Assembling and auditing a comprehensive DNA barcode reference library for European marine fishes

L. M. Oliveira[1], T. Knebelsberger[2], M. Landi[1], P. Soares[1], M. J. Raupach[2], F. O. Costa[1*]

[1]Center of Molecular and Environmental Biology (CBMA), University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

[2]Senckenberg am Meer, German Center for Marine Biodiversity Research (DZMB), Suedstrand 44, 26382 Wilhelmshaven, Germany

Running headline: DNA barcode library for European marine fish

*Author to whom correspondence should be addressed:

Telephone: +351 253601527

Facsimile: +351 253678980

Email address: fcosta@bio.uminho.pt

ABSTRACT

A large-scale comprehensive reference library of DNA barcodes for European marine fishes was assembled, allowing the evaluation of taxonomic uncertainties and species genetic diversity, which were otherwise hidden in geographically restricted studies. A total of 4,118 DNA barcodes were assigned to 358 species generating 366 BINs (Barcode Index Number). Initial examination revealed as much as 141 BIN discordances (more than one species in each BIN). After implementing an auditing and 5-grade (A to E) annotation protocol, the number of discordant species BINs was reduced to 44 (13% / grade E), while concordant species BINs amounted to 271 (78% / grades A and B), and 14 other had insufficient data (grade D). Fifteen species displayed comparatively high intraspecific divergences ranging from 2.0% to 18.5% (grade C), which is biologically paramount information to be considered in fish species monitoring and stock assessment. On balance, this compilation contributed to the detection of 59 European fish species in likely need of taxonomic clarification or re-evaluation. The generalized implementation of an auditing and annotation protocol for reference libraries of DNA barcodes is recommended.

Key words

Marine fishes; DNA barcode; reference library; taxonomic reliability grade; Barcode Index Number; hidden diversity

INTRODUCTION

DNA barcoding, especially the partial sequencing of cytochrome *c* oxidase subunit I (COI), has been successfully employed as a molecular tool for the identification and discrimination of fish species in the past (Knebelsberger *et al.*, 2014). Nevertheless, given the increasing number of publications involving DNA barcodes of European marine fish, a global synthesis of these data, including the compilation and annotation of a reference library, is still lacking. Despite the frequently large distance separating samples, previous studies showed the reliability of DNA barcoding for marine fish identification independently of geographic distance (Ward *et al.*, 2008; Zemlak *et al.*, 2009).

Apart from the compilation, the main objective is to analyze the consistency of DNA barcodes obtained by independent research groups. Public databases, namely GenBank and BOLD (Barcode of Life Data System; Ratnasingham & Hebert, 2007), are susceptible to operational errors, including inaccurate taxonomic identification of the original specimens and insufficient quality of the molecular data and metadata (Knebelsberger *et al.*, 2014). Methodological control measures are imperative, including species identification by expert taxonomists and submission of compliant data according to the requirements of the Barcode Data Standards (Walters & Hanner, 2006). Post-barcoding annotation tools for libraries are vital to maintain the quality standards of the compiled data, as for example the assignment of categories of taxonomic reliability of DNA barcodes (Costa *et al.*, 2012). Such approaches combined with automated analysis tools secure the quality of the library and allows the user, either skilled or not, to use it confidently with high reliability. A reference library, in addition to its use as a robust tool for the identification of sequences from unknown organisms

(Costa *et al.*, 2012), is also essential for applications involving authentication of fishery products (Hanner *et al*., 2011), either fresh or processed (Carvalho *et al.*, 2015), and detection of illegal use of protected species for biosecurity (Armstrong & Ball, 2005; Rasmussen & Morrissey, 2008).

