Original research article

Unveiling biodiversity: The current status of marine species barcoding in Red Sea Metazoans

Carlos Angulo-Preckler [a,*], Christopher Hempel [a], Sofia Frappi [a], Kah Kheng Lim [a], Tullia Terraneo [a], Dirk Steinke [b], Lotfi J. Rabaoui [c], Francesca Benzoni [a], Carlos M. Duarte [a]

[a] Marine Sciences Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[b] Centre for Biodiversity Genomics - University of Guelph, Guelph, Canada
[c] National Center for Wildlife (NCW), Riyadh 12411, Saudi Arabia

ARTICLE INFO

ABSTRACT

Preserving biodiversity is a global challenge. Censuses of marine biodiversity are indispensable for monitoring the responses of marine life to environmental changes induced by human activities. Ongoing extinction events affect both species and populations amid unprecedented environmental changes induced by climate shifts and habitat degradation. These changes result in substantial declines in biodiversity. In addition, our understanding of oceanic life remains incomplete, especially regarding elusive, rare, delicate, or understudied organisms. One example for this is the biodiversity of the Red Sea which remains largely unexplored and poorly understood. In an attempt to evaluate the current status of known versus COI-barcoded marine animal species in the Red Sea we used online taxonomic and genetic databases to provide a comprehensive analysis of the region's described marine life, focusing on the occurrence data of marine animal species to identify disparities in COI barcoding coverage at the phylum level. Our analysis reveals that barcoding coverage varies significantly among phyla, with Nematoda, Platyhelminthes, Bryozoa, and Porifera being highly underrepresented compared to Chordata. While over 6000 metazoan species from 22 phyla are known to inhabit the Red Sea, only 49.77 % appear to be barcoded. COI barcoding helps preserve biodiversity by providing a reliable and standardized method for accurately identifying and monitoring species, including those that are cryptic or newly discovered, thereby informing and enhancing conservation efforts and guiding future research efforts toward understudied regions and organisms.

1. Introduction

Human activities such as habitat fragmentation, overpopulation, and pollution, are leading to a substantial decline of biodiversity, thereby fueling a Sixth Mass Extinction (Dirzo and Raven, 2003; Barnosky et al., 2011; Pievani, 2014). While intervention and conservation efforts have slowed-down the overall decline of global biodiversity (Duarte et al., 2020), the fact that we perhaps only know 20 % of the life in the ocean bears the risk that many undiscovered species may go extinct unnoticed. Although biodiversity can be

examined at various levels (genes, species, and ecosystem) and scales (spatial and temporal) (Rogers et al., 2022), species delimitation and identification remains fundamental for biodiversity assessments and subsequent protection and conservation efforts. Therefore, there is an urgent need to employ innovative methodologies that streamline and improve the documentation of marine biodiversity and its distribution patterns (Rogers et al., 2022).

The current inventory of marine metazoan species comprises about 230,000 valid species, with a potential further 1000,000 species awaiting discovery and characterization (Bucklin et al., 2011). The catalogue of species is growing at a rate too slow (Bouchet and Duarte, 2006; Rogers et al., 2022) to enable reliable monitoring of conservation programs meant to halt biodiversity loss. Major gaps remain, especially when it comes to small, rare, delicate, and cryptic organisms as well as those that live in poorly-explored marine habitats.

