

Gap-analysis of DNA barcoding in ERMS reference libraries for ascidians and cnidarians

Guy Paz (✉ guy@ocean.org.il)

Israel Oceanographic and Limnological Research Institute <https://orcid.org/0000-0003-2263-9019>

Baruch Rinkevich

Israel Oceanographic and Limnological Research Institute

Research

Keywords: Biological monitoring, Ascidians, Cnidarians, BoLD, Quality control, ERMS

Posted Date: October 30th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-50653/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on January 9th, 2021. See the published version at <https://doi.org/10.1186/s12302-020-00449-9>.

Abstract

Background

All-inclusive DNA barcoding libraries in the storage and analysis platform of the BOLD (Barcode of Life Data) system are essential for the study of the marine biodiversity and are pertinent for regulatory purposes, including ecosystem monitoring and assessment, such as in the context of the EU Water Framework Directive (WFD) and the Marine Strategy Framework Directive (MSFD). Here we investigate knowledge gaps in the lists of DNA barcoded organisms within two inventories, Cnidaria (Anthozoa and Hydrozoa) and Ascidiacea from the reference libraries of the European Register of Marine Species (ERMS) dataset (402 ascidians and 1200 cnidarian species). ERMS records were checked species by species, against publicly available sequence information and other data stored in BOLD system. As the available COI barcode data adequately covers just a small fraction of the ERMS reference library, it is of importance to employ quality control on existing data, to close the knowledge gaps and purge errors off BOLD.

Results

Results revealed that just 22.9% and 29.2% of the listed ascidians and cnidarians species, respectively, are BOLD barcodes of which, 58.4% and 52.3% of the seemingly barcoded species, respectively, were noted to have complete BOLD pages. Thus, only 11.44% of the tunicate and 17.07% of the cnidarian data in the ERMS lists are of high quality. Deep analyses revealed seven common types of gaps in the list of the barcoded species in addition to a wide range of discrepancies and misidentifications, discordances and errors primarily in the GenBank mined data as with the BINs assignments, among others.

Conclusions

Gap knowledge in barcoding of important taxonomic marine groups exist and in addition, quality management elements (quality assurance and quality control) were not employed when using the list for national monitoring projects, for regulatory compliance purposes and other purposes. Even though BOLD is the most trustable DNA barcoding reference library, worldwide projects of DNA barcoding are needed to close these gaps of mistakes, verifications, missing data and unreliable sequencing labs. Tight quality control and quality assurance is important to close the knowledge gaps of Barcoding of the European recommended ERMS reference library.

Background

The accurate evaluation of biodiversity for any given ecosystem is a keystone element, even imperative, in numerous biological and applied disciplines, including ecology, conservation biology, food regulatory compliance, forensics, and ecosystem monitoring and assessment [1-2]. In response to these needs, DNA-based taxon identification relying on the genetic barcode markers, mainly and above all, Cytochrome c Oxidase I (COI) gene, are commonly used to assess biodiversity, including species identification, boundaries and diversity analyses. Except for the COI, other mitochondrial genes are in used for animal barcoding such as *Cytb*, *12S* or *18S*, and while COI marker is commonly used for most animals, several other markers are in usage for different taxa, such as the RuBisCO (Ribulose-1,5-bisphosphate carboxylase/oxygenase) is used for plants, internal transcribed spacer (ITS) rRNA often used for fungi, 16S rRNA gene is widely used in identification of prokaryotes, the 18S rRNA gene is mostly used for detecting microbial eukaryotes and other markers as well, can be found under “primers list” in BOLD where a few thousands of primers are available for the different markers for the species identification. BOLD [3], a cloud-based data storage and analysis platform that can be further employed as a curation tool, currently contains (updated in April 2020) about eight million barcodes, encompassing >310,000 animal, plant and fungal species. The rationale for using the COI marker gene in species barcoding relies on the fact that intraspecific diversity for this gene is usually lower than interspecific diversity and thus is more effective in species identification, and along with the difficulties associated with the traditional morphological taxonomy [4]. The major benefit of using BOLD is immediately emerged when an unknown sequence is compared against a database to determine its closest species match, an evaluation that strictly depends on the correctness and reliability of the data stored BOLD and on the quality of the barcode libraries [5]. Yet, it should be noted that while barcodes may be excellent tools to identify species that are already in BOLD, they may have poor predictive power in identification of unknown species, however, it should be noted that they do have good predictive power if the species has close relatives in BOLD already.

