MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2014) **14**, 1210–1221 doi: 10.1111/1755-0998.12257

Spatial heterogeneity in the Mediterranean Biodiversity Hotspot affects barcoding accuracy of its freshwater fishes

M. F. GEIGER,[1] F. HERDER,[1] M. T. MONAGHAN,[2] V. ALMADA,[3] R. BARBIERI,[4] M. BARICHE,[5] P. BERREBI,[6] J. BOHLEN,[7] M. CASAL-LOPEZ,[8] G. B. DELMASTRO,[9] G. P. J. DENYS,[10] A. DETTAI,[10] I. DOADRIO,[8] E. KALOGIANNI,[4] H. KÄRST,[11] M. KOTTELAT,[12] M. KOVAČIĆ,[13] M. LAPORTE,[6,14] M. LORENZONI,[15] Z. MARČIĆ,[16] M. ÖZULUĞ,[17] A. PERDICES,[8] S. PEREA,[8] H. PERSAT,[18] S. PORCELOTTI,[19] C. PUZZI,[20] J. ROBALO,[3] R. ŠANDA,[21] M. SCHNEIDER,[11] V. ŠLECHTOVÁ,[7] M. STOUMBOUDI,[4] S. WALTER[1] and J. FREYHOF[1,2]

[1]*Zoologisches Forschungsmuseum Alexander Koenig, Leibniz Institute for Animal Biodiversity, Adenauerallee 160, 53113 Bonn, Germany,* [2]*Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 301 & 310, 12587 Berlin, Germany,* [3]*Eco-Ethology Research Unit, ISPA-Instituto Universitário, Rua Jardim do Tabaco, 34, 1149-041, Lisbon, Portugal,* [4]*Hellenic Centre for Marine Research, 46.7 km Athens Sounio ave., P.C. 19013 Anavyssos, Greece,* [5]*American University of Beirut, Riad El Solh, 1107 2020 Beirut, Lebanon,* [6]*Institut des Sciences de l'Évolution de Montpellier, Place Eugène Bataillon – Université Montpellier 2, 34095 Montpellier Cedex 05, France,* [7]*Institute of Animal Physiology and Genetics, Liběchov Rumburská 89, 277 21 Liběchov, Czech Republic,* [8]*National Museum of Natural History, José Gutiérrez Abascal 2, 28006 Madrid, Spain,* [9]*Natural History Museum, Cascina Vigna via San Francesco di Sales 188, 10022 Carmagnola, Italy,* [10]*Muséum national d'Histoire naturelle, 57 Rue Cuvier, 75005 Paris, France,* [11]*Deutsche Killifischgemeinschaft e.V., 65183 Wiesbaden, Germany,* [12]*Route de la Baroche, 2952, Cornol, Switzerland,* [13]*Natural History Museum Rijeka, Lorenzov prolaz 1, 51000 Rijeka, Croatia,* [14]*Centre de Recherche sur les Interactions Bassins Versants – Écosystèmes Aquatiques, C.P. 500, Trois-Rivières, QC G9A 5H7, Canada,* [15]*University of Perugia, Via Elce di Sotto, 06123 Perugia, Italy,* [16]*University of Zagreb, Rooseveltov trg 6, 10000 Zagreb, Croatia,* [17]*Istanbul University, 34134 Vezneciler, Istanbul, Turkey,* [18]*Université Claude Bernard, Lyon 1, 69622 Villeurbanne Cedex, France,* [19]*Associazione Ichthyos Italia, Via. A Cecchi 12, 52100 Arezzo, Italy,* [20]*Gestione e Ricerca Ambientale Ittica Acque, 21020 Varano Borghi, Italy,* [21]*National Museum, Vinohradská 1, 110 00 Prague, Czech Republic*

Abstract

Incomplete knowledge of biodiversity remains a stumbling block for conservation planning and even occurs within globally important Biodiversity Hotspots (BH). Although technical advances have boosted the power of molecular biodiversity assessments, the link between DNA sequences and species and the analytics to discriminate entities remain crucial. Here, we present an analysis of the first DNA barcode library for the freshwater fish fauna of the Mediterranean BH (526 spp.), with virtually complete species coverage (498 spp., 98% extant species). In order to build an identification system supporting conservation, we compared species determination by taxonomists to multiple clustering analyses of DNA barcodes for 3165 specimens. The congruence of barcode clusters with morphological determination was strongly dependent on the method of cluster delineation, but was highest with the general mixed Yule-coalescent (GMYC) model-based approach (83% of all species recovered as GMYC entity). Overall, genetic morphological discontinuities suggest the existence of up to 64 previously unrecognized candidate species. We found reduced identification accuracy when using the entire DNA-barcode database, compared with analyses on databases for individual river catchments. This scale effect has important implications for barcoding assessments and suggests that fairly simple identification pipelines provide sufficient resolution in local applications. We calculated Evolutionarily Distinct and Globally Endangered scores in order to identify candidate species for conservation priority and argue that the evolutionary content of barcode data can be used to detect priority species for future IUCN assessments. We show that large-scale barcoding inventories of complex biotas are feasible and contribute directly to the evaluation of conservation priorities.

Keywords: DNA barcoding, Evolutionarily Distinct and Globally Endangered score, fish, freshwater diversity, Mediterranean Biodiversity Hotspot, molecular identification

Received 14 January 2014; revision received 15 March 2014; accepted 19 March 2014

Correspondence: Matthias F. Geiger, Fax: +0049 228 9122 252;
E-mail: m.geiger.zfmk@uni-bonn.de

Introduction

The Biodiversity Hotspot (BH) concept serves to prioritize geographical regions of high conservation value (Myers 1988; Mittermeier *et al.* 1999), and the Mediterranean area was included in the first list of 25 globally important BHs (Myers *et al.* 2000). The area is geographically highly structured and includes 23 ecoregions in 20 countries (Abell *et al.* 2008). While the initial Hotspot classification was based on plant diversity, the freshwater fauna of the region is also considerably rich (Mittermeier *et al.* 1999; De Figueroa *et al.* 2013). Despite the disproportionately small global surface cover of freshwater habitats ('paradox of freshwater biodiversity', Martens 2009), the importance of freshwater biodiversity for humans is increasingly recognized (Dudgeon *et al.* 2006). Hotspots of freshwater diversity are not necessarily congruent with terrestrial hotspots (Abell *et al.* 2008), nonetheless, the species numbers reported by Mittermeier *et al.* (2004) were surprisingly low: only 216 native freshwater fishes for the Mediterranean BH with 63 endemic species. That these numbers were a gross underestimation became clear when the IUCN Red List for endemic Mediterranean freshwater fish species, published in 2006, included 253 species (Smith & Darwall 2006). Recent descriptions of new freshwater fish species demonstrate that the Mediterranean BH remains only partially explored (Kottelat & Freyhof 2007; Fontaine *et al.* 2012; Essl *et al.* 2013), which is a stumbling block for conservation planning (Freyhof & Brooks 2011). To overcome this, molecular information is increasingly used to explore biodiversity and improve biodiversity knowledge, including large-scale barcoding approaches for molecular identification systems that serve various purposes including conservation (Hebert *et al.* 2003, 2004; Hebert & Gregory 2005; Lahaye *et al.* 2008; May 2011; Mora *et al.* 2011).