The analysis of DNA sequences, particularly the mitochondrial protein-coding gene cytochrome c oxidase subunit I (COI), represents a promising possibility for accurate and reliable species identification across all metazoan groups (Rogers et al., 2022). Approximately 650 nucleotides of the DNA barcode sequence allow for rapid species identification (Hebert et al., 2003; Miller, 2007). Platforms like the Barcode of Life Datasystems (BOLD; Ratnasingham and Hebert, 2007) provide identification tools and community access to DNA barcodes, contributing to the ongoing efforts to better understand and document marine biodiversity. Metabarcoding, an approach to generate DNA barcodes for entire bulk samples, has become a prominent tool for assessing marine biodiversity (Andújar et al., 2018). Despite its widespread use, the barcode progress for marine species has been relatively slow. For instance, Bucklin et al. (2011) reported that only 9.5 % of the 192,702 known marine metazoan species had been barcoded at that time. Later, Steinke et al. (2016) estimated that around 40,000 marine metazoan species had been barcoded, reflecting gradual improvements but still highlighting a significant gap, given an estimated 1–2 million marine species of which 75–90 % remain undescribed (Rogers et al., 2023). To date, out of 210,810 globally known marine animal species (WoRMS Editorial Board, 2024), only ~14 % have COI barcodes available (Mugnai et al., 2021). The slow progress in barcoding can be attributed to the immense taxonomic diversity of marine life, difficulties to access it, and the ongoing challenge of developing consistently effective primers for all species. Taxa like Porifera, Ctenophora, and Anthozoa are particularly challenging, as their mitochondrial rate of evolution renders traditional genetic identification methods inadequate for reliable discrimination between closely related species (Bucklin et al., 2011). Furthermore, COI coverage is inconsistent, with certain taxa having greater representation in public databases than others, indicating a need for more comprehensive coverage across all taxa.

Such gaps in public DNA barcoding reference databases impede the effectiveness of biodiversity assessments, in particular when it comes to those based on metabarcoding of bulk samples or environmental DNA. They latter often result in a high number of unassigned species (Martin et al., 2021), or, when matching criteria are relaxed, false positives where the closest match may be a species not present in the sample or not occurring in the sampled region. This in turn leads to uncertainty about the actual biodiversity of surveyed areas (Valentini et al., 2016). Most DNA barcoding initiatives have focused on terrestrial and freshwater environments (Weigand et al., 2019), leaving a significant gap in our understanding of marine biodiversity, underscoring the urgent need for more comprehensive genetic research on underrepresented marine taxa. This is particularly important for poorly explored marine ecosystems with high prevalence of endemism, such as the Red Sea (Dibattista et al., 2016; Berumen et al., 2019).

The Red Sea, formed around 24 million years ago during the separation of the African and Arabian tectonic plates, is a unique marine environment shaped by desiccation, hypersalinity, and intermittent connection to the Indian Ocean (Bosworth et al., 2005; Dibattista et al., 2013). It spans 2270 kilometers from 13°N in the Gulf of Aden to 30°N in the Gulf of Suez, has a maximum depth of 3040 m and a maximum width of 355 kilometers. Its unique characteristics include substantial evaporation rates coupled with minimal freshwater influx, resulting in an exceptionally oligotrophic environment (Elshanawany and Zonneveld, 2016). Marked by perpetual year-round water column stratification and high annual solar irradiation, the Red Sea exhibits significant north-to-south gradients in salinity, sea surface temperature, and nutrient concentration (Ngugi et al., 2012; Chaidez et al., 2017). All together these characteristics shape the biodiversity and distribution of the marine fauna in the basin, which is a biodiversity hotspot for several taxa, accounting for the highest endemism rates in the Indian Ocean and harboring species potentially resilient to environmental changes and crucial for adaptation (Roberts et al., 2002; Fine et al., 2013; Dibattista et al., 2016).

Despite being one of the most diverse marine ecosystems globally, it is fragile and vulnerable to oceanic warming, with intense warming trends observed since the mid-1990s (Alawad et al., 2020). Exacerbated by global climate change, this has significantly impacted coral reef ecosystems, raising concerns for biodiversity conservation (Raitsos et al., 2011). Increased human activities such as fisheries, pollution, and coastal development, pose ongoing threats to this fragile ecosystem (Ellis et al., 2019), exacerbating the stress on marine life, leading to habitat degradation and biodiversity loss. Although coral reefs are among the most diverse ecosystems in the ocean, a significant portion their diversity eludes detection through conventional methods. Especially, the prevalence of small, rare species is often overlooked in traditional biodiversity assessments. While the Red Sea holds immense ecological importance, its biodiversity remains largely unexplored and poorly studied in comparison with other major reef systems like the Great Barrier Reef or the Caribbean (Berumen et al., 2013). Particularly, taxa such as Arthropoda and Mollusca, which are among the most diverse phyla in coral reefs, have received little attention in ecological studies to date. Despite recent efforts to explore the cryptobiome along the Saudi Arabian coastline (Villalobos et al., 2024a, 2024b), research focuses on limited areas such as the Gulf of Aqaba (Berumen et al., 2013), and more recently northern Egypt and central Saudi Arabia (Cochran et al., 2024), highlighting the need for more comprehensive studies across the entire Red Sea to better understand and conserve its rich marine biodiversity.

