Multilocus species delimitation analyses show junior synonyms and deep-sea unknown species of genus Gaidropsarus (Teleostei: Gadiformes) in the North Atlantic/Mediterranean Sea area.

David Barros-García (dbarros@ciimar.up.pt)
UP CIIMAR: Universidade do Porto Centro Interdisciplinar de Investigacao Marinha e Ambiental
https://orcid.org/0000-0002-5283-2605
ÁNGEL SEBASTIÁN COMESAÑA
University of Vigo: Universidade de Vigo
RAFAEL BAÑÓN
Xunta de Galicia
FRANCISCO BALDÓ
IEO: Instituto Espanol de Oceanografia
JUAN CARLOS ARRONTE
IEO: Instituto Espanol de Oceanografia
ELSA FROUFE
UP CIIMAR: Universidade do Porto Centro Interdisciplinar de Investigacao Marinha e Ambiental
ALEJANDRO DE CARLOS
University of Vigo: Universidade de Vigo

Research Article

Keywords: Bayesian, Coalescent, Fish, Likelihood, Taxonomy

Posted Date: April 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1497892/v1

Abstract

Gaidropsarus Rafinesque, 1810 is a genus of marine teleost characterized by a high ecological diversity and by species inhabiting from the intertidal zone to the deep-sea. Several taxonomic conundrums have been historically present in this taxon due to its conservative morphology and the lack of available specimens. Species delimitation analyses were carried out in multiple datasets combining both mitochondrial (*COI*, *CytB*, *ND2*) and nuclear (*Rho*, *ZIC1*) genetic markers. Despite some incongruence between mitochondrial and nuclear data, the analyses supported the synonymy between *Gaidropsarus biscayensis* – *Gaidropsarus macrophthalmus* and the existence of a putative undescribed deep-sea *Gaidropsarus* species in the North Atlantic Ocean. Furthermore, recent speciation events can explain the close relationships among several *Gaidropsarus* species, including *Gaidropsarus granti* – *Gaidropsarus vulgaris* and *Gaidropsarus argentatus* – *Gaidropsarus ensis*. These results support previous findings highlighted through DNA Barcoding analyses. The first evidence of a complex evolutionary scenario has arisen between *Gaidropsarus guttatus* and *Gaidropsarus mediterraneus*, but further analyses will be necessary to unravel the phenomena related to it.

1. Introduction

1.1 *Species delimitation*

Accurate species delimitation analyses are a keystone for most aspects of biological sciences including the recognition of biodiversity hotspots, detecting declines in populations, assessing environmental impacts, developing effective management strategies and monitoring the trade of endangered species (Krishnamurthy and Francis, 2012). Furthermore, a correct understanding of species diversity and distributions is also fundamental in the study of speciation mechanisms and revealing broader macroevolutionary patterns (Tan et al. 2021). Hence, species is the basic unit of analysis in many scientific disciplines in which proper identification of specimens is the crucial first step in a large number of biological analyses (da Silva et al. 2018; De Queiroz, 2007).

Traditionally, specimen identification has been based primarily on the description and comparison of morphological characters to define and address species (Vitecek et al. 2017). However, an integrative approach based on independent sources of information (anatomical, ecological, geographical, molecular, morphological, physiological, etc.) has been proposed as an improved method for delineation and description of taxa (Dayrat, 2005). On this basis, morphology-based approaches have been complemented in the last years with the arrival of molecular techniques, such as the use of a fragment of the mitochondrial protein-coding gene cytochrome oxidase subunit I (*COI*) as a DNA barcode to identify specimens when compared with an available dataset (DNA Barcoding) (Hebert et al. 2003). Nevertheless, the use of single-locus analyses can lead to uncertainties or inaccuracies due to the biological nature of the marker used (events of introgression, hybridization, etc.) or differences between single-gene and species trees (Krishnamurthy and Francis, 2012). Nowadays, DNA-based species delimitation analyses rely mostly on the study of the branching pattern of multi-locus phylogenetic trees to define one and/or several thresholds to group the specimens in species-like clusters, such as the Poisson Tree Processes (PTP) (Zhang et al. 2013). This method is based on the hypothesis of a higher number of nucleotide substitutions between species than within species and has been improved into a multi-rate Poisson Tree Processes (mPTP), to consider possible variations in the substitution ratio due to differences in the evolutionary process among branches (Kapli et al. 2017). These analyses are available in a web server which includes another modification: adding bayesian support (BS) values to delimited species thanks to Markov Chain Monte Carlo sampling (bPTP); where a higher BS value on a node indicates that all descendants from that node are more likely to belong to one species (Zhang et al. 2013).

In addition, the coalescent theory has taken a place in species delimitation analyses, and its use has grown in the last years (Fujita et al. 2012). One of the most popular algorithms is General Mixed Yule Coalescent (GMYC), which

differentiates species by estimating the point where the branch pattern exchanges from the coalescent process to pattern expected of the arising of species (Pons et al. 2006). An improvement of this approach is the use of several points in a phylogenetic tree, the multiple GMYC (mGMYC), that allows for variable transitions, from coalescent to speciation, among different lineages (Monaghan et al. 2009).

Other strategies for DNA-based species delimitation are available, such as relying on the existence of a gap among the genetic distances calculated from the data, as in the software Automatic Barcode Gap Discovery (ABGD) (Puillandre et al. 2012). In ABGD, several thresholds are tested to obtain distinct species-like clusters assuming the existence of a gap among the genetic distances calculated from the data, when values within and between groups of individuals are compared (Puillandre et al. 2012).

1.2 *Genus Gaidropsarus*

The genus *Gaidropsarus* Rafinesque, 1810 is characterized by an impressive ecological diversity despite its otherwise preserved morphology and its moderate number of species. Their main characteristics are an elongated and slender body, one barbel on the chin and two at each anterior nostril in the snout, and a first dorsal fin with a ray followed by a row of small fleshy filaments (Cohen et al. 1990). The position of the genus *Gaidropsarus* in the gadiforms phylogeny has been a point of debate for a long time, with several hypotheses proposed over the years, being alternatively placed in the families Lotidae (Froese and Pauly 2016) and Gadidae (Nelson, 2016), or even creating a specific family called Gaidropsaridae, which seems to be definitive (Roa-Varón et al. 2020).

