Regional Studies in Marine Science 36 (2020) 101307

Contents lists available at ScienceDirect

Regional Studies in Marine Science

journal homepage: www.elsevier.com/locate/rsma

Gap-analysis and annotated reference library for supporting macroinvertebrate metabarcoding in Atlantic Iberia

B.R. Leite [a,b,c,*], P.E. Vieira [a,b], M.A.L. Teixeira [a,b], J. Lobo-Arteaga [d,e], C. Hollatz [d], L.M.S. Borges [f], S. Duarte [a,b], J.S. Troncoso [c,g], F.O. Costa [a,b,**]

[a] CBMA - Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[b] IB-S - Institute of Science and Innovation for Bio-Sustainability, University of Minho, Portugal
[c] UVIGO Marine Research Centre (CIM-UVIGO), ECIMAT Marine Station, Toralla Island, s/n, Vigo, Spain
[d] IPMA – Instituto Português do Mar e da Atmosfera, Avenida de Brasília 6, 1449-006 Lisboa, Portugal
[e] MARE – Marine and Environmental Sciences Centre, Universidade Nova de Lisboa, Campus de Caparica, 2829-516 Caparica, Portugal
[f] L3 Scientific Solutions, Runder Berg 7e, Geesthacht, Germany
[g] Department of Ecology and Animal Biology, Marine Sciences Faculty, University of Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Spain

ARTICLE INFO

Article history:
Received 3 March 2020
Received in revised form 20 April 2020
Accepted 21 April 2020
Available online 28 April 2020

Keywords:
Macroinvertebrates
DNA barcoding
Reference library
Cytochrome c oxidase subunit I
Atlantic iberia

ABSTRACT

DNA metabarcoding provides a rapid and effective identification tool of macroinvertebrate species. The accuracy of species-level assignment, and consequent taxonomic coverage, relies on comprehensive DNA barcode reference libraries, which, due to incompleteness, are currently a recognized limitation for metabarcoding applications. In this study, we assembled a comprehensive reference library of DNA barcodes for Atlantic Iberia marine macroinvertebrate species, assessed gaps in species coverage and examined data ambiguities. Initially, an Iberian species checklist for the three dominant groups of marine macroinvertebrates was compiled, comprising 2827 species (926 Annelida, 638 Crustacea and 1263 Mollusca). A total of 18162 DNA sequences of the cytochrome c oxidase subunit I barcode region (COI-5P) matching the species checklist were compiled in a BOLD dataset, where taxonomic discordances were evaluated and cases of deep intraspecific divergence flagged. Gap-analysis showed that 63% of the Iberian macroinvertebrate species still lack a DNA barcode. Coverage gaps varied considerably across taxonomic groups with Mollusca displaying the highest sequence representation in the dataset (427 species, 49% of the total number of sequences), and Crustacea the highest species coverage with 338 species barcoded (53% of the checklist). In contrast, Polychaeta displayed the lower levels of completion (288 species, 16% of the total number of sequences). In total, 1545 Barcode Index Numbers (BINs) were assigned to 1053 barcoded species, of which 66% were taxonomically concordant, 26% displayed multiple BINs and 8% were discordant. Overall, results show that there is still a large portion of marine invertebrate taxa in this region of Europe pending barcode coverage, even considering only the dominant groups. However, the most notable finding was the relevant proportion of species flagged for significant intraspecific divergence and possible hidden diversity. The annotated reference library and gap-analysis here provided can therefore contribute to prioritize marine macroinvertebrate taxa for future research efforts and barcode coverage.

1. Introduction

The Atlantic coast of the Iberian Peninsula occupies a central geographic position in the North east Atlantic, spreading along intermediate latitudes (between 43° 47'N and 36° 00'N), connecting the north temperate and warm sub-tropical waters (Spalding et al., 2007; Briggs and Bowen, 2012). This Atlantic region is right at the core of the Lusitanian biogeographic province (Spalding et al., 2007; Briggs and Bowen, 2012) that harbors a diverse marine fauna, enriched by the faunas from the various adjacent regions, such as the Mediterranean, the sub-tropics and Macaronesia, as well as the faunas from further north and western Atlantic. Many invertebrate species have their northern or southern range limits in this area (e.g. Boaventura et al., 2002; Pereira et al., 2006; Lima et al., 2007), which makes it a particularly suitable region to monitor the impact of global change in marine

* Corresponding author at: CBMA - Centre of Molecular and Environmental Biology, Department of Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal.
** Corresponding author. Centre of Molecular and Environmental Biology (CBMA), University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal.
E-mail addresses: barbaradrl.bio@gmail.com (B.R. Leite), fcosta@bio.uminho.pt (F.O. Costa).

https://doi.org/10.1016/j.rsma.2020.101307

species ranges in the NE Atlantic. Furthermore, offshore Atlantic Iberia waters constitute one of the largest routes of the globe for maritime traffic (Nunes et al., 2014; Pejovic et al., 2016), which, together with major commercial ports and numerous recreational marinas in Portugal and Spain, make this region highly susceptible to exposure to non-indigenous marine species (Chainho et al., 2015; Rubal et al., 2018). Monitoring of coastal fauna in the Iberian Atlantic Peninsula is therefore of prime importance for early assessment of impacts and changes in marine communities and ecosystems that may have repercussions in other areas of the NE Atlantic (Araújo et al., 2009; Pascual et al., 2010; Miralles et al., 2016; Múrria et al., 2019).

Due to their rapid and sensitive response to environmental and human pressures, marine macroinvertebrates have been widely used as bioindicators of ecological status in marine ecosystems (Aylagas et al., 2014). Up to now, long term monitoring of coastal ecosystems in the Iberian Atlantic Peninsula has been carried out through morphology-based community assessments (e.g. Guerra-García et al., 2011; Veiga et al., 2016), including monitoring of macroinvertebrate assemblages in the scope of EU's Water Framework Directive (WFD, Directive 2000/60/EC) and Marine Strategy Framework Directive (MSFD, Directive 2008/56/EC) (European Commission, 2000, 2008). Morphology-based diagnosis is particularly challenging in Atlantic Iberia, considering the diversity of faunal assemblages, including species coming from various adjacent regions, and the lack of dedicated taxonomic keys for this region (especially when compared with other marine regions such as Mediterranean — Ruffo, 1982; the British Isles — Lincoln, 1979; Naylor, 1972 or Northwest Europe — Hayward and Ryland, 1995). Moreover, reports displaying growing evidence for the existence of a sizeable proportion of hidden or cryptic diversity among marine invertebrates (Hupało et al., 2019; Teixeira et al., 2019; Vieira et al., 2019), including many taxa that occur in this region (e.g. Borges et al., 2016; Lobo et al., 2016, 2017), further call into question the accuracy of morphology-based assessments.

The recent development of DNA metabarcoding approaches for species assessments (Hajibabaei et al., 2012; Cristescu, 2019) provides an opportunity to complement morphology-based procedures, thereby globally improving the accuracy, throughput and efficiency of marine monitoring, including macroinvertebrate communities (Bourlat et al., 2013; Cowart et al., 2015; Aylagas et al., 2018; Pearman et al., 2018). In addition to constituting the single available method to diagnose cryptic species (Lindeque et al., 2013), strengths of DNA-based approaches include reduced ambiguity and greater accuracy, identification of small taxa, immature or larval stages, and possibility of direct comparison among sites and studies and future verification of the identifications (Leese et al. 2018, 2016). It also enables higher spatial–temporal frequency in monitoring due to higher throughput (Bush et al., 2019). However, the usefulness and efficiency of DNA metabarcoding depends heavily on the extent of the taxonomic coverage and the quality of records available in the reference libraries of DNA barcodes that underpin the method (Siddall et al., 2009; Leray and Knowlton, 2016). The development of several biomonitoring programs associated with high-throughput biodiversity data prompts the necessity to provide quality assurance for DNA barcodes (Leese et al., 2018, 2016; Oliveira et al., 2016; Weigand et al., 2019). In fact, important taxonomic gaps in the reference libraries of DNA barcodes of marine invertebrates have been recently reported for the European marine regions, which are typically much larger than their freshwater counterparts. Some records are flagged as doubtful barcodes and inadequate quality standardization of reference barcodes can affect the reliability of a reference library. This includes identification errors, sequence contamination, incomplete reference data without trace files or primer information and inadequate data management (Weigand et al., 2019).