Molecular information is also increasingly used to evaluate conservation priorities, complementing the IUCN classification (Rolland *et al.* 2012; Abellán *et al.* 2013). Such data are of special importance for freshwater fishes in the Mediterranean BH, where a large number of threatened species occur (56% of 253 evaluated species, Smith & Darwall 2006; 17 endemics already extinct Freyhof & Brooks 2011). One approach to assess priority species from a given list of threatened species is the Evolutionarily Distinct and Globally Endangered (EDGE) score, which combines estimators for unique evolutionary history with the formal IUCN conservation status (Isaac *et al.* 2007). In this approach, globally threatened species that represent isolated and phylogenetically old lineages receive highest conservation priority, because their loss would mean a disproportionate loss of unique evolutionary history (Isaac *et al.* 2007; Safi *et al.* 2013).

The EDGE approach appears compelling, especially considering the growing availability of DNA barcode data (Hebert *et al.* 2003, 2004; Hebert & Gregory 2005; Lahaye *et al.* 2008; May 2011; Mora *et al.* 2011; April *et al.* 2013); however, its meaningful application depends on correctly identified material underlying barcode reference libraries. Moreover, a recent study in aquatic insects concluded that the reliability of species determination by DNA barcoding may be scale dependent (Bergsten *et al.* 2012). This means that large-scale databases may be problematic for species identification, because evolutionarily close relatives may occur in distant localities. Thus, as the geographical and taxonomic size of the database increases, the resolution may decrease because of the inclusion of ever more closely related species. This may be particularly problematic in geographically highly structured areas, where allopatric speciation likely accounts for a considerable proportion of the species diversity.

Here, we present a DNA barcode database of freshwater fishes of the Mediterranean BH that covers 98% of the species known from the area. This comprehensive database is verified by taxonomic experts and allows us to address the following questions: How closely do sequence-based diversity estimates of this highly geographically structured area mirror the estimates based on morphology? How do analytical, intrinsic biological (e.g. introgressions) and extrinsic geographical (scale) factors affect accuracy and practical implementation of DNA barcoding? Can phylogenetic information from DNA barcode data be used to assist selection of EDGE species for conservation?

Materials and methods

The Mediterranean BH of Mittermeier *et al.* (1999) and Myers *et al.* (2000) includes all areas of the Mediterranean floral zone, therefore including Portugal and the Atlantic parts of Spain and Morocco, as well as the Macaronesian islands. The geographical area considered for this study (Fig. 1) is the same used for the IUCN Red List assessment (Smith & Darwall 2006). The Macaronesian islands and Libya were not considered because of the absence of endemic fish species. A total of three endemic species from the Egyptian Mediterranean area were included, and all others belong to the largely afro-tropical fish fauna (Roberts 1975) and are excluded.

The analysed material was collected with the aid of numerous colleagues, or stems from available collections. When possible, multiple individuals (2–38) from single species from different drainages were included

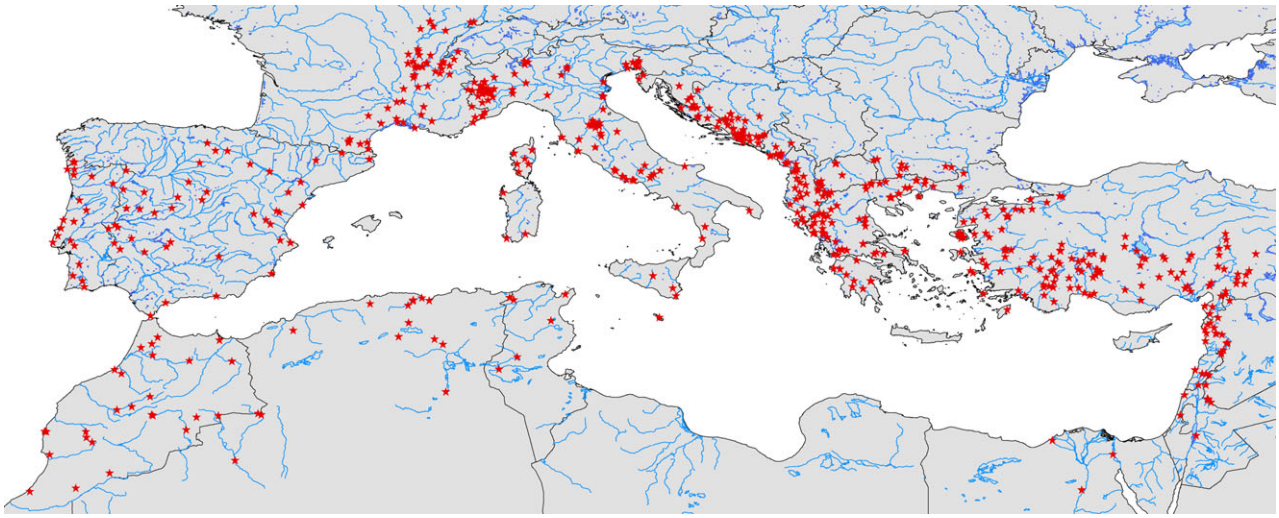

Fig. 1 Sample locations (n = 657) for materials obtained from within the Mediterranean Biodiversity Hotspot with major rivers and country borders.

to estimate intraspecific genetic variation (Appendix S1, Table S1, Supporting information). As the barcode (cytochrome *c* oxidase subunit 1: COI) data published here will serve as a reference database, we applied a strict policy to ensure data quality and considered only material reliably determined by taxonomic specialists. With a few exceptions (see Results section), all sequences were taken from individual voucher specimens deposited in publicly available collections (Table S1). Data associated with each specimen (taxonomy, collection sites and voucher catalogue numbers) are available via the respective GenBank accession numbers (see Data accessibility below, Table S1) and will also be made available via the BOLD data portal. As reference to which we compare our molecular-based diversity estimates, we use the species listed as valid in Eschmeyer's *Catalogue of Fishes* (Eschmeyer & Fong 2013). We evaluated the performance of different analytical approaches (distance- versus tree-based species delimitation) in terms of congruence to traditional taxonomy and compare the standard barcoding metrics of the Mediterranean Hotspot to other large-scale DNA barcoding studies. Our sampling scheme allowed us to assess the recently demonstrated impact of geographical scale on DNA barcoding accuracy (Bergsten *et al.* 2012; Lou & Golding 2012). We did so by separately analysing the fish faunas from three Mediterranean drainages in a focal approach, which constitutes a practical and realistic monitoring scenario on a local scale. For this, we chose the Po in Italy (48 species, 229 individuals), the Vardar in Greece (36 species, 159 individuals) and the Orontes in Turkey and Syria (34 species, 126 individuals).