Traditionally, this genus has comprised a total of 14 species of which eight are found in the North Atlantic Ocean and the Mediterranean Sea (Svetovidov 1986a,b): *Gaidropsarus argentatus* (Reinhardt, 1837), *Gaidropsarus biscayensis* (Collett, 1890), *Gaidropsarus ensis* (Reinhardt, 1837), *Gaidropsarus granti* (Regan, 1903), *Gaidropsarus guttatus* (Collett, 1890), *Gaidropsarus macrophthalmus* (Günther, 1867), *Gaidropsarus mediterraneus* (Linnaeus, 1758) and *Gaidropsarus vulgaris* (Cloquet, 1824). In reality, true diversity of species in this genus is not yet fully known, as evidenced by the recent description of new species, such as *Gaidropsarus pakhorukovi* (Shcherbachev, 1995) and *Gaidropsarus mauli* (Biscoito and Saldanha, 2018). In fact, it is sometimes difficult to differentiate some *Gaidropsarus* species from each other because of their well-preserved morphology (Svetovidov, 1986b). Moreover, there is a general lack of knowledge of the morphological variability of this genus due to the absence of representative specimens in the museums (Balushkin, 2009). Thus, it is not surprising that when the morphology-based available knowledge has been compared with molecular data, discrepancies have arisen (Barros-García et al. 2018). The use of genetic distances among 171 *COI* sequences indicated: i) a possible synonymy between the Atlantic *G. macrophthalmus* and the Mediterranean endemism *G. biscayensis*, ii) a possible synonymy between *G. mediterraneus*, whose habitat are the European continental waters, and *G. guttatus*, a Macaronesian endemism, and iii) a low genetic divergence between two boreal deep-sea species (*G. argentatus* and *G. ensis*), and between *G. granti* and *G. vulgaris*, and iv) the existence of several putative unknown deep-sea *Gaidropsarus* species in the North Atlantic Ocean (Barros-García et al. 2018).

1.3 *Objectives*

This investigation aims to apply single and multi-locus species delimitation analyses on a dataset obtained from several species of *Gaidropsarus* from the North Atlantic Ocean and the Mediterranean Sea, in an attempt to resolve certain taxonomic discrepancies previously found.

2. Material And Methods

2.1 *Sampling*

The specimens (n = 44) were captured in several locations of the North Atlantic Ocean and the Mediterranean Sea (Fig. 1). As a general rule, specimens were immediately frozen onboard and a muscle sample was obtained in the laboratory and preserved in 95% ethanol at -20 ºC. Specimens were identified to the species level by examining morphological features following the available taxonomic literature (Svetovidov, 1986a, 1986b). Moreover, *COI* nucleotide sequences from the same specimens were obtained and compared in the available repositories (BOLD systems and GenBank) to cross-check the morphological identification. The specimens were deposited in the “Museo de Historia Natural da Universidade de Santiago de Compostela” (Santiago de Compostela, Spain) and the “Colección de Fauna Marina del Centro Oceanográfico de Málaga” (CFM-IEOMA; Málaga, Spain). All the sequences of the different genetic markers used in the present study are publicly available in GenBank (Table 1).

Table 1

Species-level identification based on morphology, specimen ID, and accession numbers for each specimen of *Gaidropsarus* used in this study. Outgroup individuals used for the phylogenetic analyses are also included.

		Accession Numbers				
Taxon	Specimen ID	*COI*	*CytB*	*ND2*	*Rho*	*Zic1*
Gaidropsarus ensis	GDE001	KY250213	MZ234311	MZ234355	MZ234399	MZ234443
Gaidropsarus ensis	GDE002	KY250214	MZ234312	MZ234356	MZ234400	MZ234444
Gaidropsarus ensis	GDE003	KY250215	MZ234313	MZ234357	MZ234401	MZ234445
Gaidropsarus ensis	GDE004	KY250216	MZ234314	MZ234358	MZ234402	MZ234446
Gaidropsarus ensis	GDE005	KY250217	MZ234315	MZ234359	MZ234403	MZ234447
Gaidropsarus argentatus	GDT001	KY250179	MZ234301	MZ234345	MZ234389	MZ234433
Gaidropsarus argentatus	GDT002	KY250169	MZ234302	MZ234346	MZ234390	MZ234434
Gaidropsarus argentatus	GDT003	KY250192	MZ234303	MZ234347	MZ234391	MZ234435
Gaidropsarus argentatus	GDT004	KY250191	MZ234304	MZ234348	MZ234392	MZ234436
Gaidropsarus argentatus	GDT005	KY250190	MZ234305	MZ234349	MZ234393	MZ234437
Gaidropsarus granti	GGA001	KY250239	MZ234316	MZ234360	MZ234404	MZ234448
Gaidropsarus granti	GGA002	KY250238	MZ234317	MZ234361	MZ234405	MZ234449
Gaidropsarus granti	GGA003	KY370533	MZ234318	MZ234362	MZ234406	MZ234450
Gaidropsarus mediterraneus	GGD001	KY250282	MZ234329	MZ234373	MZ234417	MZ234461
Gaidropsarus mediterraneus	GGD002	KY250285	MZ234330	MZ234374	MZ234418	MZ234462
Gaidropsarus mediterraneus	GGD003	KY250284	MZ234331	MZ234375	MZ234419	MZ234463
Gaidropsarus mediterraneus	GGD004	KY250283	MZ234332	MZ234376	MZ234420	MZ234464
Gaidropsarus mediterraneus	GGD005	KY250296	MZ234333	MZ234377	MZ234421	MZ234465
Gaidropsarus macrophthalmus	GGR007	KY250270	MZ234324	MZ234368	MZ234412	MZ234456

