

Chapter IV

Section 1

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

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

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

Lisandra Benítez-Álvarez¹, Laia Leria¹, Rosa Fernández², Eduardo Mateos³, Nard Bennis⁴, Younes el Ouanighi⁴, Majida el Alami⁴, Mohamed Yacoubi-Khebiza⁵, Houssam Ayt Ougougdal⁵ & Marta Riutort^{1,*}

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

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

³Departament 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.

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

⁵Water, Biodiversity and Climate Change Laboratory, Semailia Faculty of Sciences, Cadi Ayyad University. BP 2390, Avenue Le Prince Moulay Abdellah, 400 10 Marrakech, Morocco

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

Highlights:

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

Abstract

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

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

1. Introduction

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

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

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

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

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

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

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

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

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

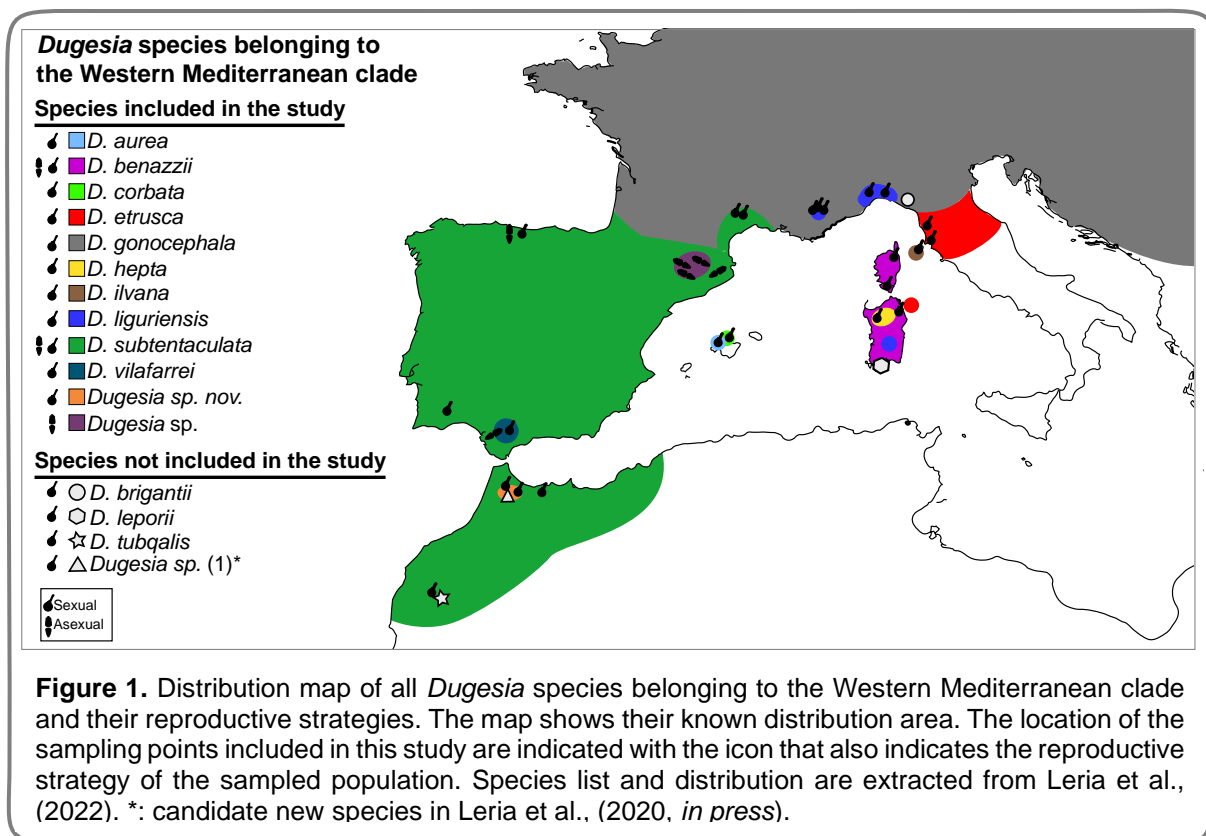
Here, for the first time, we use transcriptomic data to carry out a phylogenetic study in freshwater planarians, focusing on *Dugesia* species belonging to the West Mediterranean clade. We include representatives of most species known to date from the area (11 out of 13 described species, plus two outgroups.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

2. Material and methods

2.1. Taxon sampling

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

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



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

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

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

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

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

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

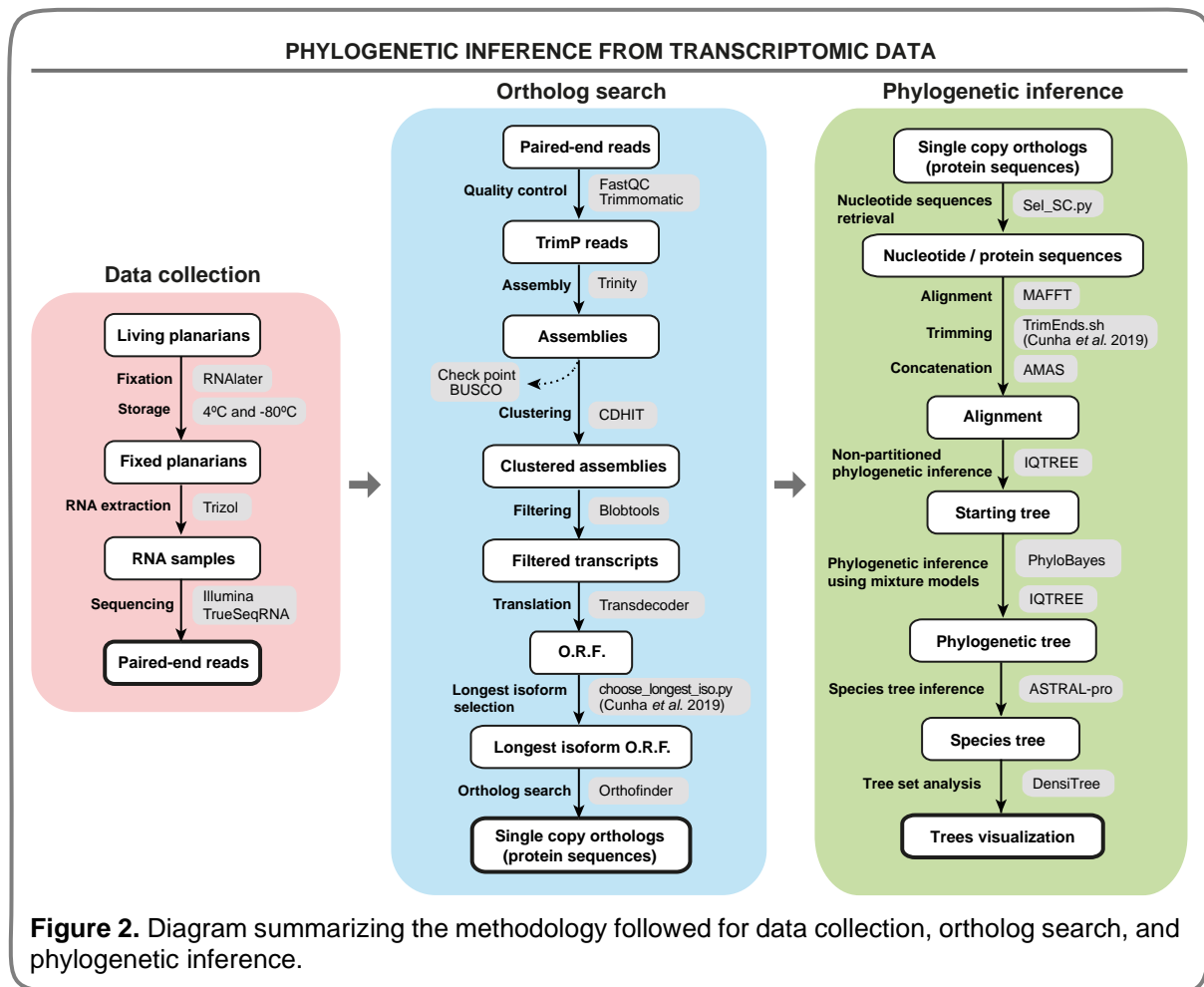
The sexuality or asexuality of the individuals was assessed observing them under the stereomicroscope and taking into account previous information about known populations. Sexual individuals were recognized by the presence of the gonopore, the external aperture of the copulatory apparatus, and asexual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

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

2.2. Nucleic acids extraction and RNA library preparation

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

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



2.3. DNA-Barcoding identification

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

A fragment of approximately 800 bp was amplified by Polymerase Chain Reaction (PCR), using 0.4 μ M of the BarT (M Álvarez-Presas, Carbayo, Rozas, & Riutort, 2011) and COIR (Lázaro et al., 2009) primers in 25 μ l of final reaction volume with MgCl₂ (2.5 mM), dNTPs (30 μ M), and 0.75 U of Go Taq® DNA polymerase enzyme (Promega Madison, Wisconsin, USA) with its buffer (1X). The amplification conditions

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

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

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

2.4. Bioinformatic Workflow for transcriptomic analysis

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

2.4.1 Quality control and trimming

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

2.4.2. Assembly and clustering

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

2.4.3. Transcript filtering

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

2.4.4. Ortholog search

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

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

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

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

2.4.5. Alignment and concatenation

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

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

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

2.4.6. Phylogenetic Inference

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

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

Datasets		Number of samples	Number of SC	Number of Positions		Analyses					
Abbreviation	Explanation*			Protein	Nucleotide	ML Exp. Prot/Nuc	ML-Mixt Prot/Nuc	BI-Mixt Prot/Nuc	ASTRAL Prot/Nuc	DensiTree Prot/Nuc	ARC Nuc
Dataset 1	All samples retained after filtering	82	717	274960	839555	X	X				
Dataset 2	Reduced Dataset 1 to infer tree 1	29	717	276745	841425	X	X	X	X	X	
Dataset 3	Reduced Dataset 1 to infer tree 2	13	717	-	831918						
Dataset 4	Subtentaculata group	23	4175	176432	5376224	X	X				X
Dataset 5	Etrusca-liguriensis group	36	1984	742966	2283579	X	X			X	
Dataset 6	Dataset 5 without samples from Berga	31	1984	744252	2255464	X	X			X	

Abbreviation

ML Exp.

ML-Mixt

BI-Mixt

ASTRAL

DensiTree

ARC Nuc

SC

***Further Explanation**

Prot/Nuc Maximum Likelihood without partitions. Protein and nucleotide data

Prot/Nuc Maximum Likelihood using Mixture Models. Protein and nucleotide data

Prot/Nuc Bayesian Inference using Mixture Models. Protein and nucleotide data

Prot/Nuc Species Tree using individual gene trees. Protein and nucleotide data

Prot/Nuc Visualization of individual gene trees. Protein and nucleotide data

Ancestral Reconstruction Character. Nucleotide data

Single Copy Orthologs

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

Maximum Likelihood Trees: one partition

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

Maximum Likelihood Trees: Mixture models

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

Bayesian Inference Trees: Mixture models

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

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

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

Reconciling gene trees with species tree: Multispecies Coalescent Model

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

2.4.7. Individual gene trees visualisation

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

2.4.8. Ancestral character reconstruction (ACR)

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

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

3. Results

3.1. Sampling

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

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

3.2. Ortholog searches and datasets

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

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

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

3.3. Phylogeny of *Dugesia* in the Western Mediterranean

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

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

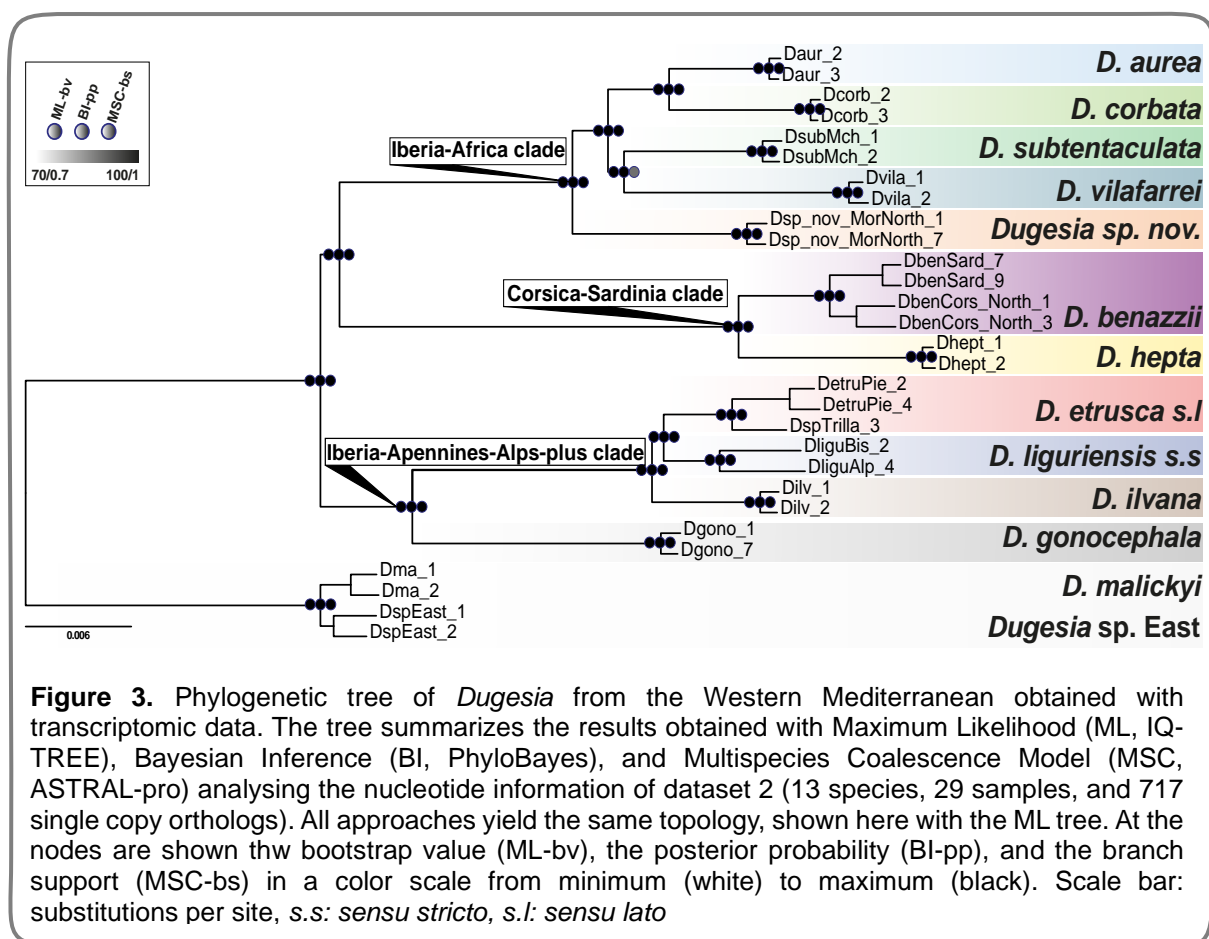
ML, BI and MSC analyses of dataset 2 yielded the same topology for nucleotide and protein data. However, regarding the information content, some support values were lower when using protein data in ML and MSC methods, but not for BI, for which the support values were maximum ($pp = 1$) for nucleotide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

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

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

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

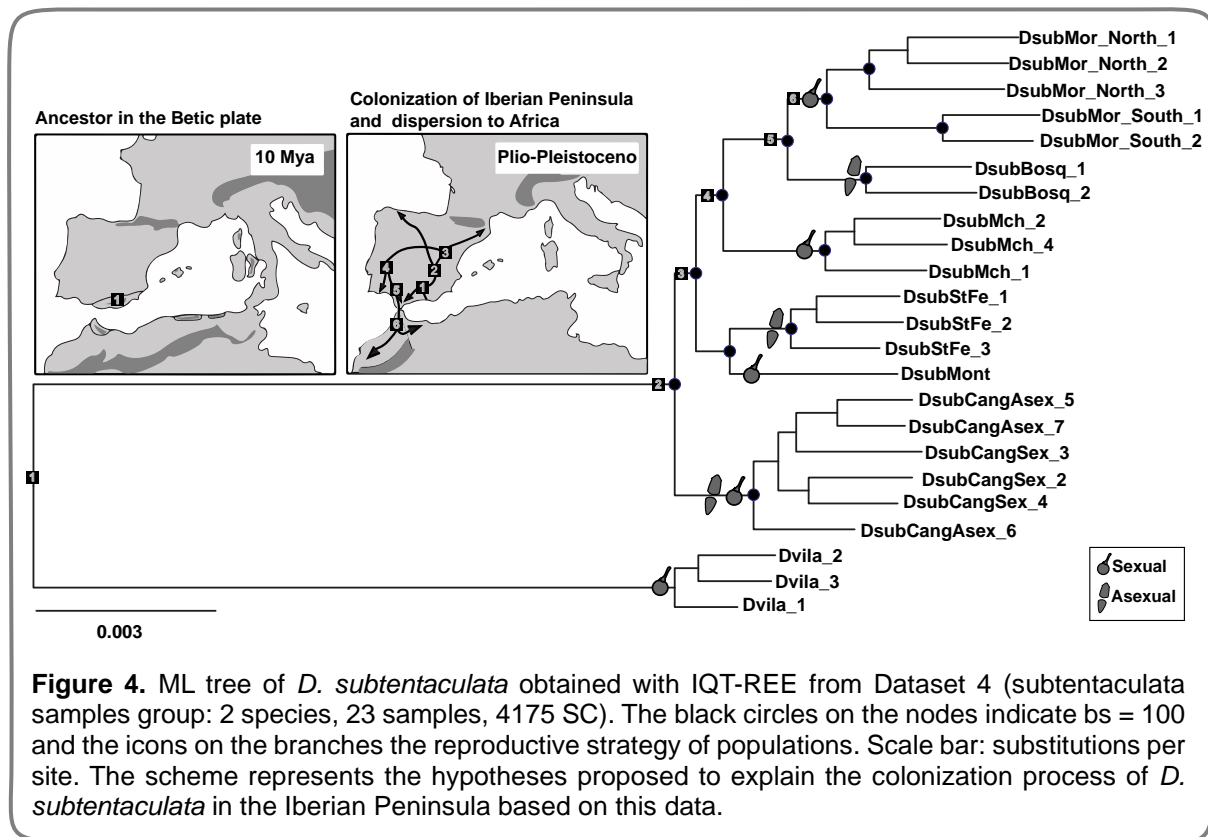
All methods showed high support values for all nodes, except for the node grouping *D. subtentaculata* and *D. vilafarrei*, where the branch support of MSC is relatively low (0.88), but the final normalised quartet score (a measure of support for the entire topology) is high (Fig. 3, Appendix A).



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

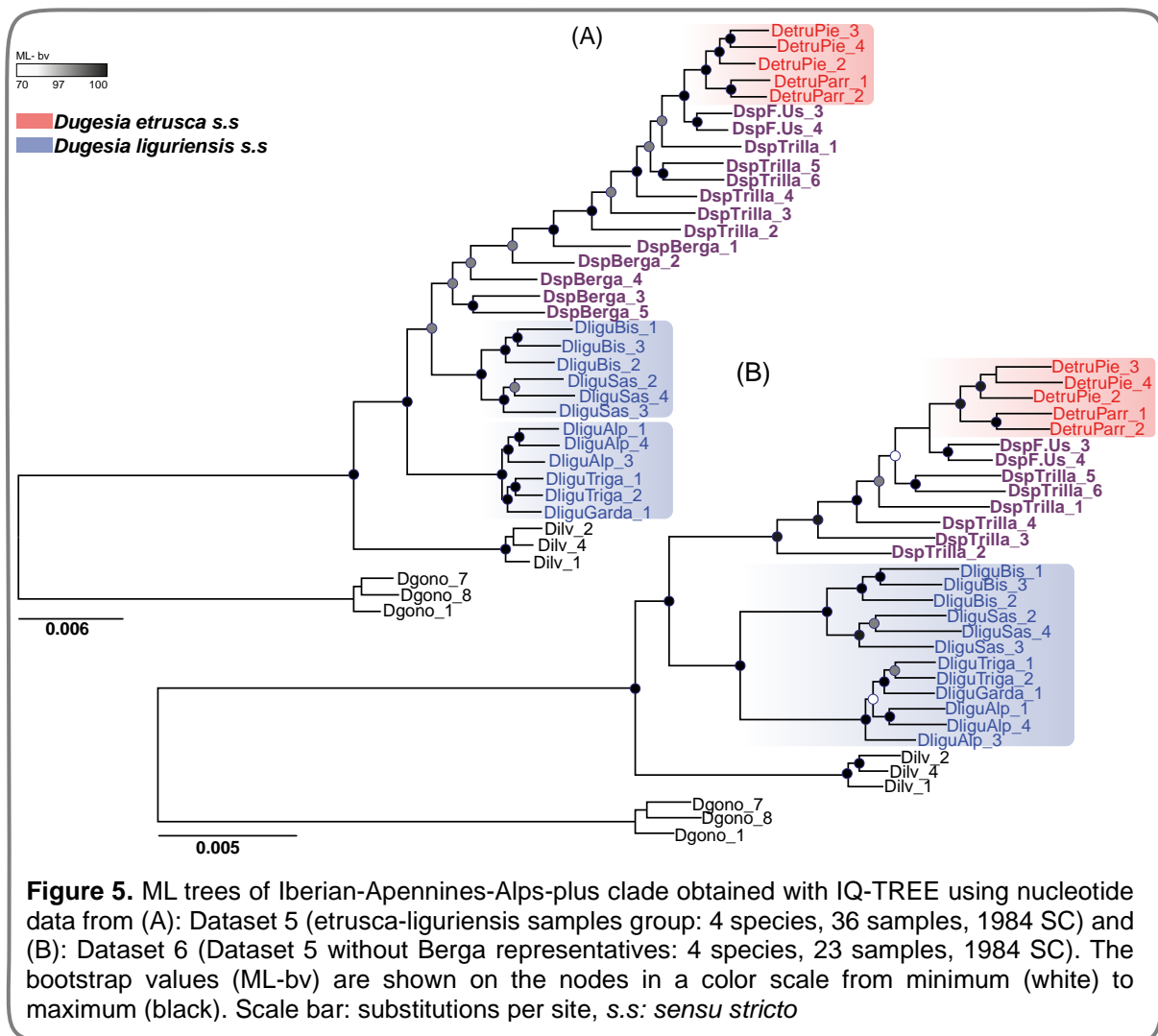
The ML tree obtained from the analysis of dataset 4 (Table 2) shows a clear structure in this species (Fig. 4). The population of Cangas, in the North of the Iberian Peninsula is the first to diverge. Although this population is represented by sexual and asexual individuals, no differentiation by reproductive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