In the specific case of European species, such reference library is valuable to assist the identification and management of fish stocks, frequently shared between the member states (Landi *et al.*, 2014), either through the detection of mixed fisheries containing mislabeled species, or through the assessment of regional biodiversity of a given species or by enabling tools for authenticity of fish stocks (Mariani *et al.*, 2015). The objective of this work is to assemble for the first time a large-scale comprehensive reference library of DNA barcodes for European marine fishes, based on all publicly available DNA barcodes, in order to examine and annotate the consistency and reliability of records obtained independently from multiple regions and studies.

MATERIAL AND METHODS

DATA GENERATION AND COMPILATION

A dataset (DS-EUROFISH, doi:dx.doi.org/10.5883/DS-EUROFISH) was created on BOLD, including samples previously generated by the research groups authoring the current manuscript, encompassing samples from the Atlantic, Mediterranean Sea, North Sea and Baltic Sea, as well as sequences obtained from BOLD projects, and GenBank sequences associated with publications, focusing on European marine fishes. The compilation effort followed the previously suggested

quality criteria for COI sequences (Walters & Hanner, 2006). In addition, new COI barcode sequences were obtained from specimens collect on the Portuguese coast and in the Baltic Sea, following published protocols (Costa *et al.*, 2012). The sequences were submitted to GenBank (Accessions KX586190-KX586232) and added to DS-EUROFISH, where the respective metadata can be consulted. The final dataset is summarized in Table I.

DATA ANALYSES AND ANNOTATION

All sequences listed in Table I were aligned using MAFFT version 7 (Katoh & Standley, 2013). Bayesian Inference (BI) was used to create a phylogenetic tree in order to visualize the sequence clustering pattern. The software MrBayes, version 3.2 (Ronquist *et al.*, 2012) was used to produce the BI tree, using the best fit substitution model GTR + G + I, which was determined using IQ-TREE, version 1.3.0 (Nguyen *et al.*, 2014). The analysis was run for 2 million iterations in two parallel runs with 4 chains each, and with tree sampling every 500 iterations (4000 trees sampled). A burn-in of 25% was used, discarding the first 1000 sampled trees.

The Barcode Index Number (BIN) system (Ratnasingham and Hebert, 2013) was used for the assignment of molecular operational taxonomic units (MOTUs). BINs were examined for the whole DS-EUROFISH library using the 'BIN Discordance Report' analysis tool available on BOLD. Average pairwise distances between BINs were estimated using the Kimura 2-parameter (K2P) model (Kimura, 1980), implemented in the "Distance summary" tool in BOLD. This model was selected because of its generalized use in the barcoding literature, therefore facilitating comparison of reported distances between studies.

In order to assess the level of taxonomic reliability in the library, species-specific DNA barcode subsets were ranked from Grade A to E as described before (Costa *et al.*, 2012; Borges *et al.*, 2016). The basis of such rating systems is that taxonomic reliability is greater if barcode sequences from independent researchers cluster unambiguously and consistently for a given species. Following the procedure illustrated in Figure 1, species-specific DNA barcodes were ranked as:

Grade A: External concordance: unambiguous BIN match between specimens of the same morphospecies from independent BOLD projects or published sequences.

Grade B: Internal concordance: species' BIN congruent within one dataset, with at least 3 specimens of the same species examined but no matching sequences found from independent studies.

Grade C: Suboptimal concordance (putative intraspecific genetic structure): at least 3 specimens of the same morphospecies are available within the library and split among more than one nearest neighbouring BIN.

Grade D: Insufficient Data: low number of specimens analysed (1 or 2 individuals) and no matching sequences available in BOLD.

Grade E: Discordant species assignments: sequences for a given species in our data set did not match with the BIN (or BINs) for the same species in BOLD. The specimen may match with a BIN of a different species or was assigned to a separate non-neighbouring BIN.