		Accession Numbers				
Gaidropsarus macrophthalmus	GGR008	KY250271	MZ234325	MZ234369	MZ234413	MZ234457
Gaidropsarus macrophthalmus	GGR009	KY250272	MZ234326	MZ234370	MZ234414	MZ234458
Gaidropsarus macrophthalmus	GGR010	KY250273	MZ234327	MZ234371	MZ234415	MZ234459
Gaidropsarus macrophthalmus	GGR011	KY250274	MZ234328	MZ234372	MZ234416	MZ234460
Gaidropsarus guttatus	GGT002	KY250241	MZ234319	MZ234363	MZ234407	MZ234451
Gaidropsarus guttatus	GGT003	KY250242	MZ234320	MZ234364	MZ234408	MZ234452
Gaidropsarus guttatus	GGT004	KY250243	MZ234321	MZ234365	MZ234409	MZ234453
Gaidropsarus guttatus	GGT005	KY250244	MZ234322	MZ234366	MZ234410	MZ234454
Gaidropsarus guttatus	GGT006	KY250245	MZ234323	MZ234367	MZ234411	MZ234455
Gaidropsarus vulgaris	GGU001	KY250302	MZ234340	MZ234384	MZ234428	MZ234472
Gaidropsarus vulgaris	GGU002	KY250301	MZ234341	MZ234385	MZ234429	MZ234473
Gaidropsarus vulgaris	GGU003	KY250315	MZ234342	MZ234386	MZ234430	MZ234474
Gaidropsarus vulgaris	GGU004	KY250300	MZ234343	MZ234387	MZ234431	MZ234475
Gaidropsarus vulgaris	GGU005	KY250303	MZ234344	MZ234388	MZ234432	MZ234476
Gaidropsarus biscayensis	GGY011	KY250202	MZ234306	MZ234350	MZ234394	MZ234438
Gaidropsarus biscayensis	GGY012	KY250201	MZ234307	MZ234351	MZ234395	MZ234439
Gaidropsarus biscayensis	GGY013	KY250200	MZ234308	MZ234352	MZ234396	MZ234440
Gaidropsarus biscayensis	GGY014	KY250199	MZ234309	MZ234353	MZ234397	MZ234441
Gaidropsarus biscayensis	GGY015	KY250198	MZ234310	MZ234354	MZ234398	MZ234442
Gaidropsarus sp.	ROL001	KY250298	MZ234334	MZ234378	MZ234422	MZ234466
Gaidropsarus sp.	ROL002	KY250297	MZ234335	MZ234379	MZ234423	MZ234467
Gaidropsarus sp.	ROL003	MZ198255	MZ234336	MZ234380	MZ234424	MZ234468

		Accession Numbers				
Gaidropsarus sp.	ROL004	MZ198256	MZ234337	MZ234381	MZ234425	MZ234469
Gaidropsarus sp.	ROL005	MZ198257	MZ234338	MZ234382	MZ234426	MZ234470
Gaidropsarus sp.	ROL006	MZ198258	MZ234339	MZ234383	MZ234427	MZ234471
Gadus morhua	Outgroup	NC_002081	NC_002081	NC_002081	XM_030353516	XM_030349483
Lepisosteus oculatus	Outgroup	NC_004744	NC_004744	NC_004744	KX146025.1	XM_006637702.2

2.2 *DNA extraction, PCR amplification and sequencing*

Total DNA was purified from 25 mg of muscle tissue taken from each specimen according to the spin column protocol of the Tissue DNA Extraction Kit (Omega-Biotek). Parts of three mitochondrial: Cytochrome Oxidase Subunit I (*COI*), Cytochrome B (*CytB*) and NADH Dehydrogenase 2 (*NAD2*), and two nuclear markers: Rhodopsine (*Rho*) and Zic Family Member 1 (*ZIC1*) were amplified by PCR employing specific conditions and primers (Supplemental Material-Table 1). Amplicons were sequenced with the BigDye Terminator v3.1 Cycle Sequencing Kit, and the resulting products were resolved in an ABI3130 Genetic Analyzer at the "Centro de Apoyo Científico y Tecnológico a la Investigación" (CACTI, University of Vigo). Forward and reverse chromatograms were visually inspected and finally assembled with SEQSCAPE v. 2.5 (Applied Biosystems).

For the subsequent analyses, firstly each marker was considered independently, and then clustered in three independent groups of datasets; "mtDNA" (*COI + CytB + NAD2*), "nDNA" (*Rho + ZIC1*), and "concatenated" (the sum of mtDNA + nDNA).

2.3 *Phylogenetic analyses*

The final dataset comprised nine putative *Gaidropsarus* species and two outgroup species; cod (*Gadus morhua*) and the spotted gar (*Lepisosteus oculatus*) for a total of 220 *Gaidropsarus* sequences plus the outgroups mined from Genbank.

Since most of the species delineation analyses applied rely on the phylogenetic tools used, trees obtained by Bayesian Inference (BI) were used for GMYC analyses, and Maximum Likelihood (ML) trees for bPTP analyses, as previously described (Tang et al. 2014).

The best partitioning scheme was estimated with PartitionFinder2 (Lanfear et al. 2016), spanning the range from a single partition for the entire alignment to each gene and codon position treated as a partition (supplemental material-Table 2). For every partition, the most appropriate nucleotide substitution model was selected using jModelTest 2.1.8 (Darriba et al. 2012), based on the Akaike Information Criterion (AIC; Akaike, 1973).

The ML analysis was performed using RAxMLv.8.2.10, conducting 1000 rapid bootstrap replicates, and the best-scoring ML tree was evaluated under the GTRGAMMA model implemented on the CIPRES Science Gateway portal (Miller et al. 2010).

Bayesian Inference was used to build a phylogenetic tree using BEAST 2.6.2 (Bouckaert et al. 2019). BEAUti was used to assemble the XML files with the following settings: The models used were those obtained from jModelTest 2.1.8 following the partitions suggested by PartitionFinder2 with empirical frequencies, Yule model was selected as tree prior while all other variables were not modified from their default values. Markov Chain Monte Carlo (MCMC) chain length was set to 5×10^7, logging every 1000 samples. Three different runs were carried out for a final number of 50 000 sampled trees/run. In order to check the convergence of the analyses, the resulting trace files were inspected with

TRACER.1.6.0 (Rambaut et al. 2018). The runs were merged using Logcombiner, discarding the first 25% of each run as burn-in. A maximum clade credibility tree with posterior probability limit set at 0.90 and mean node heights was constructed with TreeAnnotator v.2.4.5 and visualized with FigTree v.1.4.3 (http://tree.bio.ed.ac.uk/software/ figtree/).

Species delineation analyses

The online version of Automatic Barcode Gap Discovery software (ABGD; available at https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) was used, based on calculated *p*-distances using the default priors for the relative gap width (1.5), Pmin (0.001) and Pmax (0.1) (Puillandre et al. 2012). All analyses were carried out with 10 steps and 20 bins of distance distribution.

To estimate the number of putative species in the data, a variation of the Poisson Tree Processes (PTP) was used. The bPTP analysis adds Bayesian support values to delimited species on the input tree. The analyses were carried out in the webserver (https://species.h-its.org/ptp/) with default settings removing the outgroup to optimize the delimitation results (Zhang et al. 2013).

The BI trees obtained were analysed in the webserver (https://species.h-its.org/gmyc/) for either single and multiple threshold approaches for the generalized mixed Yule-coalescent (GMYC) method, which identifies the time threshold that defines coalescent or speciation processes on the branching patterns of ultrametric trees (Pons et al. 2006).