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



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

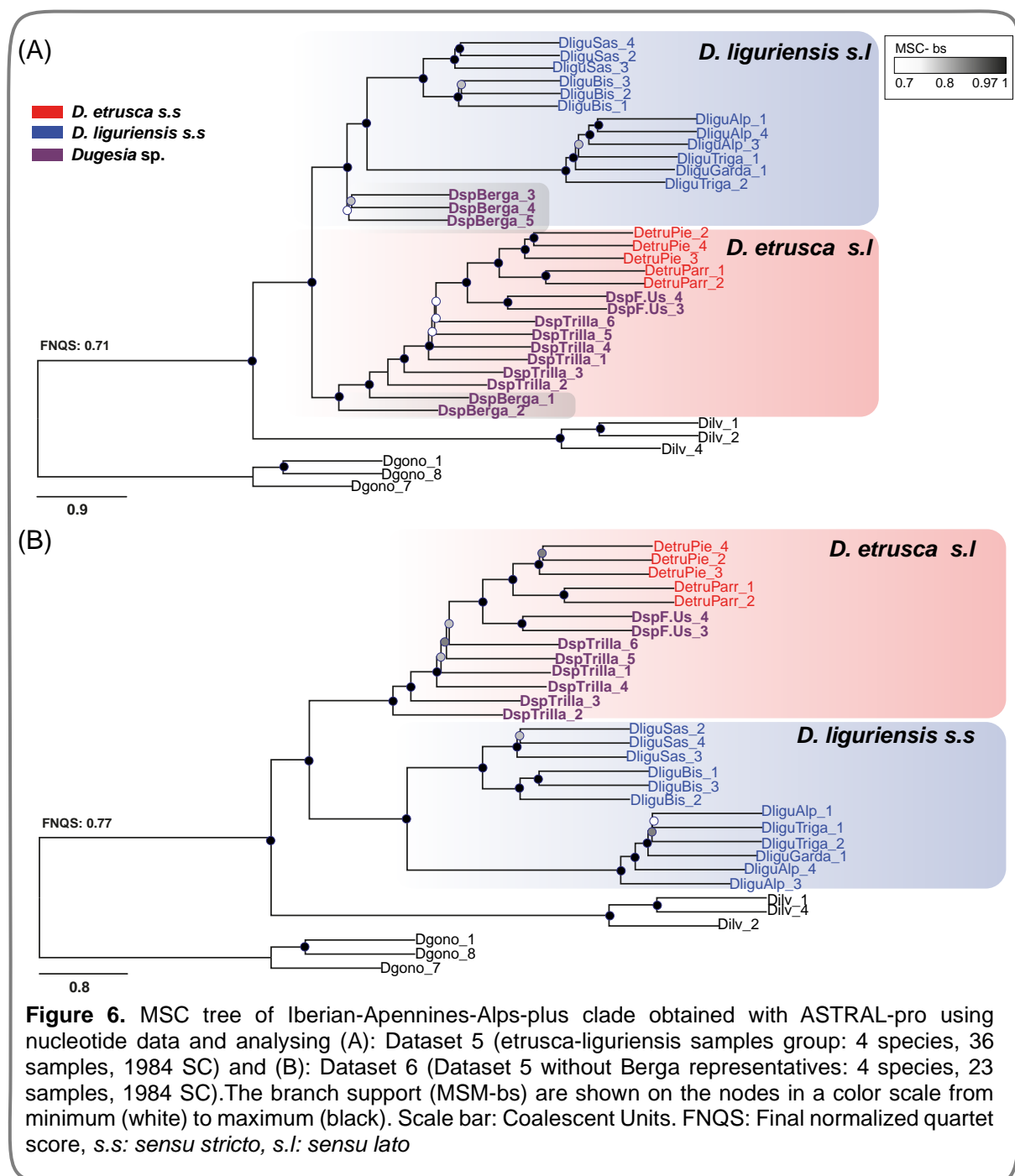
The ML tree obtained from dataset 5 (Table 2) shows an unexpected topology (Fig. 5A). *D. etrusca* s.s is monophyletic (red in Fig. 5). However, *D. liguriensis* s.s is divided in two clades and results in a paraphyletic group, since samples from Sasello and Bisagno group with *D. etrusca* s.l. The unclassified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



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

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

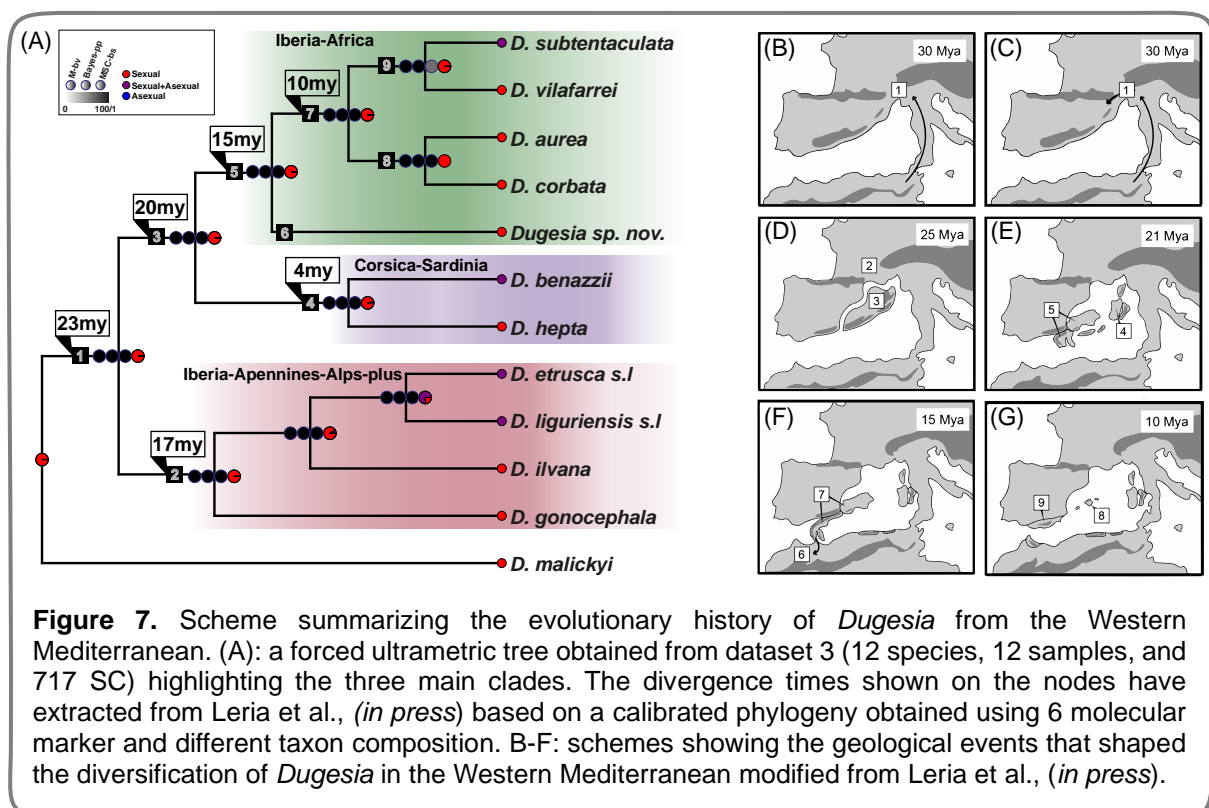
dataset 6 (Table 2). From this dataset, resolved trees were obtained, showing two monophyletic groups (Fig. 5B, Fig. 6B). Nonetheless, although the samples from Font de La Trilla grouped in *D. etrusca s.l.*, they continued showing the stranger ladder-like pattern and its grouping has low support (Fig.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



3.6. Ancestral character reconstruction

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

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



4. Discussion

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

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

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

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

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

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

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

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

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

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

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

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

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

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

4.3. The Corsica-Sardinia clade

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Final remarks

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

Data accessibility

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

Credit authorship contribution statement

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

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

Acknowledgements

Thanks to Julio Rozas for sharing with us the computational cluster of the Evolutionary Genomics & Bioinformatics research group. Thanks also to Silvia Hinojosa for guiding the first bioinformatics steps and Juan Manuel Lentijo Mondejar for his advice with the python scripts. In addition, we thanks to Ignacio Tenaguillo Arriola for the help launching some analysis to the cluster. Also thanks to Eduard Solà for the sampling campaign in Greece. Our thanks also to the Laboratoire Ecologie, Biodiversité et Environnement, Université Abdelmalek Essaâdi and the Water, Biodiversity and Climate Change Laboratory, Cadi Ayyad University by the support with the sampling and transportation permissions in Morocco. Sampling Permission in Morocco: Decision 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 FI Grant from the Generalitat de Catalunya (to Lisandra Benítez-Álvarez).

Supplementary Tables

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

Sample Code	Species	Reproductive Strategy	Region, Country	Locality	Geographic Coordinates (DMS)		Voucher ID	Biosample accession code in NCBI-SRA repository BioProject PRJNA797284.	SRA accession code in NCBI-SRA repository
					Latitude	Longitude			
Daur_1	<i>D. aurea</i>	Sexual	Majorca, Spain	Soller			MR1079.1	SAMN24979312	SRR17642743
Daur_2	<i>D. aurea</i>	Sexual	Majorca, Spain	Soller	39°45'41"	2°42'44"	MR1079.2	SAMN24979313	SRR17642732
Daur_3	<i>D. aurea</i>	Sexual	Majorca, Spain	Soller			MR1079.3	SAMN24979314	SRR17642721
DbenSard_6	<i>D. benazzii</i>	Sexual	Sardinia, Italy	Monte Albo			MR0652.6	SAMN24979320	SRR17642753
DbenSard_7	<i>D. benazzii</i>	Sexual	Sardinia, Italy	Monte Albo	40°34'30"	9°40'31"	MR0652.7	SAMN24979321	SRR17642752
DbenSard_9	<i>D. benazzii</i>	Sexual	Sardinia, Italy	Monte Albo			MR0652.9	SAMN24979322	SRR17642751
DbenCors_North_1	<i>D. benazzii</i>	Sexual	Corsica, Italy	Campile			MR1257.1	SAMN24979315	SRR17642710
DbenCors_North_2	<i>D. benazzii</i>	Sexual	Corsica, Italy	Campile	42°30'24.7"	9°22'27.8"	MR1257.2	SAMN24979316	SRR17642699
DbenCors_North_3	<i>D. benazzii</i>	Sexual	Corsica, Italy	Campile			MR1257.3	SAMN24979317	SRR17642688
DbenCors_South_5	<i>D. benazzii</i>	Sexual	Corsica, Italy	Monacia-d'Aullène	41° 32' 41.7"	09° 01' 24.1"	MR1260.5	SAMN24979318	SRR17642677
DbenCors_South_6	<i>D. benazzii</i>	Sexual	Corsica, Italy	Monacia-d'Aullène			MR1260.6	SAMN24979319	SRR17642673
Dcorb_1	<i>D. corbata</i>	Sexual	Majorca, Spain	Sa Calobra			MR1080.1	SAMN24979323	SRR17642750
Dcorb_2	<i>D. corbata</i>	Sexual	Majorca, Spain	Sa Calobra	39°49'44.6"	2°48'52.8"	MR1080.2	SAMN24979324	SRR17642749
Dcorb_3	<i>D. corbata</i>	Sexual	Majorca, Spain	Sa Calobra			MR1080.3	SAMN24979325	SRR17642748
DetruParr_1	<i>D. etrusca</i>	Sexual	Toscana, Italy	Parrana	43°32'23"	10°27'37"	MR1256.1	SAMN24979328	SRR17642745
DetruParr_2	<i>D. etrusca</i>	Sexual	Toscana, Italy	Parrana			MR1256.4	SAMN24979329	SRR17642744
DetruPie_2	<i>D. etrusca</i>	Sexual	Toscana, Italy	Pieve			MR1355.2	SAMN24979330	SRR17642742
DetruPie_3	<i>D. etrusca</i>	Sexual	Toscana, Italy	Pieve	43°29'52"	10°37'33"	MR1355.3	SAMN24979331	SRR17642741
DetruPie_4	<i>D. etrusca</i>	Sexual	Toscana, Italy	Pieve			MR1355.4	SAMN24979332	SRR17642740
Dgono_1	<i>D. gonocephala</i>	Sexual	France	Montpellier	43°43'22"	3°8'38"	MR1283.1	SAMN24979333	SRR17642739
Dgono_7	<i>D. gonocephala</i>	Sexual	France	Montpellier			MR1283.7	SAMN24979334	SRR17642738

Dgono_8	<i>D. gonocephala</i>	Sexual	France	Montpellier			MR1283.8	SAMN24979335	SRR17642737
Dhept_1	<i>D. hepta</i>	Sexual	Sardinia, Italy	Logulento			MR0980.1	SAMN24979336	SRR17642736
Dhept_2	<i>D. hepta</i>	Sexual	Sardinia, Italy	Logulento	40°49'16.682"	8°35'35.213"	MR0980.2	SAMN24979337	SRR17642735
Dhept_5	<i>D. hepta</i>	Sexual	Sardinia, Italy	Logulento			MR0980.5	SAMN24979338	SRR17642734
Dilv_1	<i>D. ilvana</i>	Sexual	Italy	Elba			MR1357.1	SAMN24979339	SRR17642733
Dilv_2	<i>D. ilvana</i>	Sexual	Italy	Elba	42° 47' 9"	10° 10' 3"	MR1357.2	SAMN24979340	SRR17642731
Dilv_4	<i>D. ilvana</i>	Sexual	Italy	Elba			MR1357.4	SAMN24979341	SRR17642730
DliguBis_1	<i>D. liguriensis</i>	Sexual	Italy	Bisagno			MR1254.1	SAMN24979345	SRR17642726
DliguBis_2	<i>D. liguriensis</i>	Sexual	Italy	Bisagno	44°26'40"	9°5'6"	MR1254.2	SAMN24979346	SRR17642725
DliguBis_3	<i>D. liguriensis</i>	Sexual	Italy	Bisagno			MR1254.3	SAMN24979347	SRR17642724
DliguAlp_1	<i>D. liguriensis</i>	Sexual	France	Alps Maritims			MR1252.1	SAMN24979342	SRR17642729
DliguAlp_3	<i>D. liguriensis</i>	Sexual	France	Alps Maritims	43°47'9"	6°38'7"	MR1252.3	SAMN24979343	SRR17642728
DliguAlp_4	<i>D. liguriensis</i>	Sexual	France	Alps Maritims			MR1252.4	SAMN24979344	SRR17642727
DliguGarda_1	<i>D. liguriensis</i>	Sexual	France	La Garda	43°49'25"	6°34'55"	MR1302.3	SAMN24979348	SRR17642723
DliguSas_2	<i>D. liguriensis</i>	Sexual	Italy	Sassello			MR1253.2	SAMN24979349	SRR17642722
DliguSas_3	<i>D. liguriensis</i>	Sexual	Italy	Sassello	44°29'19.2"	8°28'2.9"	MR1253.3	SAMN24979350	SRR17642720
DliguSas_4	<i>D. liguriensis</i>	Sexual	Italy	Sassello			MR1253.4	SAMN24979351	SRR17642719
DliguTriga_1	<i>D. liguriensis</i>	Sexual	France	Trigance	43°47'44"	6°26'40"	MR1301.1	SAMN24979352	SRR17642718
DliguTriga_2	<i>D. liguriensis</i>	Sexual	France	Trigance			MR1301.2	SAMN24979353	SRR17642717
DsubMont	<i>D. subtentaculata</i>	Sexual	France	Montpellier	43°43'22"	3°8'38"	MR1283.5	SAMN24979381	SRR17642686
DsubBosq_1	<i>D. subtentaculata</i>	Fissiparous	Andalusia, Spain	El Bosque	36°45'42"	-5°30'20"	MR1267.1	SAMN24979370	SRR17642698
DsubBosq_2	<i>D. subtentaculata</i>	Fissiparous	Andalusia, Spain	El Bosque			MR1267.2	SAMN24979371	SRR17642697
DsubCangAsex_5	<i>D. subtentaculata</i>	Facultative. Fiss.	Asturias, Spain	Cangas			MR1297.5	SAMN24979372	SRR17642696
DsubCangAsex_6	<i>D. subtentaculata</i>	Facultative. Fiss.	Asturias, Spain	Cangas			MR1297.6	SAMN24979373	SRR17642695
DsubCangAsex_7	<i>D. subtentaculata</i>	Facultative. Fiss.	Asturias, Spain	Cangas	43°21'54.65"	-5°9'6.19"	MR1297.7	SAMN24979374	SRR17642694
DsubCangSex_2	<i>D. subtentaculata</i>	Facultative. Sex.	Asturias, Spain	Cangas			MR1297.2	SAMN24979375	SRR17642693
DsubCangSex_3	<i>D. subtentaculata</i>	Facultative. Sex.	Asturias, Spain	Cangas			MR1297.3	SAMN24979376	SRR17642692
DsubCangSex_4	<i>D. subtentaculata</i>	Facultative. Sex.	Asturias, Spain	Cangas			MR1297.4	SAMN24979377	SRR17642691

DsubMor_North_1	<i>D. subtentaculata</i>	Sexual	Morocco	Magoo Timriouen	35°06'42.5"	-5°11'19.9"	MR1276.1	SAMN24979382	SRR17642685
DsubMor_North_2	<i>D. subtentaculata</i>	Sexual	Morocco	Magoo Timriouen			MR1276.2	SAMN24979383	SRR17642684
DsubMor_North_3	<i>D. subtentaculata</i>	Sexual	Morocco	Beni H'amed	35°08'37.3"	-5°06'51.8"	MR1252.2	SAMN24979384	SRR17642683
DsubMor_South_1	<i>D. subtentaculata</i>	Sexual	Morocco	Imlil	31°09'35.8"	-7°55'41.1"	MR1278.1	SAMN24979385	SRR17642682
DsubMor_South_2	<i>D. subtentaculata</i>	Sexual	Morocco	Imlil			MR1279.3	SAMN24979386	SRR17642681
DsubMch_1	<i>D. subtentaculata</i>	Sexual	Portugal	Monchique			MR1358.1	SAMN24979378	SRR17642690
DsubMch_2	<i>D. subtentaculata</i>	Sexual	Portugal	Monchique	37° 17' 20"	-8° 33' 15"	MR1358.2	SAMN24979379	SRR17642689
DsubMch_4	<i>D. subtentaculata</i>	Sexual	Portugal	Monchique			MR1358.4	SAMN24979380	SRR17642687
DsubStFe_1	<i>D. subtentaculata</i>	Fissiparous	Catalonia, Spain	Santa Fe			MR1082.1	SAMN24979387	SRR17642680
DsubStFe_2	<i>D. subtentaculata</i>	Fissiparous	Catalonia, Spain	Santa Fe	41° 46' 26.35	2° 27' 42.24	MR1082.2	SAMN24979388	SRR17642679
DsubStFe_3	<i>D. subtentaculata</i>	Fissiparous	Catalonia, Spain	Santa Fe			MR1082.3	SAMN24979389	SRR17642678
Dvila_1	<i>D. vilafarrei</i>	Sexual	Andalusia, Spain	El Bosque			MR1266.1	SAMN24979390	SRR17642676
Dvila_2	<i>D. vilafarrei</i>	Sexual	Andalusia, Spain	El Bosque	36°45'42"	-5°30'20"	MR1266.2	SAMN24979391	SRR17642675
Dvila_3	<i>D. vilafarrei</i>	Sexual	Andalusia, Spain	El Bosque			MR1266.3	SAMN24979392	SRR17642674
Dsp_nov_MorNorth_1	<i>Dugesia sp. nov</i>	Sexual	Morocco	Beni H'amed			MR1251.1	SAMN24979361	SRR17642708
Dsp_nov_MorNorth_6	<i>Dugesia sp. nov</i>	Sexual	Morocco	Beni H'amed	35°08'37.3"	-5°06'51.8"	MR1251.6	SAMN24979362	SRR17642707
Dsp_nov_MorNorth_7	<i>Dugesia sp. nov</i>	Sexual	Morocco	Beni H'amed			MR1251.7	SAMN24979363	SRR17642706
DspF.Us_3	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Font de l'Us	42°15'50"	0°59'45"	MR1263.3	SAMN24979326	SRR17642747
DspF.Us_4	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Font de l'Us			MR1263.4	SAMN24979327	SRR17642746
DspTrilla_1	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1265.1	SAMN24979364	SRR17642705
DspTrilla_2	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1265.2	SAMN24979365	SRR17642704
DspTrilla_3	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Font de la Trilla	41°55'51"	1°1'14"	MR1265.3	SAMN24979366	SRR17642703
DspTrilla_4	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1361.2	SAMN24979367	SRR17642702
DspTrilla_5	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1361.3	SAMN24979368	SRR17642701
DspTrilla_6	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Font de la Trilla			MR1361.4	SAMN24979369	SRR17642700
DspBerga_1	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Berga			MR1360.1	SAMN24979356	SRR17642714
DspBerga_2	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Berga	42°6'21.97"	1°52'48.05"	MR1360.2	SAMN24979357	SRR17642713
DspBerga_3	<i>Dugesia sp.</i>	Fissiparous	Catalonia, Spain	Berga			MR1360.3	SAMN24979358	SRR17642712

DspBerga_4	<i>Dugesia</i> sp.	Fissiparous	Catalonia, Spain	Berga			MR1360.4	SAMN24979359	SRR17642711
DspBerga_5	<i>Dugesia</i> sp.	Fissiparous	Catalonia, Spain	Berga			MR1360.5	SAMN24979360	SRR17642709
Outgroup									
Dma_1	<i>D. malickyi</i>	Sexual	Greece	Mexiates	38°53'4.09"	22°18'53.16"	MR1261.1	SAMN24979354	SRR17642716
Dma_2	<i>D. malickyi</i>	Sexual	Greece	Mexiates			MR1261.3	SAMN24979355	SRR17642715
DspEast_1	<i>Dugesia</i> sp.	Sexual	Greece	Eleonas	38°34'29.1"	22°23'38.50"	MR1262.1	SAMN24979310	SRR17642755
DspEast_2	<i>Dugesia</i> sp.	Sexual	Greece	Eleonas			MR1262.2	SAMN24979311	SRR17642754

Summary

27 localities

83 samples

10 species described for Western Mediterranean

1 not described species

2 species from Greece as outgroup

Table S2. Sequences used in the DNA-Barcoding identification

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

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

AND

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

Are available at [Riutort's Lab site](#)

Table S4. Results of the three analysis done with OrthoFinder

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

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

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

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

Supplementary Figures

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

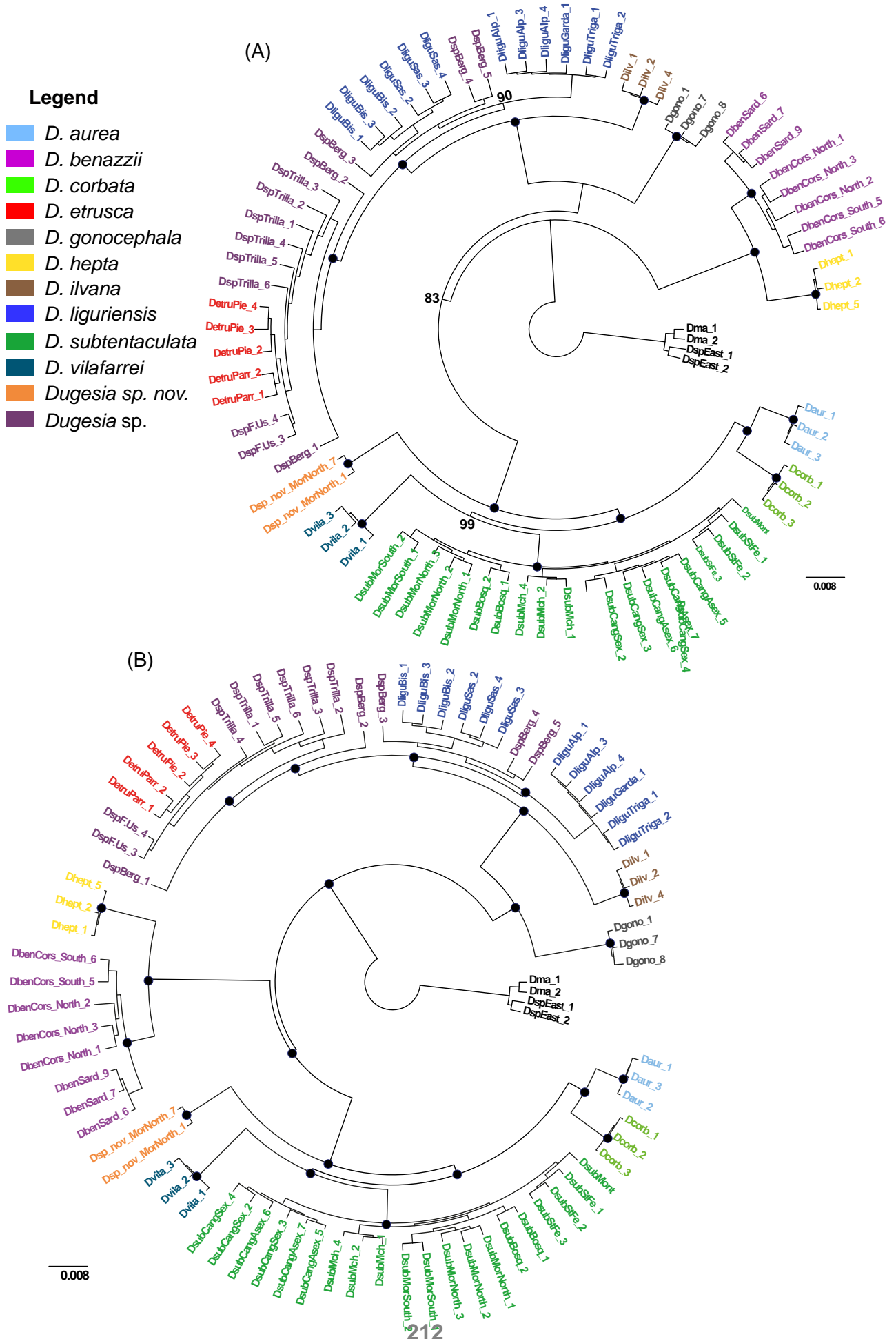


Figure S2. Phylogenetic trees obtained from dataset 2 (13 species, 29 samples, and 717 SC) with (A): ML using protein data, (B): BI using protein data, and (C): BI using nucleotide data. Supporting values are shown on the nodes, the black circles represents the maximum values (bv = 100 and pp = 1). Scale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