Considering the above described relevance of the Atlantic Iberia region and the importance of up-to-date DNA-based technologies to support macroinvertebrate monitoring in this region, we conducted a comprehensive assessment of the gaps in the regional reference library of COI barcodes for the three most prominent coastal marine taxa (Annelida, Crustacea and Mollusca) occurring in this area. We also reviewed the taxonomic congruency of the COI barcode records and provided a comprehensive and annotated reference library for the target taxa occurring in Atlantic Iberia.

2. Material and methods

2.1. Study area and checklist

A comprehensive species-level checklist of marine macroinvertebrate species occurring in Iberian Atlantic Coast was compiled using the World Register of Marine Species (WoRMS) database (http://www.marinespecies.org/) and literature records (for checklist details with species information consult Table S.1; to consult references see Table S.2). The study area comprised the marine region of Continental Atlantic Iberia, i.e., between the France–Spain Atlantic border to the strait of Gibraltar (Macaronesia not included). We assessed only selected taxonomic ranks among the three most dominant groups of marine macroinvertebrates: Crustacea (Malacostraca: Amphipoda, Decapoda and Isopoda; Thecostraca: Balanomorpha), Annelida (Polychaeta) and Mollusca (Bivalvia, Gastropoda and Polyplacophora). The validity of the species names in the final checklist and their assignment as "marine" was verified in WoRMS database with the package "worms" (Holstein, 2018), through the software R 3.6.1 (R Core Team, 2019; www.r-project.org).

2.2. Data mining and BOLD Dataset creation

All the available COI-5P sequences matching the species names in the checklist for Atlantic Iberia were mined from the Barcode of Life Data system (BOLD; Ratnasingham and Hebert, 2007) using the R package "bold" (Chamberlain, 2019). Records without information on species name, containing COI sequences with less than 500 base pairs and flagged for contamination, stop-codons or indels were subsequently removed. To this dataset, we added new sequences of specimens collected at the Iberian Atlantic coast (dataset DS-AIMARINV, which also includes records obtained by us from our past publications). A final dedicated dataset which aggregate all compiled DNA barcodes (DS-GAIMARIN — doi.org/10.5883/DS-GAIMARIN; Table 1) was created in BOLD.

The new COI barcode sequences were obtained from specimens collected on the Portuguese coast, following published protocols (Borges et al., 2016; Lobo et al., 2016, 2017), and were submitted to GenBank (accessions and specimen list are available in Table S.3).

2.3. Data processing and analyses

To conduct a global gap-analysis of the barcoded species from Atlantic Iberia, we compared the species checklist with all publicly available COI-5P sequence records in BOLD. A species was considered successfully barcoded if at least one COI-5P sequence (>500 bp) was available. The geographic origin of the specimens was also recorded. All the records sampled between the coordinates lat 42.00 and 44.00 and between long −11.00 and −02.00 (North Continental Atlantic Iberian Peninsula) or between lat 36.00 and 42.00 and between long −11.00 and −05.30 (South Continental Atlantic Iberian Peninsula) or with clear indication that were sampled in Continental Atlantic Iberia were considered

Table 1
Number of COI-5P sequences generated under this study and number of COI-5P sequences retrieved from BOLD, compiled in the dataset DS-GAIMARIN for each target taxa, with the associated BOLD project code.

Target taxa	BOLD project code	Number of COI-5P sequences
Crustacea: Decapoda	MLALE[a]	22
Crustacea: Isopoda	ISOAL[a]; WBEC[a]	3
Crustacea: Balanomorpha	FCCOM[a]	27
Annelida: Polychaeta	PCALE[a]	9
Mollusca: Bivalvia	BIPM[a]; BIV[a]; BVALN[a]; METP[a]	39
Mollusca: Gastropoda	GTALE[a]	7
Mollusca: Polyplacophora	PIPM[a]	8
Crustacea, Annelida and Mollusca[b]		18047
Total		18162

[a]New data generated on this study.
[b]Retrieved from Bold.

as "Atlantic Iberian Peninsula". All the records with clear country information outside from the delimited area were considered as "not Atlantic Iberian Peninsula". Ambiguous records indicating Iberian Peninsula or Atlantic Ocean (i.e., doubts if they were sampled in Atlantic Continental Iberia or elsewhere) were considered "unknown". For the purpose of data analyses, we first considered records of species from Atlantic Iberian Peninsula, second selected species with records indicating geographic sampling collection outside Iberian Peninsula; and lastly the "unknown" records.

All the species in the dataset which were assigned to a Barcode Index Number (BIN) (Ratnasingham and Hebert, 2013) were annotated with one of three possible taxonomic congruency grades: discordant (i.e. more than one species assigned to the same BIN), multiple BINs (i.e., one species assigned to more than one BIN) and concordant (i.e., one species assigned to only one BIN). The BINs assigned to different species (i.e. discordant BINs) were carefully inspected by checking their placement in NJ phenograms, looking for the valid species names, synonyms or contaminations, and by inspecting BINs' content on BOLD database. Namely, to further verify the congruency between BINs and morphospecies, neighbor joining (NJ) phenograms for each phylum, were built in MEGA v7, using Kimura-2-parameter (K2P) substitution model (Kimura, 1980). Node support was assessed through 1000 bootstrap replicates. Only three selected sequences from each BIN were used to construct the NJ phenograms. The selection was performed by using the following criteria in the same order: (1) without "N" and without gaps and whenever possible, select sequences with 658 bp; (2) sequences between 650 bp and 657 bp; (3) sequences between 600 bp and 649 bp; (4) sequences between 500 bp and 599 bp; (5) sequences higher than 658 bp; (6) any available sequences. When more than three sequences were compliant with the above criteria, only three were randomly selected. In the case of species without an attributed BIN, we selected three sequences from each species following the above criteria. All sequences were aligned using MAFFT v7 (https://mafft.cbrc.jp/alignment/server/; Katoh and Standley, 2013).

The bioinformatic pipeline developed to carry out these analyses is available at https://github.com/pedroemanuelvieira/Iberian-Peninsula-DNA-Reference-Library.

3. Results

The final checklist had a total of 2827 marine macroinvertebrate species occurring in the Atlantic Iberia, belonging to three major groups (926 Annelida, 638 Crustacea, and 1263 Mollusca). The distribution by taxonomic groups, the number of sequences per group and the geographic region of specimen collection are displayed in Fig. 1 (see detailed information in Tables S.3 and S.4).

The DS-GAIMARIN dataset is composed by 18162 COI-5P sequences belonging to 1053 species, assigned to 1545 BINs. One-hundred and fifteen new DNA barcodes were generated under this study, among which two species were barcoded for the first time, namely the decapod *Gilvossius tyrrhenus* (Petagna, 1792) and the polyplacophora *Leptochiton albemarlensis* A. G. Smith & Ferreira, 1977.

Mollusca was the most well represented taxon in the dataset in number of sequences (8816 sequences, 49%), and the most diverse class, in terms of species, was Gastropoda (952 species), which also displayed the highest number of sequences (5406). Crustacea also had a high proportion of species with DNA barcodes (6266 sequences, 35% of sequence representation), for which Amphipoda was the most well represented order in terms of species (268 species), although Decapoda recorded a higher number of DNA barcode sequences (3253 sequences, 52% of sequence representation). On the other hand, Polychaeta displayed very low numbers of sequences (16%), although well represented with species in the list (approximately 33% of the total number of species).