The auditing procedure here followed, assumes that automated BIN attribution and discordance flagging cannot account for all potential flaws in the DNA barcode pipeline, requiring a detailed inspection and judgment for each individual case. BINs discordances can be attributed fundamentally to 3 sets of reasons: either morphology or

molecular-based evidence do not reflect accurately the species boundaries, or a set of diverse operational failures, inaccuracies or limitations along the DNA barcoding pipeline produce misleading discordances. The latter include, among other, inaccurate morphological identifications, synonyms and misapplied species names, mislabeled specimens, cross-contamination during DNA extraction or amplification procedures, or eventually, failure of the BIN clustering algorithm to discriminate species with very low interspecific distances. The discordant BIN revision step introduced in the auditing and annotation protocol (Fig. 1), provides an opportunity for a personal evaluation by a skilled auditor in order to discard possible operational artefacts. Some artefacts were straightforwardly spotted, as in the case of synonyms or misapplied names, using FishBase (Froese & Pauly, 2015) as a reference for accepted species names. Other types of artefacts, such as contamination or mislabelling were screened out applying the majority rule (Ratnasingham and Hebert, 2013), in the cases where within a BIN with a large majority of congruent DNA barcodes, generated from various independent sources, there was one or few outstanding accompanying sequences from a taxonomically distant species originated from a single source. When BINs discordances could not be confidently ruled out, the grade E was attributed as a precautionary measure, until further evidence can help clarify the nature of the data disagreement.

For annotation purposes the “extra info” field implemented in BOLD was used to inscribe the attributed grade, followed by the auditor’s initials and date. BOLD also allows to complement this procedure with pre-established tags that can be associated with the specimen record (e.g. “contamination” or “misidentification”), or new ones that may be created at the user’s discretion.

RESULTS

A total number of 4,118 DNA barcodes distributed over 358 species of 34 orders were compiled, mined from 18 BOLD projects and 13 publications (Table I), four of them with no project associated on BOLD, only GenBank accessions (Moura *et al.*, 2008; Straube *et al.*, 2010; Serra-Pereira *et al.*, 2011; Ardura *et al.*, 2013). All of the specimens were identified down to the species level and 43 sequences are originally published under this study. The DS-EUROFISH library contains three fish classes with more than three quarters of the species belonging to the class Actinopterygii (bony fishes), followed by the Elasmobranchii class (cartilaginous fish) and the Holocephali with only two species. The distribution of samples follows similar patterns with bony fishes represented by more than 3,000 sequences and the remaining mostly from the Elasmobranchii class.

DNA barcodes were assigned to 366 distinct BINs corresponding to the before-mentioned 358 species in the library. A total of 213 concordant BINs (58%), basically indicating BINs containing records from only one species, were found, whereas 141 (39%) were discordant, displaying at least two different species within a single BIN. Furthermore, 12 BINs that include only one single sequence were also detected.

Subsequent inspection of the BIN composition revealed potential artifacts (i.e. synonyms, misidentifications) that led to an overestimation and unrealistic percentage of discordant BINs. A total of 97 in 141 discordant BINs, more than half of the putative discordant BINs, displayed further concordance following a careful inspection of the entries in the database (see auditing procedure in Fig. 1). This reveals that the discordance was due to either misidentified records or from records with incomplete taxonomy. These cases are characterized by BINs displaying a high level of taxonomic

concordance with a substantial number of records traced to independent research groups and a wide geographic range. One example of such case is the species *Boops boops* (L. 1758) which is validated by 87 records containing entries by different researchers; however, the BIN also contains two entries of *Oblada melanura* (L. 1758), a species that is found in a separate BIN with 21 concordant entries. It is very likely that records of *O. melanura* found in the *B. boops* BIN cluster are caused by misidentification. In addition, cases of incomplete taxonomy were relatively common along BINs. For example the BIN containing *Gadus morhua* L. 1758 contained 22 entries identified only to the class Actinopterygii, resulting in an erroneous classification as discordant. A few cases also included a discordant classification due to the use of synonymous and unaccepted taxonomic names, as in the case of the BIN cluster of *Chelidonichthys lucerna* L. 1758. Subsequently to the inspection of the BINs for artifacts, the number of discordant ones decreased to 44.

