Received: 2 June 2019 | Revised: 6 March 2020 | Accepted: 7 March 2020
DOI: 10.1111/zsc.12420

ORIGINAL ARTICLE

Zoologica Scripta WILEY

Multilocus phylogeny and systematics of Iberian endemic *Squalius* (Actinopterygii, Leuciscidae)

Silvia Perea[1] | **Carla Sousa-Santos**[2] | **Joana Robalo**[2] | **Ignacio Doadrio**[1]

[1]Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales - CSIC, Madrid, Spain

[2]MARE – Marine and Environmental Sciences Centre, ISPA-Instituto Universitário, Lisbon, Portugal

Correspondence
Silvia Perea, Museo Nacional de Ciencias Naturales–CSIC, José Gutiérrez Abascal, 2, Madrid 28006, Spain.
Email: sperea2@gmail.com

Funding information
Ministerio de Ciencia e Innovación, Grant/Award Number: CGL2010-15231BOS; Duero Hydrographic Confederation. Spain, Grant/Award Number: Life13 nat/es/000772

Abstract
Inferring the evolutionary history of a group of species can be challenging given the many factors involved. In recent years, the increased availability of sequences of multiple genes per species has spurred the development of new methodologies to analyse multilocus data sets. Two approaches that analyse such data are concatenated supermatrix and coalescent-based species-tree analyses. In this study, we used both of these methods to infer the phylogenetic relationships of Iberian species of the genus *Squalius* from one mitochondrial and six nuclear genes. We found mitonuclear discordance in the phylogenetic relationships of the group. According to the mitochondrial gene analysis, all species were recovered as monophyletic except *S. pyrenaicus*; besides, in the concatenated supermatrix analysis of the nuclear markers, this species resolved as polyphyletic with three divergent evolutionary lineages. The coalescent-based nuclear species-tree analysis rendered a well-resolved phylogeny compared with the supermatrix analysis, which was unable to discern between *S. carolitertii*, *S. castellanus* and one of the evolutionary lineages of *S. pyrenaicus*. This result is likely due to the better integration of population uncertainty in the coalescent approach. Furthermore, Bayesian multilocus species delimitation analyses based on a BPP approach strongly supported the distinct nuclear lineages as different species. Nevertheless, the supermatrix analysis was able to obtain well-supported relationships in the divergent lineages with low numbers of individuals. Our study highlights the usefulness of different analytical methodologies to obtain a more complete picture of the evolutionary history of taxa, especially when discordant patterns among genes are found.

KEYWORDS
coalescent species tree, mitonuclear discordance, *Squalius*, supermatrix concatenation

1 | INTRODUCTION

Understanding biodiversity patterns, and the processes generating them, is a main topic of interest for evolutionary biologists. Nevertheless, inferring the evolutionary history of a species can be challenging, particularly given the many factors involved, some of which are not usually considered in analyses, such as lineage sorting processes, population-level heterogeneity, gene duplication or loss or gene flow, among others (Harrington & Near, 2012; Kautt, Machado-Schiaffino, & Meyer, 2016; Knowles & Carstens, 2007; Knowles & Kubatko, 2011; Maddison, 1997; Maddison & Knowles, 2006; Nakhleh, 2013; Rougeux, Bernatchez, & Gagnaire, 2015). In recent years, the analysis of multiple gene and genomic sequences has become a common means to infer more robust phylogenies, especially with the advent of phylogenomics and massive sequencing technologies (Scornavacca

& Galtier, 2017; Simion et al., 2017; Sukumaran & Knowles, 2017). The large amount of data that have become available over the last two decades has spurred the development of two analytical methods developed to treat multilocus data sets: concatenated supermatrix and coalescent species-tree approaches (e.g., Gadagkar, Rosenberg, & Kumar, 2005; Gatesy et al., 2017; Simmons & Gatesy, 2015).

Concatenated multilocus analyses are thought to be advantageous over single gene ones because they increase the number of informative sites, therefore providing more robust and unbiased phylogenies (Hillis, Huelsenbeck, & Cunningham, 1994; Rokas & Carroll, 2005; Rokas, Williams, King, & Carroll, 2003). However, these methodologies do not account for the existence of different populations within a species or the stochastic variation of each gene, rather they presume that all genes have the same or very similar evolutionary histories (Edwards, Liu, & Pearl, 2007; Kubatko & Degnan, 2007; Liu, Yu, Pearl, & Edwards, 2009). To account for population history in species-tree inferences, the multispecies coalescent (MSC) model was proposed (Degnan & Rosenberg, 2009; Rannala & Yang, 2003; Yang & Rannala, 2010). This model, which generalizes the Wright–Fisher model of genetic drift to multiple populations, is based on the premise that a coalescence model can be applied to gene trees within a species-tree, thus incorporating population parameters in the inference of the species-tree (Edwards, 2009; Edwards et al., 2016). The simplest model assumes that population size is constant and that the only process affecting genealogies is incomplete lineage sorting (deep coalescence); gene flow, introgression, recombination and selection processes are not considered in the model (Degnan & Rosenberg, 2009). Obviously, this model represents a theoretically ideal situation: some of these assumptions are typically not met in real populations (Petit & Excoffier, 2009). However, more complex coalescence-based models are continually being developed to handle the restrictive conditions of the MSC model (e.g., Edwards et al., 2016; Meng & Kubatko, 2011). Coalescent species-tree approaches accommodate potential discordances among individual gene trees in species-tree estimations, especially when heterogeneity among gene trees is considerable (Kolaczkowski & Thornton, 2004; Kubatko & Degnan, 2007).

The debate about the use of coalescent species-tree versus concatenated multilocus approaches is ongoing. Some authors consider supermatrix concatenation more accurate, although with some restrictions, or that coalescence methods yield biased results for deep phylogenies (e.g., Campbell, Alfaro, Belasco, & López, 2017; Gadagkar et al., 2005; Meiklejohn, Faircloth, Glenn, Kimball, & Braun, 2016; Simmons & Gatesy, 2015; Springer & Gatesy, 2014; Wiens et al., 2008). Most, however, advocate for species-tree methods based on the coalescent model (e.g., Edwards et al., 2007; Kubatko & Degnan, 2007; Mirarab, Bayzid, & Warnow, 2016; Roch & Steel, 2015; Roch & Warnow, 2015; Xi, Liu, Rest, & Davis, 2014). Alternatively, some authors do not find the accuracy of one method or the other statistically distinguishable (Tonini, Moore, Stern, Shcheglovitova, & Orti, 2015), while some even question the ability of these models to distinguish between population structure and species boundaries (Sukumaran & Knowles, 2017). In order to obtain a more robust phylogeny based on different analytical methods, we take advantage of the strengths of both, concatenated supermatrix and coalescent-based species-tree, to phylogenetically reconstruct relationships among Iberian species of the genus *Squalius* using one mitochondrial and six nuclear genes.

Squalius has a wide distribution range along Eurasia and, with 52 described species, is one of the most diversified genera of the family Leuciscidae, although its diversity is likely an underestimation (Fricke, Eschmeyer, & Laan, 2020; Froese & Pauly, 2019; Kottelat & Freyhof, 2007). In the Iberian Peninsula, the genus is represented by ten species (Doadrio, Perea, Garzón-Heydt, & González, 2011; Sousa-Santos et al., 2016) whose phylogenetic relationships have been widely analysed over the last fifteen years (Almada & Sousa-Santos, 2010; Briolay, Galtier, Brito, & Bouvet, 1998; Brito, Briolay, Galtier, Bouvet, & Coelho, 1997; Coelho, Brito, Pacheco, Figuereido, & Pires, 1995; Doadrio & Carmona, 2003, 2006; Doadrio, Kottelat, & de Sostoa, 2007; Doadrio, Perea, & Alonso, 2007; Geiger et al., 2014; Perea, Cobo-Simón, & Doadrio, 2016; Perea & Doadrio, 2015; Sanjur, Carmona, & Doadrio, 2003; Sousa-Santos, Collares-Pereira, & Almada, 2006, 2007; Sousa-Santos, Jesus, Fernandes, Robalo, & Coelho, 2019; Waap, Amaral, Gomes, & Coelho, 2011; Zardoya & Doadrio, 1998). Despite this, there is still no consensus about the relationships among these species. Reasons for this include the lack of comprehensive studies that incorporate all known populations of the genus in the Iberian Peninsula and the small number of markers analysed in most studies. Indeed, very few studies have analysed nuclear markers or have considered the entire distribution range of the genus (Almada & Sousa-Santos, 2010; Perea et al., 2010; Sousa-Santos et al., 2019; Waap et al., 2011).

Squalius is generally subdivided into two evolutionary lineages that were probably isolated by the start of the Alpine orogeny during the Miocene: the Mediterranean and the Eurasiatic (Perea et al., 2010; Sanjur et al., 2003). In this study, only species belonging to the Mediterranean lineage are considered, namely *S. aradensis*, *S. carolitertii*, *S. castellanus*, *S. malacitanus*, S. *pyrenaicus*, *S. torgalensis* and *S. valentinus*. The *Squalius alburnoides* complex, which is of hybrid origin, has been excluded given its highly complex evolutionary history and the presence of several ploidy levels and modes of reproduction (Alves, Coelho, & Collares-Pereira, 1997; Carmona, Sanjur, Doadrio, Machordom, & Vrijenhoekm, 1997; Cunha, Coelho, Carmona, & Doadrio, 2004; Sousa-Santos, Collares-Pereira, & Almada, 2006, 2007). The only Iberian representative of the Euroasiatic lineage is *S. laietanus*, which is phylogenetically related to some of the Greek species of the genus. This species has been used as an outgroup in the phylogenetic analyses of this and other studies dealing with the Mediterranean lineage (Doadrio, Kottelat, et al., 2007; Perea et al., 2010; Sanjur et al., 2003). According to previous mitochondrial phylogenies, the most divergent clade within the Iberian *Squalius* species of the Mediterranean lineage is the one including *S. aradensis* and *S. torgalensis*, which inhabit small basins in the south-western quadrant of the peninsula in Portugal (Almada & Sousa-Santos, 2010; Doadrio & Carmona, 2003; Perea, Cobo-Simón, et al., 2016; Sanjur et al., 2003; Sousa-Santos et al., 2016, 2019). A second mitochondrial monophyletic group is composed of *S. carolitertii* and *S. castellanus*, while a third consists of *S. pyrenaicus*, *S. valentinus* and *S. malacitanus* (Doadrio, Perea, et al., 2007; Perea et al., 2010; Perea, Cobo-Simón, et al., 2016). Recent analyses, however, have found discordances between these phylogenies and those based on nuclear markers of populations from Portugal (Sousa-Santos et al., 2019; Waap et al., 2011), suggesting the presence of mitonuclear discordances across the entire distribution range of *Squalius* in the Iberian Peninsula. They also highlight the need for more robust phylogenetic hypotheses for the group.

The main goals of this study are to disentangle the evolutionary history and the phylogenetic relationships of the Iberian *Squalius* species belonging to the Mediterranean lineage across its entire distribution range by analysing (a) the complete mitochondrial cytochrome *b* gene, (b) the concatenated supermatrix of six nuclear genes and (c) the coalescent-based species-tree obtained from the six nuclear genes. The overall aim of this study is to obtain a robust phylogenetic hypothesis for the genus *Squalius* by comparing the performances of the different methodologies used for analysis.

2 | MATERIAL AND METHODS

2.1 | Sampling and laboratory procedures

A total of 134 specimens representing species of Iberian *Squalius* (Mediterranean lineage) from across its entire distribution range (63 localities belonging to 35 basins) were analysed (Figure 1 and Table S1). The species *S. laietanus* of the Euroasiatic evolutionary lineage (Sanjur et al., 2003), and one species of *Petroleuciscus* (*P. smyrnaeus*), the sister genus of *Squalius* that was formerly considered as the third evolutionary lineage of *Squalius* (Perea et al., 2010; Sanjur et al., 2003), were used as outgroups to root the phylogenetic tree of the Iberian *Squalius* species of the Mediterranean lineage. Fish were captured by electrofishing under local authority permission, fin clipped and then returned to the stream after morphological identification. Fin clips were preserved in 95% ethanol. All voucher samples were stored at the DNA collection of the Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain, and at the MARE–ISPA collection, Lisbon, Portugal.

