RESEARCH ARTICLE

Gaps in DNA sequence libraries for Macaronesian marine macroinvertebrates imply decades till completion and robust monitoring

Pedro E. Vieira[1,2] 🆔 | Ana S. Lavrador[1,2] | Manuela I. Parente[3] | Paola Parretti[3,4] | Ana C. Costa[3] | Filipe O. Costa[1,2] | Sofia Duarte[1,2]

[1]Centre of Molecular and Environmental Biology (CBMA), Department of Biology, University of Minho, Braga, Portugal

[2]Institute of Science and Innovation for Bio-Sustainability (IB-S), University of Minho, Braga, Portugal

[3]CIBIO Research Centre in Biodiversity and Genetic Resources, InBIO Associate Laboratory, University of Azores, Ponta Delgada, Portugal

[4]MARE – Marine and Environmental Sciences Centre, Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação (ARDITI), Edifício Madeira Tecnopolo, Funchal, Portugal

Correspondence
Pedro E. Vieira, Department of Biology, Centre of Molecular and Environmental Biology (CBMA), University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal.
Email: pedroefrvieira@gmail.com

Abstract

Aim: DNA metabarcoding has great potential to improve biomonitoring in island's marine ecosystems, which are highly vulnerable to global change and non-indigenous species (NIS) introductions. However, the depth and accuracy of the taxonomic identifications are mainly dependent on reference libraries containing representative and reliable sequences for the targeted species. In this study, we evaluated the gaps in the availability of DNA sequences and their accuracy for macroinvertebrates inhabiting Macaronesia's shallow marine habitats.

Location: Macaronesia (Azores, Madeira, Selvagens, Canaries).

Methods: Checklists of marine invertebrates occurring above 50 m depth were compiled using public databases and published checklists. The availability of cytochrome c oxidase subunit I (COI) and 18S rRNA (18S) gene sequences was verified in BOLD and GenBank. Finally, COI data were audited to check the congruence between morphospecies and Barcode Index Numbers (BINs).

Results: The taxonomic coverage of different phyla was greater for COI but unbalanced and variable among archipelagos. NIS were better represented in genetic databases (up to 73% and 59%, for COI and 18S, respectively) than native species (up to 47% and 31%, for COI and 18S, respectively). NIS displayed a higher number of discordant records, and native species, a higher quantity of cases of multiple BINs. Notably, DNA sequences generated from specimens collected from Macaronesia were found in less than 10% of the species. Projection of the rates of accretion of DNA sequences suggests that decades will be needed to complete these reference libraries.

Main conclusions: The level of completion of reference libraries for Macaronesia's marine macroinvertebrates is generally poor. Without a solid effort to speed up the production of sequence data (i.e. generate more DNA barcodes), the ability to employ DNA-based biomonitoring of such vulnerable fauna is compromised. The high levels

---

of suspected hidden diversity further deepen the expected gaps and reinforce the vulnerability of this endemism-rich fauna.

1 | INTRODUCTION

Although of great importance, the world's marine ecosystems and biodiversity are increasingly exposed to several threats driven by global change, over-use of natural resources, habitat loss, invasion by non-indigenous species (NIS), among other disturbances (Barbier, 2017; Molinos et al., 2016). This multitude of disturbances can severely impact ecosystems around the world, instigating the need for the densification and greater accuracy in biodiversity assessments and monitoring across the planet (Borja et al., 2020; Cardinale et al., 2012; Pereira et al., 2010).

Islands, which contain one-fifth of the world's biodiversity and a profusion of endemic species, are among the most threatened ecosystems (Kier et al., 2009; Lagabrielle et al., 2009). Endemic species often exhibit comparatively small population sizes with limited geographical distribution and habitat availability, making islands biodiversity highly vulnerable to global change, particularly to the introduction of NIS (Vitousek, 1990). Macaronesia is a group of volcanic islands composed of five archipelagos (Azores, Madeira, Selvagens, Canaries and Cape Verde), located in the Northeast Atlantic Ocean (NEA), which was established based on their flora and fauna similarities (Fernández-Palacios et al., 2011). Macaronesia has a unique and rich biodiversity and is the geographical boundary of many species, in both the terrestrial (Arechavaleta et al., 2010; Borges et al., 2008, 2010) and marine realms (Borges et al., 2010). The conservation of this valuable diversity is complex, and protection programmes have been already in place in some areas (e.g. under Natura 2000). However, further research and extensive biomonitoring programmes are still needed to assess more accurately which species are threatened and to provide a more holistic view of these ecosystems' present and changing status (Cacabelos et al., 2020; Iacarella et al., 2020). Thus, the strategic expansion of the network of protected areas and the effective allocation of resources for conservation is highly dependent on accurate and recurrent biodiversity assessments.

As above-mentioned, one of the major threats to native islands biodiversity is the introduction of NIS, and the Macaronesia archipelagos are not an exception (Arechavaleta et al., 2010; Borges et al., 2008; Moro et al., 2003). When introduced to new areas, NIS can spread rapidly and become invasive, modify habitats, compete with native fauna for resources and threaten biodiversity (Bax et al., 2003; Rilov & Crooks, 2009). Their introduction, namely ships/vessels, canals and aquaculture activities, can provoke severe

ecological, social and economic impacts (Diagne et al., 2020; Lenzner et al., 2020; Rilov & Crooks, 2009; Seebens et al., 2013; Wallentinus & Nyberg, 2007). Worldwide, it is predicted that NIS will increase one-third until 2050, with strong rises projected for Europe (Seebens et al., 2020). To determine the state of NIS introductions and their impact on ecosystems, and to implement measures to prevent biodiversity loss, data compilation is mandatory, and to that end, regulations were created. Most of the marine waters of the NEA ocean fall under the jurisdiction of the European Union and its member states, including those surrounding the Azores, Madeira, Selvagens and Canary Islands, and thus, they are targeted by the European Marine Strategy Framework Directive (EU-MSFD) (European Commission, 2008; Tsiamis et al., 2019). This directive includes assessment of NIS occurrence and has led to an increase in monitoring programmes and inventories in Europe over the last decades (e.g. Afonso et al., 2020; Chainho et al., 2015; Micael et al., 2014; Tsiamis et al., 2019).

Until recently, biodiversity assessments have been conducted almost exclusively through morphology-based species identifications. However, this approach has several drawbacks, being highly expertise-demanding and time-consuming, and delivering lower taxonomic resolution (Hering et al., 2018; Leese et al., 2016, 2018). With the exponential rise in the power of both DNA sequencers and computational technology, molecular techniques constitute an effective alternative or complement to morphology-based identifications, in particular DNA barcoding (short standardized DNA sequences amplified from a single specimen and used for species identification) and DNA metabarcoding. In the latter, DNA is extracted from bulk organismal samples or directly from the environmental sample matrix such as seawater or sediment (in this case, designated as "environmental DNA" or eDNA). Subsequently, amplicon libraries for target gene regions are generated, high-throughput-sequenced and compared to reference sequences to deliver a taxonomic identification (Duarte, Leite, et al., 2021; Fais et al., 2020; Leese et al., 2016; Steyaert et al., 2020). DNA metabarcoding offers potential benefits over morphological assessments, such as (i) increased sensitivity, (ii) discrimination of cryptic species, (iii) identification of species regardless of the life stage (e.g. eggs, larvae), (iv) enables assessments covering a wide range of taxa, and (v) allows high-throughput assessments leading to a higher spatial–temporal density of taxa occurrence data (Holman et al., 2019; Leduc et al., 2019; Schroeder et al., 2020; Suarez-Menendez et al., 2020). The taxonomic composition of hundreds of samples

can be assessed quickly and at a relatively low cost, facilitating the implementation of more extensive monitoring programmes and providing a more comprehensive view of the present and changing status of island ecosystems. This fast and reliable approach can be highly efficient for the early detection of NIS (Schroeder et al., 2020; Zaiko et al., 2015). Major caveats still include the inability to quantify species abundances and distinguish between life stages (Duarte, Vieira, et al., 2021).

The depth and accuracy of DNA metabarcoding-based identifications are mainly dependent on the availability of reference libraries containing representative and accurate sequences for the targeted species. The existence of gaps and unequal representation of taxonomic groups in reference databases may compromise the accuracy of the DNA-based biodiversity assessments (Ardura, 2019; Duarte et al., 2020; Leite et al., 2020; Weigand et al., 2019). Thus, assessing these gaps and the quality of sequence data in reference databases is mandatory for the successful implementation of DNA-based tools in biodiversity assessments.

In this study, we evaluated the gaps in the availability of DNA sequence data and their accuracy to assess macroinvertebrate diversity through DNA-based tools in Macaronesia's shallow marine habitats. As the DNA barcode region (cytochrome c oxidase subunit I—COI) and the gene encoding the nuclear 18S rRNA (18S) have been the most widely used genetic markers in metabarcoding studies targeting marine invertebrates, including NIS (Duarte, Leite, et al., 2021; Duarte, Vieira, et al., 2021), the sequence availability was assessed for both. The incorporation of cutting-edge biomonitoring tools is essential for efficient management of islands biodiversity and to develop mitigation strategies to deal with increasing environmental change in these highly vulnerable ecosystems.