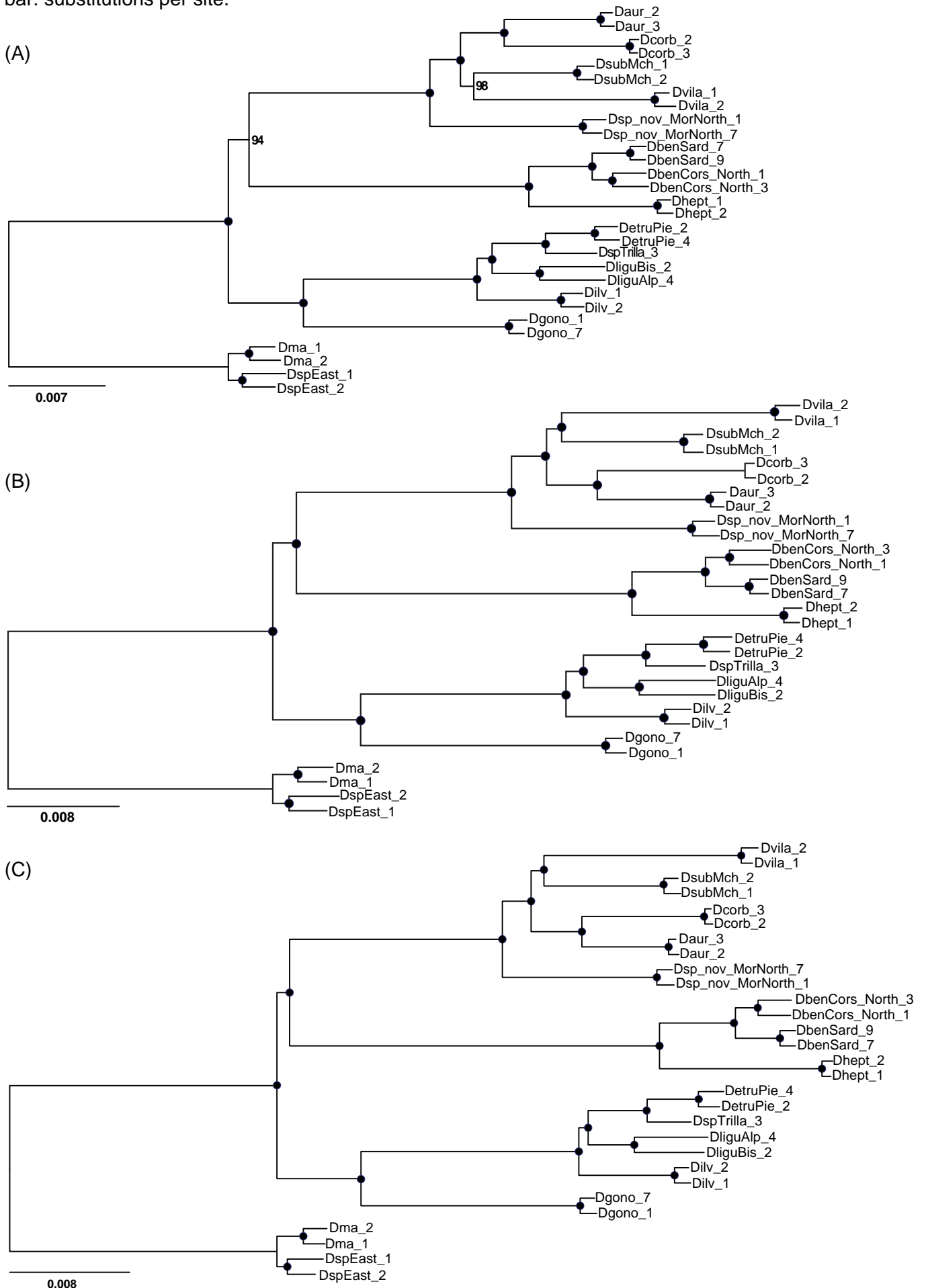
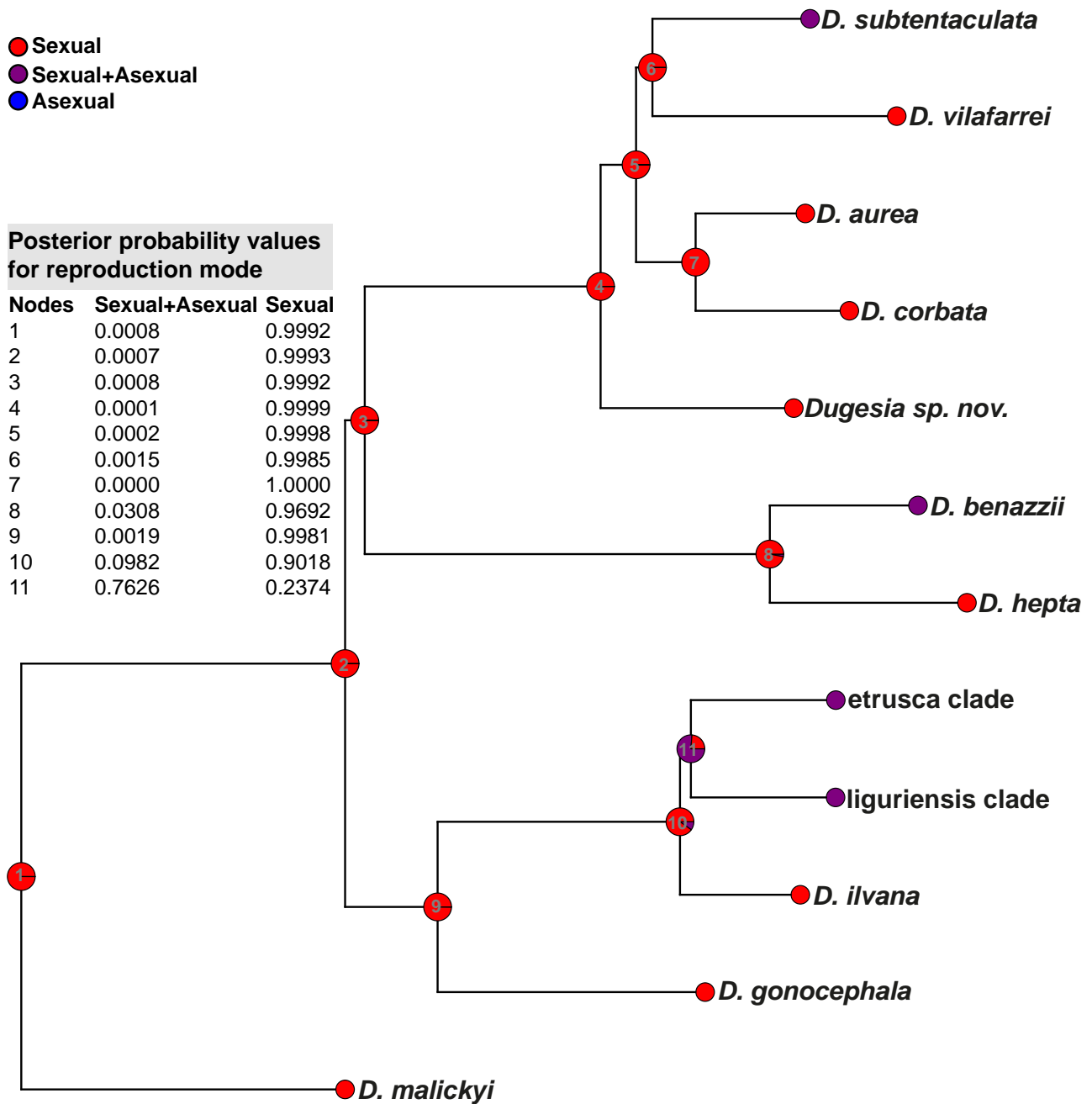


Figure S3. Ancestral character reconstruction using dataset 3 (12 species, 12 samples, and 717 SC). Pie chart of posterior probability (pp) for reproduction mode (sexual, sexual+asexual, and asexual) are shown on the nodes. The pp values are shown on the table



References

- Álvarez-Presas, M., Carbayo, F., Rozas, J., & Riutort, M. (2011). Land planarians (Platyhelminthes) as a model organism for fine scale phylogeographic studies: understanding patterns of biodiversity in the Brazilian Atlantic Forest hotspot. *Journal of Evolutionary Biology*, *24*, 887–896. <https://doi.org/10.1111/j.1420-9101.2010.02220.x>
- Álvarez-Presas, Marta, & Riutort, M. (2014). Planarian (Platyhelminthes, Tricladida) Diversity and Molecular Markers: A New View of an Old Group. *Diversity*, *6*, 323–338. <https://doi.org/10.3390/d6020323>
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Baguñà, J., Carranza, S., Pala, M., Ribera, C., Arnedo, M. A., Ribas, M., & Riutort, M. (1999). From morphology and karyology to molecules. New methods for taxonomical identification of asexual populations of freshwater planarians. A tribute to Professor Mario Benazzi. *Italian Journal of Zoology*, *66*, 207–214. <https://doi.org/10.1080/11250009909356258>
- Benazzi, M. (1946). Sopra una nova planaria d'acqua dolce. *Estratto Dall' Archivio Zoologico Italiano*, *XXXI*.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *29*, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, *4*:e1660. <https://doi.org/10.7717/peerj.1660>
- Bortolotti, V., Fazzuoli, M., Pandeli, E., Principi, G., Babbini, A., & Corti, S. (2001). Geology of central and eastern Elba island, Italy. *Ofioliti*, *26*(2 A), 97–105. <https://doi.org/10.4454/ofioliti.v26i2a.137>
- Bortolotti, V., Pandeli, E., & Principi, G. (2001). The geology of the Elba Island: An historical introduction. *Ofioliti*, *26*(2 A), 79–96. <https://doi.org/10.4454/ofioliti.v26i2a.136>
- Bouckaert, R. R., & Heled, J. (2014). DensiTree 2: Seeing Trees Through the Forest. *BioRxiv*, 012401. <https://doi.org/10.1101/012401>

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(421). <https://doi.org/10.1186/1471-2105-10-421>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. <https://doi.org/doi:10.1093/bioinformatics/btp348>
- Challis, R., & Paulini, M. (2021). *blobtoolkit/blobtools2: v2.6.1*. <https://doi.org/10.5281/zenodo.5032179>
- Chen, X., Chen, Y., Wu, C., & Wang, A. (2015). A new species of the genus *Girardia* (Tricladida: Dugesiidae) from China. *Zoological Systematic*, *40*(2), 166–178. <https://doi.org/10.11865/zs.20150202>
- Cheon, S., Zhang, J., & Park, C. (2020). Is Phylotranscriptomics as Reliable as Phylogenomics? *Molecular Biology and Evolution*, *37*(12), 3672–3683. <https://doi.org/10.1093/molbev/msaa181>
- Chris Venditti, Andrew Meade, M. P. (2008). Phylogenetic Mixture Models Can Reduce Node-Density Artifacts. *Systematic Biology*, *57*(2), 286–293. <https://doi.org/10.1080/10635150802044045>
- Cunha, T. J., & Giribet, G. (2019). A congruent topology for deep gastropod relationships. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1898). <https://doi.org/10.1098/rspb.2018.2776>
- Dapporto, L., & Cini, A. (2007). Faunal patterns in Tuscan archipelago butterflies: The dominant influence is recent geography not paleogeography. *European Journal of Entomology*, *104*(3), 497–503. <https://doi.org/10.14411/eje.2007.070>
- De Vicente, G., Cloetingh, S., Van Wees, J. D., & Cunha, P. P. (2011). Tectonic classification of Cenozoic Iberian foreland basins. *Tectonophysics*, *502*(1–2), 38–61. <https://doi.org/10.1016/j.tecto.2011.02.007>
- De Vries, E. J. (1986). On the taxonomic status of *Dugesia gonocephala* and *Dugesia subtentaculata* (Turbellaria, Tricladida, Paludicola). *Journal of Zoology*, *209*(1), 43–59. <https://doi.org/https://doi-org.sire.ub.edu/10.1111/j.1469-7998.1986.tb03565.x>

- De Vries, E. J. (1988). Further contribution to the taxonomy and biogeography of the subgenus *Dugesia* (Platyhelminthes: Tricladida: Paludicola) in the Mediterranean region and the Middle East. *Israel Journal of Zoology*, 35, 109–136.
- DeSalle, R., & Goldstein, P. (2019). Review and Interpretation of Trends in DNA Barcoding. *Frontiers in Ecology and Evolution*, 7(September), 1–11. <https://doi.org/10.3389/fevo.2019.00302>
- Dèzes, P., Schmid, S. M., & Ziegler, P. A. (2004). Evolution of the European Cenozoic Rift System: Interaction of the Alpine and Pyrenean orogens with their foreland lithosphere. *Tectonophysics*, 389(1–2), 1–33. <https://doi.org/10.1016/j.tecto.2004.06.011>
- Dèzes, P., Schmid, S. M., & Ziegler, P. A. (2005). Reply to comments by L. Michon and O. Merle on ‘Evolution of the European Cenozoic Rift System: Interaction of the Alpine and Pyrenean orogens with their foreland lithosphere’ by P. Dèzes, S.M. Schmid and P.A. Ziegler, *Tectonophysics* 389 (2004) 1-33. *Tectonophysics*, 401(3–4), 257–262. <https://doi.org/10.1016/j.tecto.2005.02.002>
- Di Nicola, M. R., & Vaccaro, A. (2020). New data on the presence of the Aesculapian snake *Zamenis longissimus* (Laurenti, 1768) (Serpentes Colubridae) on Elba Island (Tuscany, Italy). *Biodiversity Journal*, 11(2), 611–614. <https://doi.org/10.31396/biodiv.jour.2020.11.2.611.614>
- Dols-Serrate, D., Leria, L., Aguilar, J. P., Stocchino, G. A., & Riutort, M. (2020). *Dugesia hepta* and *Dugesia benazzii* (Platyhelminthes: Tricladida): two sympatric species with occasional sex? *Organisms Diversity and Evolution*, 20(3), 369–386. <https://doi.org/10.1007/s13127-020-00438-z>
- Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., ... Davis, C. C. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, 94, 447–462. <https://doi.org/10.1016/j.ympev.2015.10.027>
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(238). <https://doi.org/10.1186/s13059-019-1832-y>

- Fattorini, S. (2009). Both Recent and Pleistocene geography determine animal distributional patterns in the Tuscan Archipelago. *Journal of Zoology*, 277(4), 291–301. <https://doi.org/10.1111/j.1469-7998.2008.00540.x>
- Feng, Y. Y., Shen, T. T., Shao, C. C., Du, H., Ran, J. H., & Wang, X. Q. (2021). Phylotranscriptomics reveals the complex evolutionary and biogeographic history of the genus *Tsuga* with an East Asian-North American disjunct distribution. *Molecular Phylogenetics and Evolution*, 157, 107066. <https://doi.org/10.1016/J.YMPEV.2020.107066>
- Fernández, R., Edgecombe, G. D., & Giribet, G. (2016). Exploring Phylogenetic Relationships within Myriapoda and the Effects of Matrix Composition and Occupancy on Phylogenomic Reconstruction. *Systematic Biology*, 65(5), 871–889. <https://doi.org/10.1093/SYSBIO/SYW041>
- Fernández, R., & Gabaldón, T. (2020). Gene gain and loss across the metazoan tree of life. *Nature Ecology & Evolution*, 4, 524–533. <https://doi.org/10.1038/s41559-019-1069-x>
- Fernández, R., Laumer, C. E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P. P., ... Giribet, G. (2014). Evaluating Topological Conflict in Centipede Phylogeny Using Transcriptomic Data Sets. *Molecular Biology and Evolution*, 31(6), 1500–1513. <https://doi.org/10.1093/MOLBEV/MSU108>
- Fernández, R., Sharma, P. P., Tourinho, A. L., & Giribet, G. (2017). The Opiliones tree of life: shedding light on harvestmen relationships through transcriptomics. *Proceedings of the Royal Society B: Biological Sciences*, 284(1849). <https://doi.org/10.1098/RSPB.2016.2340>
- Filipe, A. F., Araújo, M. B., Doadrio, I., Angermeier, P. L., & Collares-Pereira, M. J. (2009). Biogeography of Iberian freshwater fishes revisited: The roles of historical versus contemporary constraints. *Journal of Biogeography*, 36(11), 2096–2110. <https://doi.org/10.1111/j.1365-2699.2009.02154.x>
- Foley, S., Lüddecke, T., Cheng, D. Q., Krehenwinkel, H., Künzel, S., Longhorn, S. J., ... Piel, W. H. (2019). Tarantula phylogenomics: A robust phylogeny of deep theraphosid clades inferred from transcriptome data sheds light on the prickly issue of urticating setae evolution. *Molecular Phylogenetics and Evolution*, 140, 106573. <https://doi.org/10.1016/J.YMPEV.2019.106573>

- Fu, L., Niu, B., Zhu, Z., Wu, S., & Weizhong, L. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, 29(7), 644. <https://doi.org/10.1038/NBT.1883>
- Guijarro-Clarke, C., Holland, P. W. H., & Paps, J. (2020). Widespread patterns of gene loss in the evolution of the animal kingdom. *Nature Ecology & Evolution*, 4(4), 519–523. <https://doi.org/10.1038/s41559-020-1129-2>
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Haas, B., & Papanicolaou, A. (2019). *TransDecoder 5.5.0*. Retrieved from <https://github.com/TransDecoder/TransDecoder/wiki>
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, 41, 95–98.
- Harrath, A. H., Sluys, R., Merzoug, D., Yacoubikhebiza, M., Alwasel, S., & Riutort, M. (2012). Freshwater planarians (Platyhelminthes, Tricladida) from the Palearctic section of the African continent: New records, with the description of a new species. *Zootaxa*, 15(3182), 1–15. <https://doi.org/10.11646/zootaxa.3182.1.1>
- Hasegawa, M., Kishino, H., & Yano, T. aki. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 1985 22:2, 22(2), 160–174. <https://doi.org/10.1007/BF02101694>
- Hoang, D. T., Chernomor, O., Haeseler, A. von, Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.*, 35, 518–522. <https://doi.org/10.1093/molbev/msx281>
- Holder, M., & Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*, 4(4), 275–284. <https://doi.org/10.1038/NRG1044>

- Jukes, T. H., & Cantor, C. R. (1969). Evolution of Protein Molecules. *Mammalian Protein Metabolism*, 21–132. <https://doi.org/10.1016/B978-1-4832-3211-9.50009-7>
- Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics*, 26(13), 1669–1670. <https://doi.org/10.1093/bioinformatics/btq243>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kobayashi, K., Maezawa, T., Nakagawa, H., & Hoshi, M. (2012). Existence of Two Sexual Races in the Planarian Species Switching between Asexual and Sexual Reproduction. *Zoological Science*, 29(4), 265. <https://doi.org/10.2108/zsj.29.265>
- Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. *F1000Research*, 6(1287). <https://doi.org/10.12688/f1000research.12232.1>
- Lanfear, R., Hua, X., & Warren, D. L. (2016). Estimating the effective sample size of tree topologies from Bayesian phylogenetic analyses. *Genome Biology and Evolution*, 8(8), 2319–2332. <https://doi.org/10.1093/gbe/evw171>
- Lartillot, N., & Philippe, H. (2004). A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. *Molecular Biology and Evolution*, 21(6), 1095–1109. <https://doi.org/10.1093/molbev/msh112>
- Lázaro, E. M., Sluys, R., Pala, M., Stochino, G. A., Baguñà, J., & Riutort, M. (2009). Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes,

- Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution*, 52, 835–845. <https://doi.org/10.1016/j.ympev.2009.04.022>
- Lázaro, E. M., & Riutort, M. (2013). *Dugesia sicula* (Platyhelminthes, Tricladida): the colonizing success of an asexual Planarian. *BMC Evolutionary Biology*, 13(1), 268. <https://doi.org/10.1186/1471-2148-13-268>
- Le, S. Q., & Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, 25(7), 1307–1320. <https://doi.org/10.1093/MOLBEV/MSN067>
- Lemer, S., Bieler, R., & Giribet, G. (2019). Resolving the relationships of clams and cockles: Dense transcriptome sampling drastically improves the bivalve tree of life. *Proceedings of the Royal Society B: Biological Sciences*, 286(1896). <https://doi.org/10.1098/rspb.2018.2684>
- Leria, L., Riutort, M., Romero, R., Ferrer, X., & Vila-Farré, M. Microplate tectonics and the environment as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography*, *In Press*.
- Leria, L., Vila-Farré, M., Álvarez-Presas, M., Sánchez-Gracia, A., Rozas, J., Sluys, R., & Riutort, M. (2020). Cryptic species delineation in freshwater planarians of the genus *Dugesia* (Platyhelminthes, Tricladida): Extreme intraindividual genetic diversity, morphological stasis, and karyological variability. *Molecular Phylogenetics and Evolution*, 143, 106496. <https://doi.org/10.1016/j.ympev.2019.05.010>
- Leria, L., Vila-Farré, M., Solà, E., & Riutort, M. (2019). Outstanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. *BMC Evolutionary Biology*, 19(1). <https://doi.org/10.1186/s12862-019-1440-1>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>

- Li, Y., Shen, X. X., Evans, B., Dunn, C. W., & Rokas, A. (2021). Rooting the Animal Tree of Life. *Molecular Biology and Evolution*, 38(10), 4322–4333. <https://doi.org/10.1093/molbev/msab170>
- Liu, L., Anderson, C., Pearl, D., & Edwards, S. V. (2019). Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model. In M. Anisimova (Ed.), *Evolutionary Genomics. Methods in molecular biology* (Vol. 1910, pp. 211–239). https://doi.org/10.1007/978-1-4939-9074-0_7
- Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, 38(10), 4647–4654. <https://doi.org/10.1093/molbev/msab199>
- Martinez Rica, J. P., & Montserrat Recoder, P. (1990). Biogeographic features of the Pyrenean range. *Mountain Research and Development*, 10(3), 235–240.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Haeseler, A. von, & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, 37, 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Mirarab, S., Nakhleh, L., & Warnow, T. (2021). Multispecies Coalescent: Theory and Applications in Phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 52, 247–268. <https://doi.org/10.1146/annurev-ecolsys-012121-095340>
- Nei, M., & Kumar, S. (2000). *Molecular Evolution and Phylogenetics* (M. Nei & S. Kumar, Eds.). New York: Oxford University Press, Inc.
- Ninot, J. M., Carrillo, E., & Ferré, A. (2017). The Pyrenees BT - *The Vegetation of the Iberian Peninsula: Volume 1* (J. Loidi, Ed.). https://doi.org/10.1007/978-3-319-54784-8_8
- Nishimura, O., Hosoda, K., Kawaguchi, E., Yazawa, S., Hayashi, T., Inoue, T., ... Agata, K. (2015). Unusually Large Number of Mutations in Asexually Reproducing Clonal Planarian *Dugesia japonica*. *PLOS ONE*, 10(11), e0143525. <https://doi.org/10.1371/JOURNAL.PONE.0143525>

- Phillips, A. J. L., Hyde, K. D., Alves, A., & Liu, J. K. (Jack). (2018). Families in Botryosphaerales: a phylogenetic, morphological and evolutionary perspective. *Fungal Diversity*, 94(1), 1–22. <https://doi.org/10.1007/S13225-018-0416-6>
- Quang, L. S., Gascuel, O., & Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*, 24(20), 2317–2323. <https://doi.org/10.1093/bioinformatics/btn445>
- Rambaut, A., & Drummond, A. J. (2007). *Tracer v1.4*. Available at: beast.bio.ed.ac.uk/Tracer.
- Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223. <https://doi.org/10.1111/J.2041-210X.2011.00169.X>
- Riutort, M., Álvarez-Presas, M., Lázaro, E., Solà, E., & Paps, J. (2012). Evolutionary history of the Tricladida and the Platyhelminthes: an up-to-date phylogenetic and systematic account. *Int. J. Dev. Biol.*, 56, 5–17. <https://doi.org/10.1387/ijdb.113441mr>
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539–542. <https://doi.org/10.1093/sysbio/sys029>
- Rosenbaum, G., Lister, G. S., & Duboz, C. (2002). Reconstruction of the tectonic evolution of the Western Mediterranean since the Oligocene. *Journal of the Virtual Explorer*, 8, 107–130. <https://doi.org/10.3809/jvirtex.2002.00053>
- Schrempf, D., Lartillot, N., & Szöllősi, G. (2020). Scalable Empirical Mixture Models That Account for Across-Site Compositional Heterogeneity. *Molecular Biology and Evolution*, 37(12), 3616–3631. <https://doi.org/10.1093/molbev/msaa145>
- Schwander, T., Henry, L., & Crespi, B. J. (2011). Molecular evidence for ancient asexuality in timema stick insects. *Current Biology*, 21(13), 1129–1134. <https://doi.org/10.1016/j.cub.2011.05.026>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with