The complete mitochondrial cytochrome *b* gene (*MT-CYB*; 1,140 bp) was amplified from at least two individuals of all analysed populations of Iberian *Squalius* species. This mitochondrial DNA marker has proven useful to solve phylogenetic relationships within this genus (e.g., Almada & Sousa-Santos, 2010; Doadrio & Carmona, 2003). Six nuclear genes (*RAG1*, *S7*, *EFA1α*, *EGR2b*, *RHO* and *ACTB*) were amplified from 105 individuals. Gene lengths, PCR primers and references are presented in Table 1. The final alignment length (4,871 bp including gaps) included all nuclear data for the 105 *Squalius* individuals and specimens of the two outgroups. PCR conditions used followed Perea et al. (2010) and Waap et al., (2011). Dream Taq Green PCR Master Mix 2× (Thermo Fisher) was used in PCRs. PCR products were checked on 1% agarose gels, purified using ExoSAP-IT (USB) and then directly sequenced by MACROGEN SPAIN (Madrid; http://www.macrogen.com/esp). All new sequences (mitochondrial haplotypes and nuclear alleles) were deposited in GenBank (accession numbers: *MT-CYB*: MT008486–MT008603; *RAG1*: MT008604–MT008704, *S7*: MT00855–MT008805, *EFA1α*: MT008910-MT009018, *EGR2b*: MT051740–MT051843, *RHO*: MT008806–MT008909, *ACTB*: MT051635–MT051739; more detailed in Table S1).

2.2 | Phylogenetic analyses based on the cytochrome b gene and the concatenated nuclear supermatrix

Mitochondrial and nuclear sequences were aligned using the default MUSCLE parameters in Geneious R10.3 (Kearse et al., 2012). Heterozygous positions of nuclear genes were first coded using the IUPAC ambiguity code and then, to resolve alleles, the nuclear data were phased with the PHASE algorithm (Stephens & Donnelly, 2003; Stephens, Smith, & Donnelly, 2001) in DNAsp v.5.0 (Librado & Rozas, 2009) using a probability threshold of 0.9. Phases of heterozygous individuals by indels were reconstructed using the method of Sousa-Santos, Robalo, Collares-Pereira, and Almada (2005).

The mitochondrial (*MT-CYB*) phylogeny was based on a subset of 134 individuals of Iberian *Squalius* species from all basins. Nuclear phylogenetic inference was also performed for 105 individuals on the basis of the six nuclear genes analysed. The incongruence length difference (ILD) test was carried out in PAUP* (Swofford, 2003), not rejecting the null hypothesis of homogeneity among partitions for nuclear genes ($p > .05$). Bayesian inference (BI) was performed using MrBayes v.3.2 (Ronquist et al., 2012).

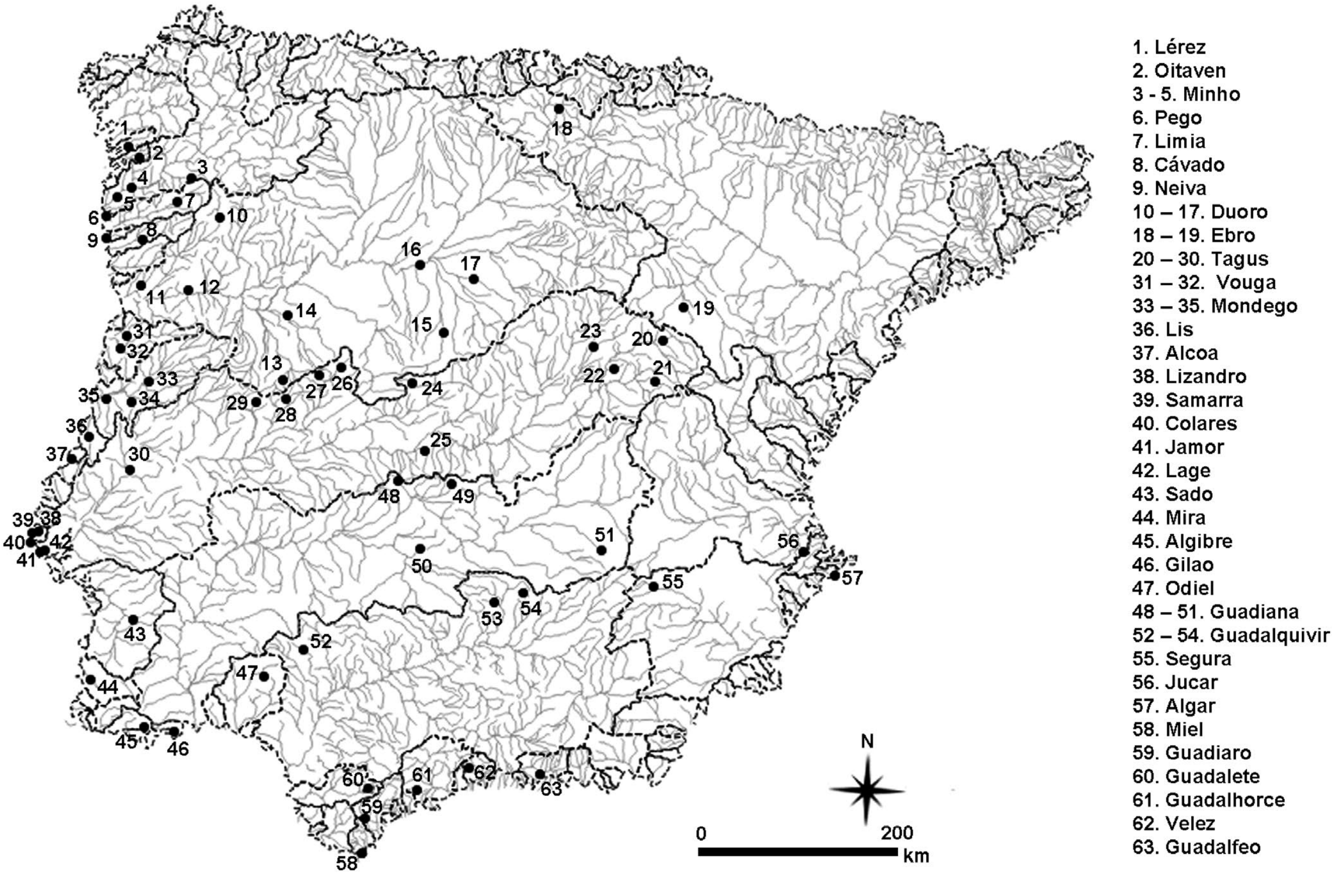

FIGURE 1 Sampling localities of the analysed *Squalius* populations in the Iberian Peninsula. Numbers correspond to the species and locality information provided in Table S1. Hydrological drainages to which sampling localities belong are indicated to the right of the map

TABLE 1 Genes analysed in Iberian *Squalius* species. The fragment length refers to the final alignment including gaps

Gene	Length (bp)	PCR primers used (Reference)
Cytochrome *b* (*MT-CYB*)	1,140	Glu F/Glu R (Zardoya & Doadrio, 1998)
Recombination activating gene–1 (*RAG1*), third exon	832	RAG1F/RAG9R (Quenouille, Bermingham, & Planes, 2004)
S7 Ribosomal protein gene (*S7*), first intron	907	S71F/S72R (Chow & Hazama, 1998)
Elongation factor 1-alpha (*EFA1α*)	660	Ef1af2/Ef1aR2 (Waap et al., 2011)
Rhodopsin (*RHO*)	792	RhForward/RhReverse (Waap et al., 2011)
Beta actin (*ACTB*)	933	B-actin-Fwd/B-actin-Rev (Robalo et al., 2006)
Early growth response 2b (*EGR2b*)	747	Egr2bF/Egr2bR (Waap et al., 2011)

For the *MT-CYB* analysis, the best-fit model of molecular evolution was obtained with jModeltest v.2.0 (Darriba, Taboada, Doallo, & Posada, 2012). For the concatenated analyses, the best-fit model of molecular evolution for each nuclear gene was estimated via the Akaike information criterion (AIC) in PartitionFinder v.1.1.1 (Lanfear, Calcott, Ho, & Guindon, 2012). The best partition schemes used in the different phylogenetic analyses based on the selected evolutionary models are presented in Table 2. Two simultaneous analyses were run for five million generations, each with four MCMC chains, sampling every 100 generations. Convergence was checked with Tracer v.1.6. (Rambaut,

TABLE 2 Evolutionary model partition schemes used in the different phylogenetic analyses

Phylogenetic analysis	Best partition scheme used for Bayesian Inference
MT-CYB	TIM2 + I + G
Nuclear (concatenated supermatrix and coalescent-based species-tree)	5 partitions (***RAG1***: TVMef + G; ***S7***: TrN + I + G; ***EFA1*** **+** ***ACTB***: TVM + G; ***RHO***: TVM + G; **EGR2b**: GTR + G)

Suchard, Xie, & Drummond, 2014). After discarding the first 10% of generations as burn-in, the 50% majority rule consensus tree and posterior probabilities were obtained. Maximum likelihood (ML) phylogenetic analyses of independent *MT-CYB* and the concatenated nuclear data sets were also conducted using RAxML (Stamatakis, 2006), as implemented in the T-REX online server. The substitution model used was GTR + G + I for *MT-CYB* and GTR + G + I for the nuclear markers. The rapid bootstrap algorithm (Stamatakis, Blagojevic, Nikolopoulos, & Antonopoulos., 2007) was also used. Node confidence in the ML analyses was estimated by rapid bootstrapping using a random seed (1,000 replicates). Uncorrected *p*-distances for *MT-CYB* and the nuclear genes were estimated in MEGA v.7.0 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) in order to quantify genetic differences among species and among populations within species.

2.3 | Coalescent species-tree analyses based on nuclear data

A multilocus species-tree analysis for the nuclear data set was carried out using *BEAST, as implemented in Beast software v.1.8.3 (Drummond, Suchard, Xie, & Rambaut, 2012; Heled & Drummond, 2010). Although the mitochondrial data were excluded from this analysis due to the discordant history of the two genomes (Sousa-Santos et al., 2019; Waap et al., 2011; this study), independent mitochondrial lineages were selected as units to test species delimitation as an a priori hypothesis is needed for the partitioned data. *BEAST analysis incorporates the MSC model into the Bayesian framework of species delimitation in order to treat the stochasticity of the lineage sorting process, demography and divergence time estimation, while also providing the species-tree and underlying gene trees (Heled & Drummond, 2010). In this context, species boundaries were tested between formally described Iberian *Squalius* species along with differentiated and highly supported mitochondrial lineages within *S. carolitertii* and *S. pyrenaicus*. We considered speciation probability values higher than 0.95 as strong support for all speciation events. A cloudogram of the *BEAST species-tree analysis was created from the posterior distribution of species-trees with DensiTree (Bouckaert, 2010).