The availability of DNA barcodes for the examined taxa varied considerably across taxonomic groups (Fig. 1.A), and in total only 37% (1053 species) of the species had at least one barcode sequence deposited in BOLD. Among the three selected groups, Polychaeta had the lowest barcode coverage, with only 31% (corresponding to 288 species) of the total species represented in the checklist being barcoded. In Mollusca 34% of the species were barcoded, however among the three major classes more than 50% were still missing DNA barcode sequences. Despite Gastropoda having the highest number of representative sequences it displayed a lower level of completion (31%), than Bivalvia (42%) or Polyplacophora (52%). Yet the number of listed taxa is highly disparate for these classes: 952 for Gastropoda, 286 for Bivalvia and 25 for Polyplacophora. Overall, Crustacea had the largest coverage with 53% (338 species barcoded), but Decapoda and Balanomorpha reach more than 80% of total species barcoded, while Amphipoda and Isopoda displayed very low completion; 43% and 24%, respectively.

Although Crustacea did not display the highest number of sequences, a detailed analysis of the geographic region from where the specimens were sampled reveals that the highest representation of species from Atlantic Iberia was found for this group (57% of total number of species, Fig. 1.B). The results also showed that the majority of the sequences had the specimens sampling collection information associated, and for only 16% of barcoded species there was no data or insufficient data (i.e. "unknown"). Sixty-seven species in the list are non-indigenous for the Iberian Peninsula, of which 61% (41 species) had at least one barcode sequence deposited in BOLD, while 39% (26 species) lack a barcode sequence (Table S.5.).

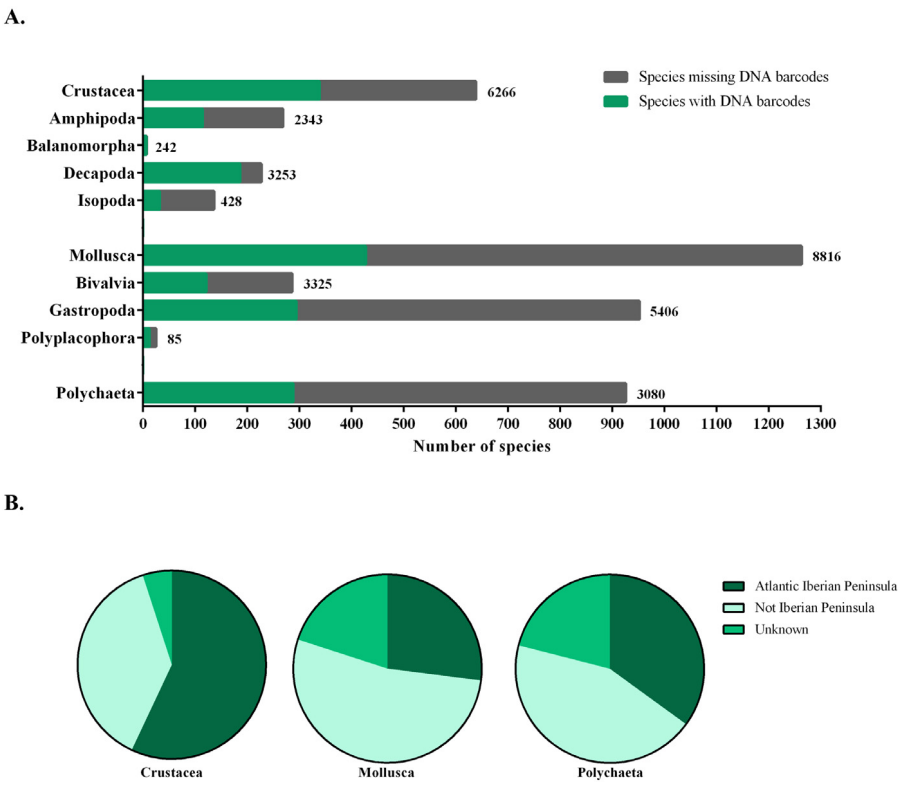

Fig. 1. A. DNA barcode coverage for marine macroinvertebrate species occurring in Atlantic Iberia. Barcode coverage with at least one COI-5P sequence per species (green bar). Numbers on the right side of each bar refer to the total number of sequences. **B.** Partitioning of the geographic origin of specimens available in the reference library: sampled in Atlantic Iberia, outside of Iberian Peninsula and uncertain geographical information (unknown). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Overall, the majority of BINs were considered concordant (i.e. one BIN = one species): 649 species corresponding to 42% of the total number of BINs (Fig. 2). A total of 284 species were assigned to more than one BIN (corresponding to 831 BINs, 27% of the species), and Mollusca displayed the highest percentage of species assigned to multiple BINs (124 species, 43% of BINs). Among Mollusca, Gastropoda had the highest number of species displaying multiple BINs (80 species, 59% of BINs), however it was also the class harboring more species and consequently more representative sequences. Only 11% of the total number of BINs were discordant (i.e. 120 BINs were shared by more than one species). Gastropods showed the highest levels of discordance (36 species), whereas Polyplacophora, Isopoda and Balanomorpha did not display any discordant BINs. However, a subsequent inspection of BINs revealed an overestimation and unrealistic percentage of discordant BINs. Following a careful inspection, 33 discordant BINs displayed concordance or can be assigned to other species, mainly due to misidentifications. Therefore, the number of discordant BINs decreased to 87 species (8%; Table S.6), and consequently the number of concordant and multiple BINs increased to 676 species (66%) and 259 species (26%; Table S.7), respectively.

Phylogenetic trees were constructed for taxonomic reliability inspection (Figs. S.1–S.8). A total of 3178 COI-5P sequences distributed over the three taxonomic groups (1010 Crustacea, 1343 Mollusca, and 825 Polychaeta) were used to construct the trees. The number of species represented by only one sequence per BIN (singletons) was 65 for Crustacea, 103 for Mollusca and 127 for Polychaeta. Furthermore, 170 species (61 Crustacea, 80 Mollusca and 29 Polychaeta) displayed a high intraspecific divergence, and the groups with the highest values were Gastropoda (46 species) and Decapoda (32 species), followed by Bivalvia (29 species) and Amphipoda (20 species).

4. Discussion

The current study highlights three main considerations: first, reference libraries still lack representative barcodes for many marine macroinvertebrate species belonging to dominant faunal groups; second, a considerable number of species apparently integrate hidden or undescribed diversity; and third, a comparatively low proportion of taxonomic incongruences were detected, which may eventually impact the accuracy of current DNA-based assessment and biomonitoring of marine ecosystems, though we partially sorted them out by auditing and annotating our compiled reference library.

Marine macroinvertebrates are among the most phylogenetically diverse communities, thereby constituting a particularly demanding component for morphology-based biomonitoring, and also a harder target to achieve a comprehensive DNA barcode reference library (Lobo et al., 2017). Yet, DNA metabarcoding's taxonomic span of detection and degree of accuracy is highly depend on reference libraries completion and reliability of the records. In the current study the gap of DNA barcodes found was considerably high for the examined taxonomic groups (63%). This result was not unexpected since a number of studies already revealed a high prevalence of gaps of DNA barcodes for specific taxa (e.g. Aylagas et al., 2014; Abad et al., 2016; Lobo et al., 2016), and other studies showed disagreements between molecular and morphological assignments which are mostly associated to incompleteness of reference sequences databases (Kelly et al., 2017; Weigand et al., 2019). Furthermore, marine biodiversity assessment is challenging due to geographical large-scale sampling effort, which has a critical impact on species assessment (Bergsten et al., 2012). These factors have negative impacts on taxonomic research, which leads to a higher proportion of

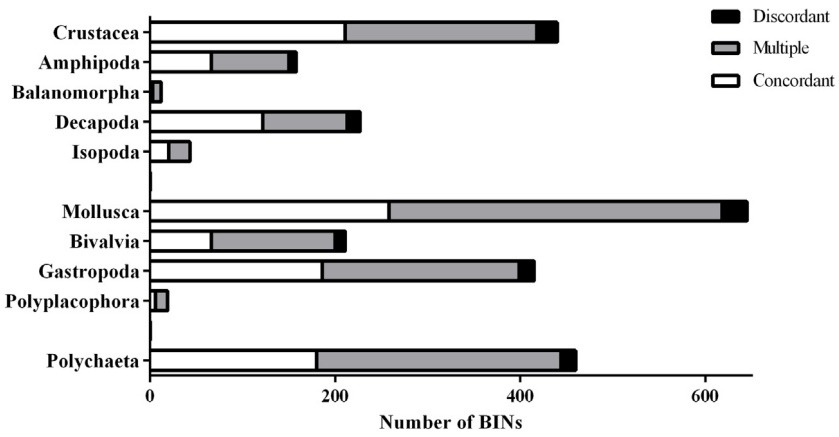

Fig. 2. Number of BINs according to taxonomic congruence annotations (concordant, multiple or discordant) for each taxonomic group on the Atlantic Iberia reference dataset.

undescribed diversity and affect the outcome of richness of a community (Pawlowski et al., 2018).

