., Stielow, B., Eberhardt, U., Verkley, G.J.M., 2019. Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. Stud. Mycol. 92, 135–154. https://doi.org/10.1016/j.simyco.2018.05.001.
Wang, H.C., Minh, B.Q., Susko, S., Roger, A.J., 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. 67, 216–235. https://doi.org/10.1093/sysbio/syx068.
Zhang, C., Scornavacca, C., Molloy, E.K., Mirarab, S., 2020. ASTRAL-pro: Quartet-based species-tree inference despite paralogy. Mol. Biol. Evol. 37 (11), 3292–3307. https://doi.org/10.1093/molbev/msaa139.