- single copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/BIOINFORMATICS/BTV351>
- Sluys, R., Solà, E., Gritzalis, K., Vila-farré, M., Mateos, E., & Riutort, M. (2013). Integrative delineation of species of Mediterranean freshwater planarians (Platyhelminthes: Tricladida: Dugesiidae). *Zoological Journal of the Linnean Society*, 169, 523–547. <https://doi.org/10.1111/zoj.12077>
- Solà, E., Leria, L., Stocchino, G. A., Bagherzadeh, R., Balke, M., Daniels, S. R., ... Riutort, M. Three dispersal routes out of Africa: The puzzling biogeographical history of *Dugesia* freshwater planarians. *Journal of Biogeography*, *In Press*.
- Solà, E., Sluys, R., Gritzalis, K., & Riutort, M. (2013). Fluvial basin history in the northeastern Mediterranean region underlies dispersal and speciation patterns in the genus *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution*, 66(3), 877–888. <https://doi.org/10.1016/j.ympev.2012.11.010>
- Stocchino, G. A., & Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology*, 80(3), 319–328. <https://doi.org/10.1080/11250003.2013.822025>
- Stocchino, Giacinta Angela, Sluys, R., & Manconi, R. (2012). A new species of *Dugesia* (Platyhelminthes, Tricladida, Dugesiidae) from the Afromontane forest in South Africa, with an overview of freshwater planarians from the African continent. *Zootaxa*, 58(3551), 43–58. <https://doi.org/10.11646/zootaxa.3551.1.3>
- Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, 17, 57–86.
- Team, R. C. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.r-project.org/>
- Vergés, J., Millán, H., Roca, E., Muñoz, J. A., Marzo, M., Cirés, J., ... Cloetingh, S. (1995). Eastern Pyrenees and related foreland basins: pre-, syn- and post-collisional crustal-scale cross-sections. *Marine and Petroleum Geology*, 12(8), 903–915. [https://doi.org/10.1016/0264-8172\(95\)98854-X](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., Gehrman, T., Stielow, B., Eberhardt, U., ... Verkley, G. J. M. (2019). Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. *Studies in Mycology*, 92, 135–154. <https://doi.org/10.1016/j.simyco.2018.05.001>
- Wang, H. C., Minh, B. Q., Susko, S., & Roger, A. J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.*, 67, 216–235. <https://doi.org/10.1093/sysbio/syx068>
- Zhang, C., Scornavacca, C., Molloy, E. K., & Mirarab, S. (2020). ASTRAL-pro: Quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution*, 37(11), 3292–3307. <https://doi.org/10.1093/molbev/msaa139>

Chapter IV

Section 2

Genomic footprints of ancient fissiparity in the evolutionary history of freshwater planarians

LISANDRA BENÍTEZ-ÁLVAREZ, LAIA LERIA, ROSA FERNÁNDEZ,
IGNACIO TENAGUILLO, AND MARTA RIUTORT

Genomic footprints of ancient fissiparity in the evolutionary history of freshwater planarians

Lisandra Benítez-Álvarez¹, Laia Leria¹, Rosa Fernández², Ignacio Tenaguillo¹, and Marta Riutort^{1, *}

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

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

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

Abstract

Reproductive strategies shape the genetic background of populations. Recombination leads to increase the genetic variability and the higher efficiency of purifying selection, while evolutionary disadvantages have always been attributed to asexuality. However, long term strict asexual lineages defy this rule. Furthermore, except for the typical consequences of asexuality, little is known about the genetic footprint of long term fissiparity. Freshwater planarians are recognized for their outstanding regeneration capability, which allows the group to reproduce asexually by fission. The genus *Dugesia* shows a great diversity of reproductive strategies, with species being strictly sexual, strictly fissiparous or combining both reproductive modes. *Dugesia* species from the Western Mediterranean have been recently studied from a biogeographic point of view using transcriptomic data. In that study, asexual populations belonging to the most ancient clade showed an unexpected phylogenetic pattern putatively associated with long term asexuality. Here, we further investigate the ladder-like topology as an effect of ancient fissiparity and not as an artefact of the phylogenetic inference. Our hypothesis proposes that under long-term asexuality, individuals belonging to ancient divergent fissiparous lineages coalesce very back in time. Then, the phylogenetic trees show the descendants of divergent asexual lineages in a ladder-like pattern. Additionally, ancient asexual populations can act as reservoirs of ancestral polymorphisms, that may result in the attraction of the sexual lineages. This information is misleading for the phylogenetic inference methods that fail to infer the true species

tree topology in this case. We propose that the ladder-like pattern can be considered a genomic footprint of long term fissiparity, which can affect the phylogenetic inference and therefore, it strongly impact the species tree estimation.

Keywords: asexual reproduction, fissiparity, planarians, *Dugesia*, phylotranscriptomics

1. Introduction

The reproductive strategy of organisms determines the genetic background of their populations. Recombination during sexual reproduction is evolutionary advantageous due to the increasing of genetic variability and a higher efficiency of purifying selection, while asexual reproduction tends to decrease the genetic variability of populations, making them clonal (Webster & Hurst, 2012).

The main predicted effect of asexuality is the irreversible accumulation of deleterious mutations, known as Muller's ratchet (Felsenstein, 1974; H. Muller, 1932; Muller, 1964), which theoretically, can lead to the extinction of asexual lineages. Additionally, under an asexual scenario, recombination does not homogenize the alleles within the population and the alleles should accumulate mutations independently of each other (Birky, 1996; Welch & Meselson, 2000). This effect of asexuality is named Meselson effect and is recognized because the intraindividual alleles show high levels of divergence. Therefore, an allele might be more closely related to a homolog allele in another individual than their sister alleles within the same individual (Schwander, Henry, & Crespi, 2011). This effect is detectable in populations under long periods of time without sexual reproduction (Normark, Judson, & Moran, 2003). However, processes such as hybridization or gene duplication can lead to a Meselson effect-like allele divergence pattern (Schwander et al., 2011).

A variant of this effect, the Mosaic-Meselson effect, has been proposed to occur in agametic asexual organisms (Leria, Vila-Farré, Solà, & Riutort, 2019). Agametic asexual species, differing from those gametic, inherit part of the soma of the progenitor during each reproductive event. This situation would allow mutations not only to independently accumulate in the alleles within a cell but also among different cells, generating a Meselson effect at a mosaic level. This effect was investigated based on the information of two molecular markers using the freshwater planarian *Dugesia subtentaculata* as a model (Leria et al., 2019). However, evidence of this effect at the genomic level is still missing.

On the other hand, recombination has other direct effects on genome evolution. One of these effects is the GC-biased gene conversion (gBGC), resulting from biased incorporation of G and C nucleotides during repair of mismatches formed during meiotic homologous recombination. gBGC is thought to be responsible for the

correlation between GC content and recombination (see Webster & Hurst, 2012 for a broader review). Thus, substantial differences in the gBGC strength have been found between low and high recombining regions (Capra, Hubisz, Kostka, Pollard, & Siepel, 2013; Duret & Galtier, 2009; Galtier, 2021).

The percent of GC at the third codon position (GC3%) has been used to characterize the genomic GC-content in different phylogenetic groups, especially the mammals (Romiguier & Roux, 2017). In *Timema* species (stick insects) the GC3% has been used to study the effect of long-term asexuality on gBGC strength. Using transcriptomic data, these authors analysed the GC3% of pairs of sexual/asexual species with different divergence times. They found higher GC3 content and higher variance among genes from sexual species than asexual species within each pair. In addition, the proportion of genes in which GC3% was higher in the sexual than in the asexual species was greater in older pairs than in youngest pairs (Bast et al., 2018).

Freshwater planarians are broadly recognized by its outstanding regeneration capability. Genera such as *Schmidtea* and *Dugesia* (family Dugesidae) lead the list of favourite model organisms in regeneration studies (Rink, 2013). This regeneration capability allows the group to reproduce asexually by fission, which can be alternated in different combinations with the sexual mode (Stocchino & Manconi, 2013). Sexual reproduction in planarians occurs by cross fertilization between two hermaphrodite individuals. The asexual reproduction by fission happens when one individual divides and each part of the body regenerates the missing structures (Reddien & Alvarado, 2004). Generally, asexual individuals do not develop the reproductive systems and although the sexual ones maintain the regenerative capabilities in response to injuries, they are not able to regenerate the entire body so efficiently as in the case of asexual individuals (Saló et al., 2009).

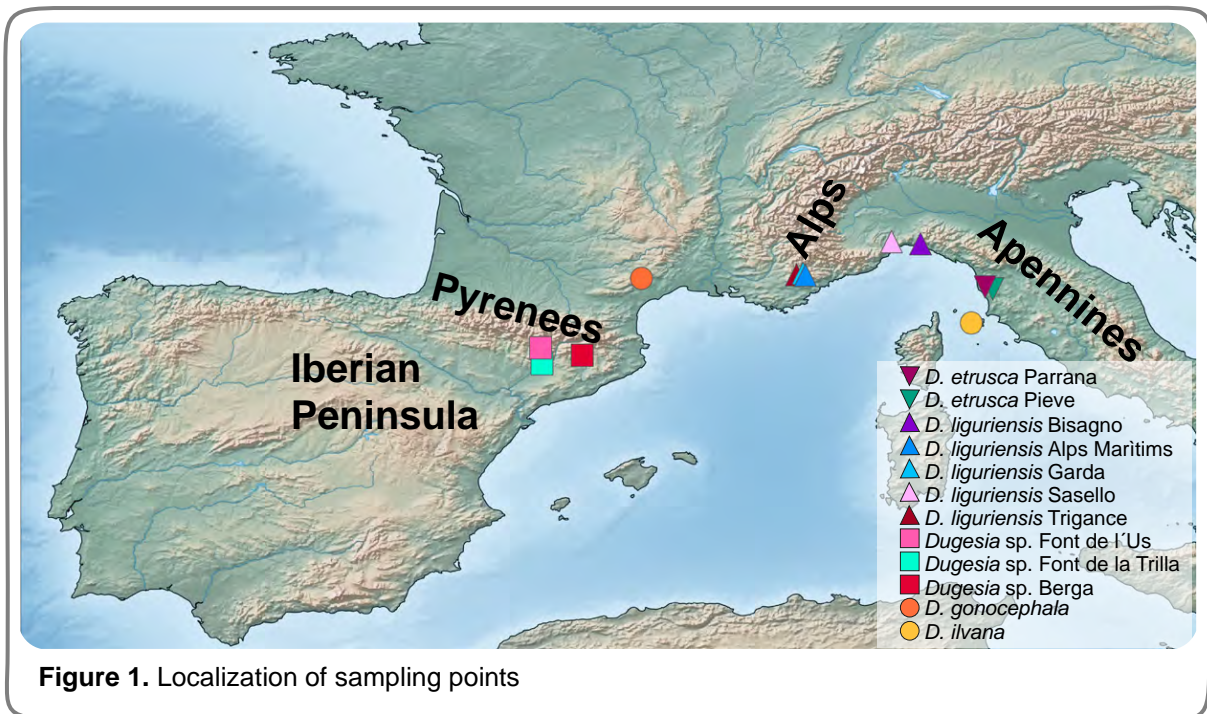
The genus *Dugesia* constitutes one of the most diverse groups of freshwater planarians in the Palearctic region (Solà et al., *in press*). Moreover, *Dugesia* species show a great diversity of reproductive strategies, with species being strictly sexual, strictly fissiparous or combining both reproductive modes (Leria et al., 2019; Stocchino & Manconi, 2013). Among all *Dugesia* species, those from the Western Mediterranean region have been recently studied from a biogeographic point of view using transcriptomic data (Chapter IV.1). The Western Mediterranean species are divided into three main clades: the Iberia-Apennines-Alps-plus clade, the Corsica-Sardinia

clade and the Iberia-Africa clade. Interestingly, in Chapter IV.1 was found out that the species from the Iberia-Apennines-Alps-plus region show an unexpected phylogenetic pattern, putatively associated to a long term asexuality.

The Iberia-Apennines-Alps-plus clade includes three populations from the Iberian Peninsula still not assigned to a species, *D. etrusca* from the Apennines region, *D. liguriensis* from the Alps, *D. ilvana* from Elba island, and *D. gonocephala* distributed in almost all continental Europe (Fig. 1). One of the objectives of Chapter IV.1 was the species assignation of the Iberian populations from Berga, Font de l'Us and Font de La Trilla to *D. etrusca* or *D. liguriensis* species, but it was not possible because an unexpected topology and several controversial nodes in the phylogenies (Fig. S1). Even though the samples from Font de l'Us and Font de La Trilla could had been assigned to *D. etrusca* with certain doubts, the assignment of samples from Berga was impossible due to their incongruent groupings in the trees obtained with different methods. Therefore, the clades including the asexual individuals are referred as *sensu lato* (s.l) species. In addition, the unexpected topology consisting of a ladder-like pattern of samples from Font de La Trilla and Berga in the trees (Fig. S1) raised suspicions on fissiparity affecting the evolutionary history of these populations. The authors outlined the idea that ancient asexuality could be underlying the odd pattern observed, since the Iberians populations constitute the only extant asexual lineages in the Iberia-Apennines-Alps-plus clade. Moreover, in that study the ladder-like pattern was shown only in the Iberia-Apennines-Alps-plus clade, but not in other asexual lineages from the Western Mediterranean, namely *D. subtentaculata* populations. In consequence, the authors proposed the idea that the populations from Berga, Font de la Trilla and Font de l'Us are the remnants of a very ancient asexual lineage, while the other extant asexual lineages belonging to *Dugesia* from Western Mediterranean are more recent and, moreover, it is known they combine fissiparity with occasional sex (Leria et al., 2019).

The main objective of the present study is to validate the ladder-like topology as an effect of ancient asexuality and not an artifact of the phylogenetic inference. For that, we first validated the methodology of ortholog search using a different ortholog inference method and also a set of paralogs. Second, we used recognized effects of the lack of recombination to demonstrate that these populations have maintained asexuality for a long time without alternating with sexual stages. Finally, we analysed

how the inclusion of ancient asexual lineages can affect the inference of the phylogenetic relationships with closest groups.



2. Methods

2.1. RNA data acquisition and pre-processing

To carry out the present study, we used data previously processed in Chapter IV.1 (Table S1). This data was obtained from RNAseq sequencing using TruSeq libraries and cured to perform phylogenetic analyses.

In Chapter IV.1 were established three groups of samples to perform different ortholog searches. A first group formed by 82 samples belonging to 12 taxa from the Western Mediterranean and 2 from Eastern Mediterranean clade (outgroup) was used to obtain 717 Single Copy orthologs (SC). The second group of samples was named the subtentaculata group and included 20 samples belonging to *D. subtentaculata* and 3 samples belonging to *D. vilafarrei* (outgroup) (Table S1). Finally, the etrusca-liguriensis group included 36 samples; 5 samples belonging to *D. etrusca sensu stricto* (s.s) from The Tuscany, 12 samples belong to *D. liguriensis* (s.s) from the Ligurian region, 3 belonging to *D. ilvana* from Elba Island, 3 belonging to *D. gonocephala* from Montpellier (used as outgroup), and the 13 samples from The Iberian Peninsula,

whose classification is uncertain (Table S1; Fig. 1). The ortholog search in this sample group yielded 1984 SC. For a better understanding in the text, here we defined the two groups of orthologs as WM_717 for the ortholog search using all samples from Western Mediterranean, and etru-ligu_1984 for the search using *etrusca-liguriensis* samples group.

For the present work, we retained the sequences of the two groups of SC mentioned above. In the case of the WM_717, we used the nucleotide sequences belonging to *etrusca-liguriensis* and *subtentaculata* sample groups. Further, we used the nucleotide and protein sequences of the etru-ligu_1984.

In addition to the SC, here we extracted a sequence set of paralog candidate genes. We selected orthogroups with more than one sequence per sample, as paralog candidates based on OrthoFinder output. Using the Orthogroups.GeneCount.tsv file, the custom python script `select_OG_all_sp.py`, and simple bash commands, we generated a list of paralogs candidates (par_158) and extracted their sequences.

2.2. Ortholog search with OMA

A new ortholog search with OMA (Altenhoff et al., 2019) was performed using the same set of longest isoforms previously analyzed in Chapter IV.1 with OrthoFinder. The protein sequences of the SC were extracted from the OMA output using simple bash commands and the nucleotide sequences were extracted from Transdecoder output as is described in Chapter IV.1.

2.3. RNA datasets

Dataset 1 includes all samples belonging to *etrusca-liguriensis* samples group (Table S1) and the nucleotide sequences of 1984 SC shared by them. To obtain the datasets 2 and 3 we removed the samples from Berga, and all samples from Iberian Peninsula respectively from the dataset 1, keeping the same number of SC (Table 1).

The dataset 4 includes 1984 SC and 158 paralogs, while the dataset 5 only includes the paralogs. For the viability of the work, we avoided the step of nucleotide sequence extraction, and used protein data to build these datasets.

The dataset 6 was built using 717 SC and two groups of samples: the *etrusca-liguriensis* samples group, and the *subtentaculata* samples group (Table S1).

Table 1. Datasets used in the study, and analyses performed

Datasets	Composition	Number of samples	Data Type	Species Tree (Astral-pro -t2)	Species Tree (Astral-pro default)	DensiTree	ML tree	Topology test	GC3% content
Dataset 1	etrusca-iguriensis samples group / 1984 SC	36	Nucleotide	X		X			
Dataset 2	etrusca-iguriensis samples group without samples from Berga / 1984 SC	31	Nucleotide	X		X			
Dataset 3	etrusca-iguriensis samples group without samples from Iberian Peninsula / 1984 SC	23	Nucleotide	X	X	X	X	X	
Dataset 4	etrusca-iguriensis samples group / 1984 SC +158 paralogs	36	Protein		X				
Dataset 5	etrusca-iguriensis samples group / 158 paralogs	36	Protein		X				
Dataset 6	etrusca-iguriensis + subtentaculata samples groups / 717 SC	59	Nucleotide						X
Dataset 7	etrusca-iguriensis samples group / 2570 SC from OMA	36	Protein				X		
Dataset 8	etrusca-iguriensis samples group / 2570 SC from OMA	36	Nucleotide				X		

Finally, two datasets were obtained using the SC discovered by OMA; the dataset 7 for protein sequences and the dataset 8 for nucleotide ones.

In summary, 8 datasets were built from RNAseq data combining three groups of SC (WM_717, etru-ligu_1984, and OMA_2570), one group of paralogs (par_158), and different samples composition (Table 1); all of them with 100% of gene occupancy.

2.4. Data analyses

2.4.1. Maximum Likelihood trees

Maximum Likelihood (ML) trees were obtained in IQ-TREE (Wang, Minh, Susko, & Roger, 2018). For nucleotide data, we used three substitution models as components; JC (Jukes & Cantor, 1969), HKY (Hasegawa, Kishino, & Yano, 1985), and GTR (Tavaré, 1986), with four Gamma categories (+G4), and 1,000,000 ultrafast bootstrap replicates. For the protein data, we obtained a starting tree from non-partitioned data using MFP (-m option) and 10,000 replicates of ultrafast bootstraps. After, we ran the mixture model analysis with the following parameters: LG model (Le & Gascuel, 2008) with 20 categories (C20), Gamma rate heterogeneity calculation (+G), site-specific frequency profile inference (+F), and 1,000,000 ultrafast bootstrap replicates. ML trees were obtained from datasets 3, 7, and 8 taking into account the nature of the data to follow one or another strategy (Table 1).

2.4.2. Multispecies Coalescent Model trees

We used the Multispecies Coalescent Model (MSC) approach implemented in ASTRAL-pro (Zhang, Scornavacca, Molloy, & Mirarab, 2020) to obtain the species tree based on individual gene trees, leveraging both single copy genes and genes with paralogs. The analyses were performed using -t2 option to recover the full branch annotation of the trees obtained from datasets 1-3. The obtained information was summarized in pie charts using the AstralPlane package (Hutter, 2020) in R. Species trees from datasets 4-5 were obtained with default options.

The individual gene trees were obtained in IQ-TREE using the parameters described for ML trees, but using 10,000 ultrafast bootstrap replicates for nucleotide data, and the model LG+G+F without bootstrap replicates for protein data. The visualization of individual trees in DensiTree (Bouckaert, 2010; Bouckaert & Heled, 2014) was performed for datasets 1-3, just as is described in Benítez-Álvarez et al (in preparation).

2.4.3. Topology test

A topology test was carried on with dataset 3 using IQ-TREE. We tested the topology obtained from ML analysis and the alternative one regarding the position of *D. ilvana* in the tree. The same model (MIX{JC,HKY,GTR}+G4) described previously for nucleotide data was used. Through the -zb and -au options, several topology tests were performed: the Bootstrap proportion (BP), KH (Kishino & Hasegawa, 1989), SH (Shimodaira & Hasegawa, 1999), and expected likelihood weights (ELW) (Strimmer & Rambaut, 2002), and the approximately unbiased (AU) test (Shimodaira, 2002). All tests performed 10,000 resamplings using the RELL method (Kishino, Miyata, & Hasegawa, 1990).

2.4.4. GC content calculation

GC3% was calculated for dataset 6 (Table 1). We used MACSE (Ranwez, Harispe, Delsuc, & Douzery, 2011) to align the nucleotide coding regions without losing the codon structure. The third positions were extracted using the tool extract-codon-alignment (https://github.com/linzhi2013/extract_codon_alignment) and the percent of GC by length calculated with the fx2tab module from seqkit program (Shen, Le, Li, & Hu, 2016). A tabulate file (samples by rows and genes by columns) was analysed in R

(Team, 2021) to obtain and graph the average (AVE) and standard deviation (STD) of GC3% content by sample.

2.4.5. **Dunnuc12 validation**

To validate Dunuc12 as a marker for intraindividual genetic diversity studies, we extracted the exonic region of this marker from all cloned sequences and blasted (Camacho et al., 2009) them against the corresponding transcriptome assembly. The transcripts were identified through blasting against the NCBI nucleotide database. The coverage of this region in every transcriptome was obtained using samtools (Li et al., 2009) and visualized in IGV (Robinson, Thorvaldsdóttir, Wenger, Zehir, & Mesirov, 2017). We looked for the identified transcripts inside the WM_717 and the etru_ligu_1984 SC groups. Finally, ML trees were obtained from the identified orthogroups (For a more detailed description of the methodology used see Text S1).

2.4.6. **Intraindividual genetic diversity**

DNA of eight samples (DspBerga_2, DspBerga_3, DspBerga_5, DspTrilla_3, DspTrilla_5, DspTrilla_6, DetruPie_2, and DliguGarda_1) were used to amplify a region of approximately 680pb of the Dunuc12 marker, a fragment of TMED9 (nuclear gene Transmembrane p24 trafficking protein 9). The new specific primers D12LigEt-F2 (CGGCAAATCTACAAAATAT) and D12LigEt-R1 (CATCTTGAAATTCAAACAG), were used at 25 μ M with MgCl₂ (25 mM), dNTPs (0.5 mM), 0.25 U of Go Taq® DNA polymerase enzyme (Promega Madison, Wisconsin, USA), and the Taq buffer (5X) in a final reaction volume of 25 μ l. The amplification conditions were as follows: 1) 2' at 94 °C, 2) 45" at 94 °C, 3) 50" at 54 °C, 4) 50" at 72 °C, and 5) 3' at 72 °C, with steps 2, 3, and 4 ran for 35 cycles. PCR products purification, cloning procedure, and sequence analyses were performed as described in Leria et al., (2019).