Our results are comparable to those obtained with other checklists compiled for marine species: the AZTI Marine Biotic Index (AMBI) list (Borja et al., 2000; Aylagas et al., 2014; Weigand et al., 2019) and the European Register of Marine Species (ERMS) list (Weigand et al., 2019), both targeting European marine occurring taxa, but that differ in the taxonomic composition. Although for the three taxonomic groups our checklist had similar number of taxa as the AMBI list (2827 and 2560, respectively) and much lesser than the ERMS list (14207 species), the gap found in the current study was more similar to the one found previously for the ERMS' list (70% for the ERMS vs 63% for ours vs 50% for AMBI) (Weigand et al., 2019).

The number of DNA barcodes available on public databases can be somehow related with the number of dedicated studies, and consequently barcoding projects associated (Weigand et al., 2019). For example, many projects and studies were dedicated to complete the reference sequences databases for fishes (Costa et al., 2012; Keskin and Atar, 2013; Oliveira et al., 2016; Cariani et al., 2017), while the number of macroinvertebrate barcoding dedicated projects are much lower. Indeed, the obtained differences on DNA barcode completion among the three taxonomic groups can be explained by the frequency that specific taxonomic groups and/or species are targeted in barcoding studies (Barco et al., 2016; Lobo et al., 2016). We found many barcoding-based studies dedicated to crustaceans (Costa et al., 2007; da Silva et al., 2013; Raupach et al., 2015; Lobo et al., 2017), which can highly increase the representativeness of sequences belonging to this group in genetic databases, and is probably the best explanation for the largest coverage of Crustacea found in the current study. Furthermore, most of these studies based on crustaceans were developed in Atlantic Iberia, which can explain the highest representation of this group with records from Atlantic Iberia. Our results reveal the need to increase the projects and studies dedicated to marine macroinvertebrate species, in particular for Annelida and Mollusca. However, while reference libraries are far from being complete, the generation of DNA barcodes for the most frequent species occurring at a particular site or region may overcome databases incompleteness and consequently improve taxonomic assignment using DNA-based tools in local studies (Aylagas et al., 2014; Abad et al., 2016). Although we registered a relatively low proportion of sequences (16%) with no geographic data or insufficient data (i.e. "unknown" sequences represented on Fig. 1.B.), it is important to stress the relevance of metadata in public databases, particularly the geographic origin of the specimens, which is especially critical for a library still with considerable gaps and numerous poorly-represented species.

Globally, the comparative analysis between morphology-based taxonomic identification and BINs assignments exposed a sizeable amount of discordances. Notably, it also suggested that species diversity assessed through morphology can be currently underestimated by as much as 50% of the target taxa, with suspected hidden diversity recorded on over 5% of the examined morphospecies. These findings are transversal to all taxonomic groups and can be explained by different reasons. In some cases, the species names have not yet been updated in the BOLD database, and sequences misidentified combined with taxa absent in the databases will generate incorrect taxonomic identifications. For example, two specimens of gastropods were morphologically identified to genus-level as *Nassarius* sp. and *Ocenebrina* sp. However, based on barcodes and phylogenetic trees construction, both cases can be now identified as *Tritia incrassata* (Strøm, 1768) and *Ocenebra edwardsii* (Payraudeau, 1826). Moreover, two decapods *Melicertus kerathurus* (Forskål, 1775) and *Penaeus kerathurus* (Forskål, 1775), were attributed to the same BIN (BOLD:AAB4142), but a confirmation on WoRMS of the taxonomic status of the scientific names revealed that *M. kerathurus* is currently unaccepted and was updated to *P. kerathurus*. A closer look to the phylogenetic tree generated in the current study, suggested that other BINs discordances can be related to misidentifications. For example, the BIN BOLD:AAW8076 had sequences identified as *Caprella acanthifera* Leach, 1814 and *Caprella danilevskii* Czerniavski, 1868, which grouped on the same clade with low divergence. However *C. danilevskii* shall be the correct taxonomic identification, since there is another BIN (BOLD:AAY5434) identified as *C. acanthifera* grouped in another clade which is recognized as a species-complex since 1998 (Krapp-Schickel and Vader, 1998). Overall, our careful inspection of the composition of the discordant BINs revealed that most were related to misidentifications or synonyms.

Consistently, all taxonomic groups analyzed displayed a fair amount of cases of high intraspecific divergence, probably related with hidden or cryptic diversity, of which most of them were already reported in previous studies (Best and Stachowicz, 2013; Layton et al., 2014; Leray and Knowlton, 2015; Trickey et al., 2016; McCarthy et al., 2019; Teixeira et al., 2019; Vieira et al., 2019). For example, for the gastropod genus *Doto* represented in the reference library by 8 distinct species, has been pointed out as an extremely challenging group for taxonomic identifications due to their small body size, similar color patterns and lack of distinctive morphological characters (Morrow et al., 1992; Pola and Gosliner, 2010). In our results, *Doto coronata* (Gmelin, 1791) and *Doto koenneckeri* Lemche, 1976 were each one assigned to two different BINs and grouped in different clades with high

divergence (>8%). Some researchers described these lineages as a complex (Korshunova et al., 2016; Shipman and Gosliner, 2015), but more taxonomic and molecular work are still needed to solve this issue. Another example of an observed cryptic complex is the polychaete *Syllis gracilis* Grube, 1840. DNA barcodes for this species were sorted into multiple lineages with an unbalanced representation: 34 sequences assigned to 11 BINs. This cryptic complex has been already disclosed, however a combination of different interactions among environmental features and biogeographical factors have been hindering its full interpretation (Langeneck et al., 2020).

Although previous studies on Amphipoda revealed a majority of monophyletic clades consistent with consolidated morphospecies (Raupach et al., 2015), there is still considerable taxonomic instability in particular species, which display among the highest levels of intraspecific divergence here recorded. This is the case, for example of the *Apohyale stebbingi* Chevreux, 1888 complex (Desiderato et al., 2019), which was assigned to two different BINs (BOLD:AAI8298 and BOLD:ACX2700) diverging over 13% K2P.

One of the salient advantages of metabarcoding compared to morphology-based monitoring is the ability to detect and document the occurrence of cryptic species. However, because metabarcoding procedures typically use shorter fragments then the full COI-5P, one question that may arise is if there will be still enough resolution to discriminate cryptic species in such conditions. At least for COI metabarcoding of marine invertebrates, studies have shown that there is very little loss of discrimination ability for segments of COI-5P down to 200 base pairs (Hollatz et al., 2017). However, this may not hold for other markers, such as for example 18S rDNA sequences, which have been documented to have little discrimination ability at the species level (Tang et al., 2012; Lejzerowicz et al., 2015; Danovaro et al., 2016). On the other hand, instances of hybridization or mitochondrial introgression between closely related species will fail detection by metabarcoding (like through regular DNA barcoding), but such phenomena will be very likely overlooked by morphology-based monitoring as well (Cowart et al., 2015; Pawlowski et al., 2018). There are of intrinsic pitfalls both in morphology- and barcode-based identifications. Thus, the combination of morphological identification with DNA barcoding in an integrative approach for monitoring biodiversity contribute to significantly facilitate comparative studies of genetic diversity in different species. In addition, this integrative approach also facilitates comprehensive analyses of a given taxonomic assemblage and provides insights into the patterns of genomic diversity within species.