2 | METHODS

2.1 | Checklist compilation

The four European Macaronesian archipelagos were used in this study: Azores, Madeira, Selvagens and Canaries (Figure 1). As recent studies based on marine biota suggest that Cape Verde's community structure and biogeographic relationships differ significantly from the remaining Macaronesian islands (Cunha et al., 2014; Freitas et al., 2019; Wirtz et al., 2013), we opted not to include this archipelago in our analysis. A list of native species (marine invertebrates) was compiled (14 May 2020) based on GBIF (gbif.org, 2020) and WoRMS (WoRMS Editorial Board, 2020) databases.

The following search criteria were applied for retrieving the species lists from GBIF: only Animalia, without doubtful records, only taxa identified to species level and species recorded between 0 and 50 m depth. Data from the Azores (https://doi.org/10.15468/dl.ckd9sw) were mined selecting the polygon composed by the coordinates (long lat: −31.51977 36.55884, −24.62036 36.55884, −24.62036 39.97192, −31.51977 39.97192, −31.51977 36.55884),

from Canaries (https://doi.org/10.15468/dl.ubajg9) within the polygon (long lat: −18.37463 27.42004, −13.29163 27.42004, −13.29163 29.58069, −18.37463 29.58069, −18.37463 27.42004), from Madeira (https://doi.org/10.15468/dl.z58kpz) within the polygon (long lat: −17.34146 32.35245, −16.22818 32.35245, −16.22818 33.15445, −17.34146 33.15445, −17.34146 32.35245) and from Selvagens (https://doi.org/10.15468/dl.dab8eb) within the polygon (long lat: −16.13777 29.87343, −15.77614 29.87343, −15.77614 30.22865, −16.13777 30.22865, −16.13777 29.87343). Data were mined from WoRMS using the geounits for the Azores: "Azores (Archipelago)," for the Canaries: "Canaries (Archipelago)," for Madeira: "Madeira (island)" plus "Porto Santo(island)" and for Selvagens: "Selvagens (Archipelago)." The search was limited to marine and extant animal species and accepting only valid names. Species with the annotation "alien" were added to the list of NIS; see below. Then, the lists obtained from GBIF and WoRMS were merged, duplicate entries were removed, and only marine invertebrates were accepted (taxonomy was confirmed on WoRMS).

The list of NIS (marine invertebrates) was compiled (14 May 2020) for each archipelago based on multiple sources: WoRMS (species with "alien" designation), the Information System on Aquatic Non-indigenous and Cryptogenic species (AquaNIS—http://www.corpi.ku.lt/databases/index.php/aquanis; Olenin et al., 2014), Castro et al. (2020), Chainho et al. (2015), ICES (2019), MITECO database (Ministerio para la Transición Ecológica y el Reto Demográfico, Catálogo Español de Especies Exóticas Invasoras) and Ramalhosa et al. (2019) (https://nutelapat.wixsite.com/canningclode/copia-madeira-nis-2 and references therein). Lists were merged, duplicate entries were removed, and only marine invertebrates were kept (taxonomic validity was confirmed in WoRMS). Depth of occurrence was verified on GBIF, and all species occurring only below 50 m depth were discarded.

2.2 | Genetic data mining and analyses

For each list (native and NIS), COI and 18S genetic data were mined from BOLD (Ratnasingham & Hebert, 2007) and GenBank (Sayers et al., 2019) using the R 3.6.0 software (R Core Team, 2019; www.r-project.org) with the package "bold" (Chamberlain, 2019) and "rentrez" (Winter, 2017), respectively. In BOLD, the following terms were used to filter the sequences: for COI—"COI-5P"; for 18S—"18S" or "18Sa". In GenBank, the terms used were as follows: for COI—"COI[Gene] OR CO1[Gene] OR COXI[Gene] OR COX1[Gene] OR complete genome [All Fields] OR mitochondrial genome[All Fields]"; for 18S—"18S ribosomal RNA[Title] OR 18S rRNA[Title] OR 18S small subunit ribosomal RNA[Title] OR 18S ribosomal RNA[Gene] OR 18S rRNA[Gene] OR 18S small subunit ribosomal RNA[Gene]". Only sequences with more than 500 base pairs were considered as this is the minimum length required for a sequence to meet Barcode Compliance standards (Ratnasingham & Hebert, 2007) and which has also been used in other gap analysis studies of European aquatic invertebrates (Duarte et al., 2020; Leite et al., 2020; Weigand
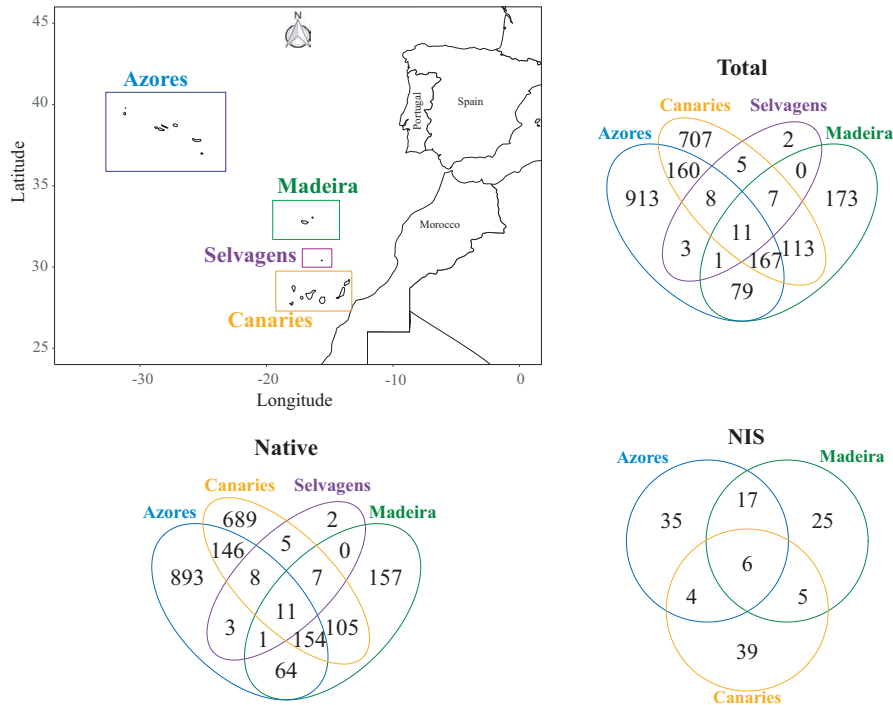
FIGURE 1 Top left: Map of Macaronesian archipelagos studied: Azores, Madeira, Selvagens and Canaries. Top right and bottom: Venn diagrams showing the shared and unique total, native and non-indigenous species (NIS) among all Macaronesia archipelagos

et al., 2019). To avoid duplicate records, GenBank records were detected in BOLD through the presence of the tag "Mined from GenBank, NCBI" or if the record had a GenBank's accession number. All GenBank accession numbers detected in BOLD were then manually confirmed on GenBank to double-check the duplicated status. The geographic origin of specimens and year of submission of the sequences were verified through BOLD metadata.

The number of Barcode Index Numbers (BINS) (Ratnasingham & Hebert, 2013) for each species within each taxon was retrieved from BOLD, based on the COI marker. Then, to verify the reliability of the genetic data for each species, the auditing and grading software BAGS was used (Fontes et al., 2020; https://github.com/tadeu95/BAGS). This tool relies on COI, the BIN system and the number of records to annotate and grade species according to the quality of their available public sequences. Grade A and grade B are considered concordant (one species = one BIN), grade C indicate multiple BINs for a given species (one species = two or more BINs), grade D is insufficient data (less than three records), and grade E indicates discordances, that is more than one species is assigned to a single BIN (two or more species = one BIN). More details can be found in Fontes et al. (2020). All the scripts used in this study can be consulted at https://github.com/pedroemanuelvieira/NGB_Macaronesia.

3 | RESULTS

3.1 | Taxonomic composition

Final lists consisted of 1,342, 550, 1,177 and 36 species for Azores, Madeira, Canaries and Selvagens (Figure 1), comprising 13, 11, 15 and 5 phyla, respectively (Figure 2; Table S1). Madeira was the archipelago

with the highest percentage of NIS (9.5% of the total number of species), while Azores and Canaries displayed similar percentages (4.6 and 4.5%, respectively). No NIS were reported to occur in Selvagens (Table S1).

Mollusca was the most well-represented phyla in the Azores, Madeira and Canaries (31.9% 63.0% of the total number of species) (Figure 2; Table S1). Other dominant phyla in these regions included Arthropoda (12.1% to 31.0%) and Cnidaria (8.4% to 15.3%). Selvagens comprised distinct phyla, dominated by Cnidaria (52.8%), Mollusca (22.2%) and Brachiopoda (13.9%) (Figure 2; Table S1). Dominant phyla containing NIS included Chordata in the Azores (25.8% of the number of NIS) and Bryozoa (up to 39.6%) in Madeira and Canaries (Figure 2; Table S1).