As a curator tool, it is inferred that all barcode sequences stored in the BOLD database are backed by vouchered specimens and thoroughly identified by taxonomy experts. Yet, being a public database, it is inevitable that BOLD, as any other similar curation tool, might accrue erroneous data, sometimes significantly [2,6]. Taxonomic misidentifications and/or taxonomic conundrums, cryptic species complexes, delimiting cryptic species, technical faults, such as deficient DNA extraction, PCR-based errors and foreign DNA contaminations, including bacterial sequences, especially COI sequences, are just some of the causes that may unavoidably generate erroneous data and inaccurate sequences [2,6-10]. The above difficulties may affect dramatically the accuracy of barcoding. For example, the Barcode Index Number (BIN) is used as a system persistent registry for animal OTUs and is recognized through sequence variation in the COI DNA barcode region, and aid for the taxonomy of species by flagging possible cases of synonyms for specimens that are likely to belong to the same species, however, it has been claimed that it can lead to the lack of an unambiguous species level identification in the BOLD system, and to taxonomic conflicts by the assignment of more than a single species name per BIN [11].

The European Register of Marine Species (ERMS [12]) is an authoritative taxonomic check-list of species that are found in all European marine environments (the all-taxon marine species inventory from the Canaries and Azores to Greenland and north west Russia, towards the Mediterranean

sea and the Baltic Sea), from the deep sea, all continental shelf areas and up to the splash zone above the high tide mark, and in estuaries, down to 0.5 psu salinity. During 1997–1999, ERMS was published on the internet and subsequently as a book, containing a list of about 30,000 marine species of the kingdoms Animalia, Plantae, Fungi and Protoctista, occurring in the European marine environment [13]. It is projected that this marine species inventory will be used as the standard reference and technological tool for marine research and for management of the marine environment in Europe.

Until recently, the standardized methodologies available for biological monitoring and management in the marine environments, primarily for practitioners, were restricted to traditional morphological taxonomy, tedious and time-consuming methodologies that require the involvement of expert taxonomists with skills that can only be attained via years of practice. This line of analyses is currently being complemented and may be even replaced in the future by molecular approaches such as DNA barcoding and metabarcoding of bulk or environmental DNA (eDNA) [5, 14-17]. The success of these approaches is strictly dependent of complete and reliable DNA barcode reference libraries. Thus, it is of special interest to identify gaps in the current existing or developing DNA barcode reference libraries, primarily those that are pertinent in the context of the EU Water Framework Directive (WFD) and the Marine Strategy Framework Directive (MSFD). A recent global study on this perspective [5] has revealed that the barcoding coverage varies strongly among taxonomic groups, and among geographic regions, pointing to many missing species and unreliable data (e.g., errors in species identification, discordance among taxonomists) that are relevant to monitoring and highlighted the needs for improving quality assurance of the barcode reference libraries.

Following Weigand et al. [5] global analysis, we aim here to investigate potential gaps in already DNA barcoded organisms (based on publicly available data in BOLD database) listed in the ascidians and cnidarian (Anthozoa and Hydrozoa)- reference libraries of the ERMS inventory. We discuss the necessity of quality control (QC) when building and curating a barcode reference library, and provide recommendations for filling the gaps in the barcode library of European aquatic taxa.

Methods

Each BOLD page consists of six sections: 1. A short taxon description with a link to a species-specific website; 2. Statistics data, including: the number of records, specimens with sequences, specimens with barcodes, subspecies, subspecies with barcodes, public records, public available subspecies and public BIN clustering; 3. Worldwide specimen repositories. 5. Collection sites including countries; 4. Origin of sequences (GenBank ID numbers or sequencing laboratories); 6. Species images gallery.