Following the ranking system for taxonomic reliability (Costa *et al.*, 2012), a total of 242 species (70% of a total of 344 morphospecies with attributed BIN) can be classified with with the highest level of reliability (Grade A), meaning that each species was allocated consistently with a single BIN providing for the user an unequivocal identification of a given species. Grade B was assigned to 29 species (8%) with concordant BINs but limited to a single study and no matching sequences in BOLD, whereas 15 species (4%) showed suboptimal concordance and were graded C. Their divergence into neighbouring BINs was mostly associated with geographical clustering. Fourteen of the species examined (4%) had a low number of sequences available (<3), and therefore were assigned to grade D. A considerable percentage of species in the reference library – 13% (44 species) – showed taxonomic ambiguity. This includes also economically important species, which were allocated into BINs containing several

species but from the same genus. Table II lists the species, or groups of species, which were attributed grade E, together with an annotation about the reasons for discordance and possible justification.

The 15 species graded C showed distances between BINs higher than 2% and reaching 18.5% in one case. These 15 species were assigned to a total of 36 BINs, from 2 to 4 BINs per species. Results are displayed in Table III. In most cases the records of a species were assigned in two different BINs, and the specimens were sorted among BINs according to their geographical origin. The most common geographic splits were obtained between the Atlantic and the East Mediterranean (4 species), between the Atlantic and the North Sea (2 species) and between the west and east Mediterranean (2 species). Examples of intraspecific structure and geographically sorted monophyletic clusters in three of the C-graded fish species can be visualized in a section of the BI phylogenetic tree displayed in Fig. 2 (full tree available in Fig. S1). Three species within the Atlantic Ocean were divided into 2 BINs, independently of no geographical separation. The remaining species contained two or more BINs where specimens were not geographically sorted. Further investigation on the status of these species as a unit is warranted.

DISCUSSION

The relevance of the implementation of a post-barcode auditing and annotation procedure to the European fish reference library was illustrated in the present paper by the significant reduction of discordant BINs reported after individual inspection and judgment (from 141 to 44). In addition to the examples of BIN discordance artifacts provided in the methods and results, there were examples of the occasional inadequacy

of the BIN clustering algorithm to discriminate species with very low interspecific distances. Such is the case of the genus *Trachurus*, namely *Trachurus mediterraneus* (Steindachner 1868), *Trachurus picturatus* (Bowdich 1825), and *Trachurus trachurus* (L. 1758), three well-established species, each one holding its exclusive set of DNA barcode haplotypes and forming neighboring monophyletic clusters. Yet, species which were finally attributed with grade E, still represent a fair proportion of the total (13%). Although there will be cases of species which cannot be resolved with DNA barcodes, as for example the shad species *Alosa alosa* and *Alosa fallax* due to mtDNA introgression (Alexandrino *et al.*, 2006; Faria *et al.*, 2012), the status of other grade E species may be eventually clarified as additional data become available and detailed studies are performed (e.g. gobies, Knebelsberger & Thiel, 2014).

The library compilation and auditing procedure here followed was also crucial in the detection of some species exhibiting comparatively high levels of intraspecific genetic distances (grade C species). Extensive data on COI barcode variation in thousands of fish species shows that the vast majority of well-established species have average intraspecific COI distances below 2% (Ward *et al.*, 2009; Ward, 2012). The 15 cases listed in Table III, therefore require additional investigation and verification of their species status, ideally entailing a morphological and multi loci revision of specimens from populations across the distribution range. Independently of the conclusions of such revisions, the occurrence of highly divergent and geographically segregated intraspecific mitochondrial lineages is a strong indication of population isolation that should be considered for stock management and conservation purposes. An annotated DNA barcode library can be of great utility to help mapping such lineages in greater detail, and to provide a basis for lineage (or eventually stock) identification in

fisheries landings and, consequently, improving lineage or stock-specific catch statistics.