A total number of 11 species were shared among all archipelagos (Figure 1). The highest number of shared species was found between the Azores and Canaries (346), although Selvagens is the archipelago sharing the highest proportion of species with the other regions (between 19 and 31 of a total of 36 species). The Azores (913) and Canaries (708) harboured the highest number of unique species (i.e. not present in other archipelagos). Six NIS (*Amathia verticillata* (delle Chiaje, 1822), *Botryllus schlosseri* (Pallas, 1766), *Bugula neritina* (Linnaeus, 1758), *Bugulina stolonifera* (Ryland, 1960), *Caprella scaura* (Templeton, 1836) *and Spirorbis (Spirorbis) marioni* Caullery & Mesnil, 1897) were found in the Azores, Madeira and Canaries. While the Azores had the highest number of NIS (62), Canaries was the archipelago with the highest number of unique NIS (39) (Figure 1). Thirteen Azorean NIS are native in the other regions, mostly in Canaries (10) (Figure S1). Of the 54 NIS present in the Canaries, 10 are native in the other archipelagos (most from the Azores: 7), while of the 53 NIS present in Madeira, 12 are native in the other archipelagos (10 from the Canaries) (Figure S1).

FIGURE 2 Marine invertebrate's taxonomic composition by native and by non-indigenous species (NIS) occurring in each archipelago. The numbers in the top of each bar indicate the total number of species per phylum. Chordata only includes Ascidians
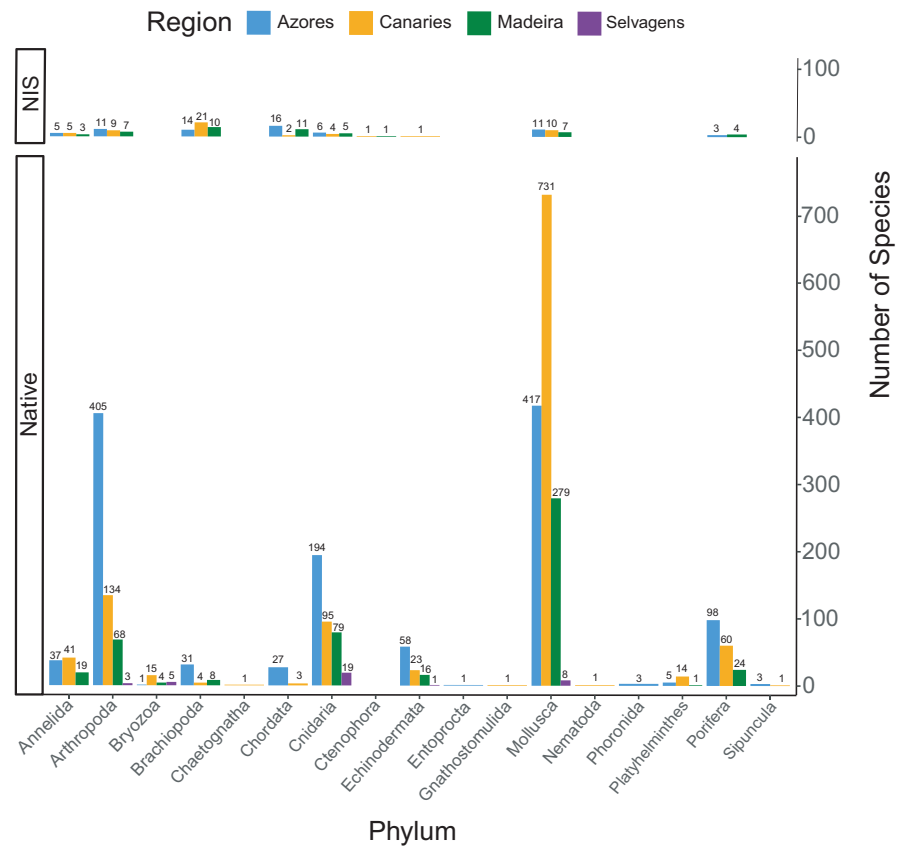
3.2 | Gap analysis and grading system

More records were found on GenBank than on BOLD, with more sequences available for COI than 18S (Tables 1, S2, S3). When merging the information of both databases and excluding duplicated records, the Azores was the archipelago with the highest number of native (COI: 46.8% and 18S: 30.6%) and NIS represented in the genetic databases (COI: 72.6% and 18S: 54.8%), while Madeira displayed the lowest, for both native (COI: 39.8% and 18S: 26.5%) and NIS (COI: 61.5% and 18S: 40.4%). In general, NIS displayed higher percentages (65.4%–79.0%) of at least one of the genetic markers (COI or 18S) when compared with native species (38.1%–51.6%). However, the percentage of species with both genetic markers was much lower for native (17.8%–25.8%) and NIS (36.5%–52.8%).

In the Azores, Phoronida was the phylum with the highest coverage for COI (100.0%), followed by Arthropoda (61.5%) and Annelida (57.1%), while for 18S, Entoprocta and Brachiopoda were the most well-represented with sequences (both with 100.0%). In the Canaries, Ctenophora and Chordata had full coverage for both markers (100.0%), with 100.0% of Sipuncula species having COI sequences, but none with 18S. On the other hand, Gnathostomulida species are fully covered for 18S, but none is represented in genetic databases for COI. In Madeira, Platyhelminthes and Ctenophora had 100% coverage for both markers. In Selvagens, Echinodermata and Arthropoda species had 100.0% coverage for COI, and Arthropoda and Cnidaria had 66.7% and 42.1% for 18S, respectively (Figure 3).

When considering only COI sequences generated from specimens collected in any of the four Macaronesian archipelagos here studied (according to BOLD; Figure S2), the availability of sequences on genetic databases decreased drastically. For instance, only 64 of the 1,280 native species of the Azores met this criterion. These values were much lower for the other archipelagos, with only 19 out of 1,124 native species from the Canaries, 17 out of 498 native species from Madeira and five out of 37 native species from Selvagens. Among NIS, only *Phorcus sauciatus* (Koch, 1845) had records generated from specimens collected in Macaronesia, although these records were from Madeira and Canaries where is considered a native species. *Percnon gibbesi* (H. Milne Edwards, 1853), which has been considered a NIS in the Azores, despite the present uncertainty of its current status, also has COI sequences in genetic databases obtained from specimens collected in the region.

Considering all archipelagos, the grading system BAGS classified a higher percentage of NIS (~50%) as discordant (grade E) when compared with native species (~41%) (Figure 4). In general, around one-quarter of the species had insufficient records (grade D), and less than 10% were concordant species (grades A and B). More cases of Multiple BINs (grade C) were detected in native species (between 12.6% and 36.4%) than in NIS (between 7.7% and 18.5%) (Figure 4). When excluding discordant and insufficient cases (grades D and E), for the native species, all phyla displayed more BINs than species (except Porifera), which for NIS was only observed in half of the phyla (Figure S3). For native species, Arthropoda (228) and Mollusca (128) displayed the highest number of

TABLE 1 Overall DNA sequence coverage for COI and 18S genetic markers on GenBank and BOLD for native and non-indigenous species (NIS) occurring on each Macaronesian archipelago

| Archipelago | | No. of species | Marker | No. of records | | No. of species with DNA sequences | | |
				GenBank	BOLD	GenBank (% of coverage)	BOLD (% of coverage)	GenBank +BOLD (% of coverage)
Azores	Native	1,280	COI	14,470	12,370	559 (43.7)	559 (43.7)	599 (46.8)
			18S	1,551	95	385 (30.1)	10 (0.8)	391 (30.6)
			COI+18S			316 (24.7)	10 (0.8)	330 (25.8)
			At least one marker			628 (49.1)	559 (43.7)	660 (51.6)
	NIS	62	COI	3,174	2,845	44 (71.0)	41 (66.1)	45 (72.6)
			18S	155	0	34 (54.8)	0 (0.0)	34 (54.8)
			COI+18S			29 (46.8)	0 (0.0)	30 (48.4)
			At least one marker			49 (79.0)	41 (66.1)	49 (79.0)
Canaries	Native	1,124	COI	11,229	10,682	353 (31.4)	349 (31.1)	384 (34.2)
			18S	853	151	235 (20.9)	9 (0.8)	244 (21.7)
			COI+18S			187 (16.6)	9 (0.8)	200 (17.8)
			At least one marker			401 (35.7)	349 (31.0)	428 (38.1)
	NIS	53	COI	3,809	3,095	34 (64.2)	29 (54.7)	34 (64.2)
			18S	174	0	31 (58.5)	0 (0.0)	31 (58.5)
			COI+18S			28 (52.8)	0 (0.0)	28 (52.8)
			At least one marker			37 (69.8)	29 (54.7)	37 (69.8)
Madeira	Native	498	COI	5,310	4,954	181 (36.4)	182 (36.6)	198 (39.8)
			18S	508	44	130 (26.1)	5 (1.0)	132 (26.5)
			COI+18S			108 (21.7)	5 (1.0)	111 (22.3)
			At least one marker			203 (40.8)	182 (36.6)	219 (44.0)
	NIS	52	COI	2,089	1,839	31 (59.6)	28 (53.9)	32 (61.5)
			18S	141	0	21 (40.4)	0 (0.0)	21 (40.4)
			COI+18S			18 (34.6)	0 (0.0)	19 (36.5)
			At least one marker			34 (65.4)	28 (53.9)	34 (65.4)
Selvagens	Native	36	COI	402	421	13 (36.1)	13 (36.1)	15 (41.7)
			18S	151	0	11 (30.6)	0 (0.0)	11 (30.6)
			COI+18S			8 (22.2)	0 (0.0)	9 (25.0)
			At least one marker			16 (44.4)	13 (36.1)	17 (47.2)
	NIS	0	COI	–	–	–	–	–
			18S	–	–	–	–	–
			COI+18S	–	–	–	–	–
			At least one marker	–	–	–	–	–

BINs, but Cnidaria had the highest ratio BIN/species, with five times more BINs than species. For NIS, Chordata displayed the highest number of BINs (9), with almost three times the number of BINs per species. More than 40% of the native species displayed two or more BINs, with nine species displaying six or more BINs, while one-third of NIS were single BINs, with only two species having more than two BINs (Figure S3).