A set of 9 alignments (690 pb) of Dunuc12 cloned sequences were built to construct the haplotype networks; one alignment for every individual (8-14 sequences per individual), plus the alignment including all sequences (89) obtained from all individuals (Table S2). The networks were constructed using the median-joining method (Bandelt, Forster, & Röhl, 1999). In addition, the haplotype diversity (Hd), and the nucleotide diversity (Pi) were calculated in the DnaSP software from a total alignment length of 690 positions (Librado & Rozas, 2009).

To compare the results obtained in asexual populations from Iberia-Apennines-Alps-plus clade with the asexual populations belonging to *D. subtentaculata* clade, we downloaded the data of individuals from Estella (3), Santa Fe (3), and Truchas (3) (Leria et al., 2019) from DRYAD database. We used an alignment of 691 positions including 129 cloned sequences of Dunuc12 from asexual individuals to recalculate the intraindividual diversity parameters in *D. subtentaculata*.

Additionally, proportion of Non-synonymous (Ka) and Synonymous (Ks) mutations were calculated from the same coding region (120 pb) in both *D. subtentaculata* and *etrusca-liguriensis* lineages.

2.4.7. Genetic distance calculation

Using the dataset 1 (Table 1), we calculated the genetic distance by pair using the TN93 model (Tamura & Nei, 1993) in the `dist.dna` function from `ape` package (Paradis & Schliep, 2019) in R.

3. Results

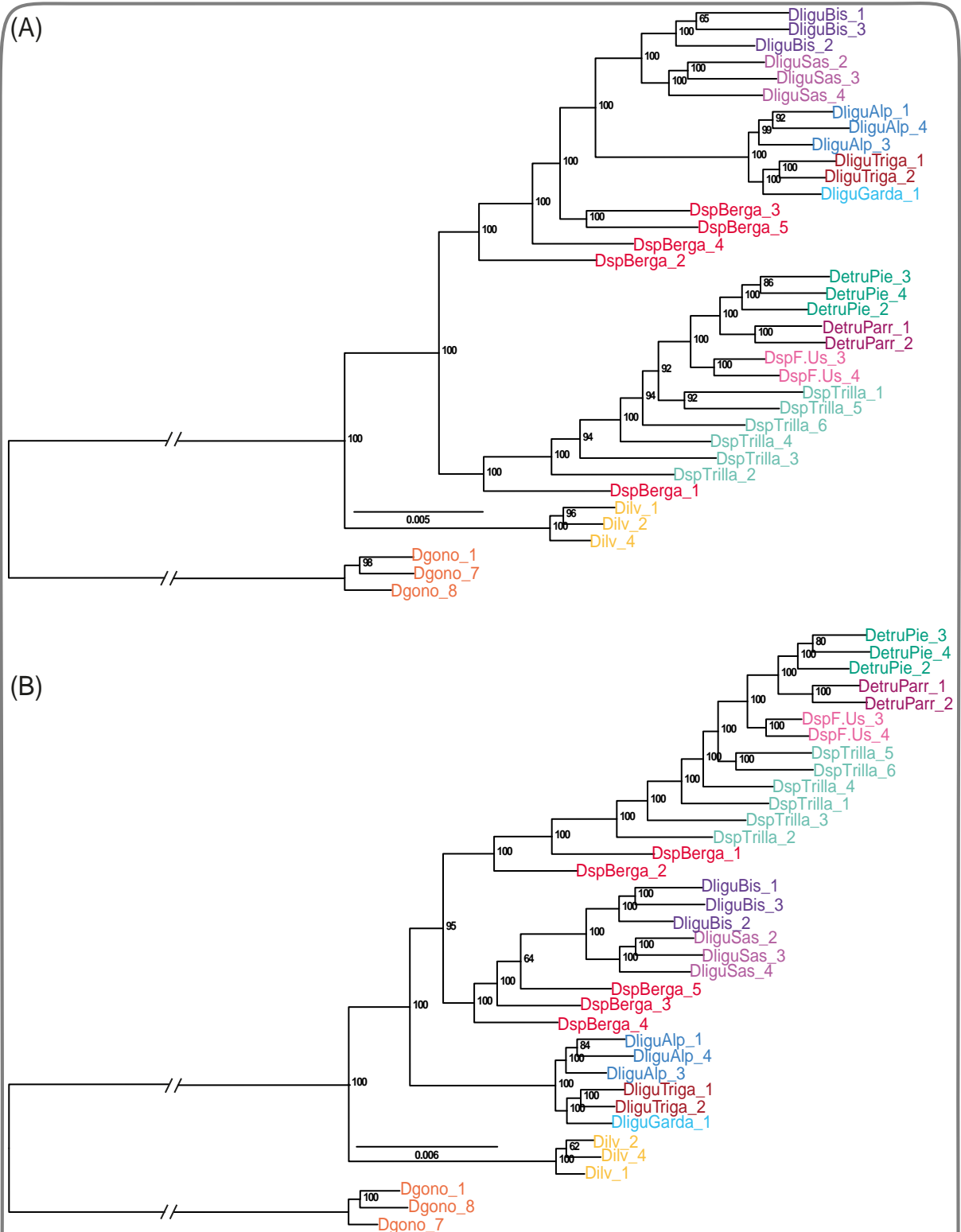
3.1. Validating the ladder-like topology

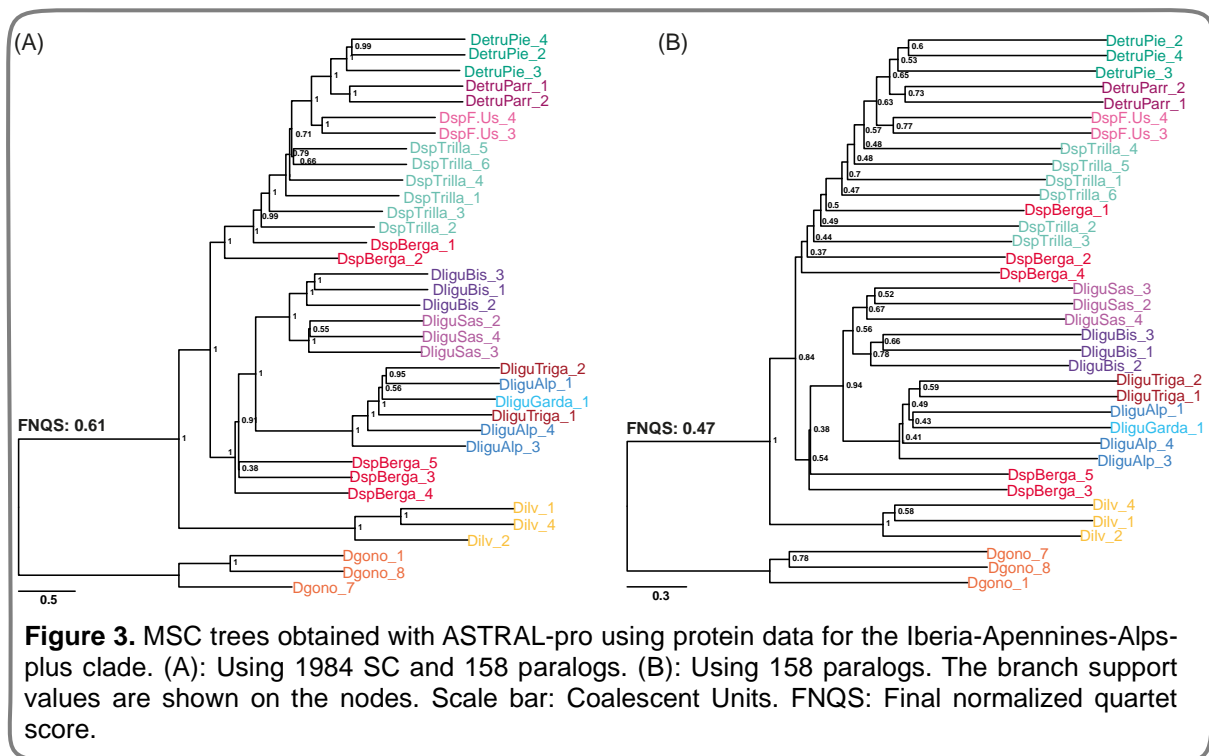
3.1.1. Ortholog search with OMA

A total of 548,739 orthogroups were obtained with OMA, of which 2,570 were defined as single copy (only one sequence per sample with all samples present). The ML trees obtained with the protein and nucleotide sequences of OMA_2570 recovered the previous topology, showing the ladder-like pattern of the samples from Berga and Font de la Trilla (Fig. 2).

3.1.2. Paralogs information

We extracted 158 paralog candidates for the Iberia-Apennines-Alps-plus clade. The individual trees from these paralogs were inspected by eye, to confirm that the gene tree showed duplication events. Taking advantage of the capability of ASTRAL-pro to handle paralogs information, we analysed our paralogs set together with the SC (Fig. 3A), and alone (Fig. 3B). These trees showed the same topology previously obtained. However, the values of branch length, branch support, and final normalized quartet score were low, especially in the tree obtained only with the 158 paralogs (Fig. 3B).





3.2. Ancient asexuality in the Iberia-Appennines-Alps-plus clade

3.2.1. GC content in third positions

The GC3% was calculated in sexual and asexual individuals belonging to the *etrusca-liguriensis* samples group and the *subtentaculata* samples group. *D. subtentaculata* was included because despite their great number of asexual populations, the ladder-like pattern has not been seen in this clade ever.

We analyzed the WM_717 SC set. The average of GC3% among all genes by sample was low and not much different between sexual and asexual samples (Fig. S2). In the case of Iberia-Appennines-Alps-plus clade, non-noticeable differentiation is shown when the standard deviation is plotted against the GC3% average. On the other hand, *D. vilafarrei* (sexual) shows a slight differentiation from the *D. subtentaculata*, whose sexual and asexual populations appear mixed. However, this graph has to be analyzed with caution, since the scale is very large and, in fact, the values are very close.

3.2.2. Dunuc12 validation

Using the transcriptomic data, we validated Dunuc12 as SC gene (Text S1). An exonic region of 120 positions (Dunuc12-exo) was identified in every cloned sequence. All the Dunuc12-exo from the same individual, matched against the same transcript in its individual transcriptome assembly. In addition, all identified transcripts (one by individual) mapped against sequences from *Dugesia* identified as TMED9 in the nucleotide database of NCBI. We analysed two groups of SC used here, finding that all the transcripts formed only one orthogroup; OG0007336 in WM_717, and OG0010797 in etru-ligu_1984. Although neither was defined as SC by OrthoFinder (see Supporting Information Text S1 for a broader explanation), the ML trees obtained from OG0007336 and OG0010797 showed the topology of a region that has not suffered any duplication event in the evolutionary history of the group.

3.2.3. Intraindividual genetic diversity

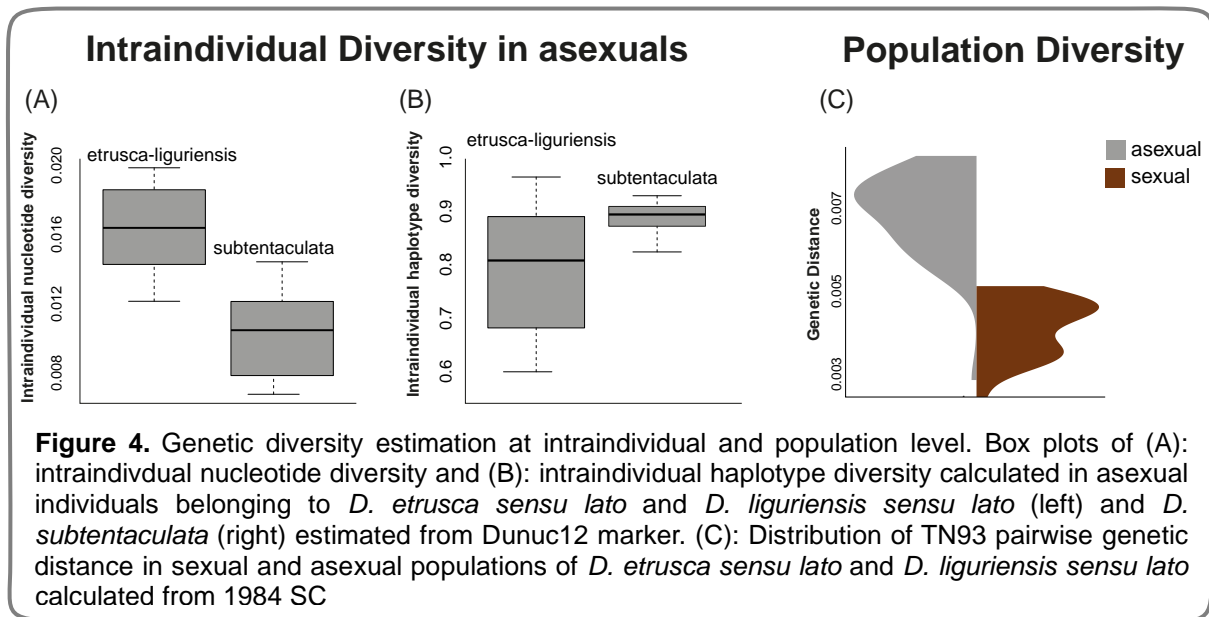
From the eighth selected individuals a total of 89 sequences of Dunuc12 were obtained (Table S2). The length of the sequences was 690bp, with a total of 138 variable sites. After the correction of the sequences eliminating the putative amplification errors, the analysis in DnaSP software revealed a total of 35 different haplotypes. For the two sexual individuals selected, one from Pieve (*D. etrusca*) and other from La Garda (*D. liguriensis*), 6 and 4 different haplotypes were found respectively, while for the 6 fissiparous individuals from Iberian Peninsula the number of intraindividual haplotypes ranged between 3 and 8. All analyzed individuals, except DspTrilla_3, presented a higher number of haplotypes than its ploidy would suggest, an indication of mosaicism (Table S2, S3).

Intraindividual haplotype networks of the asexual individuals (Fig. S3A) showed a pattern in which the haplotypes were highly divergent from one another with similar frequencies. However, the sexual individuals (Fig. S3B) showed a star-like pattern, with a predominant haplotype in the center and some closely related haplotypes at a distance of one or few mutations surrounding it.

The haplotype network with all cloned sequences (Fig. S3C) shows that sexual individuals do not share any haplotype with any other population. However, fissiparous individuals shared haplotypes among populations, forming clusters of closely related

haplotypes, but those clusters are diverging among them, a clear Meselson-effect pattern.

Regarding the intraindividual genetic diversity parameters; the nucleotide diversity was higher in asexual individuals of etrusca-liguriensis group (Fig. 4A, Table S3). However, the haplotype diversity (H_d) was higher in *D. subtentaculata* individuals than in the etrusca-liguriensis ones (Fig. 4B, Table S3).



Surprisingly, the proportion of synonymous mutations (K_s) in the exonic region (120 pb) was nule for all individuals belonging to the etrusca-liguriensis group, except for DspBerga_3 ($K_s=0.0107$), and only individuals from Berga showed non-synonymous mutations ($K_a>0$). With this data was impossible to calculate the K_a/K_s ratio and obtain information regarding the Muller's ratchet effect. On the other hand, no signals of this effect were observed in the subtentaculata lineage, since K_a values were lower than K_s (Table S3).

3.2.4. Genetic distance between individuals

A genetic distance matrix between all sexual individuals from Alps and Apennines regions, and asexual individuals from Iberian Peninsula was built using 1984 SC and the TN93 model. The genetic pair-wise distances between asexual individuals from Berga and Font de la Trilla are higher than the genetic pair-wise distances between individuals from sexual population (Fig. 4C; Fig. S4).

3.3. Effect of asexual taxa in species tree inference

We obtained the per branch quartet support and the branch support (local posterior probability) in the species tree obtained with the MSC method from datasets 1-3. Figure 5 shows the quartet support and local posterior probability for the three possible topologies around the internal branches of the Iberia-Apennines-Alps-plus clade. The red arrows indicate the most controversial nodes in the phylogeny. The values of quartet support indicate the amount of gene tree conflict around a branch. Values close to 0.33 indicate a very high level of discordance around the branch (Fig. 5).

In the case of the tree obtained from dataset 1 (Fig. 5A), two discordant nodes are shown; the node that joins samples from Font de la Trilla and Berga (nodes 49 and 60 in Supplementary Figure S5A). In addition, the final normalized quartet support of this tree is low (FNQS=0.71) and the DensiTree is very blurred.

When the samples from Berga were removed, the nodes around the samples from Font de la Trilla continued showing discordance (Fig. 5B; Fig. S5B). The individual trees show a less blurred pattern and the FNQS is slightly greater (0.77).

From the dataset 3, a different topology was obtained (Fig. 5C). In this analysis were eliminated all asexual samples from the Iberian Peninsula. Unlike datasets 1 and 2, in this case *D. ilvana* and *D. etrusca* s.s form a monophyletic group. None discordant-nodes are shown in this tree. The quartet supports and the branch supports on the internal branches are high (Fig. S5C). Furthermore, the individual trees show a more defined pattern and the FNQS is higher than in the previous topologies (0.82). In addition, the ML tree obtained from this dataset, shows the same topology and high bootstrap values (Fig. S6, Table S4).

3.3.1. Topology test

Taking into account the discordant results regarding the position of *D. ilvana*, we tested the topologies obtained with dataset 3 [Tree 1 (*D. gonocephala*, (*D. liguriensis sensu stricto*, (*D. ilvana*, *D. etrusca sensu stricto*)));] and the previously ones [Tree 2 (*D. gonocephala*, (*D. ilvana*, (*D. liguriensis sensu stricto*, *D. etrusca sensu stricto*)));]. All tests showed a significantly better adjustment to the data than to the tree 1 (Table 2), supporting the topology where *D. ilvana* is sister group of *D. etrusca sensu stricto* (Fig. 3C, Fig. S4) when the Iberian populations were removed from the analyses.

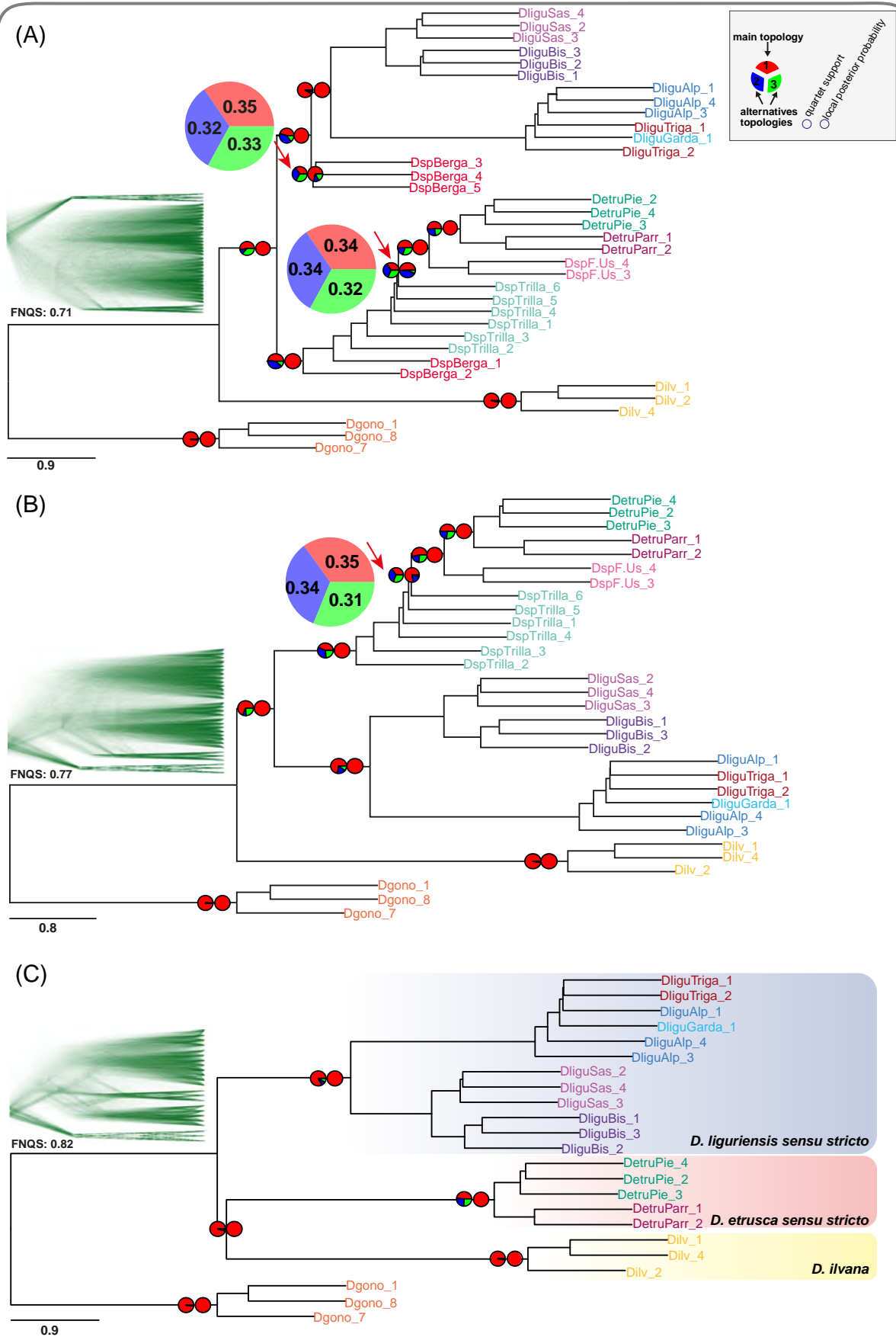


Figure 5. Species trees obtained using Multispecies Coalescent Model (MSC) from three data sets of single copy orthologs and different sample composition. *The figure caption continues in the next page*

Figure 5 cont. (A): MSC tree from data set 1 containing all samples belonging to the Iberian-Apennines-Alps-plus clade. (B): MSC tree from dataset 2 excluding the samples from Berga locality. (C): MSC tree from dataset 3 excluding all samples from the Iberian Peninsula. The per branch quartet support and the branch support (local posterior probability) values for the main and the alternative topologies are shown on the branch in pie charts. The red arrows indicate discordant topologies around a branch and the big pie charts the corresponding quartet support values. Scale bar: Coalescent Units (MCM). FNQS: Final normalized quartet score. See Supplementary Table S4 and Figure S5 for more information about the support values of the branches. The individual trees are represented in DensiTree schemes in green.

Table 2. Results of topology test between Tree 1 and Tree 2, using the Dataset 3.

Tree	LogL	deltaL	Bp-RELL	P-KH	P-SH	C-ELW	P-AU
1	4466502,48	0	1+	1+	1+	1+	1+
2	4466713,34	210,86	0-	0-	0-	-1,97E-42	-1,41E-61

deltaL: logL difference from the maximal logl in the set.
 bp-RELL: bootstrap proportion using RELL method
 p-KH: p-value of one sided Kishino-Hasegawa test
 p-SH: p-value of Shimodaira-Hasegawa test
 c-ELW: Expected Likelihood Weight
 p-AU: p-value of approximately unbiased (AU) test
 Plus signs denote the 95% confidence sets.
 Minus signs denote significant exclusion.

Tree 1

Tree 2

4. Discussion

Phylogenies describe the diversification process through bifurcated branches. Thus, dichotomous nodes with high support values are always the expected topology to validate any phylogenetic relationship. In the same way, monophyly constitutes a qualitative measure to validate taxa, and it is expected that taxonomic categories always correspond with monophyletic groups. When this pattern disappears, the first doubts hover around the methodologies used and the systematic errors associated with them.

Here, we studied an unexpected topological pattern associated to the inclusion of asexual individuals in phylogenetic analyses. In addition, we analysed how the genetic information carried by asexual taxa affect the tree inference.

4.1. The ladder-like topology is not an artifact of phylogenetic reconstruction

The species *D. etrusca s.s* and *D. liguriensis s.s* were described for the first time based on morphological characters of the reproductive apparatus, just as is common in freshwater planarians (Benazzi, 1946; De Vries, 1988). Asexual populations from the Iberian Peninsula have been reported without being possible its accurate identification to one species or another using specific molecular markers (Baguñà et al., 1999; Lázaro et al., 2009). Using transcriptomic data, a larger sampling size, and different inference methods were obtained odd topologies in the phylogenies (Chapter IV.1). The trees obtained using ML approach showed an unexpected ladder-like pattern of asexual individuals and a paraphyletic topology for *D. liguriensis s.l* (Fig. S1A). When asexual representatives from Berga population were removed from the dataset, ML tree recovered a monophyletic topology for the two species, but asexual samples grouping with *D. etrusca s.s* continued showing the ladder-like topology (Fig. S1C). On the other hand, the MSC method recovered two monophyletic groups, but it cast doubt on the assignment of Berga's samples, since different samples were placed with different species (Fig. S1B and D). Additionally, all trees showed low support values on the nodes joining the asexual samples from Font de la Trilla and Berga.