Overall, the annotation of the reference library of European marine fish produced a clear majority of species with a high level of data congruence and taxonomic reliability (70% and 8% A and B grades, respectively), meaning that DNA barcode-based identifications of those species are very robust. Furthermore, attribution of grade C to a species does not preclude its robust DNA barcode-based identification, but, on the contrary, may enable gathering more detailed geographic or stock-specific data on that species. As new DNA barcode data are generated and made available for more species and populations, additional auditing and annotation must be carried out regularly. Through such regular reviewing, grades may be changed and, by means of an iterative process, the expected trend is that species move progressively to upper grades due to the continuous refinement of the data and the auditing process: grade D and E species will tend to be re-assigned to upper grades, and grade B species will be re-assessed in light of new data from independent sources confirming or refuting initial congruence. Grade A species are also subject to re-assignment, but much less likely to change.

A global appraisal of the completeness of the reference library for European marine fish, reveals that the available COI barcode data only covers a small fraction of the reported ichthyofauna, notably only about 28% of the species reported for the Portuguese EEZ and extended continental platform area (Carneiro *et al.*, 2014), or even a lower proportion (26.5 %) considering all ichthyofauna listed for Europe in the European Register of Marine Species (Costello *et al*., 2006). Hence, substantial research commitment is still required to complete the reference library for European marine fish, although the existing core library already covers the majority of the most abundant and

289 commercially relevant species. The availability of a comprehensive reference library,
290 dully audited and annotated, for European ichthyofauna provides a crucial framework
291 for a DNA-based identification system of fish species, with far-reaching applications
292 and benefits for fish biology, ecology, fisheries and fisheries products quality control
293 Costa & Carvalho, 2007). The emergence of second and third generation sequencing
294 technologies further expanded the potential of DNA-based identification systems,
295 particularly by enabling species identification from community or environmental
296 samples, rather than from individual specimens sequentially (Bohmann *et al*, 2014;
297 Creer *et al*., 2016). Supported by this technology, ecosystem-based approaches to
298 ichthyofaunal ecology and fisheries can be applied which incorporate analysis of
299 different trophic levels and biotic interactions. Among other applications, it can be used
300 for high-throughput species identification in ichthyoplankton surveys (Bucklin *et al*.,
301 2016), gut content analyses and trophic web research (Leray *et al.*, 2013; Leray *et al.*,
302 2015), facilitation of species identification in processed food, commercial markets and
303 food industry (Shokralla *et al*., 2016) as well as for non-invasive monitoring of fish
304 species in environmental DNA (eDNA) obtained from seawater (Bohmann *et al*., 2014;
305 Thomsen & Willerslev, 2015).

306 This study clearly demonstrates that only by integrating data from multiple
307 sources it is possible to unravel pertinent cases of taxonomic uncertainties and hidden
308 species diversity that otherwise would have remain unnoticed. The cases of deep within-
309 species divergences detected constitute biologically meaningful information that should
310 be considered in fish species monitoring and stock assessment. The geographically
311 focused assembly and auditing of DNA barcodes is therefore essential to assure the
312 robustness and consistency of the reference libraries To this end, it is strongly
313 recommended that an auditing an annotation framework, such as the one here applied, is

adopted by the research community to fully substantiate the potential of the reference libraries, and to improve their accuracy and utility to the various end-users.

The authors would like to thank the three independent reviewers for their helpful reviews of this manuscript. This work was supported by national funds through the FCT I.P. and by the ERDF through the COMPETE2020 – Program Operacional Competitividade e Internacionalização (POCI) in the scope of the strategic programme UID/BIA/04050/2013 (POCI-01-0145-FEDER-007569). PS is supported by FCT, ESF, POPH and the FCT Investigator Programme (IF/01641/2013). ML was supported by an FCT fellowship with reference SFRH/BPD/45246/2008. We thank R. Martins and M. Carneiro (IPMA, Portugal), as well as A. Kijewska and R. Wenne (Institute of Oceanology, Polish Academy of Sciences, Poland) for providing specimens from Portugal and from the Baltic Sea, respectively, used to generate the new sequences here reported.