Recognizing the urgent need to assess and preserve the biodiversity of the Red Sea, this study aims to address critical knowledge gaps by exploring the its biodiversity and its representation in public barcoding libraries. By integrating advanced molecular data with established biodiversity information systems, our goal is to evaluate the present status of reported versus barcoded marine animal species in the Red Sea based on online taxonomic and genetic databases to provide a comprehensive analysis of the current state of the

region's marine life.

2. Material and methods

We conducted a comprehensive review of prominent biodiversity databases with a primary emphasis on global-scale geographic distributions across various taxonomic groups. Our aim was to synthesize information on species occurrence in the Red Sea. A total of 23 databases were initially selected, however, upon careful evaluation, only seven were considered relevant and reliable for our analyses (see Table 1 for details). As some databases lack the functionality to perform searches based on geographic areas, we removed them from our investigation. Additionally, some databases were found to be interconnected with primary global databases, thereby only contributing redundant records that were subsequently excluded from our analyses to avoid duplication.

We narrowed the search parameters to include only Metazoans, excluding databases that predominantly cataloged algae and bacteria. This methodological approach ensured a targeted and relevant exploration of metazoan biodiversity in the Red Sea, leveraging the most pertinent data sources for our comprehensive analysis. When the option was available, we followed latitude 29.999751°N as the northern limit and latitude 12.535624°N as the southern limit to search for species in the databases.

We also extracted every specimen entry from the Barcode Of Life Data System (BOLD) that was collected in the Red Sea. Therefore, we searched for all entries that fell within the coordinate limits of the Red Sea (latitude: 12.5–30°N, longitude: 32–44°E) and removed all entries that fell onto land by doing a spatial join between coordinates of collected specimens and continental boundaries using the geopandas python module v0.14.3 in Python v3.12.2. The majority of organisms had no species information assigned, so we removed

Table 1
List of databases revised to compile a Red Sea Metazoan species list.

Database	Status	Species retrieved	Species unique to database	Link to dataset	% contribution to the total number of species
World Register of Marine Species (Worms)	Data collected	4129	1996	https://www.marinespecies.org	66.6
Marine species identification portal	Not available				
Integrate Taxonomic Information System (ITIS)	Red Sea filter not available			https://www.itis.gov	
Catalogue of life	Red Sea filter not available			https://www.catalogueoflife.org	
Encyclopedia of life	Red Sea filter not available			https://eol.org	
Encyclopedia of lifeTraitBank	Not available				
European Register of Marine Species (ERMS)	Irrelevant				
Census of marine life	Connected to other databases			http://www.coml.org	
Ocean Biodiversity Information System (OBIS)	Data collected	3388	1355	https://obis.org/area/40033	54.6
Global Biodiversity Information Facility (GBIF)	Data collected	1480	166	https://www.gbif.org	23.9
ReefBase	Not available				
International Union for Conservation of Nature (IUCN)	Red Sea filter not available			https://www.iucnredlist.org/search/list	
Fishbase	Data collected	1465	488	https://www.fishbase.se/search.php	23.6
AlgaeBase	Data collected	2	2	https://www.algaebase.org	0.03
Safhos cpr	Not available				
Copepod	Red Sea filter not available			https://www.st.nmfs.noaa.gov/copepod/about/databases.html	
PlanktonPortal	Red Sea filter not available			https://www.zooniverse.org/projects/kelseyswieca/plankton-portal/	
EcoTaxa	Red Sea filter not available			https://ecotaxa.obs-vlfr.fr	
Plankton Net	Connected to other databases			https://planktonnet.awi.de	
International Institute for Sustainability (IIS)	Irrelevant				
Larvalbase	Irrelevant				
Global Invasive Species Database (GISD)	Data collected	8	1	http://www.iucngisd.org/gisd/	0.1
Global Registerer of Introduced and Invasive Species (GRIIS)	Connected to other databases			https://griis.org	
Corals in-house	Data collected	326	74	Received from collaborators in-house	5.3