Species checklist of two distant taxa (the Anthozoa and Hydrozoa of the phylum Cnidaria and the class Ascidiacea of the phylum Chordata) were downloaded from the European Register of Marine Species dataset (ERMS [12]). These taxa were not analysed in Weigand et al. [5] and are used here for deep analyses on quality control (QC) of the barcoded species from these two lists. The conformity of taxonomy and assurance of correct spelling were performed using BOLD toolbox and manually, species by species, for each species BOLD pages, against the World Register of Marine Species database (WoRMS [18]) and assessed following the recommendations by Costello et al. [19]. Finalized species-level checklists were re-ordered and compared with the BOLD list. For the analyses on the barcoded species (the COI marker) we used the checklist on BOLD created by Dirk Steinke, titled 'Marine Animals Europe' (BOLD checklist code: CL-MARAE; last updated on March 20th, 2017). The full list contains 27,634 records of marine species belonging to 10 phyla, Annelida, Arthropoda (class Decapoda and superorder Peracarida), Brachiopoda, Chordata (class Ascidiacea - subphylum Tunicata and class Pisces), Cnidaria (classes Anthozoa and Hydrozoa), Echinodermata, Mollusca (classes Bivalvia and Gastropoda), Nemertea, Priapulida, and Sipuncula. Datasets were generated for the two checklists (the cnidarians and the ascidians; updated 18 July 2019) that were compared, species by species, to all publicly available sequence information in BOLD system and to the other data stored in BOLD pages. Working species by species allows the discrimination and the analyses of records in BOLD, the number of sequences/species, BIN numbers/species, specimen in public depositories, the number of barcodes publicly stored in BOLD, including those mined from the GenBank database at the National Center for Biotechnology Information (NCBI [20]) and the number of privately stored barcodes/species in BOLD. Geographical data were not considered for the taxa analysed. The analyses were performed on all the available BOLD records at three levels: the statistics level, the repository level and at the sequence level. At the statistic level, data for each species was individually inspected for the number of BOLD specimen, records that hold validated sequences, i.e. containing trace files which are necessary to qualify the barcode status and to provide quality control for sequences (both reverse and forward directions of the trace files and/or final edited sequences), records with just barcodes, and the numbers of public records and public BINs. At the repository level, we searched for number of records mined from GenBank and the number of repository facilities where voucher specimens were deposited. At the sequence level we investigate the open to the public sequences and scored the sequence trace files quality, and the mark (good, medium or low) which is scored by the BOLD system (Tables 1, Supporting Information Tables 2-3).

Results

The analysis was performed on the BOLD systems (version v4) [3] database, for the 1602 species extracted from the ERMS taxonomic checklists, including 402 ascidians species and 1200 hydrozoans and anthozoans species (Supporting Information: Table 3). Checking against the BOLD database (July 18th, 2019 inventory) we found only 88 (22.9%) of the ascidians species and just 351 (29.2%) cnidarians species in the list of BOLD's barcoded species (Table 1).

Ascidians (Tables 1, Supporting Information Table 1)

Of the 88 ERMS species referenced in the BOLD database, only 50 species (12.4% of the whole ERMS list of ascidians) BOLD had more than five specimens barcoded. COI gene sequences were found for 81 species of this list and just 78 of the species (19.4% of total species) have full descriptions in the BOLD pages, including number of specimen records, sequences, specimens with barcodes, species names, public records, public species and public BINS. As for the COI sequences, we assigned three types of records: records with no sequences, records containing sequences downloaded from GenBank (hence with no trace files, without a reliable curation), and records containing BOLD related new sequences (with trace files). Thus, in 77 species, the COI gene was sequenced and contains trace files, while for 74 species, the COI sequences were mined from GenBank. A total of 68 public BINS are assigned to 88 species and 32 species contained more than a single BIN (Table 1).

Cnidarians (Tables 1, Supporting Information Table 2)

For only 153 species (12.7% of the whole 1200 ERMS list), of the 351 species found in the BOLD database, contained more than five specimens barcoded. COI gene sequences were found for 310 species of this list and just 297 species (6.5% of total species) had full descriptions in the BOLD pages, including number of specimen records, sequences, specimens with barcodes, species names, public records, public species and public BINS. As for the COI sequences, we assigned three types of records- records with no sequences, records containing sequences downloaded from GenBank (hence with no trace files, without a reliable curation), and records containing BOLD related new sequences (with trace files). In 278 species COI was sequenced and contained trace files, however for 205 species, sequences were mined from GenBank. A total of 231 public BINS were found, out of which 65 species contained more than a single BIN (Table 1).