The addition of publicly available sequences obtained from specimens collected in other geographic regions allowed to understand patterns of concordance/discordance between BINs. For example, the polychaete *Cirriformia tentaculata* (Montagu, 1808) was assigned to two distinct BINs which grouped in two clades: BOLD:ACI3598 corresponding to samples originated from China, and BOLD:ACI2312 corresponding to samples collected from Portugal. This can be an evidence of possible cryptic polymorphism in this species, already pointed out in a previous study (George, 1967). However, BINs were composed by a low number of sequences (less than 3 sequences), which are not enough to reach a strong conclusion.

Considering the diversity of faunal assemblages combined with the introduction of non-indigenous Species (NIS) in the Iberian Atlantic coast, it is especially important to early assess and monitor the impacts and changes in marine species range, identifying possible biological invasions and enable the development of mitigation strategies (Briski et al., 2016; Rey et al., 2019; Viard et al., 2019). In order to use metabarcoding as a tool to early detect and improve monitoring of NIS in coastal and marine ecosystems, it is extremely important to complete the number of missing barcode sequences for NIS (Briski et al., 2016; Ardura, 2019), as well as to solve problems of multiple or discordant BINs associated to NIS, since in this case species-level identification is mandatory and wrong identifications can trigger action or inaction when not required.

The detection of a reasonable number of marine macroinvertebrates still missing DNA barcodes and the presence of hidden or undescribed diversity in the reference library compiled in this study, highlight the urgent need to complete and curate reference sequences databases for such important marine groups. The reference library compiled, audited and annotated in the current study is ought to be a valuable support to improve the precision of taxonomic assignments in metabarcoding studies in Atlantic Iberia and to overcome under- or overestimation of species richness.

5. Conclusions

The reference library compiled in this study covers the most dominant groups for marine macroinvertebrate species occurring in the Iberian Atlantic coast, which are the most commonly used species in biomonitoring programs. To our best knowledge, this is the first study to assemble a barcode reference library for these dominant groups of marine macroinvertebrate species from this important region of the Atlantic. However, we are still far from having a representative reference library for such diverse taxonomic groups, with prevalence of large gaps in the library. Furthermore, other important marine taxa (e.g. echinoderms or ascidians) should be included in forthcoming studies to improve the completion of reference libraries and broader integration in ecological assessments of marine species, namely through DNA metabarcoding. A significant finding emerging from our analysis was the circa 50% higher number of species delimited through molecular data (i.e. BINs) compared to described morphospecies occurring in this Atlantic Iberia only. Implications of such exceptional levels of suspected hidden diversity should be taken into consideration in upcoming macroinvertebrate-based ecosystem monitoring and research. The continuous growth of reference libraries with comprehensive sampling strategies, ranging from different regions and a broad range of specimens, combined with morphological taxonomy and molecular phylogenetic techniques will probably allow to better understand the diversity and deep genetic structure within species, in order to solve the observed discrepancies and incongruences, most of them probably associated with undescribed or cryptic diversity.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

B.R. Leite: Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing. **P.E. Vieira:** Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing. **M.A.L. Teixeira:** Data curation, Writing - review & editing. **J. Lobo-Arteaga:** Methodology, Data curation, Writing - review & editing. **C. Hollatz:** Methodology, Writing - review & editing. **L.M.S. Borges:** Methodology, Data curation, Writing - review & editing. **S. Duarte:** Conceptualization, Methodology, Data curation, Writing - review & editing. **J.S. Troncoso:** Supervision, Writing - review & editing. **F.O. Costa:** Conceptualization, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgments

This study was supported by the project The NextSea [NORTE-01-0145-FEDER-000032], under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). This work was supported by the "Contrato-Programa" UIDB/04050/2020 funded by national funds through the FCT I.P. SD and PEV work was supported by national funds through the Portuguese Foundation for Science and Technology (FCT, I.P.) in the scope of the project NIS-DNA [PTDC/BIA-BMA/29754/2017]. BRL benefitted from an FCT fellowship PD/BD/127994/2016. MALT benefitted from an FCT fellowship co-financed by European Social Fund (ESF) SFRH/BD/131527/2017.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.rsma.2020.101307.

References

Abad, D., Albaina, A., Aguirre, M., Laza-Martínez, A., Uriarte, I., Uriarte, A., Villate, F., Estonba, A., 2016. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Mar. Biol. 163 (7), 149. http://dx.doi.org/10.1007/s00227-016-2920-0.

Araújo, R., Bárbara, I., Tibaldo, M., Berecibar, E., Tapia, P.D., Pereira, R., Santos, R., Pinto, I.S., 2009. Checklist of benthic marine algae and cyanobacteria of northern Portugal. Bot. Mar. 52 (1), 24–46. http://dx.doi.org/10.1515/BOT.2009.026.

Ardura, A., 2019. Species-specific markers for early detection of marine invertebrate invaders through eDNA methods: gaps and priorities in GenBank as database example. J. Nat. Conser. 47, 51–57. http://dx.doi.org/10.1016/j.jnc.2018.11.005.

Aylagas, E., Borja, A., Muxika, I., Rodríguez-Ezpeleta, N., 2018. Adapting metabarcoding based benthic biomonitoring into routine marine ecological status assessment networks. Ecol. Indic. 95, 194–202. http://dx.doi.org/10.1016/j.ecolind.2018.07.044.

Aylagas, E., Borja, A., Rodríguez-Ezpeleta, N., 2014. Environmental status assessment using DNA metabarcoding: towards a genetics based marine biotic index (gAMBI). PLoS One 9 (3), e90529. http://dx.doi.org/10.1371/journal.pone.0090529.

Barco, A., Raupach, M.J., Laakmann, S., Neumann, H., Knebelsberger, T., 2016. Identification of North Sea molluscs with DNA barcoding. Mol. Ecol. Resour. 16 (1), 288–297. http://dx.doi.org/10.1111/1755-0998.12440.

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G.N., Ribera, I., Nilsson, A.N., Barraclough, T.G., Vogler, A.P., 2012. The effect of geographical scale of sampling on DNA barcoding. Syst. Biol. 61, 851–869. http://dx.doi.org/10.1093/sysbio/sys037.

Best, R.J., Stachowicz, J.J., 2013. Phylogeny as a proxy for ecology in seagrass amphipods: which traits are most conserved? PLoS One 8 (3), http://dx.doi.org/10.1371/journal.pone.0057550.

Boaventura, D., Ré, P., Cancela da Fonseca, L., Hawkins, S.J., 2002. Intertidal rocky shore communities of the continental Portuguese coast: analysis of distribution patterns. Mar. Ecol. 23 (1), 69–90. http://dx.doi.org/10.1046/j.1439-.0485.2002.02758.

Borges, L.M.S., Hollatz, C., Lobo, J., Cunha, A.M., Vilela, A.P., Calado, G., Coelho, R., Costa, A.C., Ferreira, M.S.G., Costa, M.H., Costa, F.O., 2016. With a little help from DNA barcoding: investigating the diversity of Gastropoda from the Portuguese coast. Sci. Rep. 6 (20226), http://dx.doi.org/10.1038/srep20226.

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://dx.doi.org/10.1016/S0025-326X(00)00061-8.

Bourlat, S.J., Borja, A., Gilbert, J., Taylor, M.I., Davies, N., Weisberg, S.B., Griffith, J.F., Lettieri, T., Field, D., Benzie, J., Glöckner, F.O., Rodríguez-Ezpeleta, N., Faith, D.P., Bean, T.P., Obst, M., 2013. Genomics in marine monitoring: new opportunities for assessing marine health status. Mar. Pollut. Bull. 74 (1), 19–31. http://dx.doi.org/10.1016/j.marpolbul.2013.05.042.

Briggs, J.C., Bowen, B.W., 2012. A realignment of marine biogeographic provinces with particular reference to fish distributions. J. Biogeography 39, 12–30. http://dx.doi.org/10.1111/j.1365-2699.2011.02613.x.