all specimens without species assignment to generate a list of all specimens with species assignment on BOLD collected in the Red Sea.

We integrated all species extracted from each of the seven investigated databases into one dataset. We found some species in the BOLD database that were collected in the Red Sea but not present in the summarized list, we added those species to the list as well. We then manually curated the list. Specifically, we removed taxa names in brackets (e.g., *Apogon* (*Nectamia*) *taeniatus* was converted to *Apogon taeniatus*) and removed all entries that only contained a genus name, or ended with "*sp*" or "*cf*". Additionally, we excluded subspecies, variety and forma names and retained only genus and species for each entry. Lastly, we replaced non-alphanumeric characters, such as umlauts and hyphens, with their appropriate equivalents. That way, we compiled a curated list of all species present in the Red Sea and assessed the number of available COI reference sequences both on National Center for Biotechnology Information (NCBI) GenBank and BOLD for each species. For each species in the list, we counted all COI sequences available on NCBI GenBank as of the 20st October, 2024 using the search string '[respective species name] [Organism] AND (COI[All Fields] OR COX1 [All Fields] OR "cytochrome oxidase subunit 1"[All Fields] OR CO1[All Fields] OR COXI[All Fields])'. In addition we uploaded the list to BOLD as a checklist on the 21nd October, 2024. BOLD's Checklist System allows users to retrieve up-to-date reports from BOLD on the barcoding progress of a user-defined list of species. We generated a report based on BOLD's Full Database to count all COI sequences available on BOLD for the species in our list. The BOLD API was additionally used to retrieve geographic (latitude/longitude) and depth information for barcoded species. Due to the low number of species with geographic information, we complement this list with data retrieved from the MetaZoo Gene Database (MZGdb) (https://metazoogene.org/database/;O'Brien et al., 2024). Using the provided latitude and longitude coordinates, we generated a density map in ArcGIS Pro (version 2.6.0) to highlight the primary sampling areas.

Since the list contained synonym species names, we standardized all species using the Global Biodiversity Information Facility (GBIF) database as the reference. Therefore, we matched the list against GBIF and replaced all synonyms with the accepted species names on GBIF in R (v4.3.2; R Core Team 2021) using the package bdc (v1.1.4; Ribeiro et al., 2022). We then aggregated the COI sequence counts of accepted species names and their synonyms to include reference sequences uploaded to NCBI GenBank under both accepted or synonym names. Furthermore, we manually cross-checked all species that were neither identified as accepted nor synonyms in GBIF. This resolved conflicts related to names representing the synonyms for multiple accepted species and ensured that no recently uploaded species on GBIF were missed, since the R package uses a GBIF version from December 2022. Some of those manually cross-checked species also represented synonyms of accepted species in the standardized list, and their COI sequence counts were added to the counts of the standardized, accepted species as described previously. All remaining species that were not manually found on GBIF were removed from the analysis. We then retrieved the kingdom, phylum, class, order, and family information for all accepted species from GBIF using the GBIF API. Lastly, we removed all species that were assigned to the class "Aves", "Insecta", and "Reptilia" (excluding marine turtles), since birds and insects were classified as "marine" in some of the investigated databases but out of the scope of this project. The code for the manual curation of the initial species list, the counting of COI sequences on NCBI GenBank, and the taxonomic standardization against GBIF are available on GitHub (https://github.com/hempelc/SIREN_project) and involve Python (v3.10; Available at http://www.python.org) and the Python modules Pandas (v2.0.3; Mckinney, 2010), Plotly (v5.19.0; Plotly Technologies Inc. https://plotly.com), and biopython (v1.81; Cock et al., 2009). The final Red Sea metazoan species list is available as Supplemetary Material (Table S1) and includes information on the presence or absence of COI reference sequences in NCBI Genbank or BOLD as well as information on the presence or absence of geographic (latitude and longitude) and depth data in BOLD and MZGdb.