Molecular data analysis

Details for DNA extraction, PCR and standard barcode data preparation are listed in the Supporting information (Appendix S2). Clustering with Kimura 2-corrected distances (K2P) in SpeciesIdentifier (Meier *et al.* 2006) was used to cluster sequences at 2%, 1% and an optimized, data-derived threshold. Taking into account the number of true and false positives, and true and false-negative identifications at a given threshold with cumulative error (false negative + false positive), the optimized threshold was derived from each data partition with the SPIDER package (Brown *et al.* 2012) in R v 2.15.1 (R Development Core Team 2011). We then counted the number of clusters in agreement with existing taxonomy, that is, containing only sequences of one morphological species ('perfect match'), as well as the total number of clusters (Hendrich *et al.* 2010). We treated clusters containing all COI sequences of one species together with one or more sequences that were determined to genus level only as a 'perfect match'. As the simplest, but most widely used distance-based approach has been criticized for being arbitrarily chosen without sound biological background (Meier *et al.* 2006; Srivathsan & Meier 2012; Collins & Cruickshank 2013), and the threshold varies between species from different groups (Hebert *et al.* 2003), we applied additional methods to analyse the data. First, the species criterion of monophyly was applied to test for congruence, requiring the grouping of all COI haplotypes of a given species as a monophyletic unit in a given phylogram; here, we screened ML trees for those groups generated under the model

assumptions as derived below. Singletons were considered as match, if they did not cluster unresolved within another species, and we counted every monophyletic clade without defining a certain *a priori* support. Second, the general mixed Yule-coalescent (GMYC) approach was used, which identifies clusters by fitting models that predict the inter- and intraspecific divergence rates and threshold times differentiating these processes to multispecies coalescent trees (Monaghan *et al.* 2009; Powell 2012). The model assumes that branching patterns within genetic clusters reflect neutral coalescent processes and occur within species (Kingman 1982), whereas branching between clusters can reflect the timing of a speciation event (Yule 1924). Thus, it identifies a species boundary by identifying independently evolving lineages and the transition from coalescent to speciation branching patterns on a phylogenetic tree. Although not uncontroversial (Lohse 2009; Esselstyn *et al.* 2012), the theory behind the model is very attractive and it has proven to deliver reliable and biological sound species number estimates (Fujita *et al.* 2012; Powell 2012; Puillandre *et al.* 2012; Talavera *et al.* 2013). Details for the generation of ultrametric trees for the GMYC analysis are given in the Supporting information (Appendix S2). Finally, we applied an extension to the GMYC approach (Powell 2012) that takes into account additional models with slightly lower likelihoods and uses an Akaike information criterion (AIC) of all single- and multiple-threshold models (Powell 2012). The probabilities that two haplotypes belong to the same entity are then based on the weights associated with each model, and the variance is estimated from model averaging. Thus, uncertainty in species boundaries can directly be incorporated into diversity estimates. Based on the probabilities associated with each genetic cluster, we derived the number of morphological species supported by the GMYC approach, analogous to the 'perfect match' above, also including singletons.

Evolutionary distinctiveness (ED) was calculated from an ultrametric tree with all endemic species only with the CAPER package in R v 2.15.1 (R Development Core Team 2011). The ED estimates where then used to calculate the EDGE scores as outlined in Isaac *et al.* (2007).

Results

DNA barcode library

Of the 526 species of freshwaters fishes currently recognized for the Mediterranean BH (Eschmeyer & Fong 2013), 17 are considered to be extinct by IUCN (www.redlist.org) or local experts (Appendix S1, Supporting information). We analysed DNA barcode data (mitochondrial COI) of 98% (498 species) of the remaining 509 extant species, from a total of 20 countries (Fig. 1, Table 1). We newly sequenced DNA barcodes of 2899 individuals from 487 species, added 89 sequences from 21 species from GenBank and 183 sequences from 58 species from other DNA barcoding studies (Appendix S1, Supporting information). Eleven species were represented with material from outside the Mediterranean due to their rarity, or because they are thought to be extirpated in the BH (Appendix S1, Supporting information). Altogether, 2809 COI sequences belonged to material unambiguously identified to species level, and an additional 318 sequences from 30 genera where species membership was uncertain *a priori* (e.g. juveniles or unclear species boundaries). After re-examination of the vouchers from which DNA was extracted, literature research, and taking into account their haplotype clustering (Fig. S1, Supporting information), 115 of the 318 unidentified specimens could be assigned to 18 species currently recognized as synonyms (Eschmeyer & Fong 2013) (Table S2, Supporting information).

Introgression and hybrids

Hybridization and introgression are common in freshwater fishes, in particular in the Cyprinidae (Scribner *et al.* 2000; Freyhof *et al.* 2005), which comprised 56% of species in our data set. Seven hybrids identified in the field were excluded from subsequent analyses; all but one of these belonged to the Cyprinidae, including two cases of intergeneric crosses (Table S2, Supporting information). In addition, 44 individuals (1.4%) from 26 species had COI haplotypes not matching expectations from

Table 1 Summary of basic barcode statistics with ranges or subsets analysed in parentheses

Basic statistic	
Total individuals	3165
Species total number (analysed)	526 (498)
Number of genera	113
Species assessed as extinct	18
Extant endemics (barcoded)	382 (372)
Alien species	37
Mean number of individuals per species	5.4 (1–38)
Number of singletons	33
Mean sampling events per species	2 (1–18)
Mean sequence length in base pairs	646 (454–652)
Mean intraspecific distance	0.59% (0.02–12.5%)
Mean smallest interspecific distance	4.10% (0.1–16.8%)
95% intraspecific variance ≤	1.98%
95% smallest interspecific distance ≥	0.21%

morphological identification (Table S2, Supporting information). Fourteen of these were in the Cyprinidae, mostly in species for which introgression has been reported elsewhere, but to our knowledge, some cases are reported here for the first time. One case of presumed introgression was observed in *Oxynoemacheilus seyhanicola* with two distinct COI haplotype groups (>12% K2P), which can however not be attributed to a recent introgression event, because neither haplotype occurred in any other species.

Cytochrome c oxidase subunit 1 divergence in the full data set

Mean K2P nucleotide divergence increased from species (m = 0.59%, SE = 0.183) to genera (m = 2.89%, SE = 0.300) to families (m = 6.44%, SE = 1.000). Mean distance to the closest congener across all genera was 4.10% (SE = 0.500), seven times higher than the mean intraspecific divergence. A lack of variation (i.e. identical COI sequences) was observed in 28% of all species with multiple sequences (2–12). For intraspecific distances, 4.9% of the 390 species with multiple individuals and nonzero variation had values over 2%, while 11.0% of those species showed above 1% COI divergence. A refined barcode gap [difference between the maximum intraspecific and minimum interspecific divergence (Meyer & Paulay 2005; Meier *et al*. 2008)] was present in 74.8% of all species. The distribution of intra- and interspecific congeneric K2P-corrected distances displayed considerable overlap (Fig. 2). Among all individual intraspecific comparisons, 91.5% were below 2% sequence divergence. Interspecific distances were below 2% sequence divergence in 11.3% of all pairwise comparisons (Table 1).