Supporting Information

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

Fig. S1. Bayesian inference (BI) tree constructed using COI barcode sequence data from 4118 sequences assigned to 358 marine fish species. A best-fit substitution model (GTR+I+G) was applied.

337

338

339 **References**

340 Alexandrino, P., Faria, R., Linhares, D., Castro, F., Le Corre, M., Sabatié, R.,
341 Baglinière, J. L. & Weiss, S. (2006). Interspecific differentiation and intraspecific
342 substructure in two closely related clupeids with extensive hybridization, *Alosa*
343 *alosa* and *Alosa fallax*. *Journal of Fish Biology* **69**, 242-259. doi: 10.1111/j.1095-
344 8649.2006.01289.x.

345 Ardura, A., Planes, S. & Garcia-Vazquez, E. (2013). Applications of DNA barcoding to
346 fish landings: authentication and diversity assessment. *ZooKeys* **365**, 49-65. doi:
347 10.3897/zookeys.365.6409.

348 Armstrong, K. F. & Ball, S. L. (2005). DNA barcodes for biosecurity: invasive species
349 identification. *Philosophical Transactions of the Royal Society B: Biological*
350 *Sciences* **360**, 1813-1823. doi: 10.1098/rstb.2005.1713.

351 Bohmann, K., Evans, A., Gilbert, M. T., Carvalho, G. R., Creer, S., Knapp, M., Yu,
352 D.W. & De Bruyn, M. (2014). Environmental DNA for wildlife biology and
353 biodiversity monitoring. *Trends in Ecology and Evolution* **29**, 358-367. doi:
354 10.1016/j.tree.2014.04.003.

355 Borges, L. M. S., Hollatz, C., Lobo, J., Cunha, A. M., Vilela, A. P., Calado, G., Coelho,
356 R., Costa, A. C., Ferreira, M. S. G., Costa, M. H. & Costa, F. O. (2016). With a
357 little help from DNA barcoding: investigating the diversity of Gastropoda from
358 the Portuguese coast. *Scientific Reports* **6**, 20226. doi: 10.1038/srep20226.

359 Bucklin A., Lindeque P. K., Rodriguez-Ezpeleta N., Albaina, A. & Lehtiniemi, M.
360 (2016). Metabarcoding of marine zooplankton: prospects, progress and pitfalls.
361 *Journal of Plankton Research* **38**, 393-400. doi:10.1093/plankt/fbw023.
362 Carneiro, M., Martins, R., Landi, M. & Costa, F. O. (2014). Updated checklist of
363 marine fishes (Chordata: Craniata) from Portugal and the proposed extension of
364 the Portuguese continental shelf. *European Journal of Taxonomy* **73**, 1-73. doi:
365 10.5852/ejt.2014.73.
366 Carvalho, D. C., Palhares, R. M., Drummond, M. G. & Frigo, T. B. (2015). DNA
367 Barcoding identification of commercialized seafood in South Brazil: A
368 governmental regulatory forensic program. *Food Control* **50**, 784-788. doi:
369 10.1016/j.foodcont.2014.10.025.
370 Costa, F. O. & Carvalho, G. R. (2007). The Barcode of Life Initiative: synopsis and
371 prospective societal impacts of DNA barcoding of Fish. *Genomics, Society and*
372 *Policy* **3**, 29-40. doi: 10.1186/1746-5354-3-2-29.
373 Costa, F. O., Landi, M., Martins, R., Costa, M. H., Costa, M. E., Carneiro, M., Alves,
374 M. J., Steinke, D. & Carvalho, G. R. (2012). A ranking system for reference
375 libraries of DNA barcodes: application to marine fish species from Portugal. *PLoS*
376 *ONE* **7**, e35858. doi: 10.1371/journal.pone.0035858.
377 Costello, M. J., Bouchet, P., Emblow, C. S. & Legakis, A. (2006). European marine
378 biodiversity inventory and taxonomic resources: state of the art and gaps in
379 knowledge. *Marine Ecology Progress Series* **316**, 257-268. doi:
380 10.3354/meps316257.
381 Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., Potter, C. &
382 Bik, Holly M. (2016). The ecologist's field guide to sequence-based identification

of biodiversity. *Methods in Ecology and Evolution*, doi: 10.1111/2041-210X.12574.