A Bayesian nuclear multilocus species delimitation analysis under the MSC model was also conducted using BPP 2.2 (Yang, 2015; Yang & Rannala, 2010). This methodology accommodates lineage sorting due to ancestral polymorphism in the species-tree estimation (Rannala & Yang, 2003; Yang & Rannala, 2010). Within the BPP framework, different analyses may be considered to carry out species delimitation with the use of a user-specified guide tree to represent the species phylogeny. We used the topology of the species-tree obtained in the nuclear *BEAST analysis as our user-specified tree, as a means to validate this species-tree hypothesis. Therefore, analyses applying the reversible-jump MCMC algorithm to generate marginal posterior probabilities for species delimitation models were run. Assumptions of the MSC model, as implemented in BPP, include no recombination within a locus, free recombination between loci, no migration (gene flow) between species and neutral clock-like evolution. Recombination of the nuclear genes, used to infer the genetic structure of populations, was tested using the PHI test in Splitstree v.4.13 (Huson & Bryant, 2006), and no evidence of recombination was found ($p > .05$). Migration rate among lineages was estimated by calculating the virtual number of migrants per generation (*Nm*) using Arlequin v.3.5 (Excoffier & Lischer, 2010). The values obtained were always lower than 0.1 and, therefore, had little influence on the species delimitation analysis (Yang, 2015). Neutral clock-like evolution was tested in MEGA (Tamura et al., 2013) and was rejected for the genes *S7* and *ACTB* ($p < .05$). BPP analysis was performed with and without these two genes but no differences were observed in the results (not shown). Furthermore, given that this assumption may not be as important as previously suggested (Yang, 2015), all of the nuclear genes were considered in the BPP analyses. As BPP uses the JC69 mutation model (Jukes & Cantor, 1969), putative species should not present sequence divergences higher than 10% (Yang, 2015), a criterion met by the Iberian species of *Squalius* (Almada & Sousa-Santos, 2010; Doadrio, Kottelat, et al., 2007; Perea, Cobo-Simón, et al., 2016). A Dirichlet distribution ($\alpha = 2$) was set to account for variation in mutation rates among loci. Multiple analyses with different θ (ancestral population sizes) and τ_0 (divergence time among species) priors were conducted to determine how these parameters might influence species delimitation results in the BPP analyses (Bagley et al., 2015; Leaché & Fujita, 2010). The following hypotheses were selected: (1) small ancestral population sizes, [θ: G(2,2000)], and shallow divergence among species, [τ_0: G(2,2000)]; (2) large ancestral population sizes, [θ: G(1,10)], and deep divergence among species, [τ_0: G(1,10)]; (3) small ancestral population sizes, [θ: G(2,2000)], and deep divergence among species, [τ_0: G(1,10)]; (4) large ancestral population sizes, [θ: G(1,10)], and shallow divergence among species, [τ_0: G(2,2000)]. These four hypotheses were also tested considering the three mitochondrial lineages found in the Tagus Basin (*S. castellanus*, the Zêzere population ascribed to *S. carolitertii* and the Northern lineage of *S. pyrenaicus*) as different species or the same species. Each analysis was run for 500,000 MCMC generations from different starting seeds with a burn-in period of 50,000 and a sampling interval of 5, using the default fine-tuning parameter. Similar to the *BEAST analysis, a conservative approach was taken and a speciation probability value higher than 0.95 was considered as strong support for retaining a given node (i.e., ones indicating lineage splitting or a speciation event).

3 | RESULTS

The mitochondrial phylogenetic tree based on *MT-CYB* rendered congruent topologies in both the BI and ML analyses (Figure 2). Each of the Iberian *Squalius* species was recovered as a monophyletic group with high support except for *S. pyrenaicus*, which resolved as three distinct subclades (for localities, see Figure 1 and Table S1). One of the *S. pyrenaicus* subclades was a highly supported group constituted by the Tagus Basin, including small basins close to its mouth, together with two populations from the Ebro Basin (Northern lineage) in the central Iberian Peninsula. The second subclade consisted of a moderately supported group that clustered populations from the southern half of the Iberian Peninsula, including Guadiana, Gilão, Guadalquivir, Odiel, Guadalete and Segura basins and smaller southern basins close to the Strait of Gibraltar (i.e., Guadalhorce, Velez and Guadalfeo) (Southern lineage). The third subclade, also highly supported, was comprised of samples from the Sado Basin (Sado lineage) from Portugal in the western Iberian Peninsula. However, the phylogenetic relationships among the three *S. pyrenaicus* subclades were not highly supported in either the BI or ML analyses (Figure 2). *Squalius castellanus* resolved as the sister species to some individuals from the Zêzere River, a tributary of the Tagus in Portugal, that had been ascribed to *S. carolitertii* (Almada & Sousa-Santos, 2010). The other Zêzere river individuals clustered within the Northern lineage of *S. pyrenaicus*. *Squalius valentinus* resolved as sister species to the three *S. pyrenaicus* lineages, and both of these species diverged from the clade constituted by the two independent lineages of *S. malacitanus*. Statistical support for these relationships was high. *Squalius carolitertii* grouped together in a polytomy that included almost all of the analysed populations from the Douro Basin and other basins in the north-western Iberian quadrant (Figure 1), except for the Mondego and Alcoa basins, which constituted a differentiated group within *S. carolitertii*. One Tagus basin population (Alberche R.) and one individual from the Árrago River, also in the Tagus Basin, were more closely related to *S. carolitertii* than to *S. pyrenaicus* (Figure 2 and Table S1). Furthermore, some individuals from two populations, the Ompolveda R. (Upper Tagus Basin) and the Piedra R. (Ebro Basin), were more related to *S. pyrenaicus* and others to *S. castellanus* (Figure 2).

In general, the phylogenetic relationships based on the concatenation of all nuclear genes (4,871 bp final alignment including gaps) were not congruent with the mitochondrial phylogeny. The nuclear phylogeny showed one moderately supported subclade constituted by *S. castellanus* and *S. carolitertii* along with the Northern mitochondrial lineage of *S. pyrenaicus* and a second strongly supported subclade comprised of *S. malacitanus* and *S. valentinus* together with *S. pyrenaicus* populations from the southern half of the Iberian Peninsula and from Sado Basin (Figure 3). Statistical support for the lineage comprising the southern populations of *S. pyrenaicus* was low; however, the values supporting the monophyly of *S. valentinus*, *S. malacitanus* and the Sado lineage of *S. pyrenaicus* were high (Figure 3). The Sado lineage of *S. pyrenaicus* resolved as the sister group to the one formed by *S. malacitanus*, *S. valentinus* and the Southern lineage of *S. pyrenaicus*. Thus, according to the nuclear phylogeny, *S. pyrenaicus* is polyphyletic. The nuclear coalescent-based inference of the species-tree differed slightly from the concatenated nuclear supermatrix analysis. The *BEAST analysis also rendered the three lineages of *S. pyrenaicus* as polyphyletic; however, the coalescent-based species-tree analysis was able to resolve the relationships between *S. carolitertii* and the *Squalius* populations from the Tagus Basin (i.e., *S. castellanus*, the Northern lineage of *S. pyrenaicus* and the Zêzere population) (Figure 4a). In this case, the Northern lineage of *S. pyrenaicus* clustered with all Zêzere individuals and with *S. castellanus* in a highly supported clade, which in turn was the sister clade to *S. carolitertii* (Figure 4a). The relationships among the other analysed lineages were similar to those found in the concatenated supermatrix analysis, although with lower support values (Figure 4a). The Southern lineage of *S. pyrenaicus* was sister species of *S. valentinus*, and in turn, these two species were the sister clade to *S. malacitanus*. The nuclear species-tree analysis also confirmed the divergence of the distinctive *S. pyrenaicus* Sado lineage; however, this relationship was not highly supported by the *BEAST analysis (Figure 4). The cloudogram representing the posterior distribution of all of the estimated trees from the *BEAST analysis supports these results (Figure 4b).

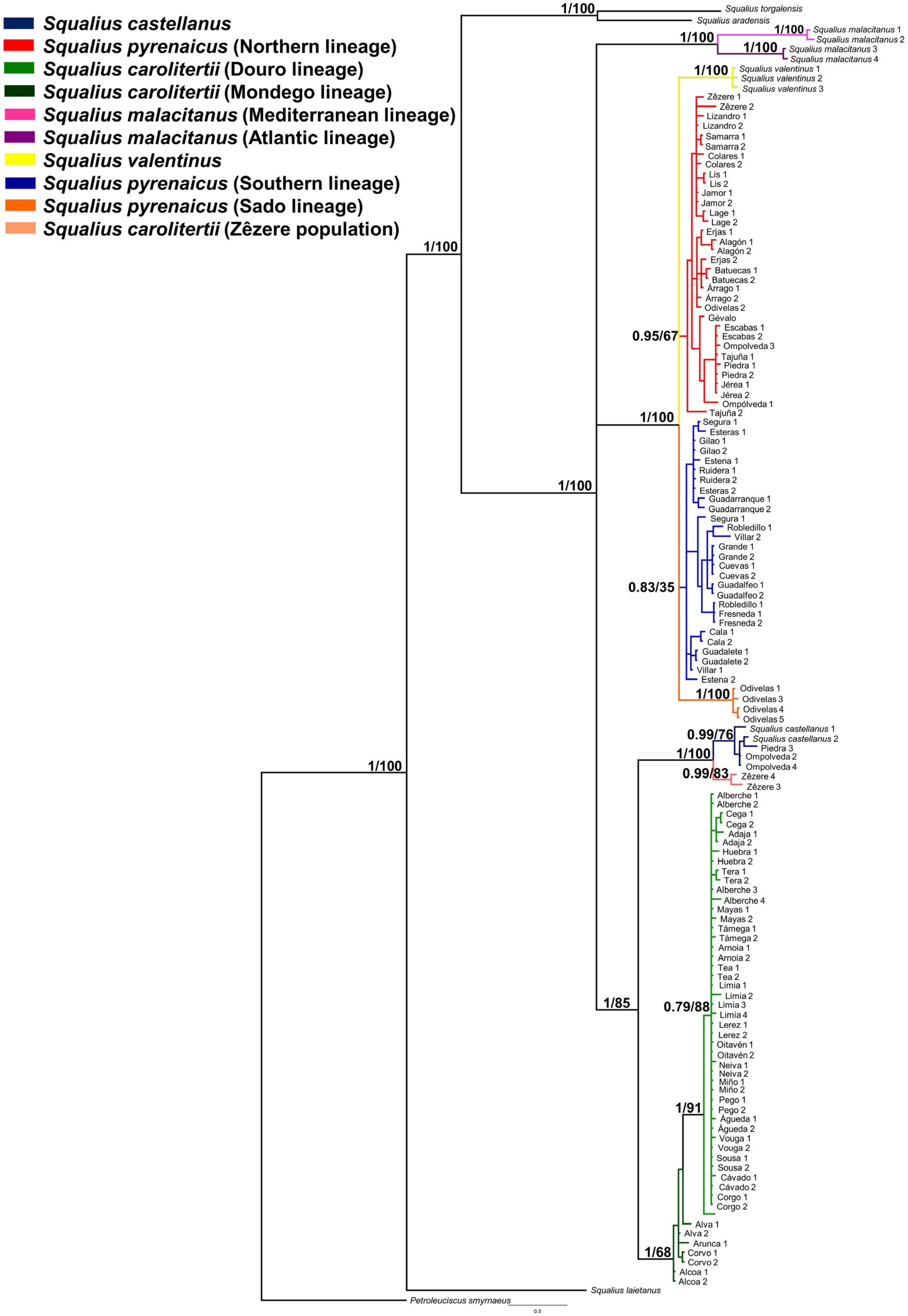

FIGURE 2 Mitochondrial phylogenetic tree (*MT-CTB*; 1,140 bp)-based Bayesian inference (BI) and maximum likelihood (ML) approaches. Bayesian posterior probabilities and ML bootstrap values are indicated before and after the dash, respectively