Types of gaps

We assigned seven common types of gaps (Table 2) in the list of barcoded species, such as: (a) records with no data available (empty pages on BOLD website) - 10 tunicate species and 41 cnidarians, (b) records with partial public data and no COI sequences- 10 tunicate species and 41 cnidarians, (c) no public available records - 17 tunicate species and 73 cnidarians, (d) records with sequences mined from the GenBank, many with gaps in sequences and all without trace files - 74 tunicate species and 205 cnidarians, (e) records containing more than a single bin for species- 29 tunicate species and 65 cnidarians, (f) species with no bin - 25 tunicates and 120 cnidarians, (g) species with sequences dispersed within several bins - 29 tunicate species and 65 cnidarians (examples for the seven knowledge gap types are detailed in Table 2). In summary, only 52.3% and 58.4% for cnidarians and tunicates, respectively, of the seemingly 'barcoded species' on the BOLD website have complete BOLD pages, which is about 11.44% of the tunicate species and 17.07% of the cnidarian species of the ERMS lists.

Discussion

In the current study, we analysed two major reference libraries for the availability of barcoding data, for the ascidians and the cnidarians from the ERMS list, which were missed from a previous gap-analysis for the monitoring of aquatic biota in Europe. By working on members of two taxa, the Anthozoa and Hydrozoa of the phylum Cnidaria and the class Ascidiacea of the phylum Chordata, our analyses showed that important reference libraries lack reliable barcodes for these dominant marine macroinvertebrate species. Results further revealed a wide range of difficulties and inconsistencies, including taxonomic congruency of the COI barcode records on the one hand and possible cryptic diversity (*sensu* Leite et al. [28]), that should be further studied, on the other hand. The above clearly affect the wholeness of the ERMS list, since only 52.3% and 58.4% of the cnidarians and tunicates, respectively, of the short list of seemingly well 'barcoded species' were noted to have complete BOLD pages, highlighting that only 11.44% of the tunicate data and 17.07% of the cnidarian data in the ERMS lists are reliable and fully supported (July 2019 status).

Further, two relevant findings emerged. The first indicated that the BINS assignments revealed a sizeable amount of discordances, many of them probably related to species misidentifications or synonyms [28] or the deficiency of the BIN clustering algorithm to correctly discriminate species [30]. The second outcome pointed towards the low control of GenBank compared to BOLD, discrepancies that were already pointed out earlier [2,31], characterized by sequence discordances and misidentifications. The BOLD and the GenBank data storage systems are highly intermingled. About 11% of COI barcode records on BOLD are mined from GenBank, while 75% of the COI barcodes on GenBank originated from BOLD [31]. Yet our results point to many weaknesses associated with the GenBank data that are less informative and do not present extended data elements such as trace electropherograms, specimen images, voucher numbers or BIN assignments, and are usually poorly curated (see also Bridge et al. [26]), compared to BOLD.

Human-made artefacts during the barcoding process, in particular for the GenBank stored data, affect the reliability of DNA barcoding to correctly assign a given specimen to a species. The overall recorded knowledge gaps in the DNA barcodes found in the present study are considerably high. Clearly, the two types of gaps; the first of records not found in BOLD, and the second the errors and missing data in for records already BOLD deposited in BOLD, may impact dramatically the accuracy of any DNA-based assessment or biomonitoring approach relying on BOLD's datasets. Many other common errors target inaccurate taxonomic identifications of specimens by nonexpert taxonomists (amplified by the lack of voucher specimen), sequence contamination, incomplete reference data and insufficient quality of the uploaded molecular data [5,30]. By working on reference libraries of DNA barcodes of marine organisms (invertebrates and fish taxa), Weigand et al. [5] recorded numerous identification errors, sequence contamination, incomplete reference (missing trace files or primer information) as well as inadequate data management. The results of the present study, as supported by other recent studies [2,5,17,28,30,32], reveal that we are still away from holding representative and reliable reference libraries for important taxonomic groups, as those that were analysed in the current study. In addition, new DNA barcode data are continuously made available for

the already barcoded and also for additional species from the reference libraries, including additional auditing and annotation processes, altogether helping in closing the gap knowledge and purging accumulated erroneous data.