Briski, E., Ghabooli, S., Bailey, S.A., MacIsaac, H.J., 2016. Are genetic databases sufficiently populated to detect non-indigenous species? Biol. Invasions 18, 1911–1922. http://dx.doi.org/10.1007/s10530-016-1134-1.

Bush, A., Compson, Z.G., Monk, W.A., Porter, T.M., Steeves, R., Emilson, E.J., Gagne, N., Hajibabaei, M., Roy, M., Baird, D.J., 2019. Studying ecosystems with DNA metabarcoding: lessons from biomonitoring of aquatic macroinvertebrates. Frontiers in Ecology and Evolution 7 (434), http://dx.doi.org/10.3389/fevo.2019.00434.

Cariani, A., Messinetti, S., Ferrari, A., Arculeo, M., Bonello, J.J., Bonnici, L., Cannas, R., Carbonara, P., Cau, A., Charilaou, C., Ouamari, N.E., Fiorentino, F., Follesa, M.C., Garofalo, G., Golani, D., Guarniero, I., Hanner, R., Hemida, F., Kada, O., lo Brutto, S., Mancusi, C., Morey, G., Schembri, P.J., Serena, F., Sion, L., Stagioni, M., Tursi, A., Vrgoc, N., Steinke, D., Tinti, F., 2017. Improving the conservation of Mediterranean chondrichthyans: The ELASMOMED DNA barcode reference library. PLoS One 12 (1), e017024. http://dx.doi.org/10.1371/journal.pone.0170244.

Chainho, P., Fernandes, A., Amorim, A., Ávila, S.P., Canning-Clode, J., Castro, J.J., Costa, A.C., Costa, J.L., Cruz, T., Gollasch, S., Grazziotin-Soares, C., Melo, R., Micael, J., Parente, M.I., Semedo, J., Silva, T., Sobral, D., Sousa, M., Torres, P., Veloso, V., Costa, M.J., 2015. Non-indigenous species in Portuguese coastal areas, coastal lagoons, estuaries and islands. Estuar. Coast. Shelf Sci. 167, 199–211. http://dx.doi.org/10.1016/j.ecss.2015.06.019.

Chamberlain, S., 2019. bold: Interface to bold systems API. R package version 0.9.0. https://CRAN.R-project.org/package=bol. (Accessed February 2020).

Costa, F.O., deWaard, J.R., Boutillier, J., Ratnasingham, S., Dooh, R.T., Hajibabaei, M., Hebert, P.D.N., 2007. Biological identifications through DNA barcodes: the case of the Crustacea. Can. J. Fish. Aquat. Sci. 64, 272–295. http://dx.doi.org/10.1139/f07-008.

Costa, F.O., Landi, M., Martins, R., Costa, M.H., Costa, M.E., Carneiro, M., Alves, M.J., Steinke, D., Carvalho, G.R., 2012. A ranking system for reference libraries of DNA barcodes: application to marine fish species from Portugal. PLoS One 7, e35858. http://dx.doi.org/10.1371/journal.pone.0035858.

Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., Arnaud-Haond, S., 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. PLoS One 10 (2), e0117562. http://dx.doi.org/10.1371/journal.pone.0117562.

Cristescu, M.E., 2019. From barcoding single individuals to metabarcoding biological communities: Towards an integrative approach to the study of global biodiversity. Trends Ecol. Evol. 29 (10), 566–571. http://dx.doi.org/10.1016/j.tree.2014.08.001.

Danovaro, R., Carugati, L., Berzano, M., Cahill, A.E., Carvalho, S., Chenuil, A., Corinaldesi, C., Cristina, S., David, R., Dell'anno, A., Dzhembekova, N., Garcés, E., Gasol, J., Goela, P., Feral, J.P., Ferrera, I., Forster, R., Kurekin, A.A., Rastelli, E., Marinova, V., Miller, P.I., Moncheva, S., Newton, A., Pearman, J., Pitois, S., Renè, A., Rodriguez-Ezpeleta, N., Saggiomo, V., Simis, S.G.H., Stefanova, K., Wilson, C., Lo Martire, M., Greco, S., Cochrane, S., Borja, A., 2016. Implementing and innovating marine monitoring approaches for assessing marine environmental status. Front. Mar. Sci. 3, 213. http://dx.doi.org/10.3389/fmars.2016.00213.

Desiderato, A., Costa, F.O., Serejo, C.S., Abbiati, M., Queiroga, H., Vieira, P.E., 2019. Macaronesian islands as promoters of diversification in amphipods: The remarkable case of the family Hyalidae (Crustacea, Amphipoda). Zool. Scripta 48 (3), 359–375. http://dx.doi.org/10.1111/zsc.12339.

European Commission, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy. https://eur-lex.europa.eu/eli/dir/2000/60.

European Commission, 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 Establishing a Framework for Community Action in the Field of Marine Environmental Policy (Marine Strategy Framework Directive). https://eur-lex.europa.eu/eli/dir/2008/56.

George, J.D., 1967. Cryptic polymorphism in the cirratulid polychaete, *Cirriformia tentaculata*. J. Mar. Biol. Assoc. UK 47 (1), 75–79. http://dx.doi.org/10.1017/S0025315400033579.

Guerra-García, J.M., Baeza-Rojano, E., Cabezas, M.P., García-Gómez, J.C., 2011. Vertical distribution and seasonality of peracarid crustaceans associated with intertidal macroalgae. J. Sea Res. 65 (2), 256–264. http://dx.doi.org/10.1016/j.seares.2010.12.001.

Hajibabaei, M., Spall, J.L., Shokralla, S., van Konynenburg, S., 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. BioMed. Cent. Ecol. 12, 28. http://dx.doi.org/10.1186/1472-6785-12-28.

Hayward, P.J., Ryland, J.S., 1995. HandBook of the Marine Fauna of North-West Europe. Oxford University press, Oxford, United Kigdom.

Hollatz, C., Leite, B.R., Lobo, J., Froufe, H., Egas, C., Costa, F.O., 2017. Priming of a DNA metabarcoding approach for species identification and inventory in marine macrobenthic communities. Genome 60, 260–271. http://dx.doi.org/10.1139/gen-2015-0220.

Holstein, J., 2018. Worms: Retriving aphia information from world register of marine species. R package version 0.2.2. https://CRAN.R-project.org/package=worms. (Accessed 2020).

Hupało, K., Teixeira, M.A.L., Rewicz, T., Sezgin, M., Iannilli, V., Karaman, G.S., Grabowski, M., Costa, F.O., 2019. Persistence of phylogeographic footprints helps to understand cryptic diversity detected in two marine amphipods widespread in the Mediterranean basin. Mol. Phylogenet. Evol. 132, 53–66. http://dx.doi.org/10.1016/j.ympev.2018.11.013.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. http://dx.doi.org/10.1093/molbev/mst010.

Kelly, R.P., Closek, C.J., O'Donnell, J.L., Kralj, J.E., Shelton, A.O., Samhouri, J.F., 2017. Genetic and manual survey methods yield different and complementary views of an ecosystem. Front. Mar. Sci. 3, 283. http://dx.doi.org/10.3389/fmars.2016.00283.

Keskin, E., Atar, H.H., 2013. DNA barcoding commercially important fish species of Turkey. Mol. Ecol. Resour. 13, 788–797. http://dx.doi.org/10.1111/1755-0998.12120.

Kimura, M., 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120. http://dx.doi.org/10.1007/BF01731581.

Korshunova, T., Sanamyan, N., Zimina, O., Fletcher, K., Martynov, A., 2016. Two new species and a remarkable record of the genus Dendronotus from the North Pacific and Arctic oceans (Nudibranchia). ZooKeys 630, 19–42. http://dx.doi.org/10.3897/zookeys.630.10397.

Krapp-Schickel, T., Vader, W., 1998. What is, and what is not, *Caprella acanthifera* Leach, 1814 (Amphipoda, Caprellidae)? Part 1: the acanthifera-group. J. Nat. Hist. 32 (7), 949–967. http://dx.doi.org/10.1080/00222939800770491.