3. Results

3.1 mtDNA data

Phylogenetic analyses of the combined mitochondrial markers showed a fully resolved topology (Fig. 2a; supplemental material Fig. S1). Both *G. mediterraneus* and *G. guttatus* were located at the basal position of the *Gaidropsarus* clade, with mixed individuals of both species in two different clades.

In addition, *G. granti* and *G. vulgaris* showed a close relationship as well as *G. argentatus* and *G. ensis*, while *G. biscayensis* and *G. macrophthalmus* are mixed (Fig. 2a). On the contrary, the six individuals assigned to *Gaidropsarus* sp. clustered together in an independent branch.

Five out of the nine putative species showed no incongruences among all the delimitation analyses: *G. argentatus*, *G. ensis*, *G. granti*, *Gaidropsarus* sp. and *G. vulgaris*. Among the others, four analyses (ABGD, bPTP and GMYC) indicated a single cluster between *G. biscayensis* and *G. macropthalmus* (Fig. 2a). Only the GMYC with the multi-threshold approach divided these species into two independent clusters, and yet each cluster contained sequences from both species (supplemental material-Fig. S1). The same results were obtained for the species *G. guttatus* and *G. mediterraneus*, with the difference that, in this case, the phylogenetic analyses divided them into two well-supported clusters combining sequences from both species (Fig. 2a). All the clusters obtained with bPTP showed a high level of confidence of bayesian support (> 0.9) except for the cluster which brings together *G. guttatus* and *G. mediterraneus*. The results for the mtDNA markers independently analysed are summarized in the supplemental material (supplemental material Fig. S2).

3.2 nDNA data

Regarding the nuclear markers, interestingly *ZIC1* showed a highly conserved nature with all substitutions found in *Gaidropsarus* being synonymous and an average genetic distance among all samples of 0.7% whereas *Rho* resulted in a more variable marker, with an average genetic distance of 3.6% among all samples.

Phylogenetic analyses of nDNA markers showed a partially resolved phylogeny with *G. guttatus* and *G. mediterraneus* splitting from the other species in the *Gaidropsarus* lineage (Fig. 2b; supplemental material-Fig. S3). *Gaidropsarus granti* and *Gaidropsarus* sp. are well supported as well as another two clades formed by the combination of *G. argentatus* / *G. ensis*, and *G. biscayensis* / *G. macrophthalmus*, respectively. The branches not supported in the tree topology included sequences from *G. vulgaris*, *G. argentatus* (GDT002) and *G. guttatus* specimens (GGT003, GGT005), and one allele from GGT002 and GGT004 (Fig. 2b; supplemental material-Fig. S3).

Species delimitation analyses showed a gradient in the results; ABGD differentiated only two groups, the first including *G. guttatus* and *G. mediterraneus* and a larger second group with the rest of the sequences (Fig. 2b). Both bPTP and GMYC analyses showed the joint grouping of *G. argentatus* and *G. ensis*, as well as the independence of *G. granti* and the division between *G. guttatus* and *G. mediterraneus*. Single threshold GMYC and bPTP agreed in one cluster for *Gaidropsarus* sp. and *G. biscayensis* / *macrophthalmus*, while mGMYC divided the former into two groups and the latter in five. The unresolved part of the consensus tree is considered a single cluster by both GMYC and mGMYC, while bPTP separated it, including *G. argentatus* and *G. guttatus* sequences. The results for the independently analysed nDNA markers are graphically summarized in the supplemental material (supplemental material-Fig. S4 and Fig. S5).

3.3 Concatenated data

The phylogenetic trees obtained with the dataset comprising all genetic markers showed a fully resolved topology, with *G. mediterraneus* and *G. guttatus* splitting firstly in the *Gaidropsarus* clade (Fig. 2c; supplemental material-Fig. S6). Sequences grouped in two independent clusters, one comprising sequences of both species and the other of *G. guttatus* specimens only, but with low statistic support. *Gaidropsarus argentatus* and *G. ensis* showed a close relationship as *G. granti* and *G. vulgaris*. All the individuals belonging to *G. biscayensis* and *G. macrophthalmus* clustered together in the same way as the specimens of *Gaidropsarus* sp. (Fig. 2c).

The species delimitation analyses were fully congruent only in the case of *Gaidropsarus* sp. (Fig. 2c). ABGD and bPTP analyses were able to differentiate between the pairs of species formed by *G. argentatus* / *G. ensis*, and *G. granti* / *G. vulgaris*, but only the latter with high confidence values in bPTP. The GMYC algorithm, in both single and multi-threshold approaches, combined them as single units. Conversely, the synonymy between *G. macrophthalmus* and *G. biscayensis* is supported by three of four analyses performed (ABGD, bPTP and GMYC), while mGMYC divided them into two clusters but mixing sequences from both species (Fig. 2c; supplemental material-Fig. 6). Both ABGD and GMYC algorithms supported the existence of a single cluster for the grouping of *G. mediterraneus* and *G. guttatus* specimens, while bPTP and mGMYC differentiated the two branches obtained in the phylogenetic analyses (Fig. 2c).

4. Discussion

4.1 Phylogenetic relationships

Phylogenetic trees obtained through single-locus mitochondrial data were almost identical since all mitochondrial markers are inherited as a single locus. In addition, and as expected, the combination of the mtDNA markers in a multi-locus dataset improved the statistical support of the phylogenetic trees obtained when compared with the single-locus ones (Janko et al. 2011).

The phylogenetic tree obtained with mtDNA showed longer branches and better statistically supported clusters than those obtained with nDNA. These results could be explained in a scenario of recent divergence among some of the species and by the higher substitution rate observed in mtDNA compared to nDNA (Brown et al. 1979). Nevertheless, both mtDNA and nDNA strongly acknowledged the basal position of *G. mediterraneus* and *G. guttatus* in the

Gaidropsarus lineage in agreement with previous studies, with the remaining species in a second group (Francisco et al. 2014).

4.1 Gaidropsarus biscayensis – Gaidropsarus macrophthalmus

Previously, it has been proposed that *G. biscayensis* could be distinguished morphologically from *G. macrophthalmus* by the number of second dorsal fin rays and by the anal fin ranges (Iwamoto and Cohen, 2016). However, in further bibliographical revisions these magnitudes were found to overlap, invalidating them as distinctive (Barros-García et al. 2018). This scenario was previously observed in the synonymy between the Atlantic *Lepidion eques* (Günther, 1887) and the Mediterranean *Lepidion lepidion* (Risso, 1810), resulting in the latter as the only valid species with an Atlantic-Mediterranean distribution (Bañón et al. 2013; Barros-García et al. 2016). In both cases, *G. biscayensis* – *G. macrophtalmus* and *L. eques* – *L. lepidion*, the slight morphological differences recorded could be explained by the different environmental conditions, particularly temperature, between the Atlantic Ocean and the Mediterranean Sea (Bañón et al. 2013).