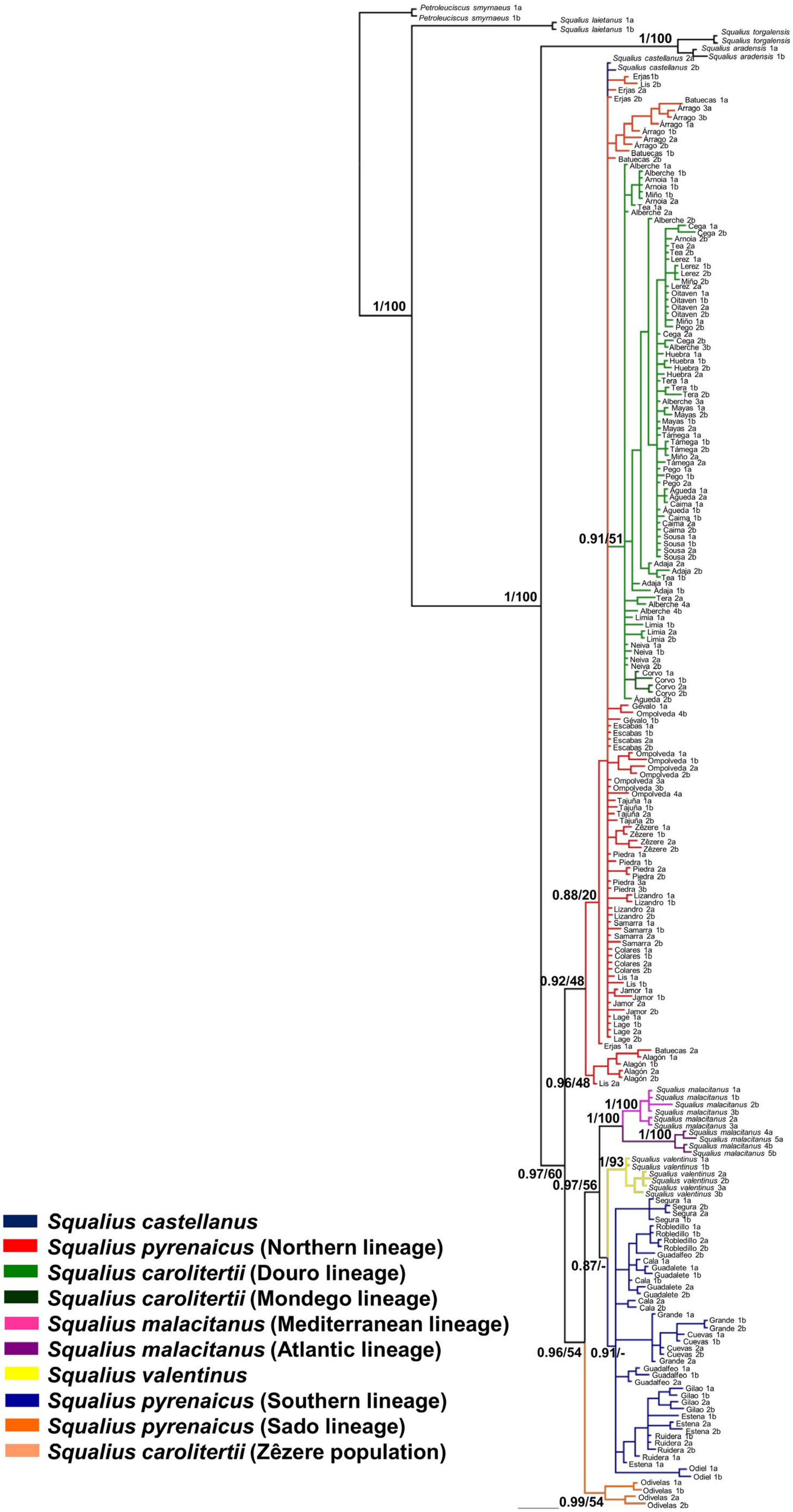

FIGURE 3 Concatenated nuclear phylogenetic tree based on six nuclear markers (4,871 bp) and Bayesian inference (BI) and maximum likelihood (ML) approaches. Bayesian posterior probabilities and ML bootstrap values are indicated before and after the dash, respectively

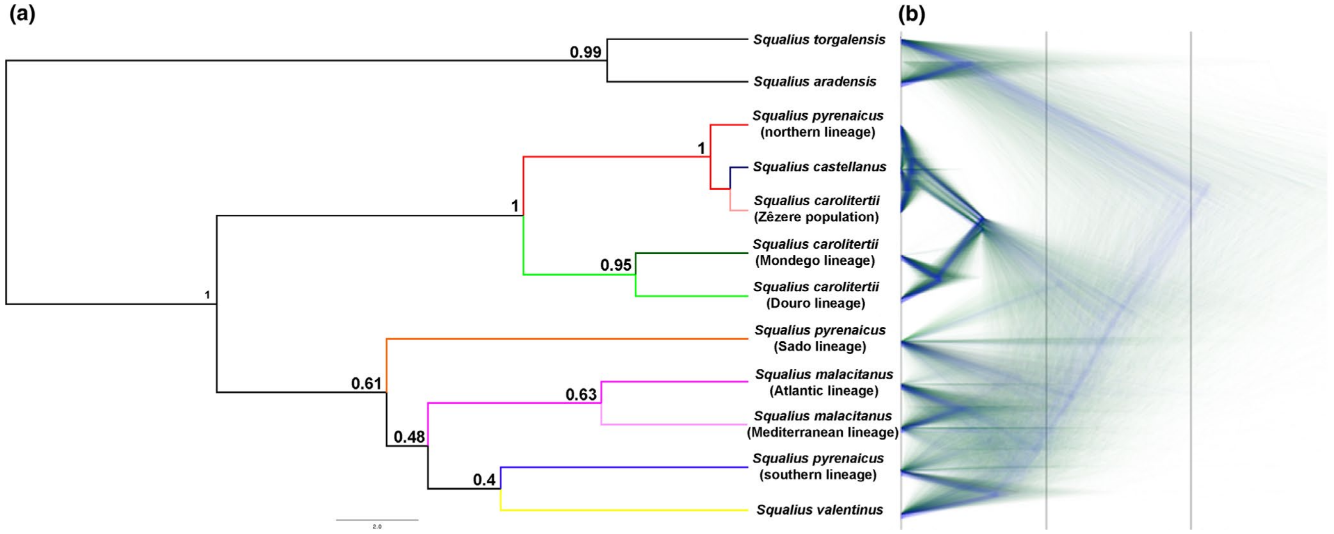

FIGURE 4 (a) Bayesian species tree delimitation for the six nuclear genes analysed. Posterior probabilities are indicated above branches. (b) Species tree consensus superimposed upon a cloudogram of the entire posterior distribution of the species-trees obtained from the *BEAST analysis. Areas in which the majority of trees agree in terms of topology and branch length are shown as darker areas (indicating well-supported clades)

BPP analyses rendered different results on the basis of the priors used for the parameters θ and τ_0. The results were especially remarkable in the delimitation of the Tagus basin populations when the three mitochondrial lineages were considered 'different species' in the models. When these same lineages were considered the 'same species' (as they share the same nuclear lineage) in the BPP analyses, the marginal posterior probability values estimated were high (pp = 1) for all four models tested (Table 3). Moreover, for some of the lineages or species, the marginal posterior probabilities were always equal to one, regardless of the model considered (i.e., *S. aradensis*, *S. torgalensis*, *S. malacitanus*, *S. valentinus*, and the Southern and Sado lineages of *S. pyrenaicus*) (Table 3).

Mitochondrial uncorrected *p* genetic distances ranged from 0.9% (between the two *S. carolitertii* lineages) to 7.5% [between *S. malacitanus* and some *S. carolitertii* populations (Zêzere R.) that resolved as the sister group to *S. castellanus* in the mitochondrial phylogeny] (Table S2). Indeed, the individuals ascribed to *S. carolitertii* from the Zêzere population in Portugal were highly differentiated from the other populations of *S. pyrenaicus* and *S. castellanus* (genetic distances ranged from 5.9%–6.2% and 3.3%–3.7%, respectively), when considering the entire Tagus Basin. The genetic distance of these Zêzere individuals from the *S. carolitertii* populations from the Douro and Mondego basins were 3.6% and 3.3%, respectively, and, similarly, the genetic distance between *S. castellanus* and the two lineages of *S. carolitertii* ranged from 3.6% to 3.7%. The genetic distance between the two *S. carolitertii* lineages was 0.9%. High genetic distances were also found between the two divergent lineages of *S. malacitanus* (3.2%), who together had even higher distances from the remaining lineages (from 5.9% to 7.5%). Genetic distances among the three divergent mitochondrial lineages of *S. pyrenaicus* ranged from 1.1% to 1.7%. The genetic distance between each of these three lineages and *S. valentinus* was 2.9% in the three pairwise comparisons. Nuclear uncorrected *p* genetic distances ranged from 0.0% to 1.0%, depending on the gene analysed, with the *S7* gene showing the highest genetic distances (Table S2).

4 | DISCUSSION

The phylogenetic relationships of the Iberian species belonging to the Mediterranean evolutionary lineage of *Squalius* were reconstructed on the basis of several mitochondrial and nuclear markers and two different analytical methodologies, concatenated matrix and coalescent-based species-tree approaches. This study includes the most comprehensive sampling of the studied species thus far, and the results obtained shed light on the phylogeny of the group and contribute to our understanding of the evolutionary history of Mediterranean Iberian *Squalius* along its distribution range.

Although the concatenated nuclear analyses did not recover *S. pyrenaicus* and *S. castellanus* as monophyletic, instead grouping them in a polytomy with the highly supported clade of *S. carolitertii* (Douro and Mondego lineages) and the Zêzere population from the Tagus Basin (ascribed to *S. carolitertii*), the mitochondrial phylogeny and the nuclear species-tree analysis based on the MSC model validated the identity of most of the Iberian *Squalius* species analysed (Coelho, Bogutskaya, Rodrigues, & Collares-Pereira, 1998; Doadrio, 1987; Doadrio & Carmona, 2006; Doadrio, Perea, et al., 2007). The only discordance with the taxonomy of Iberian *Squalius* species was the polyphyletic

TABLE 3 Marginal posterior probability values for different species delimitation models. Hypothesis (1) small ancestral population sizes, [θ: G(2,2000)], and shallow divergence among species, [τ_0: G(2,2000)]; hypothesis (2) large ancestral population sizes, [θ: G(1,10)], and deep divergence among species, [τ_0: G(1,10)]; hypothesis (3) small ancestral population sizes, [θ: G(2,2000)], and deep divergence among species, [τ_0: G(1,10)]; hypothesis (4) large ancestral population sizes, [θ: G(1,10)], and shallow divergence among species, [τ_0: G(2,2000)]. mpp, marginal posterior probabilities

Lineage	mpp Hypothesis 1	mpp Hypothesis 2	mpp Hypothesis 3	mpp Hypothesis 4
S. aradensis/S. torgalensis	1.000	1.000	1.000	1.000
S. carolitertii—Duero lineage	1.000	0.991	1.000	1.000
S. carolitertii—Mondego lineage	1.000	0.991	1.000	1.000
S. castellanus	1.000	0.994 (It estimates the Tagus Basin as the same unit)	0.43 (Tagus Basin as the same unit = 0.52)	0.48 (Tagus Basin as the same unit = 0.40)
Zêzere population	0.06 (Northern lineage of *S. pyrenaicus* + Zêzere = 0.93)	0.994 (It estimates the Tagus Basin as the same unit)	0.071 (Tagus Basin as the same unit = 0.52; Northern lineage of *S. pyrenaicus* + Zêzere = 0.36)	0.23 (Tagus Basin as the same unit = 0.40; Northern lineage of *S. pyrenaicus* + Zêzere = 0.34)
S. pyrenaicus (Northern lineage)	0.06 (Northern lineage of *S. pyrenaicus* + Zêzere = 0.93)	0.994 (It estimates the Tagus Basin as the same unit)	0.111 (Tagus Basin as the same unit = 0.52; Northern lineage of *S. pyrenaicus* + Zêzere = 0.36)	0.15 (Tagus Basin as the same unit = 0.40; Northern lineage of *S. pyrenaicus* + Zêzere = 0.34)
Tagus Basin: *S. pyrenaicus Northern lineage* + *S. castellanus* + Zêzere population	1.000	1.000	1.000	1.000
S. pyrenaicus (Southern lineage)	1.000	1.000	1.000	1.000
S. pyrenaicus (Sado lineage)	1.000	1.000	1.000	1.000
S. malacitanus (Atlantic lineage)	1.000	1.000	1.000	1.000
S. malacitanus (Mediterranean lineage)	1.000	1.000	1.000	1.000
S. valentinus	1.000	1.000	1.000	1.000

status of *S. pyrenaicus*, a consequence of the nuclear phylogenetic position of the three divergent lineages of this species.