Taking into consideration the first time (year) a native species had COI sequences publicly available on BOLD and based on a
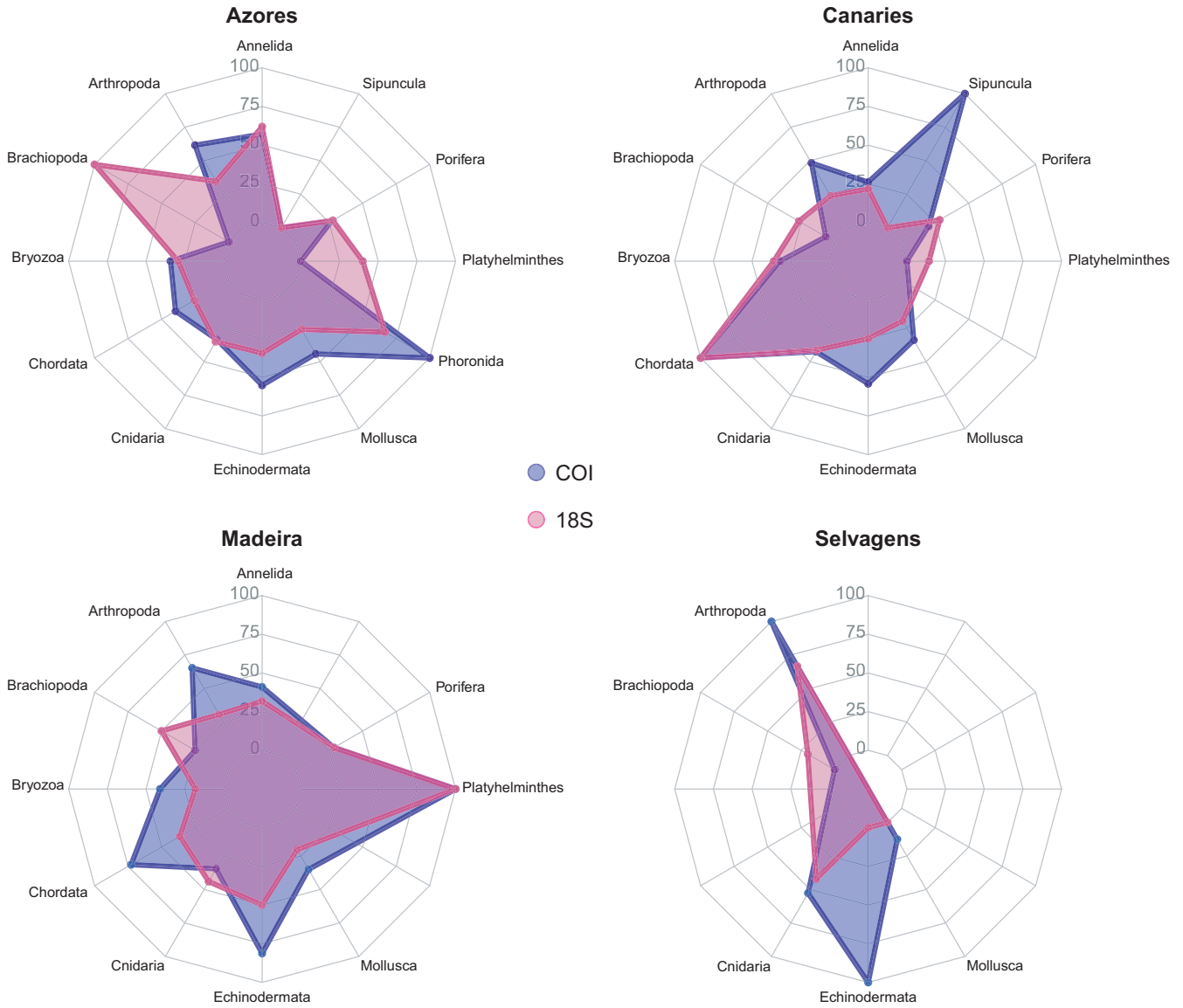
FIGURE 3 Percentage of coverage of COI and 18S for each phylum within each archipelago. Phyla with only one species were not included. To see details about these, see Tables S1 and S2. Chordata only includes Ascidians
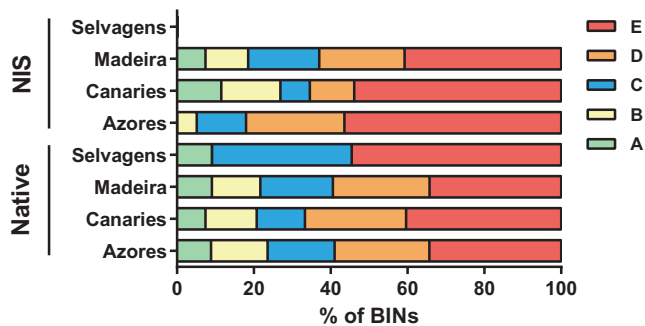

FIGURE 4 Percentage of audited and annotated species using BAGS (Fontes et al., 2020). Grades A and B—concordant species; grade C—multiple BINs; grade D—insufficient records; and grade E—discordant BINs

cumulative percentage per year (between 1.9% and 2.9% per year, that is ~0.8 to 37 new species per year), all the archipelagos still have less than half of the species with COI sequences. Assuming the constant rate of accretion of sequence records for missing species, the projection until completion (100%) of the reference libraries indicates that the Azores may be the first archipelago to achieve this goal (around 2040), followed by Madeira and Selvagens (around 2050), with the Canaries only reaching it after 2050 (Figure 5).

4 | DISCUSSION

As a result of this study, four main findings can be pointed out: (a) reference DNA sequence libraries are still highly incomplete for
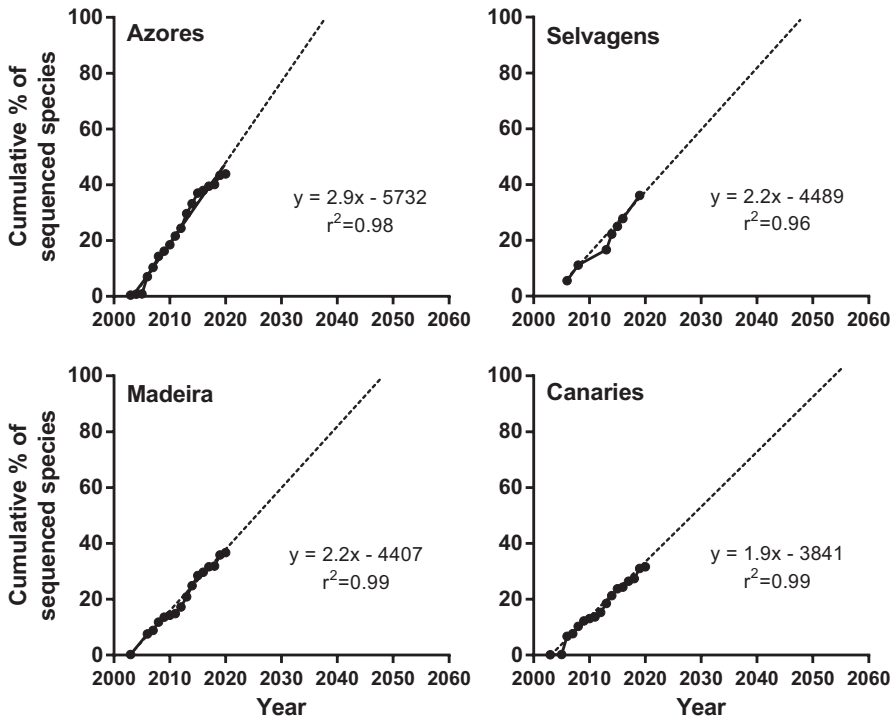
FIGURE 5 Cumulative percentage of native species with available COI sequences per year and projection until completion assuming constant rate of accretion of sequence records for missing species

Macaronesian marine macroinvertebrates; (b) non-indigenous species have higher levels of sequence completion than native species, but higher numbers of discordant records, (c) a high proportion of native morphospecies are assigned to multiple BINs, suggesting considerable hidden or cryptic diversity, which, if confirmed, can only be accurately monitored through DNA-based tools, and (d) extrapolation of the rates of accretion of DNA sequences in genetic databases suggests that some decades will be needed to complete the reference libraries for marine macroinvertebrates.