Faria, R., Weiss, S. & Alexandrino, P. (2012). Comparative phylogeography and demographic history of European shads (*Alosa alosa* and *A. fallax*) inferred from mitochondrial DNA. *BMC Evolutionary Biology* **12**, 194. doi: 10.1186/1471-2148-12-194.

Hanner, R., Becker, S., Ivanova, N. V. & Steinke, D. (2011). FISH-BOL and seafood identification: geographically dispersed case studies reveal systemic market substitution across Canada. *Mitochondrial DNA* **22**, 106-122. doi: 10.3109/19401736.2011.588217.

Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software Version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780. doi: 10.1093/molbev/mst010.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.

Knebelsberger, T. & Thiel, R. (2014). Identification of gobies (Teleostei: Perciformes: Gobiidae) from the North and Baltic Seas combining morphological analysis and DNA barcoding. *Zoological Journal of the Linnean Society* **172**, 831-845. doi: 10.1111/zoj.12189.

Knebelsberger, T., Landi, M., Neumann, H., Kloppmann, M., Sell, A. F., Campbell, P. D., Laakmann, S., Raupach, M. J., Carvalho, G. R. & Costa, F. O. (2014). A reliable DNA barcode reference library for the identification of the North European shelf fish fauna. *Molecular Ecology Resources* **14**, 1060-1071. doi: 10.1111/1755-0998.12238.

408 Landi, M., Dimech, M., Arculeo, M., Biondo, G., Martins, R., Carneiro, M., Carvalho,
409 G. R., Brutto, S. L. & Costa, F. O. (2014). DNA barcoding for species
410 assignment: the case of Mediterranean marine Fishes. *PLoS ONE* **9**, e106135. doi:
411 10.1371/journal.pone.0106135.
412 Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J.
413 T. & Machida, R. J. (2013). A new versatile primer set targeting a short fragment
414 of the mitochondrial COI region for metabarcoding metazoan diversity:
415 application for characterizing coral reef fish gut contents. *Frontiers in Zoology* **10**,
416 34. doi: 10.1186/1742-9994-10-34.
417 Leray, M., Meyer, C. P. & Mills, S. C. (2015). Metabarcoding dietary analysis of coral
418 dwelling predatory fish demonstrates the minor contribution of coral mutualists to
419 their highly partitioned, generalist diet. *PeerJ* **3**, e1047. doi: 10.7717/peerj.1047.
420 Mariani, S., Griffiths, A. M., Velasco, A., Kappel, K., Jérôme, M., Perez-Martin, R. I.,
421 Schröder, U., Verrez-Bagnis, V., Silva, H., Vandamme, S. G., Boufana, B.,
422 Mendes, R., Shorten, M., Smith, C., Hankard, E., Hook, S. A., Weymer, A. S.,
423 Gunning, D. & Sotelo, C. G. (2015). Low mislabeling rates indicate marked
424 improvements in European seafood market operations. *Frontiers in Ecology and*
425 *the Environment* **13**, 536-540. doi: 10.1890/150119.
426 Moura, T., Silva, M. C., Figueiredo, I., Neves, A., Muñoz, P. D., Coelho, M. M. &
427 Gordo, L. S. (2008). Molecular barcoding of north-east Atlantic deep-water
428 sharks: species identification and application to fisheries management and
429 conservation. *Marine and Freshwater Research* **59**, 214-223. doi:
430 10.1071/MF07192.
431 Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. (2014). IQ-TREE: A
432 fast and effective stochastic algorithm for estimating maximum likelihood