Previous studies proposed the existence of three divergent genetic groups of *S. pyrenaicus* on the basis of mitochondrial and nuclear multilocus phylogenies that also included some Portuguese populations (Almada & Sousa-Santos, 2010; Sousa-Santos et al., 2019; Waap et al., 2011). The findings presented here, which analysed populations from the entire distribution range of the Iberian species, corroborate those studies. However, differences between mitochondrial and nuclear phylogenies (both concatenated matrix and coalescent-based species-tree) were observed in our study. The main difference was the phylogenetic position of the three *S. pyrenaicus* lineages: in the mitochondrial analysis, they were grouped within the same clade, but they were polyphyletic in both of the nuclear analyses. One of the *S. pyrenaicus* lineages comprised populations from the Tagus Basin, including small basins close to its mouth, and from part of the Ebro Basin (Northern lineage). In the nuclear analyses, this lineage was phylogenetically closer to *S. carolitertii* and *S. castellanus*. Populations from the southern half of the Iberian Peninsula constituted the second *S. pyrenaicus* lineage (Southern), which was more closely related to *S. valentinus* in the nuclear analyses. The third divergent lineage, which was highly supported, was the Sado population, which resolved as the sister group of the clade constituted by *S. malacitanus*, *S. valentinus* and the Southern lineage of *S. pyrenaicus*.

Although the concatenated nuclear matrix and the coalescent-based species-tree analyses both recovered *S. pyrenaicus* as polyphyletic, concatenation did not provide a sufficiently strong signal to resolve the phylogenetic relationships among the Northern lineage of *S. pyrenaicus*, *S. carolitertii* and *S. castellanus* compared with the coalescent-based species-tree analysis. The use of several nuclear genes has proven reliable for obtaining consistent phylogenetic relationships in Iberian *Squalius* species (Sousa-Santos et al., 2019; Waap et al., 2011). However, in our case, the BI and ML phylogenies of the six nuclear genes analysed could not integrate gene tree conflicts, such as those that stem from coalescence processes, to solve the relationships among *S. castellanus*, *S. carolitertii*, the Northern lineage of *S. pyrenaicus* and the population from Zêzere R. ascribed to *S. carolitertii*, which all had short branches in the phylogenetic tree topology. Theoretical and simulation studies have shown that concatenation methods can yield misleading results when internal branches are short (Degnan & Rosenberg, 2009; Kubatko & Degnan, 2007; Lambert, Reeder, & Wiens, 2015; Liu & Edwards, 2009), as seems to be the case for these *Squalius* lineages. Contrastingly, species-tree analysis has proven to be an effective tool to solve discordances between gene trees as coalescent methods can accommodate lineage sorting and topological heterogeneity among the trees (Heled & Drummond, 2010; Knowles & Carstens, 2007; Kubatko & Degnan, 2007; Liu & Edwards, 2015; Maddison & Knowles, 2006; O'Meara, 2010; Wu, 2012; Zhong, Liu, Yan, & Penny, 2013). This is an important feature considering that, across the genome, most genes, or their alleles, are not expected to be reciprocally monophyletic among lineages, particularly in those that have recently diverged and may retain ancestral polymorphisms (Fujita, Leaché, Burbrink, McGuire, & Moritz, 2012; Knowles & Carstens, 2007; Kubatko, Gibbs, & Bloomquist, 2011). As none of the analysed nuclear genes were independently able to separate *Squalius* lineages, incomplete lineage sorting could be a common pattern found in the genus for this marker type. Compared with mitochondrial DNA, which is fully resolved for *Squalius* (Almada & Sousa-Santos, 2010; Doadrio & Carmona, 2003, 2006; Doadrio, Kottelat, et al., 2007; Doadrio, Perea, et al., 2007; Perea, Cobo-Simón, et al., 2016; Sanjur et al., 2003), nuclear DNA may be particularly susceptible to incomplete lineage sorting due to differences in effective population size, implying that mitochondrial genomes complete the lineage sorting process faster than nuclear ones (Ballard & Whitlock, 2004; Brumfield, Beerli, Nickerson, & Edwards, 2003; Rosenberg, 2003). Given this context, the coalescent-based species-tree approaches performed in this study were able to resolve nuclear relationships with high support, providing a robust phylogeny for the Iberian *Squalius* lineages and supporting the distinction of the *Squalius* lineages from the Tagus Basin from *S. carolitertii*.

The BPP analyses rendered different results depending on the θ (ancestral population size) and τ_0 (divergence among species) parameters used, particularly regarding the estimation of the Tagus Basin and whether the three mitochondrial lineages found in the basin were considered as independent or not. Indeed, the only discordance found in these analyses was the status of the Tagus basin populations; the other lineages or species always yielded a posterior probability value of one. However, the species delimitation model that considered large ancestral population sizes and deep divergence among species was more resolved than the others. Based on the coalescent theory and empirical data, priors specifying large ancestral population sizes and recent divergence times among speciation events are expected to recover fewer species in BPP analyses (Salter, Carstens, & Hedin, 2013; Yang & Rannala, 2010). Nonetheless, the high support found in all four species delimitation models (pp = 1) when the Tagus basin lineages were considered as the same unit provides strong evidence that all of these populations (*S. castellanus*, Northern lineage of *S. pyrenaicus* and Zêzere population) belong to the same evolutionary lineage, thus questioning their taxonomic status. The BPP analyses also provide strong support to consider the two divergent lineages of *S. carolitertii* as different evolutionary units.

Our results suggest that the systematic position of the Northern lineage of *S. pyrenaicus* and the Zêzere population ascribed to *S. carolitertii* is not fully clarified and that the clade grouping these lineages with *S. castellanus* constitutes a single evolutionary unit that is distinct from the *S. carolitertii* populations from the Douro and Mondego basins. The Southern and Sado lineages of *S. pyrenaicus* possibly constitute different species, as was previously proposed (Almada & Sousa-Santos, 2010; Sousa-Santos et al., 2019; Waap et al., 2011). The two independent lineages of *S. carolitertii* are highly supported by both this and previous studies (Almada & Sousa-Santos, 2010; Brito et al., 1997; Sousa-Santos et al., 2019; Waap et al., 2011; this study). However, the consideration of the distinct lineages of *S. pyrenaicus* and *S. carolitertii* as different species has to be taken with caution given that genetic distances based on *MT-CYB* between Sado and the other *S. pyrenaicus* mitochondrial lineages ranged from 1.4% to 1.7% and those between the Mondego lineage and the remaining populations of *S. carolitertii* were always less than 1%. Thus, further studies that take into account the morphology or higher genomic levels of these populations are needed to elucidate the taxonomic status of these *Squalius* species, especially in the case of the polyphyletic *S. pyrenaicus*.

In contrast to *S. castellanus*, *S. carolitertii* and *S. pyrenaicus*, the phylogenetic relationships of the other Mediterranean *Squalius* species inhabiting the Iberian Peninsula were fully resolved and concordant in the mitochondrial and nuclear analyses. For instance, two species, *S. aradensis* and *S. torgalensis*, always constituted the most divergent lineage, and *S. malacitanus* and *S. valentinus* were always phylogenetically related to the Southern and Sado lineages of *S. pyrenaicus*. These results are congruent with previous studies except for the relationship of the Sado lineage, which has been uncertain (Almada & Sousa-Santos, 2010; Brito et al., 1997; Doadrio & Carmona, 2003, 2006; Perea et al., 2010; Perea, Cobo-Simón, et al., 2016; Sanjur et al., 2003; Sousa-Santos, Collares-Pereira, & Almada, 2007; Sousa-Santos et al., 2019; Waap et al., 2011).

Species delimitation analyses also support the existence of two well-differentiated *S. malacitanus* lineages, similar to the case in *S. carolitertii*. Coalescent-based species delimitation methods have proven effective to delimit cryptic lineages within morphologically conserved species (Domingos, Colli, Lemmon, Lemmon, & Beheregaray, 2017; Kotsakiozi et al., 2018; Leaché & Fujita, 2010; Niemiller, Near, & Fitzpatrick, 2012; Welton, Siler, Oaks, Diesmons, & Brown, 2013), as is the case of the aforementioned species. Indeed, the robustness of *BEAST and BPP in species delimitation analyses of different organisms has been broadly demonstrated (e.g., Bagley et al., 2015; Kubatko et al., 2011; Leaché & Fujita, 2010; Salter et al., 2013; Yang & Rannala, 2010; Zhang, Zhang, Zhu, & Yang, 2011).

For the nuclear clade comprised of *S. malacitanus*, *S. valentinus* and two of the *S. pyrenaicus* lineages (Southern and Sado), we observed lower support values in the Bayesian coalescent species-tree approach based on *BEAST than those from the concatenated supermatrix or coalescent BPP analyses in which the phylogenetic relationships among these lineages were fully solved. Although Bayesian analysis under the MSC model may help assess species delimitation hypotheses in the case of short branches or incomplete lineage sorting, the information provided by concatenated analyses must not be discarded, especially given that some coalescent species delimitation approaches are subjected to the particularities of taxonomic and sampling strategies, uneven sampling across lineages or the lack of certain sampled taxa, all of which may lead to inaccurate results (Bagley et al., 2015; Sites & Marshall, 2003). Although all of the Iberian lineages are represented in our study, which also included a representative sampling across the entire distribution range of some of the lineages, the number of individuals analysed for some of the putative species was small, as in the case of *S. malacitanus*, *S. valentinus* and the Sado lineage of *S. pyrenaicus.* The number of individuals sampled per species is one of the factor influencing phylogenetic reconstructions based on coalescent species-tree approaches (Maddison & Knowles, 2006); therefore, the lower support values found for these lineages in the *BEAST analysis may be explained by their small sampling numbers. However, the phylogenetic and phylogeographic relationships between *S. malacitanus* and *S. valentinus* have been thoroughly analysed (Perea, Cobo-Simón, et al., 2016; Perea & Doadrio, 2015). On the other hand, concatenation assumes that population processes do not have an effect on phylogenetic estimations; therefore, this approach may be most appropriate when few loci and few individuals are analysed, as in the case of *S. malacitanus*, *S. valentinus* or the Sado lineage of *S. pyrenaicus*. In contrast, coalescent methods may prove more accurate when multiple individuals per species are analysed (Corl & Ellegren, 2013; McCormack, Huang, & Knowles, 2009). However, in contrast to the *BEAST analysis, support values for these species or lineages was high in BPP analyses (pp = 1), a methodology proven to be robust with few loci and few number of individuals per loci (Zhang et al., 2011). The guide tree is a factor that can limit the accuracy of BPP analyses (Zhang et al., 2011), and given that the one used in our analysis was obtained from *BEAST, it is conceivable that the lower support values found for those lineages in the *BEAST species-tree could have impacted the results obtained in the BPP analysis that were used to validate the different species delimitation models. Nevertheless, the same phylogenetic relationships were highly supported in the tree topology obtained from the concatenated supermatrix analysis, and hence, the topology used as the input in the BPP analyses, and the inferences postulated from the results, should be considered reliable.

A sometimes unconsidered factor in the concatenation of nuclear genes is the use of phased or unphased alleles. Each heterozygous allele may have a different coalescent history and, therefore, may greatly influence the phylogenetic reconstruction (Andermann et al., 2018; Weisrock et al., 2012). Furthermore, concatenating allele sequences from unlinked loci randomly may lead to incorrect topologies (Degnan & Rosenberg, 2009; Kolaczkowski & Thornton, 2004). Intra-individual polymorphism is a rich source of phylogenetic information; however, most of the developed algorithms treat it as ambiguous or missing data, which could decrease the resolution of the phylogenetic tree topology (Potts, Hedderson, & Grimm, 2014). Coalescence-based species-tree methodologies can avoid this issue by considering phased alleles as different 'individuals', hence reconstructing the genealogy of the whole set of alleles of a gene (Liu, Xi, Wu, Davis, & Edwards, 2015; Mirarab & Warnow, 2015). Indeed, allele phasing may improve the estimation of gene or species-trees or even divergence times (Eriksson, Blanco-Pastor, Sousa, Bertrand, & Pfeil, 2017; Lischer, Excoffier, & Heckel, 2014; Schrempf, Minh, Maio, Haeseler, & Kosiol, 2016). However, empirical studies have suggested that allele phasing does not have a clear impact on phylogenetic performance (e.g., Kates, Johnson, Gardner, Zerega, & Wickett, 2018). In this study, we phased the nuclear sequences in order to incorporate the phylogenetic information of each allele into the analyses. We confirmed that random linking of alleles for the different genes of each individual did not have a negative impact on the phylogenetic reconstruction based on the concatenated supermatrix analysis (i.e., all alleles were nested within their expected clade).