Despite the contribution of several studies to complete Macaronesia's macrozoobenthos reference libraries (Borges et al., 2016; Gargan et al., 2017; Gomes, 2014; Luz & Keskin, 2019; Silva et al., 2011; Valdés, 2017), we found that their taxonomic coverage is still incipient compared with the diversity of the region. Besides, we found that different archipelagos and taxonomic groups display different degrees of completeness (Figure 3; Tables S2, S3). Despite having more species than other archipelagos (Figure 2), the Azores had the highest percentage of species sequenced—47% for native and 73% for NIS—and Madeira the lowest—40% for native and 62% for NIS, respectively. Consistently, a higher number of records were found on GenBank than on BOLD. However, although GenBank contains reference sequences from many different genetic markers and includes all domains of life, it is more prone to errors than BOLD as it contains many non-curated data entries (López-Escardó et al., 2018). NIS displayed higher levels of completion in all archipelagos (Table 1). These species are generally the focus of a greater number of studies due to their high impact on the environment and thus may experience a higher trend of sequence deposition in genetic databases (Briski et al., 2016; Pyšek et al., 2008; Trebitz et al., 2015). The number of NIS is also much smaller than native species; therefore, levels of completion are typically higher (Duarte et al., 2020; Weigand et al., 2019).

A noteworthy observation when considering sequence records from specimens collected in Macaronesia is that the percentage of COI barcoded native species drops to less than 10% (Canaries and Madeira have less than 2%) (Figure S2). Ideally, these reference libraries should include specimens collected in Macaronesia (and from different Macaronesian islands and archipelagos) to allow the detection of possible highly divergent lineages or endemic cryptic taxa that require DNA-based tools to be recognized. Indeed, such cases have already been reported in previous studies of peracarid crustaceans in the region (Desiderato et al., 2019; Vieira, Desiderato, Holdich, et al., 2019). This may be important to assess small-scale variation of populations or newly introduced populations. This reasoning was supported by a quick search into the Web of Science (25 March 2021), which yielded only 4 published research papers dealing with DNA barcoding of Macaronesian marine invertebrate fauna. Furthermore, a look into the MarBOL (Marine Life) campaign within the BOLD database indicates no dedicated projects to DNA barcoding of Macaronesia macroinvertebrates, and very few projects dedicated to marine invertebrate fauna from the Northeast Atlantic, that may include Macaronesian specimens (e.g. BNAGB, Barcoding Northeast Atlantic Gastropods and Bivalves; BNEAC, Barcoding Northeast Atlantic Cephalopoda).

DNA-based biodiversity assessments in NEA have been limited by poor taxonomic coverage of genetic databases (Hestetun et al., 2020). These limitations are transversal across Europe (Duarte et al., 2020; Leite et al., 2020; Weigand et al., 2019), which led to the creation of national (Price et al., 2020) and international initiatives aiming to fill the reference libraries for aquatic biota (Leese et al., 2016, 2018). Nearby coasts that share many species with Macaronesia, such as the Iberian Peninsula, still have 60% of the species lacking COI barcodes (based on BOLD only), among Annelida, Crustacea and Mollusca (Leite et al., 2020). Although this value

is only slightly lower than the one here reported for Macaronesia (globally 63% merging BOLD and GenBank data), we must keep in mind that most of the sequenced specimens were not collected in Macaronesia. Because several studies indicate the occurrence of highly divergent lineages in Macaronesia, to the point of segregating in separate endemic lineages (Desiderato et al., 2019; Vieira, Desiderato, Holdich, et al., 2019; Xavier et al., 2010), various morphospecies may skip DNA-based detection even if they are present in reference libraries. Considering this possibility, we suspect that these completion levels for Macaronesia may be somewhat overestimated, though it is still unknown how much. Therefore, reference libraries must include specimens collected locally.

We also found significant differences between COI and 18S completeness (Tables 1, S2, S3). Despite many gaps in the COI library, 18S still falls behind, and more work should be conducted in populating other non-COI reference libraries. If only species sequenced for both markers are considered, these values decrease noticeably (Table 1). This may be a relevant limitation to efficiently detect some species as several studies suggest that some taxonomic groups are preferentially detected by different markers and primers (Grey et al., 2018; Lacoursière-Roussel et al., 2018; Leduc et al., 2019; Leite et al., 2019). Considering this, it has been argued that DNA metabarcoding, either to detect native species or NIS, should rely on more than one genetic region to assure detection of the widest possible spectrum of taxa (Duarte, Leite, et al., 2021; Stat et al., 2017).

More than one-third of the species still display discordant records, with higher percentages in NIS than native species (Figure 4). Incongruencies should be carefully examined to detect the sources of conflict (e.g. misidentifications, incomplete taxonomy or sequences that were deposited under different synonyms) and subsequently curated, so that DNA-based tools can reliably identify these species in bulk or environmental samples. Discordant records raise mistrust because erroneous observations derived from them may easily remain undetected through unsupervised taxonomic assignments of metabarcoding data and quickly propagate across studies. As such, quality control and quality assurance tools must be implemented to audit and curate reference libraries (Fontes et al., 2020; Leese et al., 2018; Weigand et al., 2019), as the reliability of the reference sequences is as essential as their availability, or even more.

When considering only concordant species records assigned to multiple BINs, approximately 20% of native and 10% of NIS fell under this condition (Figure 4). From a taxonomic perspective, specific phyla displayed one to five times more BINs than barcoded species (Figure S3). Therefore, it appears that a very high proportion of species from Macaronesia may incorporate undescribed or cryptic diversity. Indeed, several recent studies report the high incidence of deeply divergent endemic lineages in Macaronesia (Tavares et al., 2017; Vieira, Desiderato, Azevedo, et al., 2019; Xavier et al., 2010). Most of these highly divergent lineages have restricted distributions, frequently even limited to a single island, which makes them potentially more susceptible to global change and NIS impacts, thereby constituting a prime target for conservation measures. DNA-based approaches detect molecular entities (molecular operational taxonomic units—MOTUs), and it is important to connect the different MOTUs to their occurrence in each island/archipelago, as some may be endangered lineages or endemic species, which may only be diagnosed through DNA-based methods. Hence, it becomes imperative to generate more sequence records of specimens collected in the Macaronesia archipelagos.

Although the number of species with COI sequences available on BOLD has been increasing in the last twenty years, so far, only less than half of the native species present in these Macaronesian archipelagos have sequences available. Excluding species discovery or extinction, we estimate it will take another twenty to thirty years to exhaustively complete the reference libraries of DNA barcodes for the species present in these islands, if the rate of production of COI sequences is sustained. However, as rarer species may be harder to find, these projections are probably the best-case scenario as they do not consider the expected difficulties in the access to specimens of rarer species. Moreover, these projections do not contemplate the predicted growth of NIS introductions due to the increase in maritime traffic and the absence of legislation to prevent the involuntary transport of these species in hull fouling. More likely, it will take even longer to complete the DNA barcode libraries of all marine invertebrates present in Macaronesia (Figure 5). Many studies based on DNA metabarcoding of marine taxa may also contribute to generate sequences that can potentially match species still unavailable in the libraries, but that will remain as unknown until matching sequences are finally deposited in reference databases.

In what concerns the coastal area and the number of islands, the Azores and Canaries are the most extensive archipelagos and held the highest number of NIS and native species of marine macroinvertebrate fauna compiled in the current study (Figure 1). Furthermore, the Azores and Madeira shared the highest number of NIS, but a minimal number of species (only six) are considered NIS in all archipelagos. However, we should keep in mind that species with the non-indigenous status in one archipelago may be native to the others and vice versa, and the species status (i.e. indigenous, non-indigenous, cryptogenic) may change over the years as more knowledge is acquired. For instance, *Percnon gibbesi* has been previously considered a NIS in the Azores, but its status remains uncertain. Several NIS recorded in the Azores and Madeira are considered native species in the Canaries, whereas some NIS occurring in the Canaries are native in the Azores (Figure S1).

Madeira displayed the highest % of NIS (NIS/total number of species ratio), particularly in Arthropoda and Mollusca, which is also supported by recent data found in the literature that considers this archipelago highly impacted by bioinvasions (Bailey et al., 2020). However, we cannot discard the possibility that the highest % found in this region can be biased by the greater effort employed in conducting NIS-focused studies in Madeira (Canning-Clode et al., 2013; Parretti et al., 2020; Ramalhosa et al., 2014, 2019). To our best knowledge, no NIS were reported in Selvagens. Being a tiny remote archipelago of difficult access and with no permanent human population, it is probably less susceptible to NIS introductions, but, for the same reasons, an updated assessment of NIS may also be more challenging to accomplish.

5 | CONCLUSIONS

Unless more intense efforts are made to obtain representative DNA sequences for the macroinvertebrate species occurring in Macaronesia, either native or NIS, DNA-based monitoring will be deficient or impractical for this region's invertebrate fauna. Furthermore, an efficient curation of the available DNA sequences must be taken into consideration as, according to our results, only one-fifth of the species had concordant records, and less than 10% of the records were from specimens sampled in Macaronesia. Although completing the gaps in reference libraries is essential to make the most of the DNA-based tools, a careful compilation, verification and annotation of available sequences are fundamental to assemble large curated and reliable reference libraries that provide support for rigorous species identifications. Conventional morphology-based tools used in biomonitoring continue to provide invaluable information about the status of the populations. However, DNA metabarcoding can provide high data density over space and time, impart a unique diagnosis tool for cryptic taxa and ultimately be more responsive to environmental management needs while also enabling the early detection of NIS. Santos et al. (2016) "advocate a continuing effort to build comprehensive island data for multiple taxa, to serve the wider scientific community in the coming decades." We extend this plea, as current rates of accretion of reference DNA sequence data for Macaronesia are too slow to materialize the benefits of DNA-based monitoring for enhancing biodiversity conservation efforts in this region.