DNA barcoding accuracy

When applying the 2% K2P divergence clustering approach to the complete set of 498 freshwater fish species, a total of 391 clusters were recovered; 230 of these were in perfect match with the species entities determined by morphology. This translates to 44.8% of the expected species recovered using the simple distance-based approach. For the Cyprinidae (data set IV), the match increased from 34.8% to 44.7% when using a 1% threshold (Table 2). Highest accuracy, with 82% of the species resolved as perfect barcode cluster applying a 1%

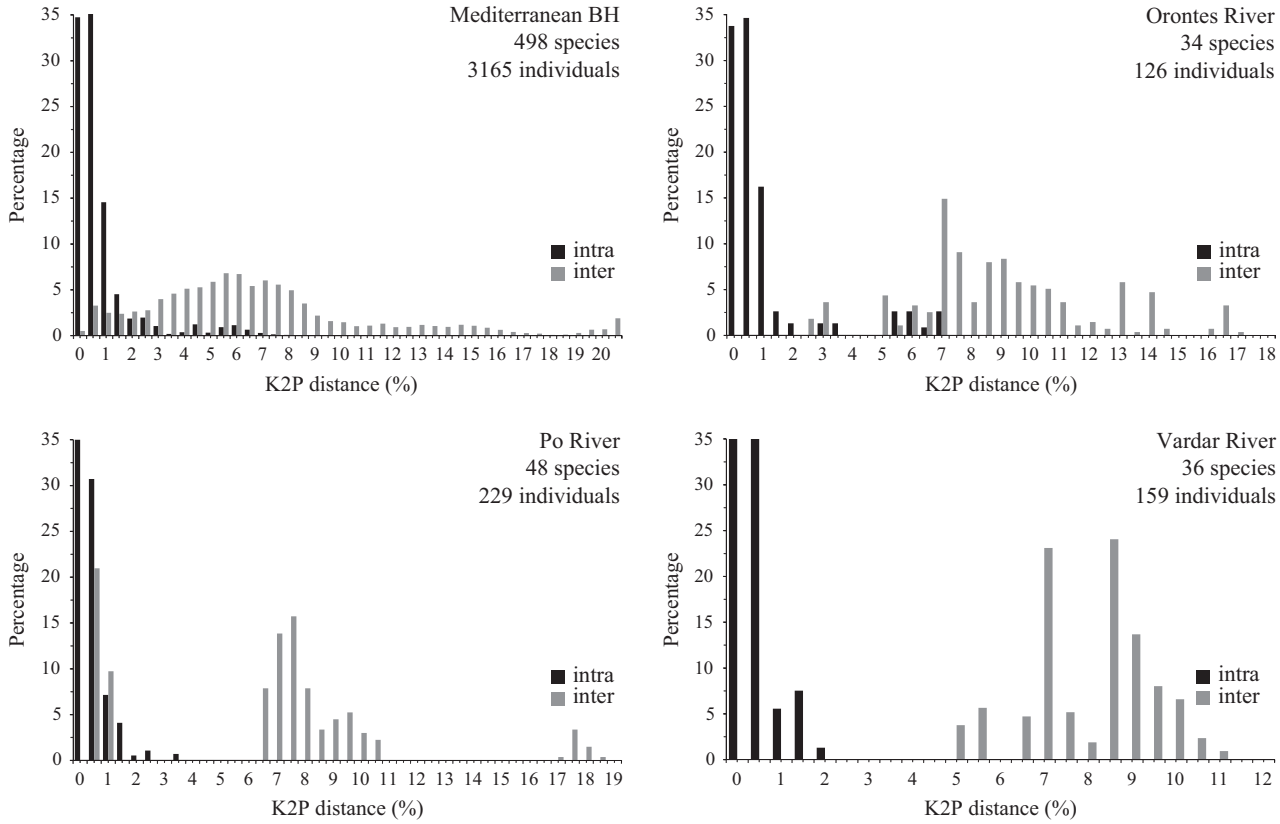

Fig. 2 Frequency distribution of the intra- and interspecific pairwise Kimura 2-corrected (K2P) distances for all 498 species (upper left), and each native fish fauna for three selected drainages of the Mediterranean Biodiversity Hotspot.

Table 2 Number of clusters obtained with different distance thresholds and identification success percentage for different analytical methods; seq: number of cytochrome *c* oxidase subunit 1 sequences analysed; spp: number of species; optimum: threshold values from cumulative error estimation in parenthesis; perfect cluster: percentage of clusters in congruence with taxonomy; general mixed Yule-coalescent (GMYC): single- and multiple-threshold entity estimation with confidence intervals (CI)

			Distance cluster			% perfect cluster					GMYC entities			
Group	Seq	Spp	2%	1%	Optimum	2%	1%	Optimum	Mono	GMYC	Single	CI	Multi	CI
I	289	62	69	77	69 (2.04)	67.7	67.7	67.7	88.7	86.9	79	73–85	77	71–91
II	331	44	53	60	57 (1.19)	80.0	82.0	80.0	94.0	93.2	62	56–70	46	42–83
III	247	34	34	37	31 (2.29)	62.9	62.9	57.1	77.1	73.5	36	32–42	38	30–51
IV	1914	289	196	268	247 (1.24)	34.8	44.7	39.4	71.7	81.9	322	305–349	347	346–363
V	215	34	31	33	31 (1.78)	69.0	71.9	69.0	90.6	90.6	7	1–88	37	13–47
VI	175	35	11	13	10 (5.5)	10.3	10.3	10.3	46.2	78.1	9	8–31	20	9–33
All	3165	498	391	490	433 (1.45)	44.8	51.2	49.2	–	85.2*	463.54 (variance 6.6)*			

Group composition: (I) *Barbatula, Cobitis, Misgurnus, Oxynoemacheilus, Paramisgurnus, Sabanejewia, Seminemacheilus*; (II) *Acipenser, Ameiurus, Anguilla, Aphanius, Chelon, Clarias, Esox, Fundulus, Gambusia, Huso, Lota, Mugil, Mystus, Poecilia, Pterygoplichthys, Silurus, Valencia, Xiphophorus;* (III) *Atherina, Cottus, Dicentrarchus, Economidichthys, Gasterosteus, Knipowitschia, Neogobius, Ninnigobius, Orsinigobius, Padogobius, Platichthys, Pomatoschistus, Proterorhinus, Pungitius, Syngnathus;* (IV) *Abramis, Acanthobrama, Achondrostoma, Alburnoides, Alburnus, Anaecypris, Aulopyge, Barbus, Blicca, Capoeta, Carasobarbus, Carassius, Chondrostoma, Clupeonella, Ctenopharyngodon, Cyprinus, Delminichthys, Garra, Gobio, Hypophthalmichthys, Iberochondrostoma, Iberocypris, Ladigesocypris, Leucaspius, Leuciscus, Luciobarbus, Mirogrex, Pachychilon, Parachondrostoma, Pelasgus, Petroleuciscus, Phoxinellus, Phoxinus, Protochondrostoma, Pseudochondrostoma, Pseudophoxinus, Pseudorasbora, Pterocapoeta, Rhodeus, Romanogobio, Rutilus, Scardinius, Squalius, Telestes, Tinca, Tropidophoxinellus, Vimba;* (V) *Amatitlania, Australoheros, Caspiomyzon, Coptodon, Eudontomyzon, Gymnocephalus, Haplochromis, Hemichromis, Lampetra, Lepomis, Micropterus, Morone, Odontesthes, Oreochromis, Perca, Petromyzon, Pseudocrenilabrus, Salaria, Sander, Sarotherodon, Tristramella, Zingel;* (VI) *Alosa, Clupeonella, Coregonus, Hucho, Ictalurus, Oncorhynchus, Salmo, Salvelinus, Thymallus;* (All) all sequences including unidentified specimens.
*Based on the average numbers over all models, calculated with the modified GMYC approach (Powell 2012).

divergence criterion, was achieved in group II, a heterogeneous group composed of 14 families with 44 species only. Lowest congruence (10.3%) between morphological species and barcode clusters detected is present in group VI (salmonids and herrings), irrespective of the distance threshold chosen. In summary, distance-based methods appear suitable to detect families and species groups, but did not support all of the species expected in the full data set.