Our assessment on the completeness of the two selected taxa from the ERMS library (i.e. Ascidiacea and Cnidaria) elucidates that the available COI barcode data may adequately covers just a small fraction of this reference library, raising an alarm for similar statuses in other reference libraries, using indexes such as the AZTI's Marine Biotic Index (AMBI; [33]).

Global change of marine ecosystems and marine biodiversity loss [21-22] pose great challenges to emerging marine and water strategy directives (such as WFD and MSFD), as well as to the development of marine policy and management approaches [23-25]. In view of these challenges, reliable barcode reference libraries are of particular importance [5], revealing the necessity of quality control (QC) when building and curating a barcode reference library. Our analysis further elucidate the necessity of filling the gaps in the barcode libraries used. It is evident that the level of taxonomic detection and degree of accuracy by newly developed molecular tools is directly contingent on the reference libraries completion and reliability of the DNA records [26-29]. Given our increased reliance on molecular taxonomy as a robust tool [4], strengthening the existing reference libraries is a need for a wide range of scientific and applied purposes, such as monitoring, eDNA and metabarcoding approaches, all targeting matched identifications and for assessments of biodiversity and abundance [5,14-17,29].

Conclusions

DNA barcoding-based approaches are superior in issues like diminished ambiguity and improved accuracy of species identification with ultimate verification of results against repository documentations [14,32]. Quality management elements (such as quality assurance and quality control) should be employed when using the list for monitoring and other purposes and for closing the knowledge gaps. Purging of errors from BOLD, the most reliable DNA barcoding reference library, will significantly contribute to future attempts in biodiversity monitoring efforts, in eDNA and metabarcoding approaches and their assessments, for various regulatory compliance purposes, forensics, among others [1-2,27]. Given the increasing use of high throughput sequencing approaches and of automated pipelines, data quality aspects of DNA barcodes should be cogitated with higher priority.

Abbreviations

BOLD: Barcode of Life Data, COI: Cytochrome c Oxidase I, CQ/QA: Quality control/Quality assurance, eDNA: environmental DNA, ERMS: European Register of Marine Species, MSFD: Marine Strategy Framework Directive, NCBI: National Center for Biotechnology Information National Library of Medicine, WFD: Water Framework Directive, WoRMS: World Register of Marine Species database.

Declarations

Acknowledgements

This study associates with the National Israeli Marine Barcoding project (BIM) at IOLR and is the outcome of activities and discussion in the COST Action CA15219 "Developing new genetic tools for bioassessment of aquatic ecosystems in Europe" (DNAqua-Net).

Authors contributions:

The authors conceived, wrote and approved the manuscript.

Funding

The authors did not receive any funding for this research.

Availability of data and materials:

All data generated or analyzed during this study are included in this published article.

Ethics approval and consent to participate

The authors declare they have no competing financial interests.

Consent for publication

Not applicable.

Competing interests

The authors declare they have no competing interests.