Table 2

Current DNA barcoding status for Red Sea metazoan species by phylum. Numbers of barcoded species refer to species with at least one COI reference sequence in either NCBI GenBank or BOLD.

Phylum	All Species	Species with reference	% of all species
Chordata	1855	1492	80.4
Mollusca	1230	516	41.8
Arthropoda	1137	534	47
Cnidaria	673	391	58.7
Annelida	362	100	27.6
Echinodermata	308	187	60.7
Platyhelminthes	284	21	7.4
Porifera	149	23	15.4
Bryozoa	92	15	16.3
Nematoda	44	4	4.5
Chaetognatha	16	11	75
Xenacoelomorpha	14	2	14.3
Sipuncula	12	7	41.7
Ctenophora	6	3	50
Nemertea	5	5	100
Brachiopoda	4	0	0
Kinorhyncha	4	0	0
Tardigrada	3	0	0
Acanthocephala	1	1	100
Gnathostomulida	1	0	0
Hemichordata	1	1	100
Priapulida	1	0	0
TOTAL	6202	3313	53.4

3. Results

Our dataset primarily relied on two main sources: the World Register of Marine Species (WoRMS Editorial Board, 2021) and the Ocean Biodiversity Information System (OBIS, Intergovernmental Oceanographic Commission of UNESCO: www.iobis.org), contributing 66.6 % and 54.6 % of the total 6202 species list, respectively (Table 1). Additionally, an Red Sea specific coral list based on the integrated systematic studies on a reference collection housed at King Abdullah University of Science and Technology (KAUST) made a significant contribution, accounting for 326 species positively identified (5.3 %). In total, our compilation included 6202 marine animal species listed for the Red Sea. Remarkably, from the 32 known marine animal phyla, the Red Sea is home to 22 (see Table 2). Several databases turned out to be either irrelevant or lacked the capability for performing geographical searches limited to the Red Sea (Table 1).

We found 3313 species with COI sequences when combining both genetic databases. Specifically, 3120 species had at least one COI sequence on Genbank (50.02 %), and 2987 species were found in BOLD (48.16 %). These percentages represent the proportion of species with COI sequences compared to the total number of species in the Red Sea species list established in this study. 2361 species had three or more sequences on GenBank, constituting 38.07 % of the total known animal species. Chordata, Mollusca, and Arthropoda showed the highest species richness (based on the number of species) of COI-sequenced, followed by Cnidaria, Annelida and Echinodermata (Fig. 1). Platyhelminthes, Porifera, Bryozoan, and Nematoda had very low COI coverage. Conversely, phyla with very low diversity, such as Hemichordata, Acanthocephala, and Nemertea, each represented by fewer than 5 species, exhibited complete coverage (100 %).