The tree-based analysis outperformed the distance-based approach in recovering patterns congruent between *a priori* species identifications and molecular-based units (Table 2). In total, 83% of all species were recovered as GMYC entities, with probabilities above 0.95 containing only conspecifics, and including singletons. In four of the six data partitions, a multiple-threshold model provided the best fit to the data, but the even distribution of the weights given to each model suggests that no single model best represented species boundaries in a given subset of the data (Appendix S2, Supporting information). We did not observe a relationship between number of haplotypes per species and assignment probabilities (GLM, $P = 0.2372$), thus we did not find an effect of sample size bias on GMYC accuracy. Of all species represented by multiple individuals ($n = 465$), 74.4% carried COI haplotypes that formed monophyla in the screened ML trees. Mean bootstrap support derived from 500 pseudoreplicates for those species was 95.4% (SD = 10.7), and 83% of these clusters had support values above 90%. As for the GMYC analysis, we did not observe a relationship between number of individuals per species and bootstrap support (GLM, $P = 0.8201$).

Cytochrome c oxidase subunit 1 divergence within selected drainages

We used our comprehensive database to test whether barcoding accuracy changes under conservation- and operator-oriented conditions on a smaller scale, using the species sets from three selected large Mediterranean drainages (Po in Italy, 48 species; Vardar in Greece, 36 species; Orontes in Turkey and Syria, 34 species). Here, the mean distance to the closest congeneric species was similar for all three drainages (6.50–6.93%) and was, on average, 19 times higher than the mean intraspecific divergence (0.26–0.55%). Compared with the ratio derived from the complete data set, this is a threefold increase. We found intraspecific divergences exceeding 2% in one species in the Orontes (*Aphanius mento*), where a cryptic species might be involved, in two cases in the Po (*Thymallus aeliani* and *Barbus plebejus*) where introgression by alien, congeneric species is likely, and in two individuals of *Orsinigobius punctatissimus*. In all species occurring in the three drainages, the maximum intraspecific divergence was smaller than the minimum interspecific divergence, indicating the presence of the

refined barcode gap (Meyer & Paulay 2005; Meier *et al.* 2008) (Fig. 2). Barcoding accuracy under the 2% distance-based approach was between 96% (Po) and 100% (Orontes and Vardar). For the Po sample, the two *Salmo* and *Alosa* species occurring within the drainage show COI sequence divergences below 2% and thus decreased overall accuracy.

Candidate species

Divergence levels between haplotypes exceeding 2% K2P remained in ten of the species analysed (Table S2, Supporting information). Examples for allopatric subclades with geographically structured haplotype pools included *A. mento* (intraspecific distance ≤8%) and *Telestes pleurobipunctatus* (intraspecific distance ≤3%). The first species is challenging, given that material from the type locality (Northern Iraq) is currently not available; the type series of *T. pleurobipunctatus* is composed of populations found to belong to different subclades and, awaiting nomenclatural clarification, cannot be applied to a particular one. Here, we define candidate species as (i) intraspecific clusters exceeding 2% K2P distance, (ii) *a priori* to genus level only identified populations with over 2% K2P distance to a described species, and (iii) the 18 species which were treated as synonym before, and which are proposed to be revalidated. All candidate species fulfilled the monophyly criterion and were detected as GMYC entities. This indicates the existence of cryptic diversity representing up to 64 candidate species (Table S2 and Fig. S1, Supporting information).

Evolutionarily Distinct and Globally Endangered score and IUCN status

We used the COI sequence data to calculate EDGE scores for 311 IUCN Red List fish species endemic to the Mediterranean Hotspot. We found no correlation between ED and IUCN ranking (Spearman's rank-order $r = 0.0965$, $P = 0.08$), but a clear correlation between EDGE scores and IUCN ranks ($r = 0.963$, $P < 0.0001$, Fig. S2, Supporting information). This correlation was expected given that the IUCN status contributes to the EDGE score (see Materials and methods). Consequently, it was not surprising that the top 10 EDGE score species are also listed as critically endangered; yet, a number of 'outlier' species with higher than expected EDGE scores for their IUCN category were identified (Fig. S2, Supporting information). In addition, the top three species, ranked according to their ED, include two listed by IUCN as data deficient (DD), the other as least concern (LC), followed by a set of species which is different from the set ranked by EDGE scores.

Discussion

The present data (with 98% species coverage) comprise the first comprehensive molecular study on the freshwater fish diversity of a complete BH and will serve as reference for future studies of this large fauna. Our finding that model-based clustering of sequences outperformed the simplest, but most widely used version of distance-based clustering is consistent with theoretical and previous empirical work (Papadopoulou *et al.* 2008; April *et al.* 2011; Bergsten *et al.* 2012). Low levels of introgression and the focal analyses of three drainages demonstrated that DNA barcoding is a powerful tool for specimen identification at the drainage scale, even with rapid distance-based methods. With drainage-specific species lists and barcoding data available, species can be identified with high accuracy. Our findings support the notion that challenges of barcode identification may increase with the complexity of the reference database and corroborate a recent empirical test of the effects of spatial scale on DNA barcoding (Bergsten *et al.* 2012). We identified several candidate populations that may be new species and require focused taxonomic research, potentially raising diversity by 12%. Phylogenetic information from COI barcode data delivered estimates of ED for the calculation of EDGE scores and allowed the identification of species for conservation prioritization. This constitutes a promising approach to practically benefit from growing DNA barcode libraries to assist conservation planning, or rapidly pre-assess data-deficient species.

Divergence levels in Mediterranean freshwater fishes

The large geographical scale of our study may explain why divergence levels within species largely met expectations from other studies of freshwater fishes (Ward *et al.* 2005; April *et al.* 2011; De Carvalho *et al.* 2011; Pereira *et al.* 2011, 2013), but divergence within genera did not (April *et al.* 2011; Pereira *et al.* 2013). The only other study of freshwater fishes of comparable scale and scope (i.e. including several ecoregions) (April *et al.* 2011) reported a mean divergence within species of 0.73% (0.59% this study), within genera of 13.67% (2.89% this study) and within families of 15.91% (6.44% this study) in North America. Comparisons with studies of marine fishes (Ward *et al.* 2005; Lakra *et al.* 2011) are of limited value because dispersal of pelagic larvae should lead to lower levels of allopatric divergence through gene flow (Helfman *et al.* 2009). It has also been proposed that differences in vagility, as in the marine environment as opposed to the freshwater realm, can be responsible for different divergence levels as well (Bergsten *et al.* 2012; Young *et al.* 2013).

The low divergence within genera and families that we observed suggests that species in Mediterranean genera and families are more closely related than in North America and has implications for the use of DNA barcodes for identification. In the BH, relatively recent (2.5–0.01 Ma) dispersal and vicariance events are thought to have had the greatest impact on species formation and contemporary diversity (Banarescu 1989; Zardoya & Doadrio 1999). The low differentiation in COI between several Mediterranean freshwater fish species might also be related to a combination of renewed species concepts in ichthyology (Kottelat 1997) and increased importance of molecular tools for species discoveries and descriptions (Doadrio & Perdices 1997). A review of the European freshwater fish fauna in 2007 (Kottelat & Freyhof 2007) recognized 546 native species, compared with only 215 species recognized in 1977 (Maitland 1977). From the Mediterranean freshwaters alone, 99 species have been described since the year 2000 (22% of the fauna recognized before), while only 76 and 25 have been described from North America and Australia, respectively, (9% and 16% of fauna recognized before) during the same period (Froese & Pauly 2013). This could also explain the difference in the proportion of candidate species suggested, which – if all would be formally described – would lead to a 28% increase in North American species diversity (April *et al.* 2011) versus up to 12% for the Mediterranean fauna presented herein.