The analyses carried out in the present study showed a general agreement in considering *G. biscayensis* and *G. macrophthalmus* as a single species, despite the different approaches and dataset combinations tested. It is true that, since mtDNA has been included in the analyses, the possibility of hybridization/introgression events must be considered (Harrison and Larson, 2014). However, no evidence of mixed alleles has been found in the nuclear markers considered, so these possibilities can be discarded. Present results are in agreement with previous results obtained by combining morphological data and DNA Barcoding (Barros-García et al. 2018). Therefore, *G. biscayensis* should be considered a junior synonym of *G. macrophthalmus*, resulting in a single species with an Atlantic-Mediterranean distribution.

4.2 Gaidropsarus guttatus – Gaidropsarus mediterraneus

During their original descriptions, the high similarity between *G. guttatus* and *G. mediterraneus* was reported (Collet, 1905). Recently, a bibliographical revision of the morphological data available showed overlaps in the main characters measurements used to distinguish between these two species (Barros-García et al. 2018). Only colour patterns and distribution areas remained as distinguishing features (Cohen and Russo, 1979). Therefore, it is not surprising that a DNA barcoding analysis was not able to differentiate them, producing a single cluster where individuals of both species were mixed (Barros-García et al. 2018).

Far from helping to clarify the taxonomic problem of these two species, the nDNA analysis showed incongruent results when compared with mtDNA, misplacing some of these individuals with other species. It is well known that nDNA possess specific characteristics such as a lower mutation rate and higher effective population sizes than compared with mtDNA which slower the time required for divergence (Torres-Hernández et al. 2022). The different responses of the two sets of DNA data could be due to a variety of biological phenomena such as introgression of the mtDNA or incomplete lineage sorting at the nDNA level. Similar results with mixed individuals in nDNA based trees have been found in other fishes like Nemacheilidae (Cypriniformes) (Dvořák et al. 2022). Taking into account the mtDNA results, the former is to be considered unlikely, but further analyses will be required to clarify the evolutionary phenomenon behind these results (Kornilios et al. 2020).

In any case, the undeniable fact is that, apart from colouring, neither the examination of morphological characters nor the mtDNA sequences comparisons can distinguish between specimens of one species or another. Without prejudice to whether this phenomenon is due to adaptive processes of population type, or even to an imminent speciation event, the data provided seem to indicate that *G. guttatus* and *G. mediterraneus* are, in fact, a single species.

4.3 Gaidropsarus argentatus – Gaidropsarus ensis

The boreal *Gaidropsarus argentatus* and *G. ensis* are deep-sea species with an overlap in their North Atlantic Ocean distribution areas. Although their colouration and morphology are similar, they can be easily differentiated by the length of the first ray of the first dorsal fin, which is longer than its head in *G. ensis* and shorter in *G. argentatus* (Svetovidov, 1986b). Previous DNA barcoding and phylogenetic studies indicated a close relationship between these two species (Francisco et al. 2014; Barros-García et al. 2018). Interestingly, species delimitation analyses based on mtDNA clearly distinguished them, whereas nDNA markers did not, since phylogenetic analyses based on nDNA grouped them similarly to *G. guttatus* and *G. mediterraneus* In a recent speciation scenario, nDNA trees may fail to differentiate between species since not enough time has passed to accumulate substitutions in nuclear markers compared to the faster-evolving mtDNA genes (Brown et al. 1979). Therefore, this discrepancy observed in the species delimitation analyses performed with mitochondrial and nuclear DNA could be interpreted as evidence of a recent speciation phenomenon between *G. argentatus* and *G. ensis*.

4.4 Gaidropsarus granti – Gaidropsarus vulgaris

While morphological examination places *G. granti* as a species close to *G. mediterraneus* and *G. guttatus* (Svetovidov 1986a), molecular data place it close to *G. vulgaris* (Francisco et al. 2014; Barros-García et al. 2018). The latter two species can be differentiated from each other by their colouration pattern, habitat and distribution. *Gaidropsarus vulgaris* is found in the European continental shelf up to 120 meters of depth, while *G. granti* is a far more uncommon species, mainly inhabiting islands and seamounts at greater depths (Bañón et al. 2020). Analyses relying on mtDNA data clearly distinguished between the two species in a similar way to those performed for the boreal *G. argentatus* and *G. ensis*. Despite clustering almost all sequences together, nDNA indicated the independence of *G. granti*, which could be indicative of a speciation event older than in the previous case of the boreal species, in a way in which nDNA markers would have had some time to begin to acquire species-level substitutions (Brown et al. 1979). Since it is a rare species, the sample size of *G. granti* in this investigation is low and its related results should be interpreted cautiously. Nevertheless, the analysis of the data set used suggests that its status as an independent species must be maintained although in close relationship with *G. vulgaris.*

4.5 Gaidropsarus sp.

Six deep-water specimens of *Gaidropsarus* captured in two independent locations in the North Atlantic Ocean did not fit with any current description of *Gaidropsarus* species, but they showed common morphological features and, therefore, were appointed as *Gaidropsarus* sp. Both mtDNA and nDNA data, when employed in the phylogenetic and species delimitation analyses, indicated that this set of individuals belongs to the same species, different from the other eight nominal species considered in this investigation. Further analyses will be required, including morphological comparisons with the relatively unknown species from the South Hemisphere, to clarify if these specimens belong to a completely unknown *Gaidropsarus* taxon or if their capture constitutes evidence of the existence of a new distribution range in the North Hemisphere for a known southern species.

4.6 Final Remarks

The aim of this study was to shed new light on several taxonomic incongruities regarding the known *Gaidropsarus* species of the North Atlantic Ocean and the Mediterranean Sea. Thus, two synonymies among the already recognised species, and the presence of an unidentified species for the northern hemisphere have been detected. *Gaidropsarus biscayensis* and *G. guttatus* were described under the premises of the existence of phylogeographic barriers between the Mediterranean Sea and the Atlantic Ocean for the former, and between European continental waters and the Macaronesian region for the latter. These assumptions have led to an overestimation of *Gaidropsarus* species in the shallow waters of the North Atlantic Ocean and the Mediterranean Sea. Conversely, the recently discovered *Gaidropsarus mauli* (Biscoito and Saldanha, 2018), and the previously unknown specimens included in this investigation indicate the presence of a certain number of yet unknown *Gaidropsarus* species in the deep-sea

ecosystems of the North Atlantic Ocean, despite the relatively good knowledge of the fish fauna of this geographic area (Barros-García et al. 2018; Bañón et al. 2021).