Another source of discordance between gene trees is when mitochondrial and nuclear topologies are different, as in the case of the Iberian *Squalius* species analysed here. The mitonuclear discordance found in this group of species suggests the presence of past hybridization events within the genus in the Iberian Peninsula, most likely as a result of the geomorphological rearrangements of Iberian paleobasins, as previously suggested by Sousa-Santos et al., (2019). Introgressive hybridization in freshwater fishes has been broadly reported (e.g., Allen, Anderson, Mee, Coombs, & Rogers, 2016; Kwan, Ko, Jeon, Kim, & Won, 2018; Machordom, Berrebi, & Doadrio, 1990; Magalhães, Ornelas-García, Leal-Cardin, Ramírez, & Barluenga, 2015) and is particularly common in *Squalius*, and can even involve different genera along its distribution range (Almodovar, Nicola, Leal, Torralva, & Elvira, 2012; Doadrio & Carmona, 1998; Freyhof, Lieckfeldt, Pitra, & Ludwig, 2005; Perea, Vukic, Sanda, & Doadrio, 2016; Sousa-Santos, Collares-Pereira, & Almada, 2006; Sousa-Santos, Matono, Da Silva, & Ilhéu, 2018). The most complex example of hybridization within the genus is found in *S. alburnoides* (e.g., Alves et al., 1997; Carmona et al., 1997; Cunha et al., 2004; Cunha, Doadrio, Abrantes, & Coelho, 2011; Robalo, Sousa-Santos, Levy, & Almada, 2006; Sousa-Santos et al., 2007).

Bayesian species delimitation may be sensitive to ancestral population size priors or gene flow between species, which may, in turn, lead to lower posterior probability values (Yu et al., 2012; Zhang et al., 2011). Although hybridization and gene flow parameters have been incorporated in some of the algorithms developed for coalescence-based species-tree estimations (Gerard, Gibbs, & Kubatko, 2011; Meng & Kubatko, 2011), some methodologies, including the ones used in this study, are not able to incorporate these parameters or other such as genetic introgression (Leaché, Harris, Rannala, & Yang, 2014; Olave, Avila, Sites, & Morando, 2016; Yu et al., 2012), which likely explains the mitonuclear discordance found in this study for the three lineages of *S. pyrenaicus*. Concatenated analyses also have their own caveats: they tend to be biased for the marker with the higher number of informative sites, usually the mitochondrial one (e.g., Mendes, Harris, Carranza, & Salvi, 2016; but see Fisher-Reid & Wiens, 2011). For this reason, mitochondrial and nuclear markers cannot be analysed together to estimate the species-tree of the Iberian *Squalius* species belonging to the Mediterranean lineage using coalescent species-tree or concatenated supermatrix approaches. Rather, their evolutionary history should be estimated independently for each genome. Overall, our results support this conclusion and the need to integrate different analytical methodologies of phylogenetic reconstruction to obtain the most robust phylogeny possible, such as the one we present here for the Iberian species of *Squalius*.

5 | CONCLUSIONS

This study emphasizes the importance of using different methodologies, such as concatenated supermatrix and coalescent-based species-tree approaches, to obtain a robust phylogenetic hypothesis for the Iberian species of the genus *Squalius* belonging to the Mediterranean lineage. Our results highlight the need for a systematics review of the group due to the polyphyletic status of some of the species. Moreover, the identification of a hybridization event between some lineages underscores the need to analyse mitochondrial and nuclear markers independently and to take into account both concatenated supermatrix and coalescent species-tree approaches. Our comparison between these two methodologies shows that coalescent-based approaches help improve our understanding of the phylogenetic relationships and systematics of the group. However, the concatenated supermatrix approach should not be discarded, as it may actually prove to be more reliable than coalescent ones when the number of specimens per lineage is low.

ACKNOWLEDGEMENTS

We thank P. Garzón, J.L. González and B. Prieto, for field assistance within the Project 'Basis for the monitoring of Spanish Freshwater Fishes' supported by MARM

(Ministerio de Medio Ambiente y Medio Rural y Marino) in which part of the samples used in this study was collected. L. Alcaraz helped with laboratory procedures. This study has been also supported by the Spanish Ministry of Science and Innovation (CGL2010-15231BOS) and by the management commission funded by the Duero Hydrographic Confederation in Spain (Análisis y estudios filogeográficos y de variabilidad genética de endemismos ibéricos de ciprínidos: Life13 nat/es/000772 'Actuaciones para la protección y conservación de ciprínidos ibéricos de interés comunitario'). S. Perea was awarded with the Angel Cabrera Prize to the best predoctoral CV in 2015 granted by the Biodiversity and Evolutionary Department of the Museo Nacional de Ciencias Naturales—CSIC, which covered part of the laboratory expenses of this study.

ORCID

Silvia Perea https://orcid.org/0000-0003-0436-8577
Carla Sousa-Santos https://orcid.org/0000-0002-4513-5553
Joana Robalo https://orcid.org/0000-0002-7470-0574
Ignacio Doadrio https://orcid.org/0000-0003-4863-9711

REFERENCES

Allen, B. E., Anderson, M. L., Mee, J. A., Coombs, M., & Rogers, S. M. (2016). Role of genetic background in the introgressive hybridization of rainbow trout (*Oncorhynchus mykiss*) with Westslope cutthroat trout (*O. clarkia lewisi*). *Conservation Genetics*, *17*, 521–531. https://doi.org/10.1007/s10592-015-0801-6

Almada, V., & Sousa-Santos, C. (2010). Comparisons of the genetic structure of *Squalius* populations (Teleostei, Cyprinidae) from rivers with contrasting histories, drainage areas and climatic conditions based on two molecular markers. *Molecular Phylogenetics and Evolution*, *57*, 924–931. https://doi.org/10.1016/j.ympev.2010.08.015

Almodovar, A., Nicola, G. G., Leal, S., Torralva, M., & Elvira, B. (2012). Natural hybridization with invasive bleak Alburnus alburnus threatens the survival of Iberian endemic calandino *S. alburnoides* complex and southern Iberian chub *Squalius pyrenaicus*. *Biological Invasions*, *14*, 2237–2242. https://doi.org/10.1007/s10530-012-0241-x

Alves, M. J., Coelho, M. M., & Collares-Pereira, M. J. (1997). The *Rutilus alburnoides* complex (Cyprinidae): Evidence for a hybrid origin. *Journal of Zoological Systematics and Evolutionary Research*, *35*, 1–10. https://doi.org/10.1111/j.1439-0469.1997.tb00398.x

Andermann, T., Fernandes, A. M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., … Antonelli, A. (2018). Allele phasing greatly improves the phylogenetic utility of ultraconserved elements. *Systematic Biology*, *68*, 32–46. https://doi.org/10.1093/sysbio/syy039

Bagley, J. C., Alda, F., Breitman, M. F., Bermingham, E., van den Berghe, E. P., & Johnson, J. B. (2015). Assessing species boundaries using multilocus species delimitation in a morphologically conserved group of neotropical freshwater fishes, the *Poecilia sphenops* species complex (Poeciliidae). *PLoS ONE*, *10*, e0121139. https://doi.org/10.1371/journal.pone.0121139

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729–744. https://doi.org/10.1046/j.1365-294X.2003.02063.x

Bouckaert, R. R. (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, *26*, 1372–1372. https://doi.org/10.1093/bioinformatics/btq110

Briolay, J., Galtier, N., Brito, R. M., & Bouvet, Y. (1998). Molecular phylogeny of Cyprinidae inferred from cytochrome *b* DNA sequences. *Molecular Phylogenetics and Evolution*, *9*, 100–108. https://doi.org/10.1006/mpev.1997.0441

Brito, R. M., Briolay, J., Galtier, N., Bouvet, Y., & Coelho, M. M. (1997). Phylogenetic relationships within the genus Leuciscus (Pisces: Cyprinidae) in Portuguese fresh waters, based on mitochondrial DNA cytochrome *b* sequences. *Molecular Phylogenetics and Evolution*, *8*, 435–442. https://doi.org/10.1006/mpev.1997.0429

Brumfield, R. T., Beerli, P., Nickerson, D. A., & Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution*, *18*, 249–256. https://doi.org/10.1016/S0169-5347(03)00018-1

Campbell, M. A., Alfaro, M. E., Belasco, M., & López, J. A. (2017). Early-branching euteleost relationships: Areas of congruence between concatenation and coalescent model inferences. *PeerJ*, *5*, e3548. https://doi.org/10.7717/peerj.3548

Carmona, J. A., Sanjur, O., Doadrio, I., Machordom, A., & Vrijenhoekm, R. C. (1997). Hybridogenetic reproduction and maternal ancestry of polyploidy Iberian fish: The *Tropidophoxinellus alburnoides* complex. *Genetics*, *146*, 983–999.

Chow, S., & Hazama, K. (1998). Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*, *7*(9), 1255–1256.

Coelho, M. M., Bogutskaya, N. G., Rodrigues, J. A., & Collares-Pereira, M. J. (1998). *Leuciscus torgalensis* and *Leuciscus aradensis*, two new cyprinids from Portuguese fresh waters. *Journal of Fish Biology*, *52*, 937–950. https://doi.org/10.1111/j.1095-8649.1998.tb00594.x

Coelho, M. M., Brito, R. M., Pacheco, T. R., Figuereido, D., & Pires, A. M. (1995). Genetic variation and divergence of *Leuciscus pyrenaicus* and *L. carolitertii* (Pisces, Cyprinidae). *Journal of Fish Biology*, *47*, 243–258. https://doi.org/10.1111/j.1095-8649.1995.tb06059.x

Corl, A., & Ellegren, K. (2013). Sampling strategies for species trees: The effects of phylogenetic inference of the number of genes, number of individuals and whether loci are mitochondrial, sex-linked or autosomal. *Molecular Phylogenetics and Evolution*, *67*, 358–366. https://doi.org/10.1016/j.ympev.2013.02.002

Cunha, C., Coelho, M. M., Carmona, J. A., & Doadrio, I. (2004). Phylogeographical insights into the origins of the *Squalius alburnoides* complex via multiple hybridization events. *Molecular Ecology*, *13*(9), 2807–2817. https://doi.org/10.1111/j.1365-294X.2004.02283.x

Cunha, C., Doadrio, I., Abrantes, J., & Coelho, M. M. (2011). The evolutionary history of the allopolyploid *Squalius alburnoides* (Cyprinidae) complex in the northern Iberian Peninsula. *Heredity*, *106*, 100–112. https://doi.org/10.1038/hdy.2010.70

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, *24*(6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009

Doadrio, I. (1987). *Leuciscus carolitertii* n. sp. from the Iberian Peninsula (Pisces: Cyprinidae). *Senckenbergiana Biologica*, *68*(4–6), 301–309.

Doadrio, I., & Carmona, J. A. (1998). Genetic divergence in Greek populations of the genus Leuciscus and its evolutionary and biogeographical implications. *Journal of Fish Biology*, *53*, 591–613. https://doi.org/10.1111/j.1095-8649.1998.tb01004.x

Doadrio, I., & Carmona, J. A. (2003). Testing freshwater Lago Mare dispersal theory on the phylogeny relationships of Iberian cyprinid genera *Chondrostoma* and *Squalius* (Cypriniformes, Cyprinidae). *Graellsia*, *59*(2–3), 457–473. https://doi.org/10.3989/graellsia.2003.v59.i2-3.260

Doadrio, I., & Carmona, J. A. (2006). Phylogenetic overview of the genus *Squalius* (Actinopterygii, Cyprinidae) in the Iberian Peninsula, with description of two new species. *Cybium*, *30*, 199–214.