Most of the species that were not genetically recognized with GMYC or distance approaches (Table S2, Supporting information) exhibit diagnosable morphological differences, which led to their formal recognition. We follow Padial *et al.* (2010) in that congruence between morphology and DNA barcode unit is preferred for the recognition of species entities, but emphasize that this criterion has clear limits, for example in cases of recent speciation. Neglecting the morphological differences to adapt strictly DNA barcode-based species leads away from a concept of species as evolving, natural, diagnosable units (Simpson 1961; Wiley 1978). On the other hand, some morphological species that were nested within larger COI haplotype clusters may not be recognized as species if taxonomic revisions were to be carried out. It is not our aim to review all these cases, but the data made available will aid future taxonomic studies (Table S2, Supporting information).

Sampling strategy

The high species coverage achieved might have influenced the performance of the nontree-based species delimitation approach. This is a general problem for large-scale barcoding studies, as demonstrated empirically by Bergsten *et al.* (2012), who concluded that the genetic distance to the closest heterospecific decreases with increasing geographical scale of sampling, linked to the larger intraspecific variance when incorporating more populations. A positive correlation between intraspecific genetic variation and geographical scale is predicted by concepts like isolation by distance (Wright 1943), where more distantly occurring subpopulations show higher divergence. Assuming that vicariance events were important in shaping the Mediterranean freshwater fish diversity, we do not expect the closest relatives to occur in the same drainage. Without doubt, this is reflected in the increased accuracy of DNA barcoding for species identification within the three river catchments analysed separately, and in general, only few cases of low interspecific divergence levels in sympatry were detected (e.g. *Lampetra planeri* and *Lampetra fluviatilis*, *Alosa alosa* and *Alosa fallax* or *Alosa algeriensis*, or various *Salmo* species from Lake Ohrid and *Salmo farioides*, *Salmo marmoratus* and *Salmo obtusirostris* in some rivers in the Adriatic). Congeneric species occurring in allopatry frequently possessed low or very low levels of molecular divergences, which hampered their detection. A practical solution to overcome this is to combine genetic and distributional data to cope with spatial scale effects in order to obtain similar identification rates at global and regional barcoding campaigns (Papadopoulou *et al.* 2008; Bergsten *et al.* 2012), or to construct and implement barcode databases on local or regional scales.

Conservation prioritization

Evolutionarily Distinct and Globally Endangered species are threatened species with few or no close relatives on the tree of life (Isaac *et al.* 2007). To date, the EDGE approach has only been applied to globally complete phylogenies of three major taxonomic groups: mammals (Isaac *et al.* 2007), amphibians (Safi *et al.* 2013) and reef corals (Huang 2012). The approach has not been used for regional data sets, such as single BHs, because excluding close relatives occurring outside the study area can lead to falsely elevated EDGE scores. Given that there are approximately 29 000 species of fishes (Lévêque *et al.* 2007), but only 5500 species of mammals (Wilson & Reeder 2005) and 7044 species of amphibians (Frost 2013), a complete global fish phylogeny remains a larger task. At present, this precludes fishes and other highly diverse groups from a global EDGE assessment. We see no reason not to apply the methodology to a regional fauna, if species with close relatives outside the studied area are excluded, here based on published comprehensive faunal assessments covering the Mediterranean BH (Lévêque & Daget 1984; Geldiay & Balık 2007; Kottelat & Freyhof 2007). As for amphibians (Safi *et al.* 2013) and mammals (Isaac *et al.* 2007), we found no correlation

between the IUCN threat category of a species and its ED. Therefore, a ranking according to EDGE scores allows prioritizing conservation efforts for species within one IUCN threat category. In the Mediterranean BH, the two highest ranked EDGE species (*Valencia* spp.) are the only members of the endemic family Valenciidae. Species with higher EDGE score than expected from their IUCN category (outliers in Fig. S2, Supporting information) could be candidates for a re-evaluation of their IUCN status. A conservation prioritization based on ED leads to a different set, the three most distinct species are not even present in the top 10 EDGE list. It seems promising to use molecular information (also obtained from DNA barcode data) to either complement a species' IUCN assessment, or to estimate ED for species that are not assessed. The latter approach would particularly benefit from large-scale barcode libraries.

Conclusions

This is the first large-scale study that uses DNA barcode data to estimate conservation priorities from a nearly complete freshwater fish fauna of a globally important BH. The barcodes and associated vouchers, from 3165 individuals from 498 Mediterranean freshwater fish species, will serve as baseline for all forthcoming related studies in the region. The freshwater fish diversity in the Mediterranean BH with its 23 different ecoregions (Abell *et al.* 2008) is indeed remarkable, but remains underestimated. This reaffirms that even well-known faunas still harbour unrecognized elements. If drainage-specific species lists are available, DNA barcoding is a powerful tool for specimen identification using rapid, distance-based methods for clustering DNA barcodes into species – an important notion for applications in local monitoring. The scale effect shown here demonstrates that species identification becomes more challenging as DNA barcode libraries grow in species numbers and geographical coverage, underpinned by the unexpected level of incongruence between morphological- and molecular-based species numbers. Some proportion of the described species might be attributable to a historical oversplitting of the European freshwater fish fauna by taxonomists. This may explain some of the observed incongruences; nonetheless, we refrain from synonymizing different taxa based solely on COI data, but argue that our findings may initiate necessary revisions. The results reveal the lack of a single molecular distance threshold for discriminating allopatric freshwater fish species and highlight the value of integrative and iterative identification strategies (Fujita *et al.* 2012; Jörger *et al.* 2012; Carstens *et al.* 2013). Greater accuracy from the refined GMYC method (multiple thresholds and model averaging) indicates a true value gain of these new developments and calls for attention when analysing large DNA barcode data sets.

Acknowledgements

We thank numerous colleagues for providing additional samples, including N. Alwan, S. Brosse, A. Dunz, M. Englund, T. Eriksson, F. Krönke, S. Kullander, N. Mesquita, D. Neumann, A. Nolte, L. Rüber, K. Rylkova, A. Scharbert, U. Schliewen, S. Sabatino, S. Valdesalici and P. Vonlanthen. Sebastian Hüllen, F.W. Miesen and J. Huylebrouck are thanked for clarifying some unclear determinations. Claudia Etzbauer, J. Schmidt, I. Stoeger, S. Güse and K. Gritzalis are acknowledged for help with logistics and wet-laboratory routines. We are grateful to S. Rutschmann, J. Powell and J. Pfaender for fruitful discussions on methodological issues.

Funding

The project FREDIE (Freshwater Diversity Identification for Europe, www.fredie.eu) is funded by the Joint Initiative for Research and Innovation (PAKT) program of the Leibniz Association. RŠ acknowledges support from the Czech Ministry of Culture (DKRVO2012 and DKRVO2013/14, National Museum, 00023272). This study is a product of the FREDIE project, financed in the SAW programme by the Leibniz Association.

References

Abell R, Thieme M, Revenga C, Bryer M (2008) Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, **58**, 403–414.

Abellán P, Sánchez-fernández D, Picazo F *et al.* (2013) Preserving the evolutionary history of freshwater biota in Iberian National Parks. *Biological Conservation*, **162**, 116–126.

April J, Mayden RL, Hanner RH, Bernatchez L (2011) Genetic calibration of species diversity among North America's freshwater fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 10602–10607.