Langeneck, J., Scarpa, F., Maltagliati, F., Sanna, D., Barbieri, M., Cossu, P., Mikac, B., Galletti, M.C., Castelli, A., Casu, M., 2020. A complex species complex: The controversial role of ecology and biogeography in the evolutionary history of *Syllis gracilis* Grube, 1840 (Annelida, Syllidae). J. Zool. Sys. Evol. Res. 58 (1), 66–78. http://dx.doi.org/10.1111/jzs.12336.

Layton, K.K., Martel, A.L., Hebert, P.D.N., 2014. Patterns of DNA barcode variation in Canadian marine molluscs. PLoS One 9 (4), e95003. http://dx.doi.org/10.1371/journal.pone.0095003.

Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Meissner, K., Mergen, P., Pawlowski, J., Piggott, J., Rimet, F., Steinke, D., Taberlet, P., Weigand, A., Abarenkov, K., Beja, P., Bervoets, L., Björnsdóttir, S., Boets, P., Boggero, A., Bones, A., Borja, A., Bruce, K., Bursić, V., Carlsson, J., Čiampor, F., Čiamporová Zatovičová, Z., Coissac, E., Costa, F.O., Costache, M., Creer, S., Csabai, Z., Deiner, K., DelValls, A., Drakare, S., Duarte, S., Eleršek, T., Fazi, S., Fišer, C., Flot, J., Fonseca, V., Fontaneto, D., Grabowski, M., Graf, W., Guðbrandsson, J., Hellström, M., Hershkovitz, Y., Hollingsworth, P., Japoshvili, B., Jones, J., Kahlert, M., Kalamujic Stroil, B., Kasapidis, P., Kelly, M., Kelly-Quinn, M., Keskin, E., Kõljalg, U., Ljubešić, Z., Maček, I., Mächler, E., Mahon, A., Marečková, M., Mejdandzic, M., Mircheva, G., Montagna, M., Moritz, C., Mulk, V., Naumoski, A., Navodaru, I., Padisák, J., Pálsson, S., Panksep, K., Penev, L., Petrusek, A., Pfannkuchen, M., Primmer, C., Rinkevich, B., Rotter, A., Schmidt-Kloiber, A., Segurado, P., Speksnijder, A., Stoev, P., Strand, M., Šulčius, S., Sundberg, P., Traugott, M., Tsigenopoulos, C., Turon, X., Valentini, A., van der Hoorn, B., Várbíró, G., Vasquez Hadjilyra, M., Viguri, J., Vitonytė, I., Vogler, A., Vrålstad, T., Wägele, W., Wenne, R., Winding, A., Woodward, G., Zegura, B., Zimmermann, J., 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. Research Ideas and Outcomes 2, http://dx.doi.org/10.3897/rio.2.e11321, e11321.

Leese, F., Bouchez, A., Abarenkov, K., Altermatt, F., Borja, A., Bruce, K., Torbjørn, E., Čiampor, F., Čiamporová Zaťovičová, Z., Costa, F.O., Duarte, S., Elbrecht, V., Fontaneto, D., Franc, A., Geiger, M.F., Hering, D., Kahlert, M., Stroil, B.K., Kelly, M., Keskin, E., Liska, I., Mergen, P., Meissner, K., Pawlowski, J., Penev, L., Reyjol, Y., Rotter, A., Steinke, D., van der Wal, B., Vitecek, S., Zimmermann, J., Weigand, A.M., 2018. Why we need sustainable networks bridging countries, disciplines, cultures and generations for aquatic biomonitoring 2.0: a perspective derived from the DNAqua-Net COST action. Adv. Ecol. Res. 58, 63–99. http://dx.doi.org/10.1016/bs.aecr.2018.01.001.

Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T.A., 2015. High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Sci. Rep. 5 (13932), http://dx.doi.org/10.1038/srep13932.

Leray, M., Knowlton, N., 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proc. Natl. Acad. Sci. 112 (7), 2076–2081. http://dx.doi.org/10.1073/pnas.1424997112.

Leray, M., Knowlton, N., 2016. Censusing marine eukaryotic diversity in the twenty-first century. Phil. Trans. R. Soc. B 371, 20150331. http://dx.doi.org/10.1098/rstb.2015.0331.

Lima, F.P., Ribeiro, P.A., Queiroz, N., Hawkins, S.J., Santos, A.M., 2007. Do distributional shifts of northern and southern species of algae match the warming pattern? Global Change Biol. 13 (12), 2592–2604. http://dx.doi.org/10.1111/j.1365-2486.2007.01451.x.

Lincoln, R., 1979. British Marine Amphipoda: Gammaridea. British Museum (Natural History), London, United Kigdom.

Lindeque, P.K., Parry, H.E., Harmer, R.A., Somerfield, P.J., Atkinson, A., 2013. Next generation sequencing reveals the hidden diversity of zooplankton assemblages. PloS One 8 (11), e81327. http://dx.doi.org/10.1371/journal.pone.0081327.

Lobo, J., Ferreira, M.S.G., Antunes, I.C., Teixeira, M.A.L., Borges, L.M., Sousa, R., Gomes, P.A., Costa, M.H., Cunha, M.R., Costa, F.O., 2017. Contrasting morphological and DNA barcode-suggested species boundaries among shallow-water amphipod fauna from the southern European Atlantic coast. Genome 60 (2), 147–157. http://dx.doi.org/10.1139/gen-2016-0009.

Lobo, J., Teixeira, M.A.L., Borges, L.M., Ferreira, M.S.G., Hollatz, C., Gomes, P.T., Sousa, R., Ravara, A., Costa, M.H., Costa, F.O., 2016. Starting a DNA barcode reference library for shallow water polychaetes from the southern European Atlantic coast. Mol. Ecol. Resour. 16 (1), 298–313. http://dx.doi.org/10.1111/1755-0998.12441.

McCarthy, J.B., Krug, P.J., Valdés, A., 2019. Integrative systematics of *Placida cremoniana* (Trinchese, 1892) (Gastropoda, Heterobranchia, Sacoglossa) reveals multiple pseudocryptic species. Mar. Biodivers. 49 (1), 357–371. http://dx.doi.org/10.1007/s12526-017-0812-2.

Miralles, L., Ardura, A., Arias, A., Borrell, Y.J., Clusa, L., Dopico, E., Rojas, A., Lopez, B., Muñoz Colmenero, M., Roca, A., Valiente, A.G., Zaiko, A., Garcia-Vazquez, E., 2016. Barcodes of marine invertebrates from north Iberian ports: Native diversity and resistance to biological invasions. Mar. Pollut. Bull. 112 (1–2), 183–188. http://dx.doi.org/10.1016/j.marpolbul.2016.08.022.

Morrow, C.C., Thorpe, J.P., Picton, B., 1992. Genetic divergence and cryptic speciation in two morphs of the common subtidal nudibranch *Doto coronata* (Opisthobranchia: Dendronotacea: Dotoidae) from the northern Irish Sea. Mar. Ecol. Prog. Ser. 84, 53–61. http://dx.doi.org/10.3354/meps084053.

Múrria, C., Väisänen, L.O., Somma, S., Wangensteen, O.S., Arnedo, M.A., Prat, N., 2019. Towards an Iberian DNA barcode reference library of freshwater macroinvertebrates and fishes. Limnetica 39 (1), 73–92. http://dx.doi.org/10.23818/limn.39.06.

Naylor, E., 1972. British Marine Isopods. Keys and Notes for the Identification of the Species– Synopses of the British Fauna. Academc Press, London-New York.

Nunes, A.L., Katsanevakis, S., Zenetos, A., Cardoso, A.C., 2014. Gateways to alien invasions in the European seas. Aquat. Invasions 9, 133–144. http://dx.doi.org/10.3391/ai.2014.9.2.02.

Oliveira, L.M., Knebelsberger, T., Landi, M., Soares, P., Raupach, M.J., Costa, F.O., 2016. Assembling and auditing a comprehensive DNA barcode reference library for European marine fishes. J. Fish Biol. 89 (6), 2741–2754. http://dx.doi.org/10.1111/jfb.13169.