References

1. Baird DJ and Sweeney BW (2011) Applying DNA barcoding in benthology: the state of the science. *J N Am Benthol Soc* 30: 122–124. 10.1899/10-154.1.
2. Meiklejohn KA, Damaso N, Robertson JM (2019) Assessment of BOLD and GenBank- Their accuracy and reliability for the identification of biological materials. *PLoS ONE* 14(6): doi.org/10.1371/journal.pone.0217084.
3. BOLD - The Barcode of Life Data System <http://v4.boldsystems.org>. Accessed 19 July 2020.
4. Blaxter ML (2004) The promise of a DNA taxonomy. *Phil Trans R Soc Lond B Biological Sciences* 359: 669–679. 10.1098/rstb.2003.1447.
5. Weigand H et al. (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Sci Total Environ* 678: 499–524 org/10.1016/j.scitotenv.2019.04.247.
6. Shen YY, Chen X, Murphy RW (2013) Assessing DNA barcoding as a tool for species identification and data quality control. *PLoS ONE* 8(2): e57125 10.1371/journal.pone.0057125.
7. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *App Environ Microbiol* 71: 7724–7736. 10.1128/AEM.71.12.7724-7736.2005.
8. Wells JD and Stevens JR (2008) Application of DNA-based methods in forensic entomology. *Annu Rev Entomol* 53: 103–120 org/10.1146/annurev.ento.52.110405.091423.
9. Sonet G, Jordaens K, Braet Y, Bourguignon L, Dupont E, Backeljau T et al. (2013) Utility of GenBank and the Barcode of Life Data Systems (BOLD) for the identification of forensically important Diptera from Belgium and France. *ZooKeys* 365: 307–328 10.3897/zookeys.365.6027.
10. Mioduchowska M, Czyż MJ, Gołdyn B, Kur J, Sell J (2018) Instances of erroneous DNA barcoding of metazoan invertebrates: Are universal cox1 gene primers too “universal”? *PLoS ONE* 13(6): e0199609 <https://doi.org/10.1371/journal.pone.0199609>.
11. Collins RA and Cruickshank RH (2013) The seven deadly sins of DNA barcoding. *Mol Ecol Resour* 13: 969-975 10.1111/1755-0998.12046.
12. ERMS- European Register of Marine Species. <http://www.marbef.org/data/erms.php>. Accessed 19 July 2020.
13. Costello MJ, Emblow C, White RJ (Ed) (2001) European register of marine species: a check-list of the marine species in Europe and a bibliography of guides to their identification. *Collection Patrimoines Naturels*, 50. Muséum national d'Histoire naturelle: Paris ISBN 2-85653-538-0, 463 pp.
14. Leese F, Altermatt F, Bouchez A et al. (2016) DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Res Ideas Outcomes* 2: e11321 <https://doi.org/10.3897/rio.2.e11321>.
15. Karahan A, Douek J, Paz G, Stern N, Kideys AK, Shaish L, Goren M, Rinkevich B (2017) Employing DNA barcoding as taxonomy and conservation tools for fish species censuses at the southeastern Mediterranean, a hot-spot area for biological invasion. *J Nat Conserv* 36: 1-9 doi.org/10.1016/j.jnc.2017.01.004.
16. Pawlowski, J Kahlert M, Kelly-Quinn M et al. The future of biotic indices in the ecogenomic era: integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. (2018) *Sci Total Environ* 637–638: 1295–1310 <https://doi.org/10.1016/j.scitotenv.2018.05.002>.
17. Paz G, Yudkovsky Y, Shaish L, Stern N, Lubinevski H, Mienis HK, Douek J, Galil B, Goren M, Rinkevich B. Initiating DNA barcoding of Eastern Mediterranean deep-sea biota. (2018) *Mediterr Mar Sci* 19/3: 416-429 <https://doi.org/10.12681/mms.14146>.
18. WoRMS- World Register of Marine Species. marinespecies.org. Accessed 19 July 2020.
19. Costello MJ, Bouchet P, Boxshall G, Fauchald K, Gordon D, Hoeksema BW, Poore GCB, van Soest RWM, Stöhr S, Walter TC, Vanhoorne B, Decock W, Appeltans W (2013) Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. *PLoS ONE* 8(1): e51629 10.1371/journal.pone.0051629.
20. NCBI- National Center for Biotechnology Information U.S. National Library of Medicine. ncbi.nlm.nih.gov/GenBank. Last accessed 19 July 2020.
21. Johnson CN, Balmford A, Brook BW, Buettel JC, Galetti M, Guangchun L, Wilmschurst JM (2017) Biodiversity losses and conservation responses in the Anthropocene Science. 356(6335): 270-275 10.1126/science.aam9317.
22. Rinkevich B (2019) Coral chimerism as an evolutionary rescue mechanism to mitigate global climate change impacts. *Glob Chang Biol* 25: 1198–1206 <https://doi.org/10.1111/gcb.14576>.
23. Creighton C, Hobday AJ, Lockwood M, Pecl GT (2016) Adapting management of marine environments to a changing climate: a checklist to guide reform and assess progress. *Ecosystems* 19: 187-219 10.1007/s10021-015-9925-2.
24. Cormier R, Elliott M, Rice J (2019) Putting on a bow-tie to sort out who does what and why in the complex arena of marine policy and management. *Sci Total Environ* 648: 293-305 doi.org/10.1016/j.scitotenv.2018.08.168.
25. Fine M, Cinar M, Voolstra CR, Safa A, Rinkevich B, Laffoley D, Hilmi N, Allemand D (2019) Coral reefs of the Red Sea- Challenges and potential solutions. *Reg Stud Mar Sci* 25: 100498 doi.org/10.1016/j.rsma.2018.100498.
26. Bridge PD, Roberts PJ, Spooner BM, Panchal G (2003) On the unreliability of published DNA sequences. *New Phytol* 160: 43–48 doi.org/10.1046/j.1469-8137.2003.00861.x.
27. Trebitz A, Sykes M, Barge J (2019) A reference inventory for aquatic fauna of the Laurentian Great Lakes. *J Great Lakes Res* 45: 1036-1046 doi.org/10.1016/j.jglr.2019.10.004.
28. Leite BR, Vieira PE, Teixeira MAL, Lobo-Arteaga J, Hollatz C, Borges LMS, Duarte S, Troncoso JS, Costa FO (2020) Gap-analysis and annotated reference library for supporting macroinvertebrate metabarcoding in Atlantic Iberia. *Reg Stud Mar Sci* 36: 101307 doi.org/10.1016/j.rsma.2020.101307.