Doadrio, I., Kottelat, M., & de Sostoa, A. (2007). *Squalius laietanus*, a new species of cyprinid fish from north-eastern Spain and southern France (Teleostei: Cyprinidae). *Ichthyological Exploration of Freshwaters*, *18*, 247–256.

Doadrio, I., Perea, S., & Alonso, F. (2007). A new species of the genus Squalius Bonaparte, 1837 (Actinopterygii, Cyprinidae) from the Tagus River Basin (Central Spain). *Graellsia*, *63*, 89–100. https://doi.org/10.3989/graellsia.2007.v63.i1.83

Doadrio, I., Perea, S., Garzón-Heydt, P., & González, J. L. (2011). *Ictiofauna Continental Española. Bases para su seguimiento*. Madrid, Spain: DG Medio Natural y Política Forestal. MARM, 616 pp.

Domingos, M. C. B. F., Colli, G. R., Lemmon, A., Lemmon, E. M., & Beheregaray, L. B. (2017). In the shadows: Phylogenomics and coalescent species delimitation unveil cryptic diversity in a Cerrado endemic lizard (Squamata: *Tropidurus*). *Molecular Phylogenetics and Evolution*, *107*, 455–465. https://doi.org/10.1016/j.ympev.2016.12.009

Drummond, A., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973. https://doi.org/10.1093/molbev/mss075

Edwards, S. V. (2009). Is a new and general theory of molecular systematics merging? *Evolution*, *63*, 1–19. https://doi.org/10.1111/j.1558-5646.2008.00549.x

Edwards, S. V., Liu, L., & Pearl, D. K. (2007). High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the USA*, *104*, 5936–5941. https://doi.org/10.1073/pnas.0607004104

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

Eriksson, J. S., Blanco-Pastor, J. L., Sousa, F., Bertrand, Y. J., & Pfeil, B. E. (2017). A cryptic species produced by autopolyploidy and subsequent introgression involving *Medicago prostrata* (Fabaceae). *Molecular Phylogenetics and Evolution*, *107*, 367–381. https://doi.org/10.1016/j.ympev.2016

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite vol 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fisher-Reid, M. C., & Wiens, J. J. (2011). What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? Lessons from *Plethodon* salamanders and 13 other vertebrate clades. *BMC Evolutionary Biology*, *11*, 300. https://doi.org/10.1186/1471-2148-11-300

Freyhof, J., Lieckfeldt, D., Pitra, C., & Ludwig, A. (2005). Molecules and morphology: Evidence for introgression of mitochondrial DNA in Dalmatian cyprinids. *Molecular Phylogenetic and Evolution*, *37*, 347–354. https://doi.org/10.1016/j.ympev.2005.07.018

Fricke, R., Eschmeyer, W. N., & van der Laan, R. (Eds.). (2020). *Eschmeyer's catalog of fishes: genera, species, references*. Electronic version. Retrieved from http://researchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.

Froese, R., & Pauly, D. (2019). *FishBase*. World Wide Web electronic publication. Retrieved from www.fishbase.org, version.

Fujita, M. K., Leaché, A. D., Burbrink, F. T., McGuire, J. A., & Moritz, C. (2012). Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution*, *27*(9), 480–488. https://doi.org/10.16/j.tree.2012.04.012

Gadagkar, S. R., Rosenberg, M. S., & Kumar, S. (2005). Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *304B*, 64–74. https://doi.org/10.1002/jez.b.21026

Gatesy, J., Meredith, R. W., Janecka, J. E., Simmons, M. P., Murphy, W. J., & Springer, M. S. (2017). Resolution of a concatenation/coalescence kerfuffle: Partitioned coalescence support and a robust family-level tree for Mammalia. *Cladistics*, *33*, 295–332. https://doi.org/10.1111/cla.12170

Geiger, M. F., Herder, F., Monaghan, M. T., Almada, V., Barbieri, R., Bariche, M., … Freyhof, J. (2014). Spatial heterogeneity in the Mediterranean Biodiversity Hotspot affects barcoding accuracy of its freshwater fishes. *Molecular Ecology Resources*, *14*, 1210–1221. https://doi.org/10.1111/1755-0998.12257

Gerard, D., Gibbs, H. L., & Kubatko, L. S. (2011). Estimating hybridization in the presence of coalescence using phylogenetic intraspecific sampling. *BMC Evolutionary Biology*, *11*, 291. https://doi.org/10.1186/1471-2148-11-291

Harrington, R. C., & Near, T. J. (2012). Phylogenetic and coalescent strategies of species delimitation in snubnose darters (Percidae: *Etheostoma*). *Systematic Biology*, *61*, 63–79. https://doi.org/10.1093/sysbio/syr077

Heled, J., & Drummond, A. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*(3), 570–580. https://doi.org/10.1093/molbev/msp274

Hillis, D. M., Huelsenbeck, J. P., & Cunningham, C. W. (1994). Application and accuracy of molecular phylogenies. *Science*, *264*, 671–677. https://doi.org/10.1126/science.8171318

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. https://doi.org/10.1093/molbev/msj030

Jukes, T. H., & Cantor, C. R. (1969). *Evolution of protein molecules*. New York, NY: Academic Press, pp. 21–132.

Kates, H. R., Johnson, M. G., Gardner, E. M., Zerega, N. J., & Wickett, N. J. (2018). Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *American Journal of Botany*, *105*, 404–416. https://doi.org/10.1002/ajb2.1068

Kautt, A. F., Machado-Schiaffino, G., & Meyer, A. (2016). Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. *PLoS Genetics*, *12*, e1006157. https://doi.org/10.1371/journal.pgen.1006157

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*, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Knowles, L. L., & Carstens, B. C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*, *56*, 887–895. https://doi.org/10.1080/10635150701701091

Knowles, L. L., & Kubatko, L. S. (2011). *Estimating species trees: Practical and theoretical aspects*. Hoboken, NJ: Wiley-Blackwell. 232 pp.

Kolaczkowski, B., & Thornton, J. W. (2004). Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, *431*, 980–984. https://doi.org/10.1038/natur

Kotsakiozi, P., Jablonski, D., Ilgaz, Ç., Kumlutaş, Y., Avcı, A., Meiri, S., … Poulakakis, N. (2018). Multilocus phylogeny and coalescent species delimitation in Kotschy's gecko, *Mediodactylus kotschyi*: Hidden diversity and cryptic species. *Molecular Phylogenetics and Evolution*, *125*, 177–187. https://doi.org/10.1016/j.ympev.2018.03.022

Kottelat, M., & Freyhof, J. (2007). *Handbook of European freshwater fishes*. Delémont, Switzerland: Cornol. xiv+646 pp.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, *56*, 17–24.

Kubatko, L. S., Gibbs, H. L., & Bloomquist, E. W. (2011). Inferring species level phylogenies and taxonomic distinctiveness using multilocus data inn Sistrurus rattlesnakes. *Systematic Biology*, *4*, 393–409. https://doi.org/10.1080/10635150601146041

Kwan, Y.-S., Ko, M.-H., Jeon, Y.-S., Kim, H.-J., & Won, Y.-J. (2018). Bidirectional mitochondrial introgression between Korean cobitid fish mediated by hybridogenetic hybrids. *Ecology and Evolution*, *9*, 1244–1254. https://doi.org/10.1002/ece3.4830

Lambert, S. M., Reeder, T. W., & Wiens, J. J. (2015). When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. *Molecular Phylogenetics and Evolution*, *82*, 146–155. https://doi.org/10.1016/j.ympev.2014.10.004

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

Leaché, A. D., & Fujita, M. K. (2010). Bayesian species delimitation in west African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society B*, *277*, 3071–3077. https://doi.org/10.1098/rspb.2010.0662

Leaché, A. D., Harris, R. B., Rannala, R., & Yang, Z. (2014). The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, *63*, 17–30. https://doi.org/10.1093/sysbio/syt049

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452. https://doi.org/10.1093/bioinformatics/byb187

Lischer, H. E., Excoffier, L., & Heckel, G. (2014). Ignoring heterozygous sites biases phylogenomic estimates of divergence times: Implications for the evolutionary history of *Microtus* voles. *Molecular Biology and Evolution*, *31*, 817–831.

Liu, L., & Edwards, S. V. (2009). Phylogenetic analysis in the anomaly zone. *Systematic Biology*, *58*, 452–460. https://doi.org/10.1093/sysbio/syp034

Liu, L., & Edwards, S. V. (2015). Comment on “Statistical binning enables an accurate coalescent-based estimation of the avian tree”. *Science*, *350*, 171. https://doi.org/10.1126/science.aaa7343

Liu, L., Xi, Z., Wu, S., Davis, C. C., & Edwards, S. V. (2015). Estimating phylogenetic trees from genome-scale data. *Annals of the New York Academy of Science*, *1360*, 36–53. https://doi.org/10.1111/nyas.12747

Liu, L., Yu, L., Pearl, D. K., & Edwards, S. V. (2009). Estimating species phylogenies using coalescence times among sequences. *Systematic Biology*, *58*, 468–477. https://doi.org/10.1093/sysbio/syp031

Machordom, A., Berrebi, P., & Doadrio, I. (1990). Spanish barbell hybridization detected using enzymatic markers: *Barbus meridionalis* Risso x *Barbus haasi* Mertens (Osteichthyes, Cyprinidae). *Aquatic Living Resources*, *3*, 295–303.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, *46*, 523–536. https://doi.org/10.1093/sysbio/46.4.523

Maddison, W. P., & Knowles, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*, *55*, 21–30. https://doi.org/10.1080/10635150500354928

Magalhães, I. S., Ornelas-García, C. P., Leal-Cardin, M., Ramírez, T., & Barluenga, M. (2015). Untangling the evolutionary history of a highly polymorphic species: Introgressive hybridization and high genetic structure in the desert cichlid fish *Herichthys minckleyi*. *Molecular Ecology*, *24*, 4505–4520. https://doi.org/10.1111/mec.13316

McCormack, J. F., Huang, K., & Knowles, L. L. (2009). Maximum likelihood estimates of species tree: How accuracy of phylogenetic inference upon the divergence history and sampling design. *Systematic Biology*, *58*, 501–508. https://doi.org/10.1093/sysbio/syp045

Meiklejohn, K. A., Faircloth, B. C., Glenn, T. C., Kimball, R. T., & Braun, E. L. (2016). Analysis of a rapid evolutionary radiation using ultraconserved elements: Evidence for a bias in some multispecies coalescent methods. *Systematic Biology*, *65*, 612–627. https://doi.org/10.1093/sysbio/syw014

Mendes, J., Harris, D. J., Carranza, S., & Salvi, D. (2016). Evaluating the phylogenetic signal limit from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees. *Molecular Phylogenetics and Evolution*, *100*, 254–267. https://doi.org/10.1016/j.ympev.2016.04.016

Meng, C., & Kubatko, L. S. (2011). Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence. A Model. *Theoretical Population Biology*, *75*, 35–45. https://doi.org/10.1016/j.tpb.2008.10.004

Mirarab, S., Bayzid, M. S., & Warnow, T. (2016). Evaluation summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology*, *65*, 366–380. https://doi.org/10.1093/sysbio/syu063

Mirarab, S., & Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, *31*, 44–52. https://doi.org/10.1093/bioinformatics/btv234

Nakhleh, L. (2013). Computational approaches to species phylogeny inference and gene tree reconciliation. *Trends in Ecology and Evolution*, *28*, 719–728. https://doi.org/10.1016/j.tree.2013.09.004

Niemiller, M. L., Near, T. J., & Fitzpatrick, B. M. (2012). Delimiting species using multilocus data: Diagnosing cryptic diversity in the southern cavefish, *Typhichthys subterraneus* (Teleostei: Amblyopsidae). *Evolution*, *66*, 846–866. https://doi.org/10.1111/j.1558-5646.2011.01480.x

O'Meara, B. C. (2010). New heuristic method for joint species delimitation and species tree inference. *Systematic Biology*, *59*, 59–73. https://doi.org/10.1093/sysbio/syp077