By our predictions, completeness will be accomplished only after 2040, considering the current rate of accretion of 1.9%–2.9% per year. Researchers must, at least, triple the current efforts if this goal is to be achieved in the next decade. To this end, initiatives such as "BIOSCAN" (Hobern, 2020), which involves more than 1,000 researchers from over 30 countries and aims to generate barcode coverage for 2.5 million species, may be decisive to fill up the gaps across the planet.

Robust monitoring will allow a more comprehensive view of the status of the island populations, helping to mitigate the ongoing pressures (e.g. climate change, fisheries) these populations experience and, therefore, contributing to preserve the invaluable ecosystem services these islands provide. If this goal cannot be reached due to lack of taxonomic expertise, sampling bottlenecks (e.g. inability to get specimens from rare species) and the high levels of cryptic and endemic diversity that are expected, other approaches based on reverse taxonomy, MOTUs/BINs or taxonomy-free methods (Cordier et al., 2017, 2018; Ratnasingham & Hebert, 2013; Weigand et al., 2019) may be an option, although far from ideal. MOTUs/BINs can be used provisionally and associated with an identification to the lowest possible rank, but always with the final goal of eventually reaching a true identification and recognition of species. If the intention is to use DNA-based tools to detect non-indigenous species, then identifications at the species level are mandatory, and consequently, populating reference libraries with DNA barcodes becomes paramount.

ACKNOWLEDGEMENTS

PEER REVIEW

DATA AVAILABILITY STATEMENT

ORCID

Pedro E. Vieira https://orcid.org/0000-0003-4880-3323

REFERENCES

Afonso, I., Berecibar, E., Castro, N., Costa, J. L., Frias, P., Henriques, F., Moreira, P., Oliveira, P. M., Silva, G., & Chainho, P. (2020). Assessment of the colonization and dispersal success of non-indigenous species introduced in recreational marinas along the estuarine gradient. *Ecological Indicators*, *113*, 106147. https://doi.org/10.1016/j.ecolind.2020.106147

Ardura, A. (2019). Species-specific markers for early detection of marine invertebrate invaders through eDNA methods: Gaps and priorities in GenBank as database example. *Journal for Nature Conservation*, *47*, 51–57. https://doi.org/10.1016/j.jnc.2018.11.005

Arechavaleta, M., Rodríguez, S., Zurita, N., García, A. (coord.) (2010). *Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres*. Santa Cruz de Tenerife, Gobierno de Canarias.

Bailey, S. A., Brown, L., Campbell, M. L., Canning-Clode, J., Carlton, J. T., Castro, N., Chainho, P., Chan, F. T., Creed, J. C., Curd, A., Darling, J., Fofonoff, P., Galil, B. S., Hewitt, C. L., Inglis, G. J., Keith, I., Mandrak, N. E., Marchini, A., McKenzie, C. H., ... Zhan, A. (2020). Trends in the detection of aquatic non-indigenous species across global marine, estuarine and freshwater ecosystems: A 50-year perspective. *Diversity and Distributions*, *26*, 1780–1797. https://doi.org/10.1111/ddi.13167

Barbier, E. B. (2017). Marine ecosystem services. *Current Biology*, *27*, R507–R510. https://doi.org/10.1016/j.cub.2017.03.020

Bax, N., Williamson, A., Aguero, M., Gonzalez, E., & Geeves, W. (2003). Marine invasive alien species: A threat to global biodiversity. *Marine Policy*, *27*, 313–323. https://doi.org/10.1016/S0308-597X(03)00041-1

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. *Scientific Reports*, *6*, 1–11. https://doi.org/10.1038/srep20226

Borges, P. A. V., Abreu, C., Aguiar, A. M. F., Carvalho, P., Jardim, R., Melo, I., Oliveira, P., Sérgio, C., Serrano, A. R. M., & Vieira, P. (Eds.) (2008). *A list of the terrestrial fungi, flora and fauna of Madeira and Selvagens archipelagos*. Direcção Regional do Ambiente da Madeira and Universidade dos Açores, Funchal and Angra do Heroísmo.

Borges, P. A. V., Costa, A., Cunha, R., Gabriel, R., Gonçalves, V., Martins, A. F., Melo, I., Parente, M., Raposeiro, P., Rodrigues, P., Santos, R. S.,

Silva, L., Vieira, P., & Vieira, V. (2010). *A list of the terrestrial and marine biota from the Azores*. Cascais, Princípia.

Borja, A., White, M. P., Berdalet, E., Bock, N., Eatock, C., Kristensen, P., Leonard, A., Lloret, J., Pahl, S., Parga, M., Prieto, J. V., Wuijts, S., & Fleming, L. E. (2020). Moving toward an agenda on ocean health and human health in Europe. *Frontiers in Marine Science*, 7, 37. https://doi.org/10.3389/fmars.2020.00037

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

Cacabelos, E., Martins, G. M., Faria, J., Prestes, A. C. L., Costa, T., Moreu, I., & Neto, A. I. (2020). Limited effects of marine protected areas on the distribution of invasive species, despite positive effects on diversity in shallow-water marine communities. *Biological Invasions*, 22, 1169–1179. https://doi.org/10.1007/s10530-019-02171-x

Canning-Clode, J., Fofonoff, P., McCann, L., Carlton, J., & Ruiz, G. (2013). Marine invasions on a subtropical island: Fouling studies and new records in a recent marina on Madeira Island (Eastern Atlantic Ocean). *Aquatic Invasions*, 8, 261–270. https://doi.org/10.3391/ai.2013.8.3.02

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67. https://doi.org/10.1038/nature11148

Castro, N., Ramalhosa, P., López, J., Costa, J., Gestoso, I., & Canning-Clode, J. (2020). Exploring marine invasions connectivity in a NE Atlantic Island through the lens of historical maritime traffic patterns. *Regional Studies in Marine Science*, 37, 101333. https://doi.org/10.1016/j.rsma.2020.101333

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., … Costa, M. J. (2015). Non-indigenous species in Portuguese coastal areas, coastal lagoons, estuaries and islands. *Estuarine, Coastal and Shelf Science*, 167, 199–211. https://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=bold

Cordier, T., Esling, P., Lejzerowicz, F., Visco, J., Ouadahi, A., Martins, C., Cedhagen, T., & Pawlowski, J. (2017). Predicting the ecological quality status of marine environments from eDNA metabarcoding data using supervised machine learning. *Environmental Science and Technology*, 51, 9118–9126. https://doi.org/10.1021/acs.est.7b01518

Cordier, T., Forster, D., Dufresne, Y., Martins, C. I. M., Stoeck, T., & Pawlowski, J. (2018). Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. *Molecular Ecology Resources*, 18, 1381–1391. https://doi.org/10.1111/1755-0998.12926

Cunha, R. L., Lima, F. P., Tenorio, M. J., Ramos, A. A., Castilho, R., & Williams, S. T. (2014). Evolution at a different pace: Distinctive phylogenetic patterns of cone snails from two ancient oceanic archipelagos. *Systematic Biology*, 63, 971–987. https://doi.org/10.1093/sysbio/syu059

Desiderato, D., Costa, F. O., Serejo, C., Abiatti, M., Queiroga, H., & Vieira, P. E. (2019). Macaronesian islands as promoters of diversification in amphipods: The remarkable case of the family Hyalidae (Crustacea, Amphipoda). *Zoologica Scripta*, 48, 359–375. https://doi.org/10.1111/zsc.12339

Diagne, C., Leroy, B., Gozlan, R. E., Vaissière, A.-C., Assailly, C., Nuninger, L., Roiz, D., Jourdain, F., Jarić, I., & Courchamp, F. (2020). InvaCost, a public database of the economic costs of biological invasions worldwide. *Scientific Data*, 7, 1–12. https://doi.org/10.1038/s41597-020-00586-z

Duarte, S., Leite, B. R., Feio, M. J., Costa, F. O., & Filipe, A. F. (2021). Integration of DNA-based approaches in aquatic ecological assessment using benthic macroinvertebrates. *Water*, 13, 331. https://doi.org/10.3390/w13030331

Duarte, S., Vieira, P. E., & Costa, F. O. (2020). Assessment of species gaps in DNA barcode libraries of nonindigenous species (NIS) occurring in European coastal regions. *Metabarcoding and Metagenomics*, 4, 35–46. https://doi.org/10.3897/mbmg.4.55162

Duarte, S., Vieira, P. E., Lavrador, A. S., & Costa, F. O. (2021). Status and prospects of marine NIS detection and monitoring through (e)DNA metabarcoding. *Science of the Total Environment*, 751, 141729. https://doi.org/10.1016/j.scitotenv.2020.141729

European Commission (2008). Directive of the European Parliament and the council establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Directive 2008/56/EC. *Official Journal of the European Union*, L 164.