433 phylogenies. *Molecular Biology and Evolution* **32**, 268-274. doi:
434 10.1093/molbev/msu300.
435 Rasmussen, R. S. & Morrissey, M. T. (2008). DNA-based methods for the identification
436 of commercial fish and seafood species. *Comprehensive Reviews in Food Science*
437 *and Food Safety* **7**, 280-295. doi: 10.1111/j.1541-4337.2008.00046.x.
438 Ratnasingham, S. & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System
439 (www.barcodinglife.org). *Molecular Ecology Notes* **7**, 355-364. doi:
440 10.1111/j.1471-8286.2007.01678.x.
441 Ratnasingham, S. & Hebert, P. D. N. (2013). A DNA-based registry for all animal
442 species: the Barcode Index Number (BIN) system. *PLoS ONE* **8**, e66213. doi:
443 10.1371/journal.pone.0066213.
444 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S.,
445 Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. (2012). MrBayes 3.2:
446 efficient Bayesian phylogenetic inference and model choice across a large model
447 space. *Systematic Biology* **61**, 539-542. doi: 10.1093/sysbio/sys029.
448 Serra-Pereira, B., Moura, T., Griffiths, A. M., Serrano Gordo, L. & Figueiredo, I.
449 (2011). Molecular barcoding of skates (Chondrichthyes: Rajidae) from the
450 southern Northeast Atlantic. *Zoologica Scripta* **40**, 76-84. doi: 10.1111/j.1463-
451 6409.2010.00461.x.
452 Shokralla, S., Hellberg R. S., Handy, S.M., King, I. & Hajibabaei, M. (2015). A DNA
453 mini-barcoding system for authentication of processed fish products. *Scientific*
454 *Reports* **5**, 15894. doi:10.1038/srep15894.
455 Straube, N., Iglésias, S. P., Sellos, D. Y., Kriwet, J. & Schliewen, U. K. (2010).
456 Molecular phylogeny and node time estimation of bioluminescent lantern sharks

(Elasmobranchii: Etmopteridae). *Molecular Phylogenetics and Evolution* **56**, 905-917. doi: 10.1016/j.ympev.2010.04.042.

Thomsen P. F. & Willerslev E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183**, 4-18. doi: 10.1016/j.biocon.2014.11.019.

Walters, C. & Hanner, R. (2006). Platforms for DNA banking. In *DNA Banks – Providing Novel Options for Gene Banks? Topical Reviews in Agricultural Biodiversity* (Vicente, M. C. & Andersson, M. S., eds.), pp. 25–35. Rome: International Plant Genetic Resources Institute.

Ward, R. D. (2012). FISH-BOL, A case study for DNA Barcodes. In *DNA Barcodes: Methods and Protocols* (Kress, J. W. & Erickson, L. D., eds.), pp. 423-439. Totowa, NJ: Humana Press. doi: 10.1007/978-1-61779-591-6_21.

Ward, R. D., Costa, F. O., Holmes, B. H. & Steinke, D. (2008). DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa, but *Zeus faber* and *Lepidopus caudatus* each probably constitute two species. *Aquatic Biology* **3**, 71-78. doi: 10.3354/ab00068.

Ward, R. D., Hanner, R. & Hebert, P. D. N. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology* **74**, 329-356. doi: 10.1111/j.1095-8649.2008.02080.x.

Zemlak, T. S., Ward, R. D., Connell, A. D., Holmes, B. H. & Hebert, P. D. N. (2009). DNA barcoding reveals overlooked marine fishes. *Molecular Ecology Resources* **9**, 237-242. doi: 10.1111/j.1755-0998.2009.02649.x.

Electronic References

482 Froese, R. & Pauly, D. (2015). FishBase. World Wide Web electronic publication
483 (www.fishbase.org, version 10/2015) (last accessed 24 June 2016).
484