The existence of shared nuclear alleles among different *Gaidropsarus* species could be evidence of a far more complex evolutionary history than expected for this lineage of teleost fishes. Increasing the number of individuals from more areas and examining more regions of their genomes using next-generation sequencing techniques may be necessary to try to clarify the evolution, and thus the taxonomy, of the North Atlantic and Mediterranean species of the genus *Gaidropsarus*.

Declarations

Acknowledgments

This research was supported by national funds by FCT -Foundation for Science and Technology through the projects UIDB/04423/2020 and UIDP/04423/2020. FCT also supported DBG (DBG 2020.04364.CEECIND) and EF (CEECIND/00627/2017). The authors acknowledge the support and the use of resources of EMBRC-ERIC, specifically of the Portuguese infrastructure node of the European Marine Biological Resource Centre (EMBRC-PT) CIIMAR – PINFRA/22121/2016 – ALG-01-0145-FEDER-022121, financed by the European Regional Development Fund (ERDF) through COMPETE2020 -Operational Programme for Competitiveness and Internationalization (POCI) and national funds through FCT/MCTES.

The authors would like to thank the staff involved in the following research surveys of the 'Instituto Español de Oceanografía (IEO-CSIC)': FLEMISH 2015, INDEMARES-BANGAL 0811, FLETÁN ÁRTICO 2015 (Project ECOPESLE) and PORCUPINE 2019. We also extend our thanks to the crews of the Research Vessels 'Miguel Oliver' and 'Vizconde de Eza' (Ministry of Agriculture, Food and Environment, Spain). Alberto Serrano (IEO-CSIC) is especially recognized. The IEO research surveys have been funded by the European Union through the European Maritime and Fisheries Fund (EMFF) within the National Program of collection, management and use of data in the fisheries sector and support for scientific advice regarding the Common Fisheries Policy. Thanks are due to João Medeiros for his help in providing specimens from the Azores.

Conflict of interest: The authors have no conflict of interest to declare that are relevant to the content of this article.

Author Contributions:

Francisco Baldó and Juan Carlos Arronte obtained the samples during oceanographic surveys. Angel Sebastián Comesaña was in charge of the laboratory protocols; DNA extraction, PCR and sequencing. David Barros-García carried out the bioinformatic analyses. Rafael Bañón, David Barros-García, Alejandro de Carlos and Elsa Froufe contributed to the study conception and design. The first draft of the manuscript was written by David Barros-García and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability: Molecular data generated during this study consisted of sequences belonging to three mitochondrial (*COI*; *Cytb*; *ND2*) and two nuclear (*Rho* and *ZIC1*) markers. All these sequences are deposited in GenBank; all GenBank accession numbers used in are listed in Table 1.

Ethical approval: The samples used in this study were obtained within the framework of the annual oceanographic campaigns carried out by the Spanish Institute of Oceanography in various locations in the North Atlantic and the Mediterranean Sea following the established protocols. No species of *Gaidropsarus* is listed in the IUCN red list as critically endangered, endangered, vulnerable, or near-threatened species.

Consent to participate: All authors gave consent to participate.

Consent for publication: All authors consent to the submission and publication of this article upon acceptance.

References

1. Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Caski (Eds.), Proceedings of the Second International Symposium on Information Theory (pp. 267–281). Budapest: Akademiai Kiado
2. Balushkin AV (2009) On the first occurrence of the rockling *Gaidropsarus pakhorukovi* Shcherbachev (Gairdropsarini, Lotidae, Gadidae) and on species diagnostics of *G. pakhorukovi* and *G. parini* Svetovidov. Jpn J Ichthyol. doi:10.1134/S0032945209090033
3. Bañón R, De Carlos A, Farias C, Vilas-Arrondo N, Baldó F (2021) Exploring Deep-Sea Biodiversity in the Porcupine Bank (NE Atlantic) through Fish Integrative Taxonomy. J Mar Sci Eng. doi:10.3390/jmse9101075
4. Bañón R, de Carlos A, Ruiz-Pico S, Baldó F (2020) Unexpected deep-sea fish species on the Porcupine Bank (NE Atlantic): Biogeographical implications. J Fish Biology. doi:10.1111/jfb.14418
5. Bañón R, Arronte JC, Vázquez-Dorado S, Del Río JL, De Carlos A (2013) DNA barcoding of the genus *Lepidion* (Gadiformes: Moridae) with recognition of *Lepidion eques* as a junior synonym of *Lepidion lepidion*. Mol Ecol Resour. doi:10.1111/1755-0998.12045
6. Barros-García D, Bañón R, Arronte JC, Fernández-Peralta L, García R, Iglésias SP, Sellos YD, Barreiros PJ, Comesaña AS, De Carlos A (2018) New insights into the systematics of North Atlantic *Gaidropsarus* (Gadiformes, Gadidae): flagging synonymies and hidden diversity. Mar Biol Res. doi:10.1080/17451000.2017.1367403
7. Biscoito M, Saldanha L (2018) *Gaidropsarus mauli* a new species of three-bearded rockling (Gadiformes, Gadidae) from the Lucky Strike hydrothermal vent field (Mid-Atlantic Ridge) and the Biscay Slope (Northeastern Atlantic). Zootaxa. doi:10.11646/zootaxa.4459.2.5
8. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A et al (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1006650
9. Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. P Natl Acad Sci USA. DOI: 10.1073/pnas.76.4.1967
10. Cohen DM, Inada T, Iwamoto T, Scialabba N (1990) FAO species catalogue. Gadiform fishes of the world (Order Gadiformes). An annotated and illustrated catalogue of cods, hakes, grenadiers and other gadiform fishes known to date. FAO Fisheries Synopsis 125:1–442
11. Cohen DM, Russo JL (1979) Variation in the four beard rockling, *Enchelyopus cimbrius*, a North Atlantic gadid fish, with comments on the genera of rocklings. United States National Marine Fisheries Service Fishery Bulletin 77:91–104
12. Collett R (1905) On some fishes from the sea off the Azores. Zool Anz 28:723–730
13. da Silva R, Peloso PLV, Sturaro MJ, Veneza I, Sampaio I, Schneider H, Gomes G (2018) Comparative analyses of species delimitation methods with molecular data in snappers (Perciformes: Lutjaninae). Mitochondr DNA. doi:10.1080/24701394.2017.1413364
14. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. doi:10.1038/nmeth.2109
15. Dayrat B (2005) Towards integrative taxonomy. Biol J Lin Soc. doi:10.1111/j.1095-8312.2005.00503.x
16. De Queiroz K (2007) Species Concepts and Species Delimitation. Syst Biol. doi:10.1080/10635150701701083