Fais, M., Duarte, S., Vieira, P. E., Sousa, R., Hajibabaei, M., Canchaya, C. A., & Costa, F. O. (2020). Small-scale spatial variation of meiofaunal communities in lima estuary (NW Portugal) assessed through metabarcoding. *Estuarine, Coastal and Shelf Science*, 238, 106683. https://doi.org/10.1016/J.ECSS.2020.106683

Fernández-Palacios, J. M., De Nascimento, L., Otto, R., Delgado, J. D., García-Del-Rey, E., Arévalo, J. R., & Whittaker, R. J. (2011). A reconstruction of Palaeo-Macaronesia, with particular reference to the long-term biogeography of the Atlantic island laurel forests. *Journal of Biogeography*, 38, 226–246. https://doi.org/10.1111/j.1365-2699.2010.02427.x

Fontes, J. T., Vieira, P. E., Ekrem, T., Soares, P., & Costa, F. O. (2020). BAGS: An automated Barcode, Audit & Grade System for DNA barcode reference libraries. *Molecular Ecology Resources*, 21, 573–583. https://doi.org/10.1111/1755-0998.13262

Freitas, R., Romeiras, M., Silva, L., Cordeiro, R., Madeira, P., González, J. A., Wirtz, P., Falcón, J. M., Brito, A., Floeter, S. R., Afonso, P., Porteiro, F., Viera-Rodríguez, M. A., Neto, A. I., Haroun, R., Farminhão, J. N. M., Rebelo, A. C., Baptista, L., Melo, C. S., … Ávila, S. P. (2019). Restructuring of the 'Macaronesia' biogeographic unit: A marine multi-taxon biogeographical approach. *Scientific Reports*, 9, 15792. https://doi.org/10.1038/s41598-019-51786-6

García Molinos, J., Halpern, B. S., Schoeman, D. S., Brown, C. J., Kiessling, W., Moore, P. J., Pandolfi, J. M., Poloczanska, E. S., Richardson, A. J., & Burrows, M. T. (2016). Climate velocity and the future global redistribution of marine biodiversity. *Nature Climate Change*, 6, 83–88. https://doi.org/10.1038/nclimate2769

Gargan, L. M., Morato, T., Pham, C. K., Finarelli, J. A., Carlsson, J. E. L., & Carlsson, J. (2017). Development of a sensitive detection method to survey pelagic biodiversity using eDNA and quantitative PCR: A case study of devil ray at seamounts. *Marine Biology*, 164, 1–9. https://doi.org/10.1007/s00227-017-3141-x

GBIF.org (2020). *GBIF Home Page*. https://www.GBIF.org [14 May 2020]

Gomes, N. (2014). *Construction of a DNA barcode reference library for marine Isopods (Crustacea - Isopoda) from Portugal and Macaronésia* (Master's thesis, Universidade do Minho, Braga). http://hdl.handle.net/1822/34768

Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., Lacoursière-Roussel, A., Leong, S. C. Y., Li, Y., Olds, B., Pfrender, M. E., Prowse, T. A. A., Renshaw, M. A., & Lodge, D. M. (2018). Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports*, 8, 8843. https://doi.org/10.1038/s41598-018-27048-2

Hering, D., Borja, A., Jones, J. I., Pont, D., Boets, P., Bouchez, A., Bruce, K., Drakare, S., Hänfling, B., Kahlert, M., Leese, F., Meissner, K., Mergen, P., Reyjol, Y., Segurado, P., Vogler, A., & Kelly, M. (2018). Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. *Water Research*, 138, 192–205. https://doi.org/10.1016/j.watres.2018.03.003

Hestetun, J. T., Bye-Ingebrigtsen, E., Nilsson, R. H., Glover, A. G., Johansen, P. O., & Dahlgren, T. G. (2020). Significant taxon sampling

gaps in DNA databases limit the operational use of marine macrofauna metabarcoding. *Marine Biodiversity*, *50*, 70. https://doi.org/10.1007/s12526-020-01093-5

Hobern, D. (2020). BIOSCAN: DNA barcoding to accelerate taxonomy and biogeography for conservation and sustainability. *Genome*, *64*(3), 161–164. https://doi.org/10.1139/gen-2020-0009

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2019). Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. *Scientific Reports*, *9*, 1–10. https://doi.org/10.1038/s41598-019-47899-7

Iacarella, J. C., Burke, L., Davidson, I. C., DiBacco, C., Therriault, T. W., & Dunham, A. (2020). Unwanted networks: Vessel traffic heightens the risk of invasions in marine protected areas. *Biological Conservation*, *245*, 108553. https://doi.org/10.1016/j.biocon.2020.108553

ICES (2019). Working Group on Introductions and Transfers of Marine Organisms (WGITMO). *ICES Scientific Reports*, *1*, 1–27. https://doi.org/10.17895/ices.pub.5569

Kier, G., Kreft, H., Lee, T. M., Jetz, W., Ibisch, P. L., Nowicki, C., Mutke, J., & Barthlott, W. (2009). A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9322–9327. https://doi.org/10.1073/pnas.0810306106

Lacoursière-Roussel, A., Howland, K., Normandeau, E., Grey, E. K., Archambault, P., Deiner, K., Lodge, D. M., Hernandez, C., Leduc, N., & Bernatchez, L. (2018). eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. *Ecology and Evolution*, *8*, 7763–7777. https://doi.org/10.1002/ece3.4213

Lagabrielle, E., Rouget, M., Payet, K., Wistebaar, N., Durieux, L., Baret, S., Lombard, A., & Strasberg, D. (2009). Identifying and mapping biodiversity processes for conservation planning in islands: A case study in Réunion Island (Western Indian Ocean). *Biological Conservation*, *142*, 1523–1535. https://doi.org/10.1016/j.biocon.2009.02.022

Leduc, N., Lacoursière-Roussel, A., Howland, K. L., Archambault, P., Sevellec, M., Normandeau, E., Dispas, A., Winkler, G., McKindsey, C. W., Simard, N., & Bernatchez, L. (2019). Comparing eDNA metabarcoding and species collection for documenting Arctic metazoan biodiversity. *Environmental DNA*, *1*, 342–358. https://doi.org/10.1002/edn3.35

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., … Zimmermann, J. (2016). DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas and Outcomes*, *2*, e11321. https://doi.org/10.3897/rio.2.e11321

Leese, F., Bouchez, A., Abarenkov, K., Altermatt, F., Borja, A., Bruce, K., & 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. *Advances in Ecological Research*, *58*, 63–99. https://doi.org/10.1016/bs.aecr.2018.01.001

Leite, B. R., Vieira, P. E., Teixeira, M. A. L., Lobo-Arteaga, J., Hollatz, C., Borges, L. M. S., Duarte, S., Troncoso, J. S., & Costa, F. O. (2020). Gap-analysis and annotated reference library for supporting macroinvertebrate metabarcoding in Atlantic Iberia. *Regional Studies in Marine Science*, *36*, 101307. https://doi.org/10.1016/j.rsma.2020.101307

Leite, B. R., Vieira, P. E., Troncoso, J. S., & Costa, F. O. (2019). Combining artificial substrates, morphology and DNA metabarcoding for investigating macrozoobenthic communities in NW Iberia. *Frontiers in Marine Science Conference Abstract: XX Iberian Symposium on Marine Biology Studies (SIEBM XX)*. https://doi.org/10.3389/conf.fmars.2019.08.00061

Lenzner, B., Latombe, G., Capinha, C., Bellard, C., Courchamp, F., Diagne, C., Dullinger, S., Golivets, M., Irl, S. D. H., Kühn, I., Leung, B., Liu, C., Moser, D., Roura-Pascual, N., Seebens, H., Turbelin, A., Weigelt, P., & Essl, F. (2020). What will the future bring for biological invasions on islands? An expert-based assessment. *Frontiers in Ecology and Evolution*, *8*, 1–16. https://doi.org/10.3389/fevo.2020.00280

López-Escardó, D., Paps, J., de Vargas, C., Massana, R., Ruiz-Trillo, I., & del Campo, J. (2018). Metabarcoding analysis on European coastal samples reveals new molecular metazoan diversity. *Scientific Reports*, *8*, 9106. https://doi.org/10.1038/s41598-018-27509-8

Luz, A., & Keskin, E. (2019). *Building reference library for marine fish species of Azores archipelago and bio-monitoring via DNA metabarcoding*. Paper presented at the Ecology and Evolutionary Biology Symposium, Turkey.

Micael, J., Parente, M. I., & Costa, A. C. (2014). Tracking macroalgae introductions in North Atlantic oceanic islands. *Helgoland Marine Research*, *68*, 209–219. https://doi.org/10.1007/s10152-014-0382-7

Moro, L., Martín, J. L., Garrido, M. J., & Izquierdo, I. (Eds.) (2003). *Lista de especies marinas de Canarias (algas, hongos, plantas y animales)*. Islas Canarias, Consejería de Política Territorial y Medio Ambiente del Gobierno de Canarias.