Pascual, S., Villalba, A., Abollo, E., Garci, M., González, A.F., Nombela, M., Posada, D., Guerra, A., 2010. The mussel *Xenostrobus securis*: a well-established alien invader in the Ria de Vigo (Spain, NE Atlantic). Biol. Invasions 12 (7), 2091–2103. http://dx.doi.org/10.1007/s10530-009-9611-4.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M.J., Filipe, A.F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Jones, J.I., Sagova-Mareckova, M., Moritz, C., Barquín, J., Piggott, J.J., Pinna, M., Rimet, F., Rinkevich, B., Sousa-Santos, C., Specchia, V., Trobajo, R., Vasselon, V., Vitecek, S., Zimmerman, J., Weigand A. M. Leese, F., Kahlert, M., 2018. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. 637, 1295–1310. http://dx.doi.org/10.1016/j.scitotenv.2018.05.002.

Pearman, J.K., Leray, M., Villalobos, R., Machida, R.J., Berumen, M.L., Knowlton, K., Carvalho, S., 2018. Cross-shelf investigation of coral reef cryptic benthic organisms reveals diversity patterns of the hidden majority. Sci. Rep. 8, 8090. http://dx.doi.org/10.1038/s41598-018-26332-5.

Pejovic, I., Ardura, A., Miralles, L., Arias, A., Borrell, Y.J., Garcia-Vazquez, E., 2016. DNA barcoding for assessment of exotic molluscs associated with maritime ports in northern Iberia. Mar. Biol. Res. 122, 168–176. http://dx.doi.org/10.1080/17451000.2015.1112016.

Pereira, S.G., Lima, F.P., Queiroz, N.C., Ribeiro, P.A., Santos, A.M., 2006. Biogeographic patterns of intertidal macroinvertebrates and their association with macroalgae distribution along the Portuguese coast. Hydrobiologia 555 (1), 185. http://dx.doi.org/10.1007/s10750-005-1115-3.

Pola, M., Gosliner, T.M., 2010. The first molecular phylogeny of cladobranchian opisthobranchs (Mollusca, Gastropoda, Nudibranchia). Mol. Phylogenet. Evol. 56, 931–941. http://dx.doi.org/10.1016/j.ympev.2010.05.003.

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/. (Accessed 2020).

Ratnasingham, S., Hebert, P.D.N., 2007. BARCODING, BOLD: The Barcode of Life Data System. Mol. Ecol. Notes 7, 355–364. http://dx.doi.org/10.1111/j.1471-8286.2006.01678.x.

Ratnasingham, S., Hebert, P.D.N., 2013. DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS One 8, e66213. http://dx.doi.org/10.1371/journal.pone.0066213.

Raupach, M.J., Barco, A., Steinke, D., Beermann, J., Laakmann, S., Mohrbeck, I., Neumann, H., Kihara, T.C., Pointner, K., Radulovici, A., Segelken-Voigt, A., Wesse, C., Knebelsberger, T., 2015. The application of DNA barcodes for the identification of marine crustaceans from the North Sea and adjacent regions. PLoS One 10 (9), e0139421. http://dx.doi.org/10.1371/journal.pone.0139421.

Rey, A., Basurko, O.C., Rodriguez-Ezpeleta, N., 2019. Guidelines and considerations for metabarcoding-based port baseline biodiversity surveys: towards improved marine non-indigenous species monitoring. http://dx.doi.org/10.1101/689307, bioRxiv.

Rubal, M., Costa-Garcia, R., Besteiro, C., Sousa-Pinto, I., Veiga, P., 2018. Mollusc diversity associated with the non-indigenous macroalga *Asparagopsis armata* Harvey, 1855 along the Atlantic coast of the Iberian Peninsula. Mar. Environ. Res. 136, 1–7. http://dx.doi.org/10.1016/j.marenvres.2018.02.025.

Ruffo, S., 1982. The Amphipoda of the Mediterranean. Memoires de l'Institut Oceanographique, Monaco.

Shipman, C., Gosliner, T., 2015. Molecular and morphological systematics of Doto Oken, 1851 (Gastropoda: Heterobranchia), with descriptions of five new species and a new genus. Zootaxa 3973 (1), 57–101. http://dx.doi.org/10.11646/zootaxa.0000.0.0.

Siddall, M.E., Fontanella, F.M., Watson, S.C., Kvist, S., Erséus, C., 2009. Barcoding bamboozled by bacteria: convergence to metazoan mitochondrial primer targets by marine microbes. Syst. Biol. 58 (4), 445–451. http://dx.doi.org/10.1093/sysbio/syp033.

da Silva, J.M., Dos Santos, A., Cunha, M.R., Costa, F.O., Creer, S., Carvalho, G.R., 2013. Investigating the molecular systematic relationships amongst selected *Plesionika* (Decapoda: Pandalidae) from the Northeast Atlantic and Mediterranean Sea. Mar. Ecol. 34, 157–170. http://dx.doi.org/10.1111/j.1439-0485.2012.00530.

Spalding, M.D., Fox, H.E., Gerald, G.R., Davidson, N., Ferdaña, Z.A., Finlayson, M., Halper, B.S., Jorge, M.A., Lombana, A.I., Lourie, S.A., Martin, K.D., McManus, E., Molnar, J., Recchia, C.A., Robertson, J., 2007. Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. BioScience 57, 573–583. http://dx.doi.org/10.1641/B570707.

Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., Fontaneto, D., 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proc. Natl. Acad. Sci. 109 (40), 16208–16212. http://dx.doi.org/10.1073/pnas.1209160109.

Teixeira, M.A.L., Vieira, P.E., Pleijel, F., Sampieri, B.R., Ravara, A., Costa, F.O., Nygren, A., 2019. Molecular and morphometric analyses identify new lineages within a large *Eumida* (Annelida) species complex. Zool. Scripta 49, 222–235. http://dx.doi.org/10.1111/zsc.12397.

Trickey, J.S., Thiel, M., Waters, J.M., 2016. Transoceanic dispersal and cryptic diversity in a cosmopolitan rafting nudibranch. Invertebr. Syst. 30 (3), 290–301. http://dx.doi.org/10.1071/IS15052.

Veiga, P., Torres, A.C., Aneiros, F., Sousa-Pinto, I., Troncoso, J.S., Rubal, M., 2016. Consistent patterns of variation in macrobenthic assemblages and environmental variables over multiple spatial scales using taxonomic and functional approaches. Mar. Environ. Res. 120, 191–201. http://dx.doi.org/10.1016/j.marenvres.2016.08.011.

Viard, F., Roby, C., Turon, X., Bouchemousse, S., Bishop, J., 2019. Cryptic diversity and database errors challenge non-indigenous species surveys: an illustration with Botrylloides spp. in the English Channel and Mediterranean Sea. Front. Mar. Sci. 6, 615. http://dx.doi.org/10.3389/fmars.2019.00615.

Vieira, P.E., Desiderato, A., Holdich, D.M., Soares, P., Creer, S., Carvalho, G.R., Costa, F.O., Queiroga, H., 2019. Deep segregation in the open ocean: Macaronesia as an evolutionary hotspot for low dispersal marine invertebrates. Mol. Ecol. 28 (7), 1784–1800. http://dx.doi.org/10.1111/mec.15052.

Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M.F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A.M., Willassen, E., Wyler, S.A., Bouchez, A., Borja, A., Čiamporová Zaťovičová Z. Ferreira, S., Dijkstra, K.B., Eisendle, U., Freyhof, J., Gadawski, P., Graf, W., Haegerbaeumer, A., van der Hoorn, B.B., Japoshvili, B., Keresztes, L., Keskin, E., Lesse, F., Macher, J.N., Mamos, T., Paz, G., Pešić V. Pfannkuchen, D.M., Pfannkuchen, M.A., Price, B.W., Rinkevich, B., Teixeira, M.A.L., Várbíró, G., Ekrem, T., 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. http://dx.doi.org/10.1016/j.scitotenv.2019.04.247.