17. Dvořák T, Šlechtová V, Bohlen J (2022) Using Species Groups to Approach the Large and Taxonomically Unresolved Freshwater Fish Family Nemacheilidae (Teleostei: Cypriniformes) *Biol.* https://doi.org/10.3390/biology11020175
18. Francisco SM, Robalo JI, Stefanni S, Levy A, Almada VC (2014) *Gaidropsarus* (Gadidae, Teleostei) of the North Atlantic Ocean: a brief phylogenetic review: phylogeny of *Gaidropsarus* in the north atlantic. J Fish Biol. doi:10.1111/jfb.12437
19. Froese R, Pauly D (2020) FishBase. World Wide Web electronic publication. Available via www.fishbase.org, Version December 2021
20. Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol Evol. doi:10.1016/j.tree.2012.04.012
21. Harrison RG, Larson EL (2014) Hybridization, Introgression, and the Nature of Species Boundaries. J Hered. doi:10.1093/jhered/esu033
22. Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of Birds through DNA Barcodes. PLOS Biol. https://doi.org/10.1371/journal.pbio.0020312
23. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. P R Soc Lond B Biol. doi:10.1098/rspb.2002.2218
24. Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. Mol Ecol Notes. https://doi.org/10.1111/j.1471-8286.2007.01748.x
25. Iwamoto T, Cohen DM (2016) Gaidropsaridae In: Carpenter KE, De Angelis N, editors. The Living Marine Resources of the Eastern Central Atlantic. Volume 3. Bony Fishes. Part 1 (Elopiformes to Scorpaeniformes). Rome: FAO, pp. 2015–2022
26. Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamakis A, Flouri T (2017) Multi-rate Poisson Tree Processes for single-locus species delimitation under Maximum Likelihood and Markov Chain Monte Carlo. Bioinformatics. doi:10.1093/bioinformatics/btx025
27. Kornilios P, Jablonski D, Sadek RA, Kumlutas Y, Olgun K, Avci A, Ilgaz C (2020) Multilocus species-delimitation in the *Xerotyphlops vermicularis* (Reptilia: Typhlopidae) species complex. Mol Phyl Evol. doi:10.1016/j.ympev.2020.106922
28. Krishnamurthy PK, Francis RA (2012) A critical review on the utility of DNA barcoding in biodiversity conservation. Biodivers Conserv. doi:10.1007/s10531-012-0306-2
29. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2016) PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol. doi:10.1093/molbev/msw260
30. Li C, Ortí G, Zhang G, Lu G (2007) A practical approach to phylogenomics: The phylogeny of ray-finned fish (Actinopterygii) as a case study. BMC Evol Biol. DOI:10.1186/1471-2148-7-44
31. Messing J (1982) New M13 vectors for cloning. Method Enzymol. 10.1016/0076-6879(83)01005-8
32. Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees in *Proceedings of the Gateway Computing Environments Workshop* (GCE), 14 Nov. 2010, New Orleans, LA pp. 1–8
33. Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJG, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG, Vogler AP (2009) Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. Syst Biol. doi:10.1093/sysbio/syp027
34. Nelson JS, Grande TC, Wilson MVH (2016) Fishes of the World, 5th edition. Hoboken, NJ: John Wiley and Sons

35. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Volger AP (2006) Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Syst Biol. doi:10.1080/10635150600852011
36. Puillandre N, Lambert A, Brouillet S, Achaz G (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol. doi:10.1111/j.1365-294X.2011.05239.x
37. Roa-Varón A, Dikow RB, Carnevale G, Tornabene L, Baldwin CC, Li C, Hilton EJ (2020) Confronting Sources of Systematic Error to Resolve Historically Contentious Relationships: A Case Study Using Gadiform Fishes (Teleostei, Paracanthopterygii, Gadiformes). Syst Biol. https://doi.org/10.1093/sysbio/syaa095
38. Sevilla RG, Diez A, Norén M, Mouchel O, Jérôme M, Verrez-bagnis V, Van Pelt H, Favre-Krey L, Krey G, Bautista JM (2007) Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. Mol Ecol Notes. https://doi.org/10.1111/j.1471-8286.2007.01863.x
39. Shcherbachev YN (1995) New species of rockling, Gaidropsarus pakhorukovi (Gadidae), from the Rio Grande Rise (Southwest Atlantic). Jpn J Ichthyol. ttps://doi.org/10.1134/S0032945209090033
40. Svetovidov AN (1986a) Review of three-bearded rocklings of the genus *Gaidropsarus* Rafinesque, 1810 (Gadidae) with description of a new species. Jpn J Ichthyol. https://doi.org/10.1134/S0032945209090033
41. Svetovidov AN (1986b) Gadidae. In: Whitehead PJP, Bauchot ML, Hureau JC, Nielsen J, Tortonese E (eds) Fishes of the North-Eastern Atlantic and the Mediterranean, vol 2. UNESCO, Paris, pp 680–710
42. Tan D, Parus A, Dunbar M, Espeland M, Willmot KR (2021) Cytochrome *c* oxidase subunit I barcode species delineation methods imply critically underestimated diversity in “common” *Hermeuptychia* butterflies (Lepidoptera: Nymphalidae: Satyrinae). Zool J Linn Soc-Lond. doi:10.1093/zoolinnean/zlab007
43. Tang CQ, Humphreys AM, Fontaneto D, Barraclough TG (2014) Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. Methods Ecol Evol. doi:10.1111/2041-210X.12246
44. Torres-Hernández E, Betancourt-Resendes I, Angulo A, Robertson DR, Barraza E, Espinoza E, Díaz-Jaimes P, Domínguez-Domínguez O (2022) A multi-locus approach to elucidating the evolutionary history of the clingfish *Tomicodon petersii* (Gobiesocidae) in the Tropical Eastern Pacific. Mol Phyl Evol. doi:10.1016/j.ympev.2021.107316
45. Vitecek S, Kučinić M, Previšić A, Živić I, Stojanović K, Keresztes L, Bálint M, Hoppeler F, Waringer F, Graf W, Pauls SU (2017) Integrative taxonomy by molecular species delimitation: multi-locus data corroborate a new species of Balkan Drusinae micro-endemics. BMC Evol Biol. doi:10.1186/s12862-017-0972-5
46. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philos T R Soc Lon B Biol. https://doi.org/10.1098/rstb.2005.1716
47. Zhang J, Kapli P, Pavlidis P, Stamakis A (2013) A general species delimitation method with applications to phylogenetic placements. Bioinformatics. doi:10.1093/bioinformatics/btt499