Olenin, S., Narščius, A., Minchin, D., David, M., Galil, B., Gollasch, S., Marchini, A., Occhipinti-Ambrogi, A., Ojaveer, H., & Zaiko, A. (2014). Making non-indigenous species information systems practical for management and useful for research: An aquatic perspective. *Biological Conservation*, *173*, 98–107. https://doi.org/10.1016/j.biocon.2013.07.040

Parretti, P., Canning-Clode, J., Ferrario, J., Marchini, A., Botelho, A. Z., Ramalhosa, P., & Costa, A. C. (2020). Free rides to diving sites: The risk of marine non-indigenous species dispersal. *Ocean and Coastal Management*, *190*, 105158. https://doi.org/10.1016/j.ocecoaman.2020.105158

Pereira, H. M., Leadley, P. W., Proença, V., Alkemade, R., Scharlemann, J. P. W., Fernandez-Manjarrés, J. F., Araújo, M. B., Balvanera, P., Biggs, R., Cheung, W. W. L., Chini, L., Cooper, H. D., Gilman, E. L., Guénette, S., Hurtt, G. C., Huntington, H. P., Mace, G. M., Oberdorff, T., Revenga, C., … Walpole, M. (2010). Scenarios for Global Biodiversity in the 21st Century. *Science*, *330*, 1496–1501. https://doi.org/10.1126/science.1196624

Price, B., Briscoe, A., Misra, R., & Broad, G. (2020). *DEFRA Centre of Excellence for DNA Methods: Evaluation of DNA barcode libraries used in the UK and developing an action plan to fill priority gaps*. Natural England Joint Publication, JP035.

Pyšek, P., Richardson, D. M., Pergl, J., Jarošík, V., Sixtová, Z., & Weber, E. (2008). Geographical and taxonomic biases in invasion ecology. *Trends in Ecology and Evolution*, *23*, 237–244. https://doi.org/10.1016/j.tree.2008.02.002

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramalhosa, P., Camacho-Cruz, K., Bastida-Zavala, R., & Canning-Clode, J. (2014). First record of *Branchiomma bairdi* McIntosh, 1885 (Annelida: Sabellidae) from Madeira island, Portugal (northeastern Atlantic Ocean). *BioInvasions Records*, *3*, 235–239. https://doi.org/10.3391/bir.2014.3.4.04

Ramalhosa, P., Gestoso, I., Duarte, B., Caçador, I., & Canning-Clode, J. (2019). Metal pollution affects both native and non-indigenous biofouling recruitment in a subtropical island system. *Marine Pollution Bulletin*, *141*, 373–386. https://doi.org/10.1016/j.marpolbul.2019.02.072

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. *PLoS One*, *8*, e66213. https://doi.org/10.1371/journal.pone.0066213

Rilov, G., & Crooks, J. A. (2009). Biological invasions in marine ecosystems. Ecological, management and geographic perspectives. *Marine Ecology*, *30*, 514. https://doi.org/10.1111/j.1439-0485.2009.00316.x

Santos, A. M. C., Field, R., Ricklefs, R. E., & Borregaard, M. (2016). New directions in island biogeography. *Global Ecology and Biogeography*, *25*, 751–768. https://doi.org/10.1111/geb.12477

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, *47*, D94–D99. https://doi.org/10.1093/nar/gky989

Schroeder, A., Stanković, D., Pallavicini, A., Gionechetti, F., Pansera, M., & Camatti, E. (2020). DNA metabarcoding and morphological analysis - assessment of zooplankton biodiversity in transitional waters. *Marine Environmental Research*, *160*, 104946. https://doi.org/10.1016/j.marenvres.2020.104946

Seebens, H., Bacher, S., Blackburn, T. M., Capinha, C., Dawson, W., Dullinger, S., Genovesi, P., Hulme, P. E., Kleunen, M., Kühn, I., Jeschke, J. M., Lenzner, B., Liebhold, A. M., Pattison, Z., Pergl, J., Pyšek, P., Winter, M., & Essl, F. (2020). Projecting the continental accumulation of alien species through to 2050. *Global Change Biology*, *00*, 1–13. https://doi.org/10.1111/gcb.15333

Seebens, H., Gastner, M. T., & Blasius, B. (2013). The risk of marine bioinvasion caused by global shipping. *Ecology Letters*, *16*, 782–790. https://doi.org/10.1111/ele.12111

Silva, J. M., Creer, S., Dos Santos, A., Costa, A. C., Cunha, M. R., Costa, F. O., & Carvalho, G. R. (2011). Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLoS One*, *6*, e19449. https://doi.org/10.1371/journal.pone.0019449

Stat, M., Huggett, M. J., Bernasconi, R., DiBattista, J. D., Berry, T. E., Newman, S. J., Harvey, E. S., & Bunce, M. (2017). Ecosystem biomonitoring with eDNA: Metabarcoding across the tree of life in a tropical marine environment. *Scientific Reports*, *7*, 12240. https://doi.org/10.1038/s41598-017-12501-5

Steyaert, M., Priestley, V., Osborne, O., Herraiz, A., Arnold, R., & Savolainen, V. (2020). Advances in metabarcoding techniques bring us closer to reliable monitoring of the marine benthos. *Journal of Applied Ecology*, *57*, 2234–2245. https://doi.org/10.1111/1365-2664.13729

Suarez-Menendez, M., Planes, S., Garcia-Vazquez, E., & Ardura, A. (2020). Early alert of biological risk in a coastal lagoon through eDNA metabarcoding. *Frontiers in Ecology and Evolution*, *8*, 1–10. https://doi.org/10.3389/fevo.2020.00009

Tavares, A. I., Cabezas, M. P., Xavier, R., Branco, M., Lima, F. P., Seabra, R., & Santos, A. M. (2017). Phylogeography and phylogeny of the genus *Acanthonyx* (Decapoda, Epialtidae) in the north-east Atlantic and Mediterranean. *Zoologica Scripta*, *46*, 571–583. https://doi.org/10.1111/zsc.12232

Trebitz, A. S., Hoffman, J. C., Grant, G. W., Billehus, T. M., & Pilgrim, E. M. (2015). Potential for DNA-based identification of Great Lakes fauna: Match and mismatch between taxa inventories and DNA barcode libraries. *Scientific Reports*, *5*, 12162. https://doi.org/10.1038/srep12162

Tsiamis, K., Palialexis, A., Stefanova, K., Gladan, Ž. N., Skejić, S., Despalatović, M., Cvitković, I., Dragičević, B., Dulčić, J., Vidjak, O., Bojanić, N., Žuljević, A., Aplikioti, M., Argyrou, M., Josephides, M., Michailidis, N., Jakobsen, H. H., Staehr, P. A., Ojaveer, H., ... Cardoso, A. C. (2019). Non-indigenous species refined national baseline inventories: A synthesis in the context of the European Union's Marine Strategy Framework Directive. *Marine Pollution Bulletin*, *145*, 429–435. https://doi.org/10.1016/j.marpolbul.2019.06.012

Valdés, A. Z. S. (2017). *Eukaryotic metabarcoding pipelines for biodiversity assessment of marine benthic communities affected by ocean acidification* (Master's thesis, Escola Superior de Turismo e Tecnologia do Mar and Instituto Politécnico de Leiria). http://hdl.handle.net/10400.8/2854

Vieira, P. E., Desiderato, A., Azevedo, S. L., Costa, F. O., & Queiroga, H. (2019). Prominent evolutionary divide between Macaronesian islands and nearby continental coasts in multiple peracarids (Crustacea): Over 60 suspected new species. *8th International Barcode of Life Conference. Genome*, *62*(6), 445. https://doi.org/10.1139/gen-2019-0083

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. *Molecular Ecology*, *28*, 1784–1800. https://doi.org/10.1111/mec.15052

Vitousek, P. M. (1990). Biological invasions and ecosystem processes: Towards an integration of population biology and ecosystem studies. *Oikos*, *57*, 7–13. https://doi.org/10.2307/3565731

Wallentinus, I., & Nyberg, C. D. (2007). Introduced marine organisms as habitat modifiers. *Marine Pollution Bulletin*, *55*, 323–332. https://doi.org/10.1016/j.marpolbul.2006.11.010

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., ... Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of the Total Environment*, *678*, 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247

Winter, D. J. (2017). rentrez: An R package for the NCBI eUtils API. *The R Journal*, *9*, 520–526. https://doi.org/10.32614/RJ-2017-058

Wirtz, P., Brito, A., Falcón, J. M., Freitas, R., Fricke, R., Monteiro, V., ... Tariche, O. (2013). The coastal fishes of the Cape Verde Islands – new records and an annotated check-list (Pisces). *Spixiana*, *36*, 113–142.

WoRMS Editorial Board (2020). *World Register of Marine Species*. http://www.marinespecies.org at VLIZ Accessed 2020-05-14. https://doi.org/10.14284/170

Xavier, J. R., van Soest, R. W. M., Breeuwer, J. A. J., Martins, A. M. F., & Menken, S. B. J. (2010). Phylogeography, genetic diversity and structure of the poecilosclerid sponge Phorbas fictitius at oceanic islands. *Contributions to Zoology*, *79*, 119–129. https://doi.org/10.1163/18759866-07903004

Zaiko, A., Samuiloviene, A., Ardura, A., & Garcia-Vazquez, E. (2015). Metabarcoding approach for nonindigenous species surveillance in marine coastal waters. *Marine Pollution Bulletin*, *100*, 53–59. https://doi.org/10.1016/j.marpolbul.2015.09.030

BIOSKETCH

The research team has been working on taxonomy, ecology, DNA barcoding and DNA metabarcoding of marine invertebrates in the Northeast Atlantic, including Macaronesia. They have been fulfilling the genetic databases for these groups and developing and optimizing DNA-based tools to be employed in biomonitoring programmes and in the detection of non-indigenous species.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.