Orthologs selection can lead to incorrect topology inference, which is one of the most sensible steps in phylogenomics (Fernández et al., 2020; Kapli, Yang, & Telford, 2020). In Chapter IV.1 was used OrthoFinder (Emms & Kelly, 2019) software for ortholog search from transcriptomic data. This method infers orthogroups across multiple species based on pairwise sequence similarity between all sequences. It defines an orthogroup as the set of genes descended from a single gene in the last common ancestor of all species under consideration (Emms & Kelly, 2015). OrthoFinder currently constitutes one of the most accurate methods for orthology inference (Deutekom, Snel, & Van Dam, 2021; Emms & Kelly, 2019). To corroborate that the ladder-like pattern observed in the previous study is not a product of ortholog error, we conducted a new ortholog search using OMA, which identifies the orthologs using a reference database with high performance (Altenhoff, Schneider, Gonnet, & Dessimoz, 2011). With OMA we recovered 2,570 orthologs, several more than with OrthoFinder, a situation that has been reported in other studies comparing both software (Altenhoff et al., 2019). The ML trees obtained with the SC discovered with OMA recovered the same ladder-like topology of asexual individuals with both protein

and nucleotide sequences (Fig. 2). Also, the paraphyly of *D. liguriensis* was recovered with nucleotide data (Fig. 2B). These results corroborate the topologies obtained with the SC discovered by OrthoFinder, indicating that the ladder-like pattern is not an artifact of ortholog inference.

Additionally, it would be logical to think that an artifact produced by an incorrect orthologs estimation would affect the whole tree, showing anomalous topologies in all groups. However, this pattern has only been shown by asexual individuals belonging to the Iberia-Appenines-Alps-plus clade (Chapter IV.1).

For a second validation, we took advantage of the approaches developed to extract phylogenetic information from paralog genes. The use of paralogs with adequate approaches, specially MSC methods, have been expanded in phylogenomics as another source of phylogenetic information since obtaining single copy gene sets can become more difficult when more divergent species are included in the analyses (Hellmuth et al., 2015; Yan, Smith, Du, Hahn, & Nakhleh, 2022). The species trees inferred including the set of paralogs showed the same ladder-like topology, with a consequent decrease of support due to the use of paralogs (Fig. 3), demonstrating that this topology is as real as ancient and can be recovered even from paralog information.

In conclusion, we have demonstrated that although the ladder-like topology is weird, it is not an artifact of the methodology used. Otherwise, we propose that this topology represents the phylogenetic footprint of ancient fissiparity in these populations, which have been for a long time without recombination.

4.2. Ancient asexuality in the Iberian-Apennines-Alps-plus clade

Using different methodologies, we have shown that the ladder-like pattern is not an effect of systematic errors but the consequence of a real underlying biological process. We suspect that this pattern can be reflecting the effect of ancient asexuality in Iberian populations that belong to the Iberian-Apennines-Alps-plus clade.

Because this pattern was more obvious in the populations from Font de la Trilla and Berga, with more individuals sampled, we only refer to these populations and seem that we leave the population from Font de l'Us in the backstage. The issue with this population is that only two individuals were sampled, and they group with *D. etrusca* with high support values. In this scenario, it is impossible to know if this population has

a “standard” monophyletic grouping, or it can show the ladder-like pattern if more individuals are included in the analyses. If from Font de la Trilla and from Berga, we had only analysed individuals DspTrilla_6, DspTrilla_5, DspBerga_3, and DspBerga_4 maybe this debate wasn't happening since they constitute monophyletic groups per population (Fig. S1A-D). Moreover, the fact is that Berga and Font de la Trilla revealed this unexpected topology, even in the trees done with *Cytochrome Oxidase I* for an initial species assignment (Chapter IV.1).

We hypothesize that contrary to *D. subtentaculata* lineage, in which asexual populations have never shown a ladder-like pattern, the Iberian-Appenies-Alps-plus clade has been under long-term asexuality without recombination events. (Leria, Riutort, Romero, Ferrer, & Vila-Farré, *in press*) dates the diversification of Iberian-Appenies-Alps-plus clade around 17 Mya and the *D. subtentaculata* diversification in Iberian Peninsula approximately 1.69 Mya. In Chapter IV.1 it was proposed that the ancestor of *D. etrusca*-*D. Liguriensis* lineage could be also asexual and colonized the Iberian Peninsula through fauna corridors in the Eastern Pyrenees, early in its diversification process along the Western Mediterranean. In addition, Leria et al., (2019) propose the shift between sexual and asexual reproduction as a strategy during *D. subtentaculata* evolution.

Here we used several approaches to confirm the ancient asexuality of Berga and Font de la Trilla populations with respect to the *D. subtentaculata* lineage. First, we used the GC content in the third codon position as a recombination measure (Fig. S2). Under higher recombination rates, it is expected an increase of GC content due to the gbGC (Webster and Hurst 2012). This calculation has been used as measure of the lack of recombination in ancient asexual lineages in *Timema* stick insects (Bast et al., 2018). However, here we did not detect a considerable change between sexual and asexual individuals. The low content of GC in freshwater planarians transcriptomes has been reported before (Abril et al., 2010; Nelson Hall et al., 2021; Resch, Palakodeti, Lu, Horowitz, & Graveley, 2012). Nevertheless, although it is suspected that the high variation of GC content in Platyhelminthes can be related to some adaptive trait (Lamolle, Fontenla, Rijo, Tort, & Smircich, 2019), there is no clear explanation for this. We hypothesize that a high AT content in planarian genomes can favour the active DNA replication processes that take place in these organisms regarding the cellular turnover and regeneration, but further studies are necessary to

corroborate it. In consequence, planarians do not show the expected shift in GC content related to recombination events.

Second, we used intraindividual diversity parameters to compare the effect of asexuality in the Iberian-Appenies-Alps-plus clade and in the *D. subtentaculata* lineage. For that, we validated the Dunuc12 marker as a good marker for intraindividual genetic diversity studies. Although OrthoFinder does not classify Dunuc12 as SC, we can corroborate that Dunuc12 is a single copy gene in the *Dugesia* genome. First, all sequences are grouped only in one orthogroup in two ortholog searches; and second, the trees did not show a paralog topology (Text S1). Additionally, these trees corroborated that the coding region of this marker is not informative, and all variability is found in the intron.

Then, we analysed the Meselson effect as a measure of long-term asexuality. We obtained a higher intraindividual nucleotide diversity in asexual populations belonging to Iberian-Appennines-Alps-plus clade than in asexual populations belonging to *D. subtentaculata* lineage (Fig. 4A). This result supports the premise that Berga and Font de La Trilla populations have been evolving without recombination for a longer period of time than the asexual populations of *D. subtentaculata*.

Although these signals also can be found in asexual species originated from recent hybridization (Lunt, 2008), here we discard this option since the hybrid origin of asexual Iberian populations is biogeographically unfeasible, and a facultative sexual ancestor of sexual and asexual lineages is a most plausible hypothesis (Chapter IV.1).

In agametic clonal organisms it is common to find genetic mosaicism due to the inheritance of somatic mutations from the parents (Gill, Chao Lin, Perkins, & Wolf, 1995). Leria et al., (2019) described the Mosaic-Meselson-effect in *D. subtentaculata* by the first time, as a consequence of asexuality that combine the Meselson effect with high mosaicism levels in agametic organisms.

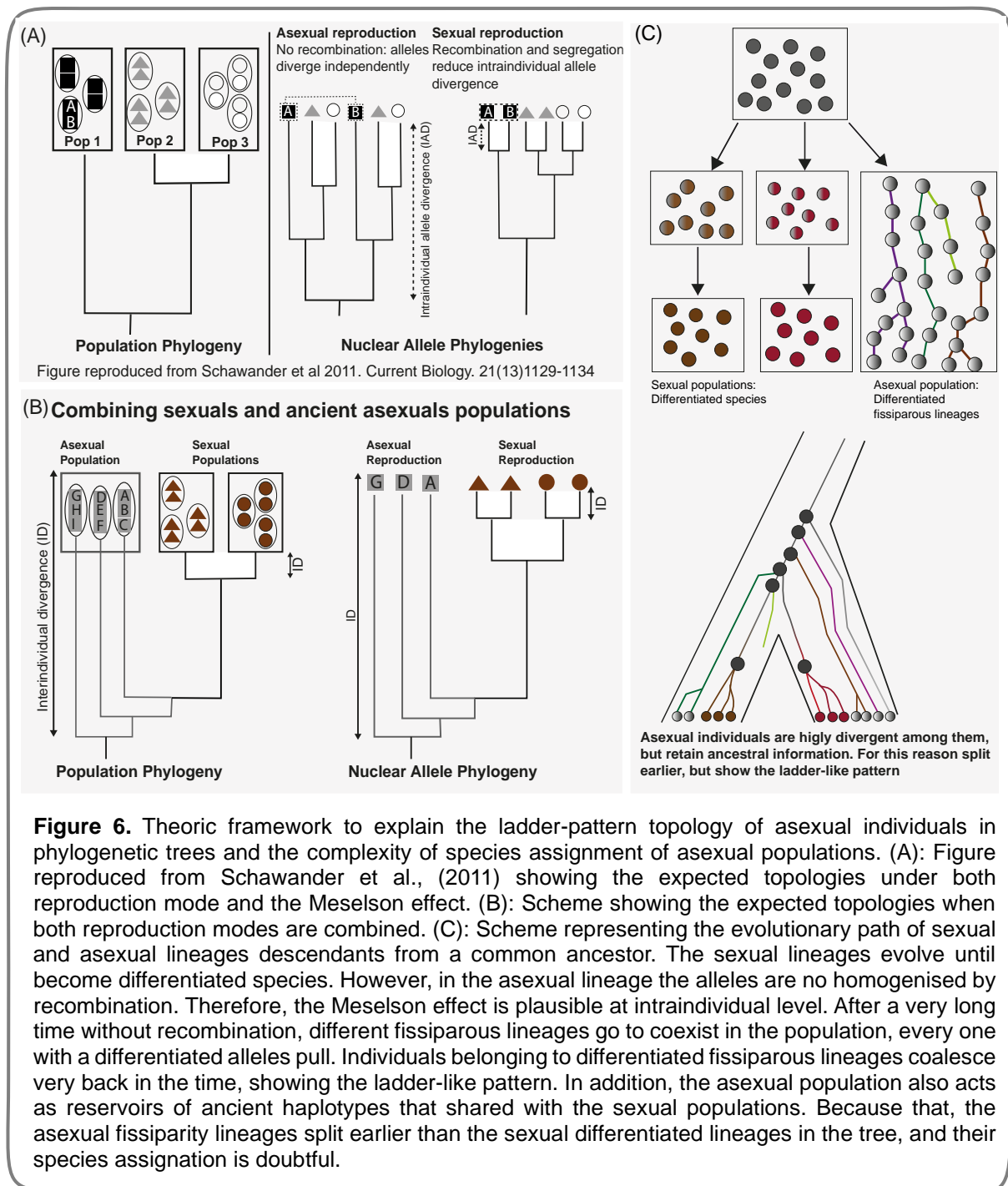
Here, we also found that asexual populations of *D. etrusca s.l* and *D. liguriensis s.l* show high levels of intraindividual haplotype diversity (Fig. 4B, Table S3), indicating that the Mosaic-Meselson effect is also occurring in these lineages. The slightly lower values of this parameter found in *D. etrusca-D.liguriensis s.l* compared to *D. subtentaculata*, likely result from a higher number of intraindividual sequences obtained in the last mentioned species (Leria et al., 2019).

Another way to show the Meselson effect is analyzing the haplotype networks. Our haplotype networks showed the typical Meselson effect pattern and are similar to those shown by Leria and collaborators for sexual and asexual individuals. Highly divergent haplotypes in asexual individuals (Fig. S3A) and star-like pattern in sexual ones (Fig. S3B). Therefore, the accentuated Meselson effect found in Berga and Font de La Trilla populations is a reliable proof of the oldest asexuality of these lineages and could explain why asexual *D. subtentaculata* populations do not show the ladder-like pattern.

Finally, our last way to test the ancient asexuality of the Iberian populations of *D. etrusca* and *D. liguriensis* was to investigate the putative accumulation of deleterious mutations in these populations (i.e., the putative existence of Müller's Ratchet). To do so, we analyzed the proportion of non-synonymous and synonymous mutations (Ka/Ks) in the same coding region of Dunuc12 in these populations and compared it with the asexual populations of *D. subtentaculata*. Surprisingly, we found that the coding region of Dunuc12 was extremely conserved in the asexual populations of *D. etrusca s.l* and *D. liguriensis s.l*, hindering the calculation of this parameter in most individuals. Nevertheless, the very few mutations found in these populations were indeed in most cases non-synonymous (Table S3), pointing to a potential existence of Müller Ratchet.

4.3. The ladder-like topology; a new genomic footprint of ancient fissiparity

Basing on the aforementioned results, here we propose the theoretic framework leading to the ladder-like phylogenetic footprint caused by ancient fissiparous reproduction. The footprint of asexuality in the evolutionary history of lineages lies on the biological characteristics of the organisms. The Meselson effect is the consequence of the lack of recombination. Under this situation the alleles in the population are not homogenized, leading to highly differentiated alleles inside the same individual, that could be more related with homologous alleles from other populations (Birky, 1996; Schwander et al., 2011; Welch & Meselson, 2000) (Fig. 6A). However, in lineages that combine both reproductive strategies, the pattern shown by sexual and asexuals is different. In this case, the intraindividual divergence (ID) of alleles in sexuals will be shorter than the ID of asexual ones, which coalesce back in time (Fig. 6B).



In the case of planarians, the fissiparity implies the expansion of clonal lineages in the population through time. After long evolutionary time without recombination, every clonal lineage will be differentiated, drawing their own genetic diversity. To demonstrate this effect under long term fissiparity, we calculated the pairwise distances based on transcriptomic data in sexual populations of *D. etrusca* and *D. liguriensis* and asexual population from Berga and Font de La Trilla. The genetic

distances, calculated from 1984 SC, between asexual individuals were higher than between sexual individuals (Fig. 4C; Fig. S4), supporting that under long term asexuality the recombination does not homogenize the pool of alleles within individuals, and neither within the population, leading to the existence of independent lineages of fissiparity.

Under the scenario where sexual and fissiparous lineages evolve from the same ancestor, the species tree will show the footprint of both reproductive strategies (Fig. 6C). While sexual lineages evolve becoming differentiated species, in asexual lineages the lack of recombination produces Meselson effect at an intraindividual level. Under longer asexuality without genetic crossing, distant fissiparous lineages accumulate high amounts of divergence between them. Therefore, individuals belonging to ancient divergent fissiparous lineages coalesce very back in the time. Then, the phylogenetic trees show the descendants of divergent ancient lineages in a ladder-like pattern. The probability to find representatives of these lineages will depend on the population size. Thus, in a random sampling it is possible to find also related individuals belonging to the same lineage, whose coalescence point is more recent in the time (Fig. 6C).

Additionally, these asexual “independent lineages” split earlier than the sexual differentiated lineages in the tree because the asexual populations also act as reservoir of ancestral polymorphisms, attracting the related sexual lineages. This ancestral relationship, conserved by asexuality masks the diversification process of the group and made unachievable the species assignation of asexual populations, since asexual lineages coalesce with those sexual ones far back in time, when the sexual lineages were not differentiated yet.

Systematics of asexual taxa constitute a challenge for taxonomist specialists (Fehrer, Krak, & Chrtek, 2009; Hörandl, 2018; Martens, Rossetti, & Baltanás, 1998), and the taxonomy and systematics of asexual lineages have been belittled due to the evolutionary constrains of asexuality (Fontaneto et al., 2007). Following the definition that establishes the species as separately evolving metapopulation lineages (De Queiroz, 2007), it could be plausible to describe the asexual populations from the Iberian Peninsula as different species. However, this group would not meet the monophyly principle to be considered a clade because the individuals are grouped in a ladder-like topology. In the case of the Berga population the species assignment is

more confused, due to the grouping with both species. Therefore, we decide to continue using *sensu stricto* as reference to sexual populations and *sensu lato* as reference to the two lineages including asexual populations. Since no morphological characters are available for systematic studies in asexual freshwater planarians, and the genetic diversity of asexual populations is shaped by the fissiparity, new theoretical frameworks are necessary to include the diversity of reproductive strategies in our taxonomic system of classification.

4.4. Ancient asexual lineages confound phylogenetic inference. A new challenge for evolutionary studies in freshwater planarians

The inclusion of asexual individuals in phylogenetic analyses also affect the relationship with closer related groups. Figure 5 shows the analysis of different datasets and how the support of species trees increases with the elimination of asexual individuals. Full dataset 1 (Fig. 5A) shows the lowest FNQS, indicating highest discordance among gene trees. Unexpectedly, when all asexual representatives are excluded, the tree shows *D. ilvana* as the sister group of *D. etrusca* s.s, instead of being sister to the clade formed by *D. etrusca* s.s and *D. ligurienses* s.s (Fig. 5C). This topology was not recovered in any analysis performed in Chapter IV, which always included representative samples of the asexual lineages in the datasets. However, Lázaro et al (2009), recovered a low supported relationship between *D. etrusca* and *D. ilvana*. Moreover, this topology is highly supported here by the MSC method (Fig. 5C; Fig. S5 C), ML method (Fig. S6) and topology tests (Table 2).

These results point to new biogeographic hypotheses to explain the diversification of the Iberia-Appennines-Alps-plus clade. Based on the results obtained in Chapter IV was proposed that the ancestor of *D. ilvana* split first from *D. etrusca*-*D. liguriensis* ancestor. However, they already pointed that the timing for this split based on the dating analysis in Leria et al (*in press*) did not support such an ancient split. Additionally, taking into account that Elba Island was formed by Appennines outcrops (Bortolotti, Fazzuoli, et al., 2001; Bortolotti, Pandeli, & Principi, 2001), it is more plausible that *D. ilvana* shared a common ancestor with the Appennines species *D. etrusca* s.s. Biogeographical studies have demonstrated that Elba Island was connected with Tuscany by a natural land bridge during the Pleistocene (Dapporto & Cini, 2007; Di Nicola & Vaccaro, 2020; Fattorini, 2009). Therefore, several animal

groups in the Island show a phylogenetic relationship with the Tuscan archipelago fauna (Fattorini, 2009 and references therein; Di Nicola & Vaccaro, 2020; Dapporto & Cini, 2007).

We suspect that the inclusion of the asexual lineages introduced ancestral polymorphisms maintained in the different fissiparous lineages shared by *D. liguriensis* and *D. etrusca*, reinforcing the ancestral relationship against the more recent splitting with *D. ilvana*. Ancestral polymorphisms can lead to incongruent phylogenetic hypothesis (Moran & Kornfield, 1993; MacEachern, McEwan, & Goddard, 2009). Moreover, the effect of ancestral shared polymorphisms in phylogeographic reconstructions based on molecular data has been demonstrated in several taxa (Boluda, Rico, Naciri, Hawksworth, & Scheidegger, 2021; Bowie, Fjeldså, Hackett, Bates, & Crowe, 2006; Du et al., 2019; B. Wang et al., 2019). Therefore, ancient asexual populations acting as reservoirs of ancestral polymorphisms will introduce noisy information that the phylogenetic inference methods incorporate as support of alternative phylogenetic hypotheses, which do not reflect the real species tree.

Summarizing, here we demonstrated that ancient asexuality in agametic organism can lead to a ladder-like topology showing the descendants of differentiated clonal lines as independent lineages that split earlier than the sexual clades in the tree. In addition, the asexual populations retain ancestral polymorphisms shared by the sexual ones, reinforcing the ancestral relationship against more recent splittings. Since the reproductive strategies shape the genetic diversity of populations and lead to differentiated genetic backgrounds between sexual and asexual lineages, the discordance between gene trees and species trees can increase and affect the phylogenetic inference. Therefore, we recommend the use of MSC methods against concatenation methods for species tree estimation with large datasets, especially in cases with putative ancient asexuals. These methods offer a framework based on Coalescent Theory that reconstruct in a better way the ancestral relationship between sexual and asexual lineages. Additionally, is highly recommendable to test different phylogenetic hypothesis with different datasets to study the effect of the inclusion of asexual representatives on the trees. Unfortunately, the inference methods and evolutionary models are designed based on sexual reproduction premises, and do not contemplate the effects of the lack of recombination or different ploidy levels. New

frameworks including these issues are necessary to reconstruct the evolutionary history of asexual lineages in the Tree of Life.

Data accessibility

RNAseq data is accessible under the BioProject accession code PRJNA797284 at the NCBI SRA database and the cloned sequences of Dunuc12 are accessible at the GenBank nucleotide database. Also, the scripts used are accessible at <https://github.com/lisy87/dugesia-transcriptome>

Acknowledgments

We are grateful to Marina Pérez for the design of Dunuc12 primers, specific for *D. etrusca* and *D. liguriensis* s.l. We thank Joan Ferrer Obiol and Julio Rozas for their valuable suggestions. We also thank our lab manager Gema Blasco for her support in all our research projects.

Authors Contributions

LBA and MR did the initial study design. LBA processed and analysed the data with the input from RF. LL and IT performed Dunuc12 cloning and analyses. LBA wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Funding

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

Supplementary Material

Supplementary tables S1, S2, and S4, are available at [Riutort's Lab site](#)

Table S3. Intraindividual diversity parameters calculated in individuals belonging to two different asexual lineages. Haplotype and nucleotide diversity were calculated from a Dnnuc12 region of 691 positions in subtentaculata lineage and 690 positions in etrusca-liguriensis lineage. The proportions of non-synonymous and synonymous mutations were calculated from the same coding region of 120 positions in both lineages.

Individual	Lineage	Haplotype Diversity (Hd)	Nucleotide Diversity (Pi)	Proportion of Non-synonymous mutations (Ka)	Proportion of Synonymous mutations (Ks)	Ka/Ks
Estella1A	subtentaculata	0,909	0,0086	0,0018	0,0281	0,0624
Estella2A	subtentaculata	0,872	0,0069	0,0016	0,0132	0,1225
Estella3	subtentaculata	0,838	0,0077	0,0028	0,0230	0,1213
SantaFe1A	subtentaculata	0,894	0,0129	0,0008	0,0458	0,0164
SantaFe2A	subtentaculata	0,885	0,0117	0,0000	0,0453	0,0000
SantaFe3A	subtentaculata	0,929	0,0123	0,0000	0,0378	0,0000
Truchas1A	subtentaculata	0,901	0,0106	0,0030	0,0346	0,0871
Truchas2	subtentaculata	1	0,0146	0,0018	0,0496	0,0353
Truchas3	subtentaculata	0,824	0,0080	0,0015	0,0188	0,0799
Berga_2	etrusca-liguriensis	0,6820	0,0165	0,0018	0,0000	NA
Berga_3	etrusca-liguriensis	0,9640	0,0221	0,0053	0,0107	0,4907
Berga_5	etrusca-liguriensis	0,8900	0,0190	0,0015	0,0000	NA
Trilla_3	etrusca-liguriensis	0,6000	0,0173	0,0000	0,0000	NA
Trilla_5	etrusca-liguriensis	0,7270	0,0152	0,0000	0,0000	NA
Trilla_6	etrusca-liguriensis	0,8890	0,0189	0,0000	0,0000	NA

Supplementary Figures

Figure S1. Phylogenetic trees obtained in Chapter IV.1 using Maximum Likelihood (ML) and Multispecies Coalescent (MSC) model from two data sets of single copy orthologs. Dataset 1 contains all samples belonging to the Iberian-Appenines-Alps-plus clade. Dataset 2 exclude the samples from Berga locality. (A) ML tree from dataset 1. (B) MSC tree from data et 1. (C) ML tree from dataset 2. (D) MSC tree from dataset 2. The bootstrap values (ML-bv) and the branch support (MSC-bs) values are shown on the nodes in a color scale from minimum (white) to maximum (black). Scale bar: substitutions per site (ML) / Coalescent Units (MSC). FNQS: Final normalized quartet score. The arrows indicate the controversial points in the phylogeny.