April J, Hanner RH, Mayden RL, Bernatchez L (2013) Metabolic rate and climatic fluctuations shape continental wide pattern of genetic divergence and biodiversity in fishes. *PLoS ONE*, **8**, e70296.

Banarescu PM (1989) Zoogeography and history of the freshwater fish fauna of Europe. In: *The Freshwater Fishes of Europe 1/II*(ed. Holcik J), pp. 88–107. AULA Verlag, Wiesbaden.

Bergsten J, Bilton DT, Fujisawa T *et al.* (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, **61**, 851–869.

Brown SD, Collins R, Boyer S *et al.* (2012) Spider: an R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, **12**, 562–565.

Carstens BC, Pelletier TA, Reid NM, Satler JD (2013) How to fail at species delimitation. *Molecular Ecology*, **22**, 4369–4383.

Collins RA, Cruickshank RH (2013) The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, **13**, 969–975.

De Carvalho D, Oliveira D, Pompeu P *et al.* (2011) Deep barcode divergence in Brazilian freshwater fishes: the case of the São Francisco River basin. *Mitochondrial DNA*, **22**(Suppl 1), 80–86.

De Figueroa JMT, Fenoglio S, Sánchez-castillo P (2013) Freshwater biodiversity in the rivers of the Mediterranean Basin. *Hydrobiologia*, **719**, 137–186.

Doadrio I, Perdices A (1997) Taxonomic study of the Iberian Cobitis (Osteichthyes, Cobitidae), with description of a new species. *Zoological Journal of the Linnean Society*, **119**, 51–67.

Dudgeon D, Arthington AH, Gessner MO *et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews of the Cambridge Philosophical Society*, **81**, 163–182.

Eschmeyer WN, Fong JD (2013) *Species By Family/Subfamily*. (http://research.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp). Electronic version Accessed 14 July 2013.

Esselstyn J, Evans B, Sedlock J, Anwarali Khan F, Heaney L (2012) Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **279**, 3678–3686.

Essl F, Rabitsch W, Dullinger S, Moser D, Milasowszky N (2013) How well do we know species richness in a well-known continent? Temporal patterns of endemic and widespread species descriptions in the European fauna (T Gillespie, Ed). *Global Ecology and Biogeography*, **22**, 29–39.

Fontaine B, van Achterberg K, Alonso-Zarazaga MA *et al.* (2012) New species in the Old World: Europe as a frontier in biodiversity exploration, a test bed for 21st century taxonomy. *PLoS ONE*, **7**, e36881.

Freyhof J, Brooks E (2011) *European Red List of Freshwater Fishes*. Publications Office of the European Union, Luxembourg.

Freyhof J, Lieckfeldt D, Pitra C, Ludwig A (2005) Molecules and morphology: evidence for introgression of mitochondrial DNA in Dalmatian cyprinids. *Molecular Phylogenetics and Evolution*, **37**, 347–354.

Froese R, Pauly D (2013) FishBase. World Wide Web electronic publication.

Frost DR (2013) *Amphibian Species of the World: An Online Reference. Version 5.6*. American Museum of Natural History, New York, NY, USA.

Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology & Evolution*, **27**, 1–9.

Geldiay R, Balık S (2007) *Türkiye Tatlısu Balıkları (Freshwater Fishes in Turkey)*. Ege Universitesi Su Ürünleri Fakültesi Yayınları, Ege Üniversitesi Basımevi, İzmir.

Hebert PDN, Gregory T (2005) The promise of DNA barcoding for taxonomy. *Systematic Biology*, **54**, 852–859.

Hebert PDN, Ratnasingham S, DeWaard JR (2003) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, S96–S99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–14817.

Helfman G, Collette BB, Facey DE, Bowen BW (2009) *The Diversity of Fishes: Biology, Evolution, and Ecology*. Wiley-Blackwell, Oxford.

Hendrich L, Pons J, Ribera I, Balke M (2010) Mitochondrial cox1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. *PLoS ONE*, **5**, e14448.

Huang D (2012) Threatened reef corals of the world. *PLoS ONE*, **7**, e34459.

Isaac NJB, Turvey ST, Collen B, Waterman C, Baillie JEM (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE*, **2**, e296.

Jörger KM, Norenburg JL, Wilson NG, Schrödl M (2012) Barcoding against a paradox? Combined molecular species delineations reveal multiple cryptic lineages in elusive meiofaunal sea slugs. *BMC Evolutionary Biology*, **12**, 245.

Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability* 27–43.

Kottelat M (1997) European freshwater fishes. An heuristic checklist of the freshwater fishes of Europe (exclusive of former USSR), with an introduction for non-systematists and comments on nomenclature and conservation. *Biologia Bratislava*, **52**, 1–271.

Kottelat M, Freyhof J (2007) *Handbook of European freshwater fishes*. Publications Kottelat, Cornol, Switzerland.

Lahaye R, van der Bank M, Bogarin D *et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2923–2928.

Lakra WS, Verma MS, Goswami M *et al.* (2011) DNA barcoding Indian marine fishes. *Molecular Ecology Resources*, **11**, 60–71.

Lévêque C, Daget J (1984) Cyprinidae. In: *Check-list of the freshwater fishes of Africa (CLOFFA)*(eds Daget J, Gosse A, van den Thys D), pp. 217–342. ORSTOM and MRAC, Paris and Tervuren.

Lévêque C, Oberdorff T, Paugy D, Stiassny M, Tedesco P (2007) Global diversity of fish (Pisces) in freshwater. *Hydrobiologia*, **595**, 545–567.

Lohse K (2009) Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). *Systematic Biology*, **58**, 439–442; discussion 442–444.

Lou M, Golding GB (2012) The effect of sampling from subdivided populations on species identification with DNA barcodes using a Bayesian statistical approach. *Molecular Phylogenetics and Evolution*, **65**, 765–773.

Maitland PS (1977) *The Hamlyn Guide to Freshwater Fishes of Britain and Europe*. Hamlyn, London, New York, NY, Sydney NSW, and Toronto, ON.

Martens K (2009) The international year of biodiversity. *Hydrobiologia*, **637**, 1–2.

May RM (2011) Why worry about how many species and their loss? *PLoS Biology*, **9**, e1001130.

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology*, **57**, 809–813.

Meyer C, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, 2229–2238.

Mittermeier R, Myers N, Mittermeier C, Robles Gil P (1999) *Hotspots: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions* (eds Mittermeier R, Myers N, Mittermeier C, Robles Gil P). CEMEX, SA, Agrupación Sierra Madre, SC, Mexico City.

Mittermeier RAA, Gil PR, Hoffman M *et al.* (2004) *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Ecoregions* (ed. CEMEX). CEMEX, Mexico.

Monaghan MT, Wild R, Elliot M *et al.* (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology*, **58**, 298–311.

Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How many species are there on earth and in the ocean? *PLoS Biology*, **9**, e1001127.

Myers N (1988) Threatened biotas: "hot spots" in tropical forests. *The Environmentalist*, **8**, 187–208.

Myers N, Mittermeier R, Mittermeier C, da Fonseca G, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Padial J, Miralles A, la Riva ID, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 1–14.

Papadopoulou A, Bergsten J, Fujisawa T *et al.* (2008) Speciation and DNA barcodes: testing the effects of dispersal on the formation of discrete sequence clusters. *Philosophical Transactions of the Royal Society of London. Series B*, **363**, 2987–2996.