Supplementary Tables

Supplemental Tables S1-S2 are not available with this version

Figures

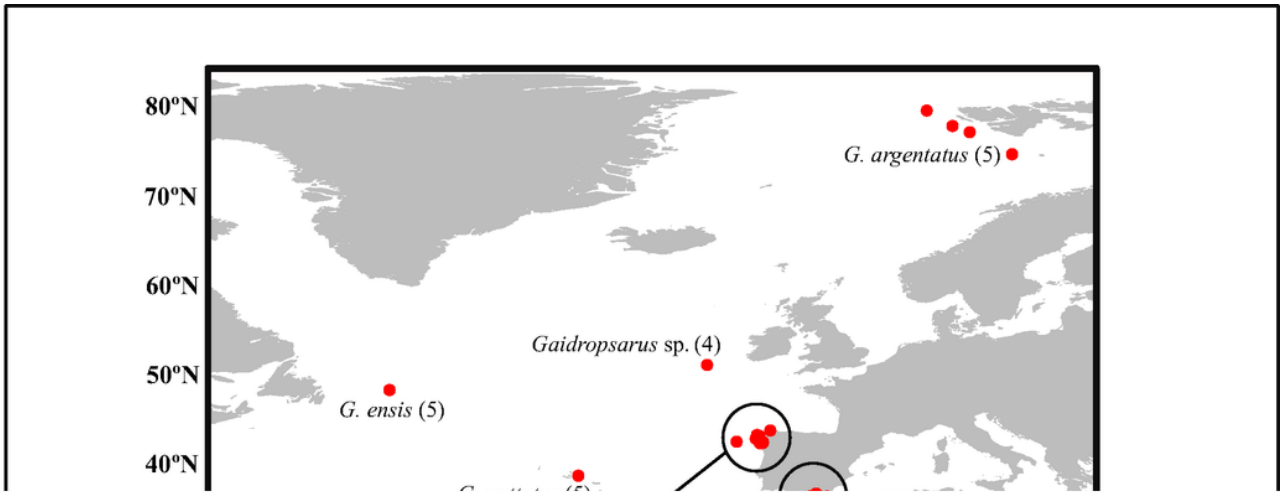

Figure 1

Map of the sampled points in the North Atlantic Ocean and the Mediterranean Sea. Each point shows the species found and the number of specimens of each one between brackets.

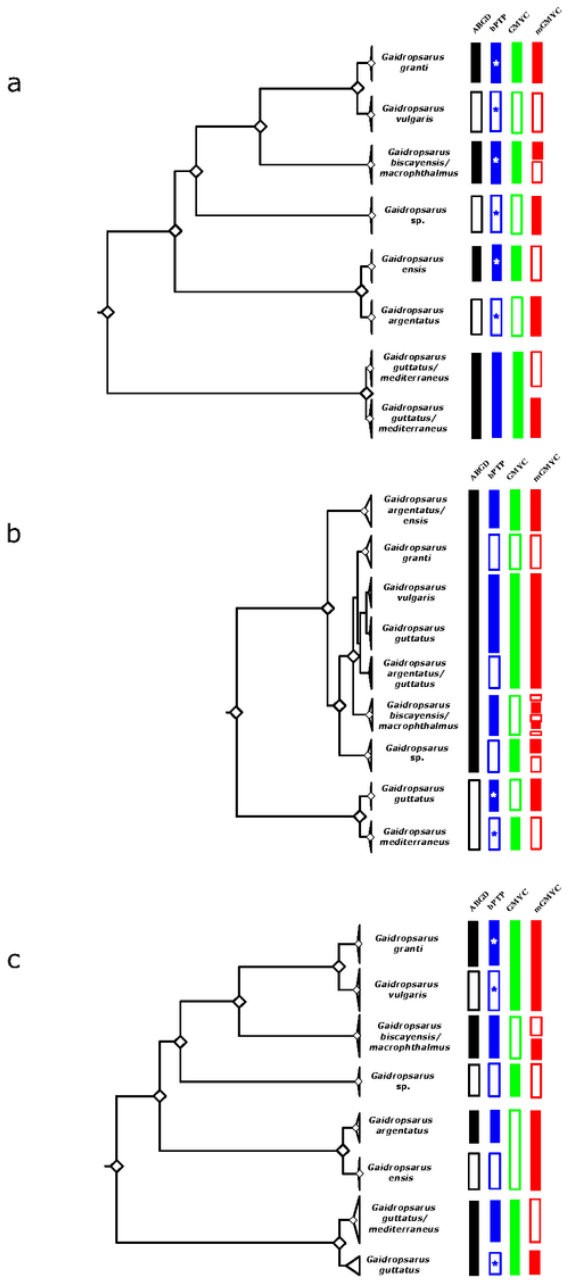

Figure 2

Collapsed consensus trees obtained with Bayesian Inference and Maximum Likelihood for three combined datasets; a: Mitochondrial DNA markers (*COI*, *CytB* and *ND2*), b: nuclear DNA markers (Rhodopsin and *ZIC1*), and c: mtDNA and nDNA concatenated. Nodes with bayesian posterior probability and bootstrap values over 0.9 are highlighted with a white diamond. Each collapsed branch show the species represented in each one. The results of the delimitation analyses are included; Automatic Barcode Gap Discovery (ABGD) in black, bPTP in blue, GMYC single threshold in green and multiple threshold GMYC in red. Clusters obtained in bPTP with a bayesian support over 0.9 are highlighted with an asterisk. The outgroup individuals have been discarded in this figure since they have been not used in the species delimitation analyses.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SFigure1mtDNAtree.pdf
- SFigure2mtDNAsinglemarkerstrees.pdf
- SFigure3nDNAtree.pdf
- SFigure4Rhosinglemarkertree.pdf
- SFigure5ZIC1singlemarkertree.pdf
- SFigure6Concatenatedtree.pdf