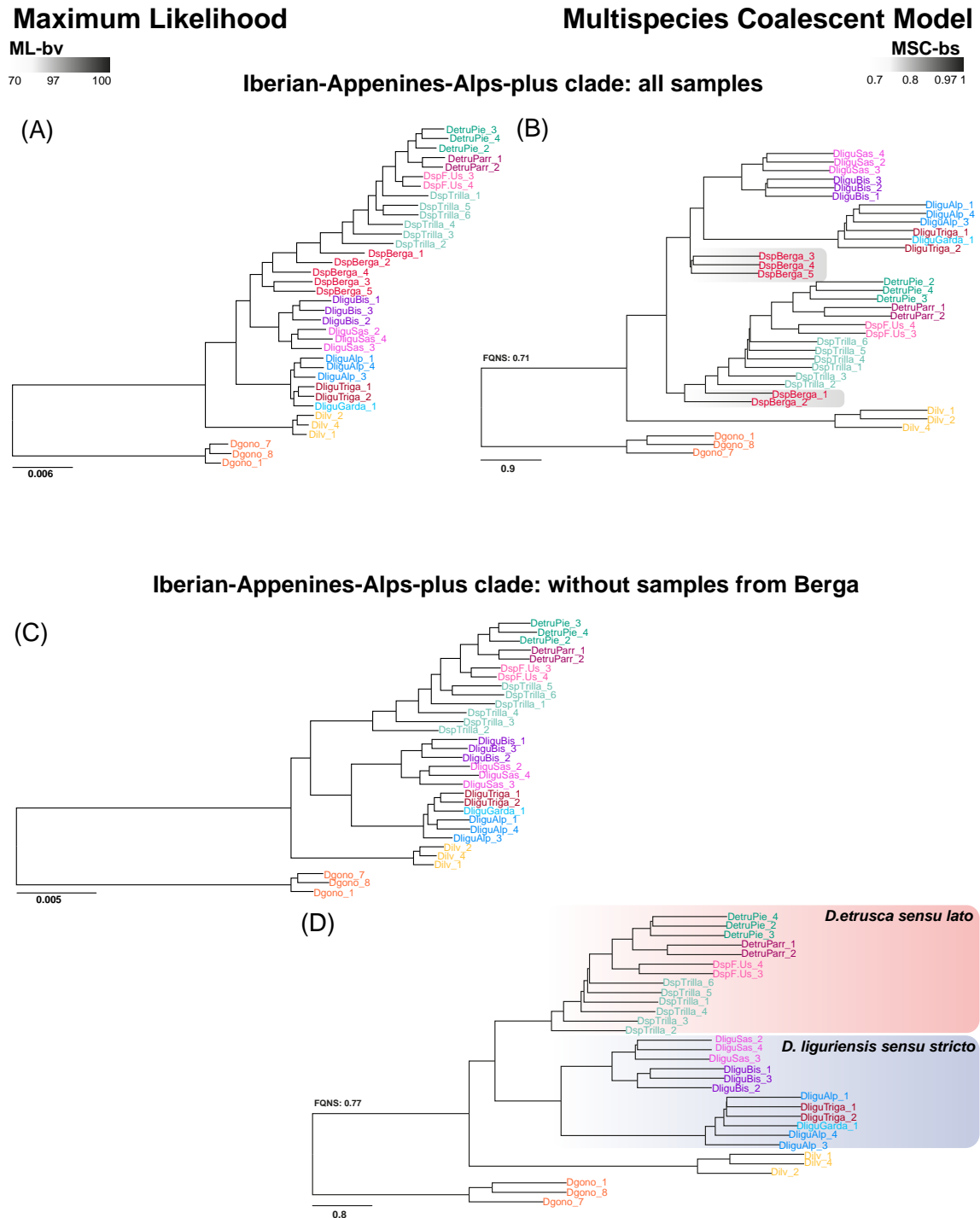


Figure S2. Average GC content in third codon positions (GC3%) and standard deviation calculated in 717 single copy orthologs of sexual (red) and asexual (green) individuals belonging to (A): Iberia-Apennines-Alps-plus clade and (B) *D. subtentaculata* and *D. vilafarrei* species.

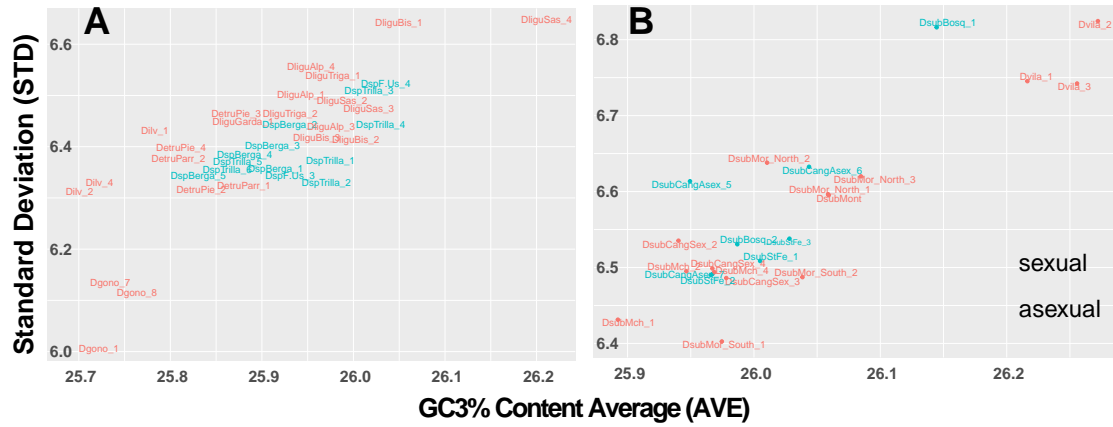


Figure S3. Haplotype Network of Dunuc12 marker. (A): Asexual individuals. (B): Sexual individuals. (C): Total haplotype network including all individuals. Each individual is indicated with a different color. Each circle represents a different haplotype and its size indicate the frequency of the haplotype in the individual. Branch lengths are proportional to the number of mutations.

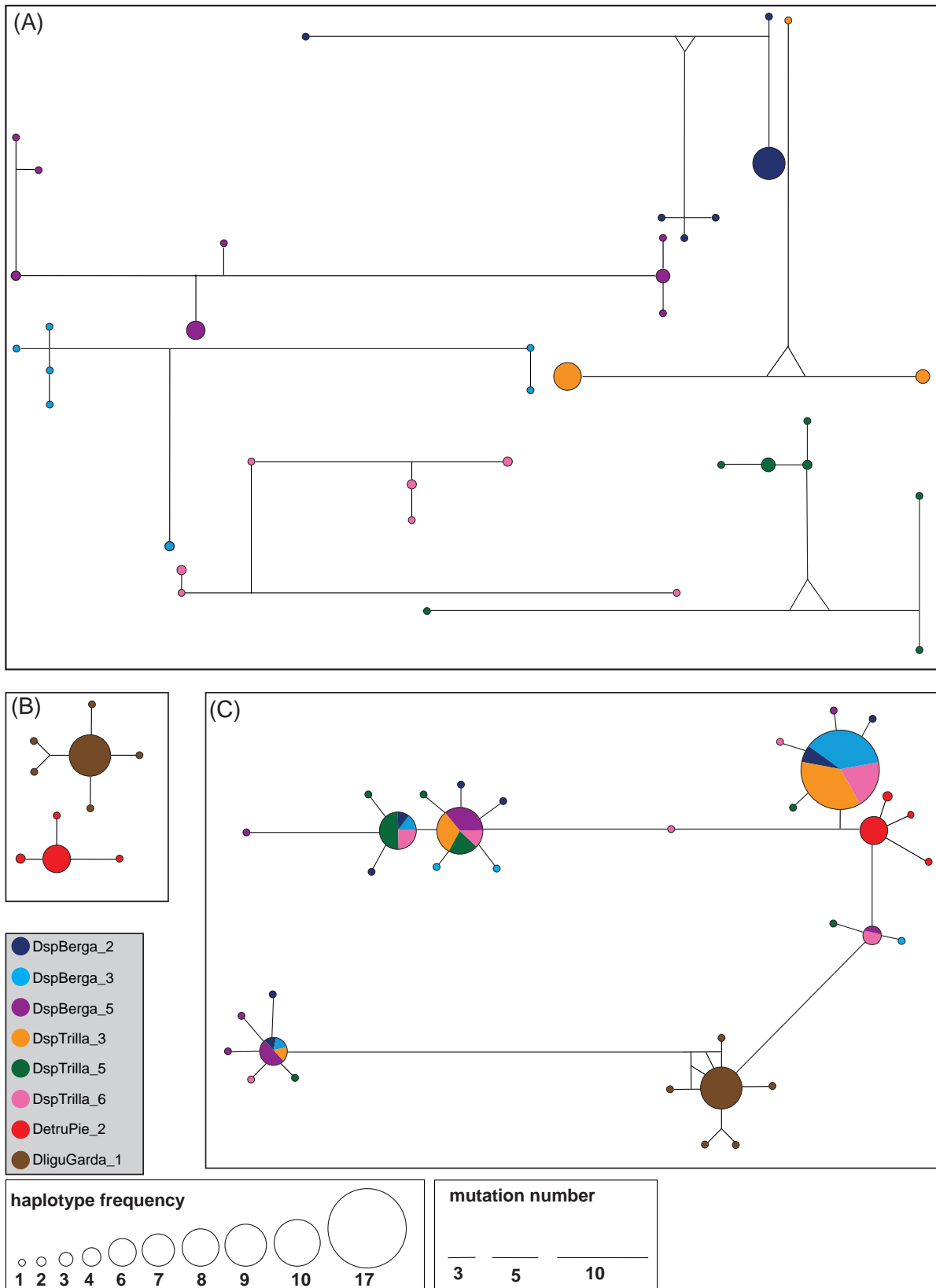


Figure S4. Pairwise distance matrix obtained from dataset 1 (1984 SC and all etrusca-liguriensis samples group) using the TN93 model. The color represent the distance values from lower (white) to higher (red).

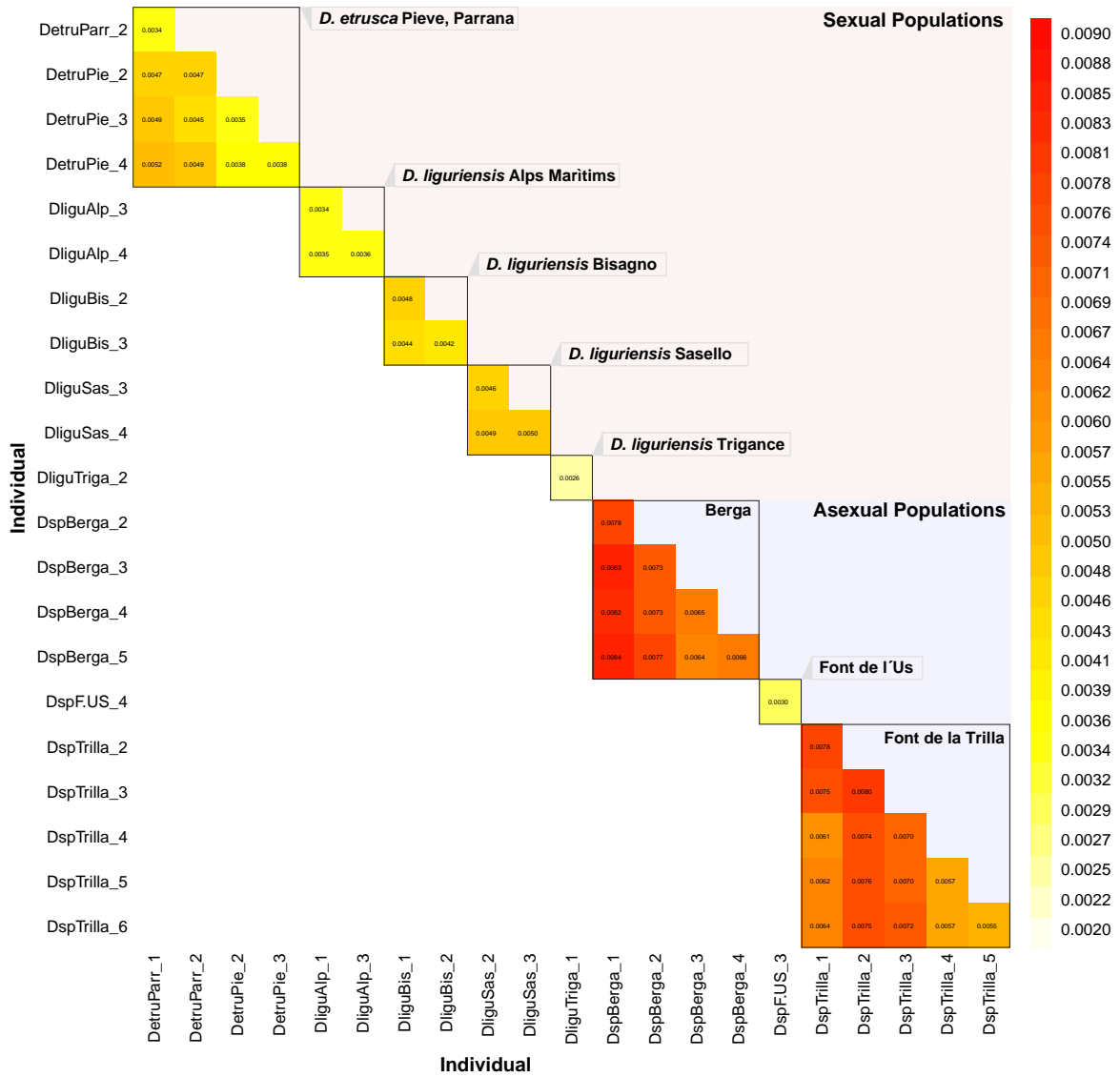


Figure S5

Per branch quartet support (q) and branch support (local posterior probability: pp) for the main (1) and the alternatives topologies (2, 3) for MSC trees obtained in ASTRAL-pro. The trees have been obtained in AstralPlane using ASTRAL-pro output and the blue circles indicate a numeric code for the nodes. The red arrows indicate discordant topologies and the correspondent values are highlighted in grey on the tables. (A): Dataset 1: contains all samples belong to the Iberian-Apennines-Alps-plus clade. (B): Dataset 2: Samples from Berga locality have been removed. (C): Data set 3: Samples from Iberian Peninsula have been removed. See Supplementary Table S4 for more information.

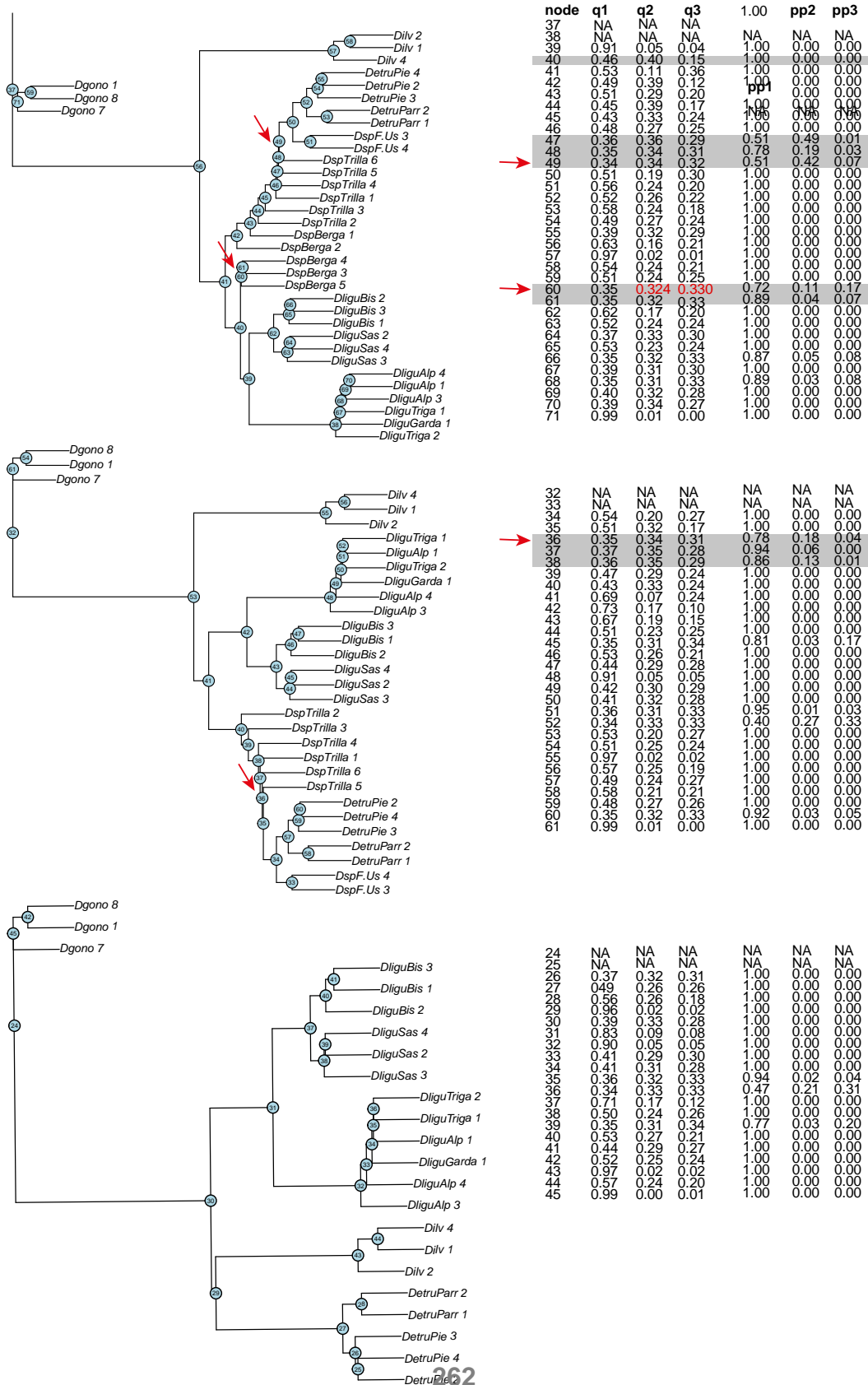
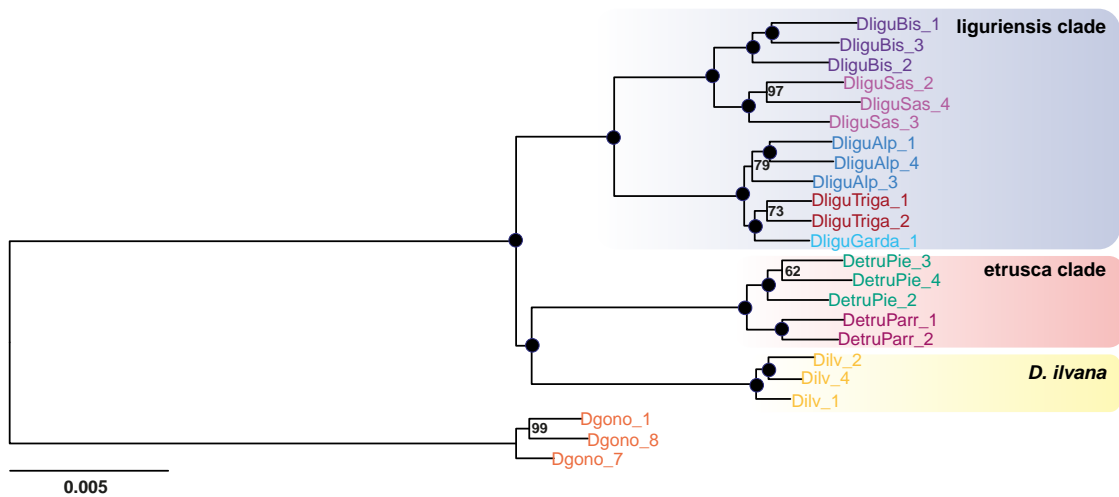


Figure S6. ML tree obtained from dataset 3 in IQ-TREE for nucleotide data. Bootstrap values are shown on the nodes. The black dots represent the maximum value (100). Scale bar: substitutions per site.



Supporting Information Text S1

Methodology to search for Dunuc12 in the transcriptomic data

Dunuc12 (*Dugesia nuclear 12*) is a fragment of the nuclear gene Transmembrane p24 trafficking protein 9 (TMED9), which is a transport protein transmembrane domain. This nuclear marker was identified from low coverage genomic sequences of *Dugesia* sp blasted against the genome of *Schmidtea mediterranea* SXI v4.0 (Leria et al., 2020) and subsequently used for species delimitation and intraindividual genetic diversity analyses in different studies on the genus *Dugesia* (Dols-Serrate, Leria, Aguilar, Stocchino, & Riutort, 2020; Leria et al., 2020; Leria, Vila-Farré, Solà, & Riutort, 2019). In this study, we cloned the Dunuc12 in 8 samples from the *etrusca-liguriensis* group to examine the intraindividual genetic diversity of sexual and fissiparous individuals. The obtained length of the Dunuc12 sequences was 680 positions, with an exonic region of 120 positions.

In this section, we describe the methodology used to find the exonic region of Dunuc12 in our transcriptomic data. We perform the steps as follow:

1. Extract the exonic region:

We aligned the entire sequences of the obtained clones of *D. etrusca* and *D. liguriensis* with sequences of Dunuc12 previously characterized and annotated (Leria et al., 2019) to find the exonic regions in the global alignment. We obtained 120 exonic positions of every cloned sequence. From the alignment of these 120 positions, we obtained a non resolved maximum likelihood tree (Fig. 1A in this text). This result is not rare, since the most variable regions in Dunuc12 have been found in the introns (Leria et al., 2019). A more resolutive tree was obtained when we included all the entire cloned sequences (Fig. 1B)

2. Search for the exonic region in the transcriptomic data:

We blasted every cloned sequence against the corresponding transcriptome using blastn with the script BLASTn_Dunuc.sh. We performed two blast searches, one against the transcriptome without filtering, just after the assembling step, and another one against the filtered transcripts after the Blobtools step (Chapter IV.1). The results

were identical in both cases, demonstrating that we retain Dunuc12, even after the filtering steps.

For every sample, the cloned sequences always matched against the same transcript, with very high identity percent, and bitscore, as well as low e-value (Table 1 in this text).

3. Transcript identification:

Taking advantage of the results obtained in the filtering steps with Blobtools, we used the Blastn of the transcriptome against the nucleotide database (Chapter IV.1) to validate the transcript identification. All the matched transcripts by the cloned sequences, yield also, hit against sequences from *Dugesia* identified as TMED9 in the nucleotide database of NCBI identified with the GeneBank accession codes MK713132-54 (Leria, et al., 2019).

4. Extract the coverage of the Dunuc12 region in the transcripts:

Working with the bam files previously obtained for Blobtools filtering, we obtained the per-base coverage of the 120 positions of Dunuc12 in the transcripts using samtools depth with the -a option, and the media coverage by region with mosdepth tool using the -n option (Fig. 2 in this text).

In addition, we visualized the transcripts in IGV, specifically the region where the cloned sequences matched. We explored these regions in IGV looking for polymorphic positions, but any position showed polymorphism; just as we expected taking into account the low variability of exonic regions of the Dunuc12.

5. Search for Dunuc12 in the orthogroups:

To check if Dunuc12 had been assembled in our transcriptome and to corroborate that this gene is a Single Copy, we looked for the matched transcripts in all orthogroups using the scripts `look_Dunuc12_in_OG.sh` and `look_Dunuc12_in_OG_all_samples.sh`. The exploration was done in the two orthologs search performed; the search with all samples (WM_717) and the search with samples from the *etrusca-liguriensis* group (*etru-ligu_1984*) (Chapter IV.1). We found that all transcripts formed only one orthogroup in the two searches: OG0010797 in etru-

ligu_1984, and OG0007336 in WM_717. Unfortunately, any of the two orthogroups was defined as SC orthogroup by Orthofinder.

In the case of etru-ligu_1984, the sample DliguSas_2 (not included in the cloning analysis) did not show any transcript into OG0010797, although one transcript matched with Temed9 in the blastn against nucleotide database of NCBI, indicating that this transcript was lost in the filtering steps. In addition, the sample Dgono_8 contributes with two transcripts to the orthogroup, being Dgono_8_DN38329_c0_g1_i1 a very rare sequence, that could be interpreted as a failure of the orthologs search.

In the case of the WM_717 search, in addition to the two problems explained above, we found more problematic sequences: DbenCors_North_3_DN37816_c0_g2_i1 was very rare, and Dvila_2_DN7972_c0_g1_i1.p2 was a partial isoform that was not eliminated with the filtering.