Olave, M., Avila, L. J., Sites, J. W. Jr, & Morando, M. (2016). Detecting hybridization by likelihood calculation of gene tree extra lineages given explicit models. *Methods in Ecology and Evolution*, *9*, 121–133. https://doi.org/10.1111/2041-210X.12846

Perea, S., Böhme, M., Zupančič, P., Freyhof, J., Šanda, R., Özuluğ, M., … Doadrio, I. (2010). Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. *BMC Evolutionary Biology*, *10*, 265. https://doi.org/10.1186/1471-2148-10-265

Perea, S., Cobo-Simón, M., & Doadrio, I. (2016). Cenozoic tectonic and climatic events in southern Iberian Peninusla: Implications for the evolutionary history of freshwater fish of the genus *Squalius* (Actinopterygii, Cyprinidae). *Molecular Phylogenetics and Evolution*, *97*, 155–169. https://doi.org/10.1016/j.ympev.2016.01.007

Perea, S., & Doadrio, I. (2015). Phylogeography, historical demography and habitat suitability modelling of freshwater fishes inhabiting seasonally fluctuating Mediterranean river systems: A case study using the Iberian cyprinid *Squalius valentinus*. *Molecular Ecology*, *24*, 3706–3722. https://doi.org/10.1111/mec.13274

Perea, S., Vukic, J., Sanda, R., & Doadrio, I. (2016). Ancient mitochondrial capture as factor promoting mitonuclear discordance in freshwater fishes: A case study in the genus *Squalius* (Actinopterygii, Cyprinidae) in Greece. *PLoS ONE*, *11*(12), e0166292. https://doi.org/10.1371/journal.pone.0166292

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology and Evolution*, *24*, 386–393. https://doi.org/10.1016/j.tree.2009.02.011

Potts, J., Hedderson, T. A., & Grimm, G. W. (2014). Constructing phylogenies in the presence of intra-individual site polymorphisms (2ISPs) with a focus on the nuclear ribosomal cistron. *Systematic Biology*, *63*, 1–16. 10-1093/sysbio/syt052

Quenouille, B., Bermingham, E., & Planes, S. (2004). Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, *31*, 66–88. https://doi.org/10.1016/S1055-7903(03)00278-1

Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *Tracer v1.6*. Retrieved from http://beast.bio.ed.ac.uk/Tracer

Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*, 1645–1656.

Robalo, J. I., Sousa-Santos, C., Levy, A., & Almada, V. (2006). Molecular insights on the taxonomic position of the paternal ancestor of the *Squalius alburnoides* hybridogenetic complex. *Molecular Phylogenetics and Evolution*, *39*, 276–281. https://doi.org/10.1016/j.ympev.2005.08.009

Roch, S., & Steel, M. (2015). Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theoretical Population Biology*, *100*, 56–62. https://doi.org/10.1016/j.tpb.2014.12.005

Roch, S., & Warnow, T. (2015). On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. *Systematic Biology*, *64*, 663–676. https://doi.org/10.1093/sysbio/syv016

Rokas, A., & Carroll, S. B. (2005). Bushes in the tree of life. *PLoS Biology*, *4*, e352. https://doi.org/10.1371/journal.pbio.0040352

Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, *425*, 798–804. https://doi.org/10.1038/nature02053

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*, 539–542. https://doi.org/10.1093/sysbio/sys029

Rosenberg, N. A. (2003). The shapes of neutral gene genealogies in two species: Probabilities of monophyly, paraphyly and polyphyly in a coalescent model. *Evolution*, *57*, 1465–1477.

Rougeux, C., Bernatchez, L., & Gagnaire, P.-A. (2015). Modeling the Multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregouns clupeaformis*). *Genome Biology and Evolution*, *9*, 2057–2074. https://doi.org/10.1093/gbe/evx150

Salter, J. D., Carstens, B. C., & Hedin, M. (2013). Multilocus species delimitation in a complex of morphologically conserved trapdoor spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). *Systematic Biology*, *62*, 805–823. https://doi.org/10.1093/sysbio/syt041

Sanjur, O. I., Carmona, J. A., & Doadrio, I. (2003). Evolutionary and biogeographical patterns within Iberian populations of the genus *Squalius* inferred from molecular data. *Molecular Phylogenetics and Evolution*, *29*, 20–30. https://doi.org/10.1016/S1055-7903(03)00088-5

Schrempf, D., Minh, B. Q., De Maio, N., von Haeseler, A., & Kosiol, C. (2016). Reversible polymorphism-aware phylogenetic models and their application to tree inference. *Journal of Theoretical Biology*, *407*, 362–437. https://doi.org/10.1016/j.jtbi.2016.07.042

Scornavacca, C., & Galtier, N. (2017). Incomplete lineage sorting in Mammalian Phylogenomics. *Systematic Biology*, *66*, 112–120. https://doi.org/10.1093/sysbio/syw082

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D., Di Franco, A., … Manuel, M. (2017). A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Current Biology*, *27*, 958–967. https://doi.org/10.1016/j.cub.2017.02.031

Simmons, M. P., & Gatesy, J. (2015). Coalescence vs. concatenation: Sophisticated analyses vs. first principles applied to rooting the angiosperms. *Molecular Phylogenetics and Evolution*, *91*, 98–122. https://doi.org/10.1016/j.ympev.2015.05.011

Sites, J. W. Jr, & Marshall, J. C. (2003). Delimiting species: A Renaissance issue in systematic biology. *Trends in Ecology and Evolution*, *18*, 462–470. https://doi.org/10.1016/S0169-5347(03)00184-8

Sousa-Santos, C., Collares-Pereira, M. J., & Almada, V. C. (2006). Evidence of extensive mitochondrial introgression with nearly complete substitution of the typical *Squalius pyrenaicus*-like mtDNA of the *Squalius alburnoides* complex (Cyprinidae) in an independent Iberian drainage. *Journal of Fish Biology*, *68*, 292–301. https://doi.org/10.1111/j.0022-1112.2006.01081.x

Sousa-Santos, C., Collares-Pereira, M. J., & Almada, V. C. (2007). Reading the history of a hybrid fish complex from its molecular record. *Molecular Phylogenetics and Evolution*, *45*, 981–996. https://doi.org/10.1016/j.ympev.2007.05.011

Sousa-Santos, C., Jesus, T. F., Fernandes, C., Robalo, J. I., & Coelho, M. M. (2019). Fish diversification at the pace of geomorphological changes: Evolutionary history of western Iberain Leuciscinae (Teleostei: Leuciscidae) inferred from multilocus sequence data. *Molecular Phylogenetics and Evolution*, *133*, 263–285. https://doi.org/10.1016/j.ympev.2018.12.020

Sousa-Santos, C., Matono, P., Da Silva, J., & Ilhéu, M. (2018). Evaluation of potential hybridization between native fishes and the invasive bleak, Alburnus alburnus (Actinopterygii, Cyprinidae). *Acta Ichthyologica Et Piscatoria*, *48*, 109–122. https://doi.org/10.3750/AIEP/02395

Sousa-Santos, C., Robalo, J. I., Collares-Pereira, M. J., & Almada, V. C. (2005). Heterozygous indels as useful tools in the reconstruction of DNA sequences and in the assessment of ploidy level and genomic constitution of hybrid organisms. *DNA Sequence*, *16*, 462–467. https://doi.org/10.1080/1042517050035665

Sousa-Santos, C., Robalo, J. I., Pereira, A. M., Branco, P., Santos, J. M., Ferreira, M. T., … Doadrio, I. (2016). Broad-scale sampling of primary freshwater fish populations reveals the role of intrinsic traits, inter-basin connectivity, drainage area and latitude on shaping contemporary patterns of genetic diversity. *PeerJ*, *4*, e1694. https://doi.org/10.7717/peerj.1694

Springer, M. S., & Gatesy, J. (2014). Land plant origins and coalescence confusion. *Trends in Plant Science*, *19*(5), 267–269. https://doi.org/10.1016/j.tplants.2014.02.012

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

Stamatakis, A., Blagojevic, F., Nikolopoulos, D., & Antonopoulos, C. (2007). Exploring new search algorithms and hadware for phylogenetics: RaxML Meets the IBM Cell. *The Journal of VLSI Signal Processing*, *48*, 271–286. https://doi.org/10.1007/s11265-007-0067-4

Stephens, M., & Donnelly, P. (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, *73*, 1162–1169. https://doi.org/10.1086/379378

Stephens, M., Smith, N., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, *68*, 978–989. https://doi.org/10.1086/319501

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences USA*, *114*, 1607–1612. https://doi.org/10.1073/pnas.1607921114

Swofford, D. L. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4*. Sunderland, MA: Sinauer Associates.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA7: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. https://doi.org/10.1093/molveb/msw054

Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., & Orti, G. (2015). Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. *PLoS Currents: Tree of Life*, *1*. https://doi.org/10.1371/currents.tol.34260cc27551a527b124ec5f6334b6be

Waap, S., Amaral, A. R., Gomes, B., & Coelho, M. M. (2011). Multi-locus species tree of the chub genus *Squalius* (Leuciscinae: Cyprinidae) from western Iberia: New insights into its evolutionary history. *Genetica*, *139*, 1009–1018. https://doi.org/10.1007/s10709-011-9692-0

Weisrock, D. W., Smith, S. D., Chan, L. M., Biebouw, K., Kappeler, P. M., & Yoder, A. D. (2012). Concatenation and concordance in the reconstruction of mouse lemur phylogeny: An empirical demonstration of the effect of allele sampling in phylogenetics. *Molecular Biology and Evolution*, *29*, 1615–1630. https://doi.org/10.1093/molbev/mss008

Welton, L. J., Siler, C. D., Oaks, J. R., Diesmons, A. C., & Brown, R. M. (2013). Multilocus phylogeny and Bayesian estimates of species boundaries reveal hidden evolutionary relationships and cryptic diversity in Southeast Asian monitor lizards. *Molecular Ecology*, *22*, 395–3510. https://doi.org/10.1111/mec.12324

Wiens, J. J., Kuczynski, C. A., Smith, S. A., Mulcahy, D. G., Sites, J. W., Townsend, T. M., & Reeder, T. W. (2008). Branch lengths, support and congruence: Testing the phylogenomic approach with 20 nuclear loci in snakes. *Systematic Biology*, *57*, 420–431. https://doi.org/10.1080/10635150802166053

Wu, Y. (2012). Coalescent-based species tree inference from gene tree topologies under incomplete lineage sorting by maximum likelihood. *Evolution*, *66*, 763–775. https://doi.org/10.1111/j.1558-5646.2011.01476.x

Xi, Z., Liu, L., Rest, J. S., & Davis, C. C. (2014). Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Systematic Biology*, *63*, 919–932. https://doi.org/10.1093/sysbio/syu055

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Current Zoology*, *61*(5), 854–865. https://doi.org/10.1093/czzool/61.5.854

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences USA*, *107*, 9264–9269. https://doi.org/10.1073/pnas.0913022107

Yu, D. W., Yinqiu, J., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, *3*, 613–623. https://doi.org/10.1111/j.2041-210X.00198

Zardoya, R., & Doadrio, I. (1998). Phylogenetic relationships of Iberian cyprinids: systematic and biogeographical implications. *Proceedings of the Royal Society of London, Series B*, *265*, 1365–1372. https://doi.org/10.1098/rspb.1998.0443

Zhang, C., Zhang, D.-X., Zhu, T., & Yang, Z. (2011). Evaluation of a Bayesian coalescent method of species delimitation. *Systematic Biology*, *60*, 77–761. https://doi.org/10.1093/sysbio/syr071

Zhong, B., Liu, L., Yan, Z., & Penny, D. (2013). Origin of land plants using the multispecies coalescent model. *Trends in Plant Science*, *18*, 492–495. https://doi.org/10.1016/j.tplants.2013.04.009

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Perea S, Sousa-Santos C, Robalo J, Doadrio I. Multilocus phylogeny and systematics of Iberian endemic *Squalius* (Actinopterygii, Leuciscidae). *Zool Scr*. 2020;00:1–18. https://doi.org/10.1111/zsc.12420