29. Stoeckle MY, Das Mishu M, Charlop-Powers Z (2020) Improved Environmental DNA Reference Library Detects Overlooked Marine Fishes in New Jersey United States. *Front Mar Sci* 7: 226 10.3389/fmars.2020.00226.
30. Oliveira LM, Knebelsberger T, Landi M, Soares P, Raupach MJ, Costa FO (2016) Assembling and auditing a comprehensive DNA barcode reference library for European marine fishes. *J Fish Biol* 89: 2741-2754 <https://doi.org/10.1111/jfb.13169>.
31. Pentinsaari M, Ratnasingham S, Miller SE, Hebert PDN (2020) BOLD and GenBank revisited – Do identification errors arise in the lab or in the sequence libraries? *PLoS ONE* 15(4): e0231814 doi.org/10.1371/journal.pone.0231814.
32. Kelly RP, Closek CJ, O'Donnell JL, Kralj JE, Shelton AO, Samhoury JF (2017) Genetic and manual survey methods yield different and complementary views of an ecosystem. *Front Mar Sci* 3: 283 doi.org/10.3389/fmars.2016.00283.
33. Borja A Franco J Pérez V (2000) A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. *Mar Pollut Bull* 40: 1100-1114 [http://doi.org/10.1016/S0025-326X\(00\)00061-8](http://doi.org/10.1016/S0025-326X(00)00061-8).

Table

Table 1. Summary of the Ascidiaceans and Cnidarians species records in BoLD.

Taxonomic group	ERMS		No. of species in BoLD					Sequences origin		No. of Public Bins*	BoLD species with more than 1 Bin
	ERMS species (n)	species in BoLD (n)	(% are calculated to the total number of species in ERMS)					Mined from Genbank	Sequencing labs		
			with sequences	without sequences	with full description	with ≥ 1 barcode	with ≥ 5 barcode				
Ascidiaceans	402	91	81 (20.1%)	88 (21.9%)	78 (19.4%)	88 (21.9%)	78 (19.4%)	74 (18.4%)	77 (19.2%)	68 (16.9%)	32 (8.0%)
Cnidarians	1200	353	310 (25.8%)	351 (29.3%)	297 (24.7%)	351 (29.3%)	297 (24.7%)	205 (17.1%)	278 (23.1%)	231 (19.2%)	65 (5.4%)

*= Bins total number of all 88 species appears in BoLD.

**= Species total number BoLD with more the one Bin in BoLD.

Supplementary Files

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

- [TableS1.docx](#)
- [TableS2.docx](#)
- [TableS3.xlsx](#)