6. Maximum likelihood trees inference from Dunuc12 orthogroups:

We eliminated the problematic sequences in OG0010797 (etru-ligu_1984), aligned the block, and obtained trees from nucleotide and protein sequences using IQ-TREE (Fig. 3 in this text). The low resolution of these trees was congruent with the previous results, which describe the most variability in the intronic regions of Dunuc12 and not in the exons, more exposed to gene regulatory mechanisms. In addition, the trees show the topology of a region that has not suffered any duplication event in the evolutionary history of the group.

Main conclusion:

Although the OrthoFinder does not classify Dunuc12 as SC (because of artifacts in the assemblies, and failures in filtering or OrthoFinder runs), we have enough proofs to corroborate that Dunuc12 is not a paralog and it is a Single Copy gene in the *Dugesia* genome. In the first place, all sequences are grouped only in one orthogroup in every orthologue search. As a second point, the trees show the typical topology of Single Copy genes. These results validate the intronic region of Dunuc12 as a good marker for studies on intraindividual genetic variability in the genus *Dugesia*.

References in this text

- Dols-Serrate et al., (2020). <https://doi.org/10.1007/s13127-020-00438-z>
 Leria, L et al., (2020). <https://doi.org/10.1016/j.ympev.2019.05.010>
 Leria, L et al., (2019). <https://doi.org/10.1186/s12862-019-1440-1>

Figures and Tables in this text

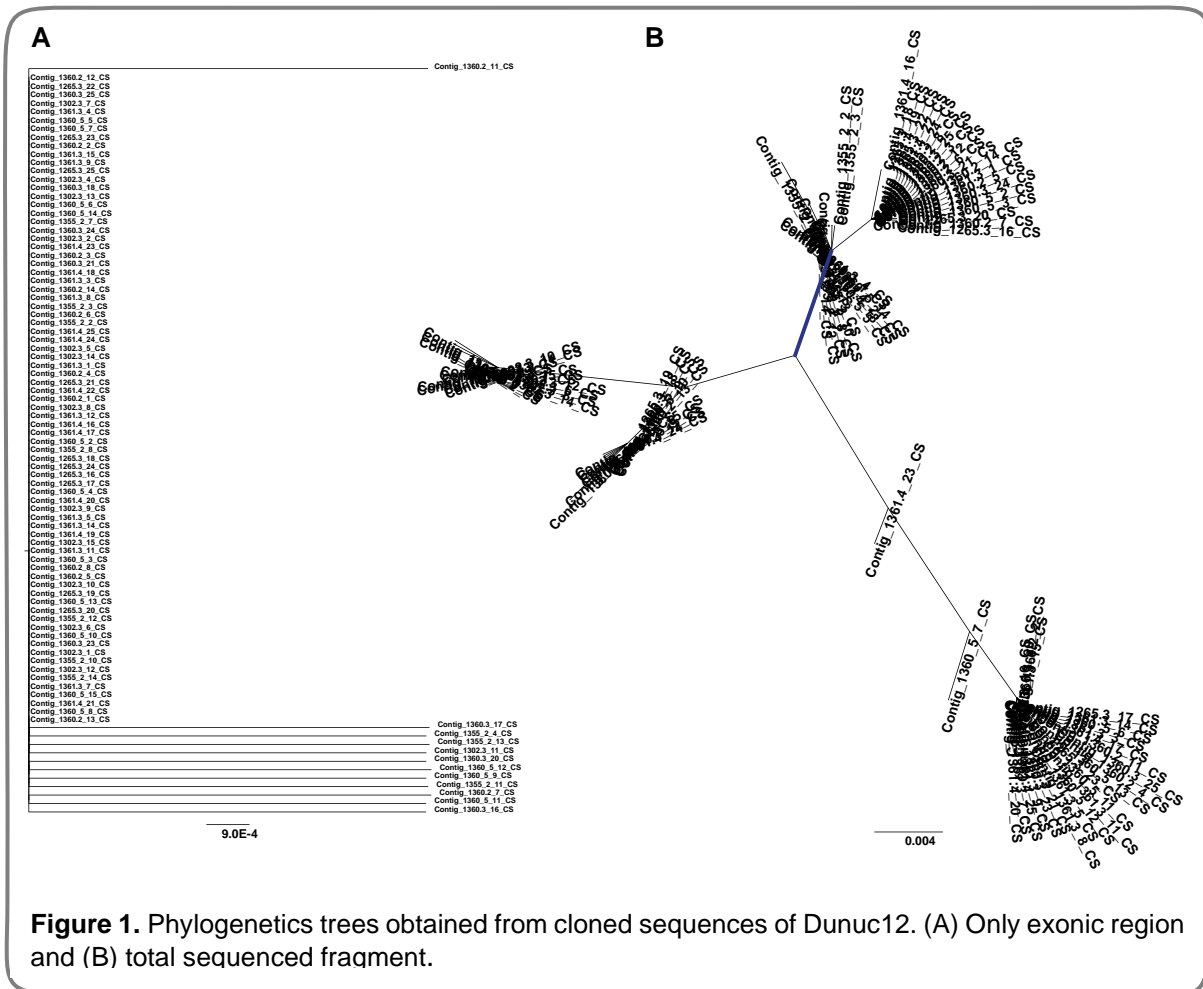


Figure 1. Phylogenetics trees obtained from cloned sequences of Dunuc12. (A) Only exonic region and (B) total sequenced fragment.

Table 1. Results of the Blastn of cloned sequences against their respective transcriptome.

Sample	MR ID	No. Clones	bitscore (lower/higher values)	Blastn results resume									
				Subject accession	% ident match	Alig length	mismatch	gapopen	query start	Query end	Subj. start	Subj. end	evaluate
DspBerga_2	1360.2	12	217/222	TRINITY_DN1077_c0_g1_i3	99.167/100	120	1	0	1	120	404	285	2.87E-56
				TRINITY_DN1077_c0_g1_i2	99.17	120	1	0	1	120	404	285	2.87E-56
				TRINITY_DN1077_c0_g1_i4	99.17	120	1	0	1	120	404	285	2.87E-56
DspBerga_3	1360.3	8	217/222	TRINITY_DN3226_c0_g1_i2	99.167/100	120	1	0	1	120	468	349	4.06E-56
				TRINITY_DN3226_c0_g1_i1	99.17	120	1	0	1	120	468	349	4.06E-56
DetruBerg_5	1360.5	14	217/222	TRINITY_DN6190_c0_g1_i1	99.167/100	120	1	0	1	120	450	569	4.6E-56
DspTrilla_3	1265.3	10	222	TRINITY_DN1320_c0_g1_i2	100.00	120	0	0	1	120	473	592	9.36E-58
DspTrilla_5	1361.3	11	222	TRINITY_DN596_c0_g1_i1	100.00	120	0	0	1	120	590	471	8.57E-58
				TRINITY_DN596_c0_g1_i4	100.00	120	0	0	1	120	590	471	8.57E-58
DspTrilla_6	1361.4	10	222	TRINITY_DN784_c0_g1_i1	100.00	120	0	0	1	120	402	283	8.6E-58
				TRINITY_DN784_c0_g1_i2	100.00	120	0	0	1	120	402	283	8.6E-58
				TRINITY_DN784_c0_g1_i3	100.00	120	0	0	1	120	402	283	8.6E-58
DetruPie_2	1355.2	10	217/222	TRINITY_DN92587_c0_g1_i1	99.167/100	120	1	0	1	120	404	285	3.76E-56
DiguGarda_1	1302.3	14	222/217	TRINITY_DN77494_c0_g1_i1	100/99.17	120	0	0	1	120	450	569	7.2E-58

References

- Abril, J. F., Cebrià, F., Rodríguez-Esteban, G., Horn, T., Fraguas, S., Calvo, B., ... Saló, E. (2010). Smed454 dataset: unravelling the transcriptome of *Schmidtea mediterranea*. *BMC Genomics*, *11*(731). <https://doi.org/10.1186/1471-2164-11-731>
- Altenhoff, A. M., Levy, J., Zarowiecki, M., Tomiczek, B., Vesztröcy, A. W., Dalquen, D. A., ... Dessimoz, C. (2019). OMA standalone: Orthology inference among public and custom genomes and transcriptomes. *Genome Research*, *29*(7), 1152–1163. <https://doi.org/10.1101/gr.243212.118>
- Altenhoff, A. M., Schneider, A., Gonnet, G. H., & Dessimoz, C. (2011). OMA 2011: Orthology inference among 1000 complete genomes. *Nucleic Acids Research*, *39*. <https://doi.org/10.1093/nar/gkq1238>
- Baguñà, J., Carranza, S., Pala, M., Ribera, C., Arnedo, M. A., Ribas, M., & Riutort, M. (1999). From morphology and karyology to molecules. New methods for taxonomical identification of asexual populations of freshwater planarians. A tribute to Professor Mario Benazzi. *Italian Journal of Zoology*, *66*, 207–214. <https://doi.org/10.1080/11250009909356258>
- Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026036>
- Bast, J., Parker, D. J., Dumas, Z., Jalvingh, K. M., Van, P. T., Jaron, K. S., ... Schwander, T. (2018). Consequences of asexuality in natural populations: Insights from stick insects. *Molecular Biology and Evolution*, *35*(7), 1668–1677. <https://doi.org/10.1093/molbev/msy058>
- Benazzi, M. (1946). Sopra una nova planaria d'acqua dolce. *Estratto Dall' Archivio Zoologico Italiano*, *XXXI*.
- Birky, C. W. (1996). Heterozygosity, Heteromorphy, and Phylogenetic Trees in Asexual Eukaryotes. *Genetics*, *144*, 427–437.
- Boluda, C. G., Rico, V. J., Naciri, Y., Hawksworth, D. L., & Scheidegger, C. (2021). Phylogeographic reconstructions can be biased by ancestral shared alleles: The

- case of the polymorphic lichen *Bryoria fuscescens* in Europe and North Africa. *Molecular Ecology*, 30(19), 4845–4865. <https://doi.org/10.1111/mec.16078>
- Bortolotti, V., Fazzuoli, M., Pandeli, E., Principi, G., Babbini, A., & Corti, S. (2001). Geology of central and eastern Elba island, Italy. *Ofioliti*, 26(2 A), 97–105. <https://doi.org/10.4454/ofioliti.v26i2a.137>
- Bortolotti, V., Pandeli, E., & Principi, G. (2001). The geology of the Elba Island: An historical introduction. *Ofioliti*, 26(2 A), 79–96. <https://doi.org/10.4454/ofioliti.v26i2a.136>
- Bouckaert, R. R. (2010). DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, 26(10), 1372–1373. <https://doi.org/10.1093/BIOINFORMATICS/BTQ110>
- Bouckaert, R. R., & Heled, J. (2014). DensiTree 2: Seeing Trees Through the Forest. *BioRxiv*, 012401. <https://doi.org/10.1101/012401>
- Bowie, R. C. K., Fjeldså, J., Hackett, S. J., Bates, J. M., & Crowe, T. M. (2006). Coalescent models reveal the relative roles of ancestral polymorphism, vicariance, and dispersal in shaping phylogeographical structure of an African montane forest robin. *Molecular Phylogenetics and Evolution*, 38(1), 171–188. <https://doi.org/10.1016/j.ympev.2005.06.001>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(421). <https://doi.org/10.1186/1471-2105-10-421>
- Capra, J. A., Hubisz, M. J., Kostka, D., Pollard, K. S., & Siepel, A. (2013). A Model-Based Analysis of GC-Biased Gene Conversion in the Human and Chimpanzee Genomes. *PLOS Genetics*, 9(8), e1003684. <https://doi.org/10.1371/JOURNAL.PGEN.1003684>
- Dapporto, L., & Cini, A. (2007). Faunal patterns in Tuscan archipelago butterflies: The dominant influence is recent geography not paleogeography. *European Journal of Entomology*, 104(3), 497–503. <https://doi.org/10.14411/eje.2007.070>
- De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, 56(6), 879–886. <https://doi.org/10.1080/10635150701701083>

- De Vries, E. J. (1988). Further contribution to the taxonomy and biogeography of the subgenus *Dugesia* (Platyhelminthes: Tricladida: Paludicola) in the Mediterranean region and the Middle East. *Israel Journal of Zoology*, 35, 109–136.
- Deutekom, E. S., Snel, B., & Van Dam, T. J. P. (2021). Benchmarking orthology methods using phylogenetic patterns defined at the base of Eukaryotes. *Briefings in Bioinformatics*, 22(3), 1–9. <https://doi.org/10.1093/bib/bbaa206>
- Di Nicola, M. R., & Vaccaro, A. (2020). New data on the presence of the Aesculapian snake *Zamenis longissimus* (Laurenti, 1768) (Serpentes Colubridae) on Elba Island (Tuscany, Italy). *Biodiversity Journal*, 11(2), 611–614. <https://doi.org/10.31396/biodiv.jour.2020.11.2.611.614>
- Du, Z., Hasegawa, H., Cooley, J. R., Simon, C., Yoshimura, J., Cai, W., ... Li, H. (2019). Mitochondrial genomics reveals shared phylogeographic patterns and demographic history among three periodical cicada species groups. *Molecular Biology and Evolution*, 36(6), 1187–1200. <https://doi.org/10.1093/molbev/msz051>
- Duret, L., & Galtier, N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annual Review of Genomics and Human Genetics*, 10, 285–311. <https://doi.org/10.1146/annurev-genom-082908-150001>
- Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(157). <https://doi.org/10.1186/s13059-015-0721-2>
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(238). <https://doi.org/10.1186/s13059-019-1832-y>
- Fattorini, S. (2009). Both Recent and Pleistocene geography determine animal distributional patterns in the Tuscan Archipelago. *Journal of Zoology*, 277(4), 291–301. <https://doi.org/10.1111/j.1469-7998.2008.00540.x>
- Fehrer, J., Krak, K., & Chrtek, J. (2009). Intra-individual polymorphism in diploid and apomictic polyploid hawkweeds (Hieracium, Lactuceae, Asteraceae): Disentangling phylogenetic signal, reticulation, and noise. *BMC Evolutionary Biology*, 9(1), 1–23. <https://doi.org/10.1186/1471-2148-9-239>

- Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, 78(2), 737–756. <https://doi.org/10.1093/genetics/78.2.737>
- Fernández, R., Gabaldon, T., Dessimoz, C., Fernández, R., Gabaldon, T., Dessimoz, C., ... Phylo-, N. (2020). Orthology: Definitions, Prediction, and Impact on Species Phylogeny Inference. In C. Scornavacca, F. Delsuc, & N. Galtier (Eds.), *Phylogenetics in the Genomic Era* (pp. 2.4:1–2.4:14). https://doi.org/10.1007/978-1-4939-9832-2_4
- Fontaneto, D., Herniou, E. A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., & Barraclough, T. G. (2007). Independently Evolving Species in Asexual Bdelloid Rotifers. *PLoS Biol*, 5(4), 914–921. <https://doi.org/10.1371/journal.pbio.0050087>
- Galtier, N. (2021). Fine-scale quantification of GC-biased gene conversion intensity in mammals. *Peer Community Journal*, 1. <https://doi.org/10.24072/pcjournal.22>. hal-03435015v2
- Gill, D. E., Chao Lin, Perkins, S. L., & Wolf, J. B. (1995). Genetic mosaicism in plants and clonal animals. *Annual Review of Ecology and Systematics*, 26(40), 423–444. <https://doi.org/10.1146/annurev.es.26.110195.002231>
- Hasegawa, M., Kishino, H., & Yano, T. aki. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 1985 22:2, 22(2), 160–174. <https://doi.org/10.1007/BF02101694>
- Hellmuth, M., Wieseke, N., Lechner, M., Lenhof, H. P., Middendorf, M., & Stadler, P. F. (2015). Phylogenomics with paralogs. *Proceedings of the National Academy of Sciences of the United States of America*, 112(7), 2058–2063. <https://doi.org/10.1073/pnas.1412770112>
- Hörandl, E. (2018). The classification of asexual organisms: Old myths, new facts, and a novel pluralistic approach. *Taxon*, 67(6), 1066–1081. <https://doi.org/10.12705/676.5>
- Hutter, C. (2020). *AstralPlane: R Package For Preparing, Running, Analyzing and Plotting from the Species Tree Program ASTRAL-III*. Retrieved from <https://github.com/chutter/AstralPlane>
- Jukes, T. H., & Cantor, C. R. (1969). Evolution of Protein Molecules. *Mammalian Protein Metabolism*, 21–132. <https://doi.org/10.1016/B978-1-4832-3211-9.50009-7>

- Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, 21(7), 428–444. <https://doi.org/10.1038/s41576-020-0233-0>
- Kishino, H., & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution* 1989 29:2, 29(2), 170–179. <https://doi.org/10.1007/BF02100115>
- Kishino, H., Miyata, T., & Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution* 1990 31:2, 31(2), 151–160. <https://doi.org/10.1007/BF02109483>
- Lamolle, G., Fontenla, S., Rijo, G., Tort, J. F., & Smircich, P. (2019). Compositional Analysis of Flatworm Genomes Shows Strong Codon Usage Biases Across All Classes. *Frontiers in Genetics*, 10(September), 1–12. <https://doi.org/10.3389/fgene.2019.00771>
- Lázaro, E., Sluys, R., Pala, M., Stochino, G. A., Baguñà, J., & Riutort, M. (2009). Molecular barcoding and phylogeography of sexual and asexual freshwater planarians of the genus *Dugesia* in the Western Mediterranean (Platyhelminthes, Tricladida, Dugesiidae). *Molecular Phylogenetics and Evolution*, 52, 835–845. <https://doi.org/10.1016/j.ympev.2009.04.022>
- Le, S. Q., & Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, 25(7), 1307–1320. <https://doi.org/10.1093/MOLBEV/MSN067>
- Leria, L., Riutort, M., Romero, R., Ferrer, X., & Vila-Farré, M. Microplate tectonics and the environment as distribution drivers in Western Mediterranean freshwater planarians. *Journal of Biogeography*, *In Press*.
- Leria, L., Vila-Farré, M., Solà, E., & Riutort, M. (2019). Outstanding intraindividual genetic diversity in fissiparous planarians (*Dugesia*, Platyhelminthes) with facultative sex. *BMC Evolutionary Biology*, 19(1). <https://doi.org/10.1186/s12862-019-1440-1>

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/BIOINFORMATICS/BTP352>
- Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451–1452. <https://doi.org/10.1093/BIOINFORMATICS/BTP187>
- Lunt, D. H. (2008). Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. *BMC Evolutionary Biology*, 8(1), 1–16. <https://doi.org/10.1186/1471-2148-8-194>
- MacEachern, S., McEwan, J., & Goddard, M. (2009). Phylogenetic reconstruction and the identification of ancient polymorphism in the Bovini tribe (Bovidae, Bovinae). *BMC Genomics*, 10, 1–17. <https://doi.org/10.1186/1471-2164-10-177>
- Martens, K., Rossetti, G., & Baltanás, A. (1998). Reproductive modes and taxonomy. In K. Martens (Ed.), *Sex and parthenogenesis: evolutionary ecology of reproductive modes in non-marine ostracods* (pp. 197–214). Leiden: Backhuys Publishers.
- Moran, P., & Kornfield, I. (1993). Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Molecular Biology and Evolution*, 10(5), 1015–1029. <https://doi.org/10.1093/oxfordjournals.molbev.a040063>
- Muller, H. (1932). Some genetic aspects of sex. *The American Naturalist*, 66(703), 118–138.
- Muller, H. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1), 2–9. [https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8)
- Nelson Hall, R., Weill, U., Khariton, M., Leal-Ortiz, S., Drees, L., Chai, C., ... Wang, B. (2021). Heterologous reporter expression in the planarian *Schmidtea mediterranea* through somatic mRNA transfection. *BioRxiv*, 1–43. Retrieved from <https://doi.org/10.1101/2021.04.20.440701>

- Normark, B. B., Judson, O. P., & Moran, N. a. (2003). Genomic signatures of ancient asexual lineages. *Biological Journal of the Linnean Society*, 79, 69–84. <https://doi.org/10.1046/j.1095-8312.2003.00182.x>
- Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.
- Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. P. (2011). MACSE: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. *PLoS ONE*, 6(9). <https://doi.org/10.1371/journal.pone.0022594>
- Reddien, P. W., & Alvarado, A. S. (2004). Fundamentals of Planarian Regeneration. *Annual Review of Cell and Developmental Biology*, 20(1), 725–757. <https://doi.org/10.1146/annurev.cellbio.20.010403.095114>
- Resch, A. M., Palakodeti, D., Lu, Y. C., Horowitz, M., & Graveley, B. R. (2012). Transcriptome analysis reveals strain-specific and conserved stemness genes in *Schmidtea mediterranea*. *PLoS ONE*, 7(4). <https://doi.org/10.1371/journal.pone.0034447>
- Rink, J. C. (2013). Stem cell systems and regeneration in planaria. *Development Genes and Evolution*, 223(1–2), 67–84. <https://doi.org/10.1007/s00427-012-0426-4>
- Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., & Mesirov, J. P. (2017). Variant review with the integrative genomics viewer. *Cancer Research*, 77(21), e31–e34. <https://doi.org/10.1158/0008-5472.CAN-17-0337/SUPPLEMENTARY-VIDEO-S1>
- Romiguier, J., & Roux, C. (2017). Analytical biases associated with GC-content in molecular evolution. *Frontiers in Genetics*, 8(FEB), 1–7. <https://doi.org/10.3389/fgene.2017.00016>
- Saló, E., Abril, J. F., Adell, T., Cebrià, F., Eckelt, K., Fernández-Taboada, E., ... Rodríguez-Esteban, G. (2009). Planarian regeneration: Achievements and future directions after 20 years of research. *International Journal of Developmental Biology*, 53(8–10), 1317–1327. <https://doi.org/10.1387/ijdb.072414es>

- Schwander, T., Henry, L., & Crespi, B. J. (2011). Molecular evidence for ancient asexuality in timema stick insects. *Current Biology*, 21(13), 1129–1134. <https://doi.org/10.1016/j.cub.2011.05.026>
- Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE*, 11(10), 1–10. <https://doi.org/10.1371/journal.pone.0163962>
- Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, 51(3), 492–508. <https://doi.org/10.1080/10635150290069913>
- Shimodaira, H., & Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution*, 16(8), 1114–1114. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026201>
- Solà, E., Leria, L., Stocchino, G. A., Bagherzadeh, R., Balke, M., Daniels, S. R., ... Riutort, M. Three dispersal routes out of Africa: The puzzling biogeographical history of *Dugesia* freshwater planarians. *Journal of Biogeography*, *In Press*.
- Stocchino, G. A., & Manconi, R. (2013). Overview of life cycles in model species of the genus *Dugesia* (Platyhelminthes: Tricladida). *Italian Journal of Zoology*, 80(3), 319–328. <https://doi.org/10.1080/11250003.2013.822025>
- Strimmer, K., & Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1487), 137–142. <https://doi.org/10.1098/RSPB.2001.1862>
- Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 10(3), 512–526. <https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040023>
- Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, 17, 57–86.
- Team, R. C. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.r-project.org/>

- Wang, B., Mojica, J. P., Perera, N., Lee, C. R., Lovell, J. T., Sharma, A., ... Mitchell-Olds, T. (2019). Correction to: Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in *Boechera stricta*. *Genome Biology*, *20*(1), 1–15. <https://doi.org/10.1186/s13059-019-1781-5>
- Wang, H. C., Minh, B. Q., Susko, S., & Roger, A. J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.*, *67*, 216–235. <https://doi.org/10.1093/sysbio/syx068>
- Webster, M. T., & Hurst, L. D. (2012). Direct and indirect consequences of meiotic recombination: Implications for genome evolution. *Trends in Genetics*, *28*(3), 101–109. <https://doi.org/10.1016/j.tig.2011.11.002>
- Welch, D. M., & Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, *288*(5469), 1211–1215. <https://doi.org/10.1126/science.288.5469.1211>
- Yan, Z., Smith, M. L., Du, P., Hahn, M. W., & Nakhleh, L. (2022). Species Tree Inference Methods Intended to Deal with Incomplete Lineage Sorting Are Robust to the Presence of Paralogs. *Systematic Biology*, *71*(2), 367–381. <https://doi.org/10.1093/sysbio/syab056>
- Zhang, C., Scornavacca, C., Molloy, E. K., & Mirarab, S. (2020). ASTRAL-pro: Quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution*, *37*(11), 3292–3307. <https://doi.org/10.1093/molbev/msaa139>