Pereira LHG, Maia GMG, Hanner R, Foresti F, Oliveira C (2011) DNA barcodes discriminate freshwater fishes from the Paraíba do Sul River Basin, São Paulo, Brazil. *Mitochondrial DNA*, **22**(Suppl 1), 71–79.

Pereira LH, Hanner R, Foresti F, Oliveira C (2013) Can DNA barcoding accurately discriminate megadiverse Neotropical freshwater fish fauna? *BMC Genetics*, **14**, 20.

Powell JR (2012) Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. *Methods in Ecology and Evolution*, **3**, 1–11.

Puillandre N, Modica M, Zhang Y *et al.* (2012) Large-scale species delimitation method for hyperdiverse groups. *Molecular Ecology*, **21**, 2671–2691.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*, Vol. 1 (ed. RDC Team). R Foundation for Statistical Computing, Vienna, Austria, pp. 409.
Roberts TR (1975) Geographical distribution of African freshwater fishes. *Zoological Journal of the Linnean Society*, **57**, 249–319.
Rolland J, Cadotte M, Davies J *et al*. (2012) Using phylogenies in conservation: new perspectives. *Biology Letters*, **8**, 692–694.
Safi K, Armour-Marshall K, Baillie J, Isaac N (2013) Global patterns of evolutionary distinct and globally endangered amphibians and mammals. *PLoS ONE*, **8**, e63582.
Scribner K, Page K, Bartron M (2000) Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. *Reviews in Fish Biology and Fisheries*, **10**, 293–323.
Simpson GG (1961) *Principles of Animal Taxonomy*. Columbia University Press, New York, NY and London.
Smith G, Darwall W (2006) *The Status and Distribution of Freshwater Fish Endemic to the Mediterranean Basin*. IUCN, Gland, Switzerland and Cambridge. v + 34 pp.
Srivathsan A, Meier R (2012) On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics*, **28**, 190–194.
Talavera G, Dincă V, Vila R (2013) Factors affecting species delimitations with the GMYC model: insights from a butterfly survey (E Paradis, Ed.). *Methods in Ecology and Evolution*, **4**, 1101–1110.
Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B*, **360**, 1847–1857.
Wiley E (1978) The evolutionary species concept reconsidered. *Systematic Biology*, **27**, 17–26.
Wilson DE, Reeder DM (2005) *Mammal Species of the World: A Taxonomic and Geographic Reference* (eds Wilson DE, Reeder DM). Johns Hopkins University Press, Baltimore.
Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.
Young MK, McKelvey KS, Pilgrim KL, Schwartz MK (2013) DNA barcoding at riverscape scales: assessing biodiversity among fishes of the genus Cottus (Teleostei) in northern Rocky Mountain streams. *Molecular Ecology Resources*, **13**, 583–595.
Yule GU (1924) A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S.. *Philosophical Transactions of the Royal Society of London. Series B*, **213**, 21–87.
Zardoya R, Doadrio I (1999) Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *Journal of Molecular Evolution*, **49**, 227–237.

M.F.G., F.H., M.T.M. and J.F. designed and performed the research. M.F.G., J.F., V.A., R.B., M.B., P.B., J.B., M.C.L., G.B.D., G.P.J.D., I.D., E.K., H.K., M.K., M.K., M.La., M.Lo, Z.M., M.Ö., A.P., S.P., H.P., S.P., C.P., J.R., R.S., M.S., V.S., M.S., and S.W. collected, provided or determined specimens and helped with logistics in the field. M.F.G. performed all analyses. M.F.G. and A.D. obtained and analysed sequence data. M.F.G., M.T.M., F.H. and J.F. wrote the article. All authors contributed to the final version of the text.

Data Accessibility

Newly generated DNA sequences: GenBank Accession nos. KJ552094-KJ554964; KC354979-KC354984; KC3550 19, KC355024, KC355025; details in Table S1.

BOLD public project-ID: FFMBH.

Externally generated DNA sequences included from GenBank: GQ328793, GQ328795, GQ328797, GQ328799, GQ328801; HQ960585-HQ960589; JQ060447; JQ060456, JQ060457, JQ060459, JQ060460, HM208835, HM208836; FN600159, JQ623947; EU392236-EU392238, JN242615-JN242620, HM180700, JQ060479-JQ060480, HM208839, HM208840; HM208833, JQ060483; HM560257; HM56 3688, HM563691, HM563692, HM563694-HM563698, HM563700, HM563703-HM563707 HM563689, HM563 690, HM563693, HM563699, HM563701, HM563702; EU524627-EU524630; FJ809715-FJ809719, FJ459499-FJ459502; HQ682696-HQ682699; HM560267; HM560268; JQ060529, JQ060532-JQ060534, JQ060536; HM989722, HM446339, HM446340, HM446342, HM446343; HM560301.

BOLD public project-IDs for additional material included: FBPIS; FFMDR; NGLF; details in Appendix S1.

Sequence alignment deposited in Dryad: doi:10.5061/dryad.6fd1n.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 (A) List of species barcoded, including those where we suggest a taxonomic update, and also including to genus level only identified specimens. In parentheses number of individuals, an 'e' marks species for which individuals where used from outside the Mediterranean region. Taxa with potential candidate species or recovered synonyms are marked with an '*'. Alien taxa in the Mediterranean BH are marked with an '$'; (B) list of Mediterranean freshwater fish species for which no material could be acquired; (Ex) indicates that the species is probably extinct; (F) indicates the failure or poor quality of COI PCR amplification; (C) list of Mediterranean freshwater fish species for which COI sequences were obtained from NCBI GenBank with their respective accession numbers; (D) DNA barcoding studies that contributed materials to the study.

Appendix S2 (A) Methodological details for DNA extraction, PCR and molecular data analysis; (B) best-fit and other GMYC models within delta AICc = 0.99 ranked by increasing delta AICc; weights: Akaike weight given to the model in the averaged parameter estimates below; method: single- or multiple-threshold GMYC model.

Fig. S1 Collection of the maximum-likelihood trees obtained from the different data partitions (see Methods), with details for calculation given for each tree.

Fig. S2 (A) For each IUCN category (LC, least concern; NT, near threatened; VU, vulnerable; EN, endangered; CR, critically endangered; number of species in parentheses), the 25–75 per cent quartiles of the EDGE scores are drawn as box, the medians given as horizontal line. Whiskers show largest and smallest data points <1.5 times the box height from the box. Outliers (stars) are values further than three times the box height from the box, values outside the inner fences are shown as circles; (B)

top 10 species listed according to decreasing EDGE score and according to decreasing ED (evolutionary distinctiveness) with their respective IUCN classification. Outliers with respect to IUCN category combined with EDGE score are marked with an (*). Only endemic species without closely related congeneric species outside the Mediterranean are considered.

Table S1 Excel spreadsheet with the name and sampling details for all newly generated individuals included in this study, museum or field accession numbers for specimens, and GenBank accession number for their DNA sequence.

Table S2 (A) List of proposed updated taxonomy and, where applicable, mean K2P distances to former synonyms, and to nearest neighbour species if different; (B) commented lists of individuals identified as hybrids in the field with their respective putative parental species and number of individuals given in parentheses; (C) species or populations with haplotype-sharing indicating the need for more systematic research; (D) divergent lineages indicating introgressions and/or the need for more systematic research.