The number of sequences per species across species ranged from 0 to 1168 sequences for GenBank and up to 1551 sequences for BOLD. Although the majority of species exhibited fewer than five sequences, we identified 22 species on GenBank and 17 in BOLD with
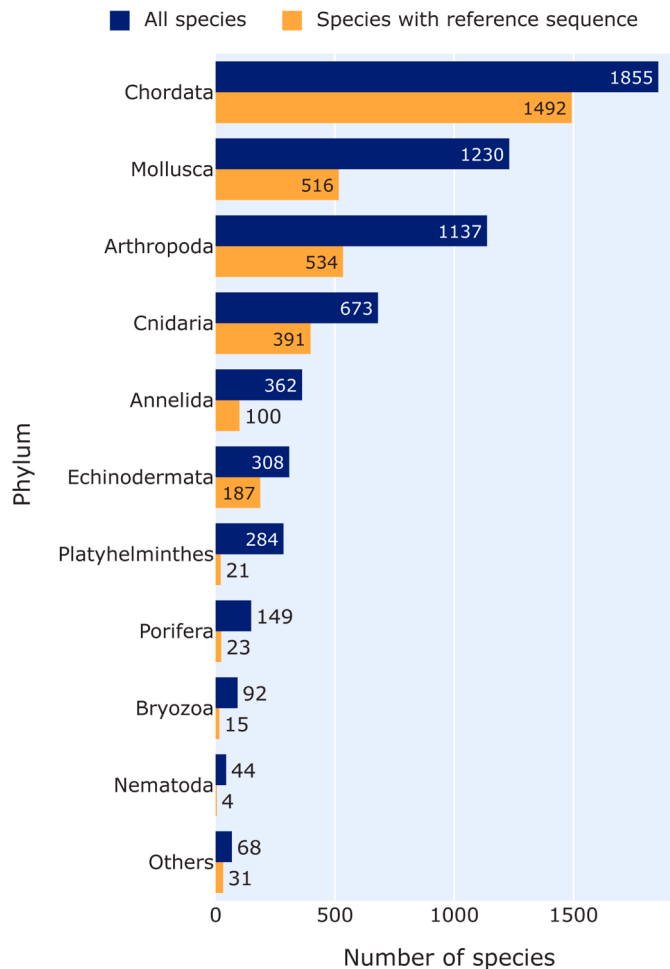

Fig. 1. Total number of known metazoan species reported from the Red Sea and number of metazoan species with at least one COI reference sequence in either NCBI GenBank or BOLD across phyla. Phyla are ordered by numbers of species. Blue bars indicate numbers of reported species; orange bars indicate numbers of barcoded species. Others includes the phyla Chaetognatha, Xenacoelomorpha, Sipuncula, Ctenophora, Nemertea, Brachiopoda, Kinorhyncha, Tardigrada, Acanthocephala, Gnathostomulida, Hemichordata, and Priapulida.

over 500 sequences each (Fig. 2). Notably, the sea star *Lickia laevigata*, the scleractinian coral *Turbinaria mesenterina*, and the giant clam *Tridacna maxima* are among the most sequenced species on GenBank, while the fishes *Oreochromis niloticus* and *Thunnus albacares* and the sea star *L. laevigata* are the most prominent on the BOLD database.

Geographic location and depth information for Red Sea species is notably limited. BOLD contains geographic data for 10,449 specimens and depth data for 3913 specimens matching species in our list of Red Sea species. These represent 1426 species with geographic data and 942 species with depth data. In addition, despite their known presence in the Red Sea, only 220 species with geographic data and 48 species with depth data were collected within the region (Fig. 3). Hence, most of the species known to occur in the Red Sea present in BOLD dataset have been sampled outside the Red Sea. However, MZGdb contains geographic data for 86,186 specimens matching species in our list of Red Sea species (O'Brien et al., 2024). These represent 2878 with geographic data. In addition, despite their known presence in the Red Sea, only 530 species with geographic data were collected within the region. The MZGdb offer improved geographic resolution, significantly enhancing the capabilities of the BOLD database for regional biodiversity mapping. Fine-scale spatial analyses of sample collections of both BOLD and MZGdb combined revealed three sampling hotspots: the main one on the central Saudi Arabia's Red Sea ranging from 19.020°N to 22.812°N, the second in the southern Saudi Arabian Red Sea (Farasan Islands, 16.798°N; 42.199°E), and the third one covering the entire Gulf of Aqaba (from 27.077°N to 29.538°N).

Some databases such as WoRMS use the International Hydrographic Organization (IHO) as a geographic unit. For the Red Sea, this results in the exclusion of the Gulf of Aqaba and the Suez Canal.
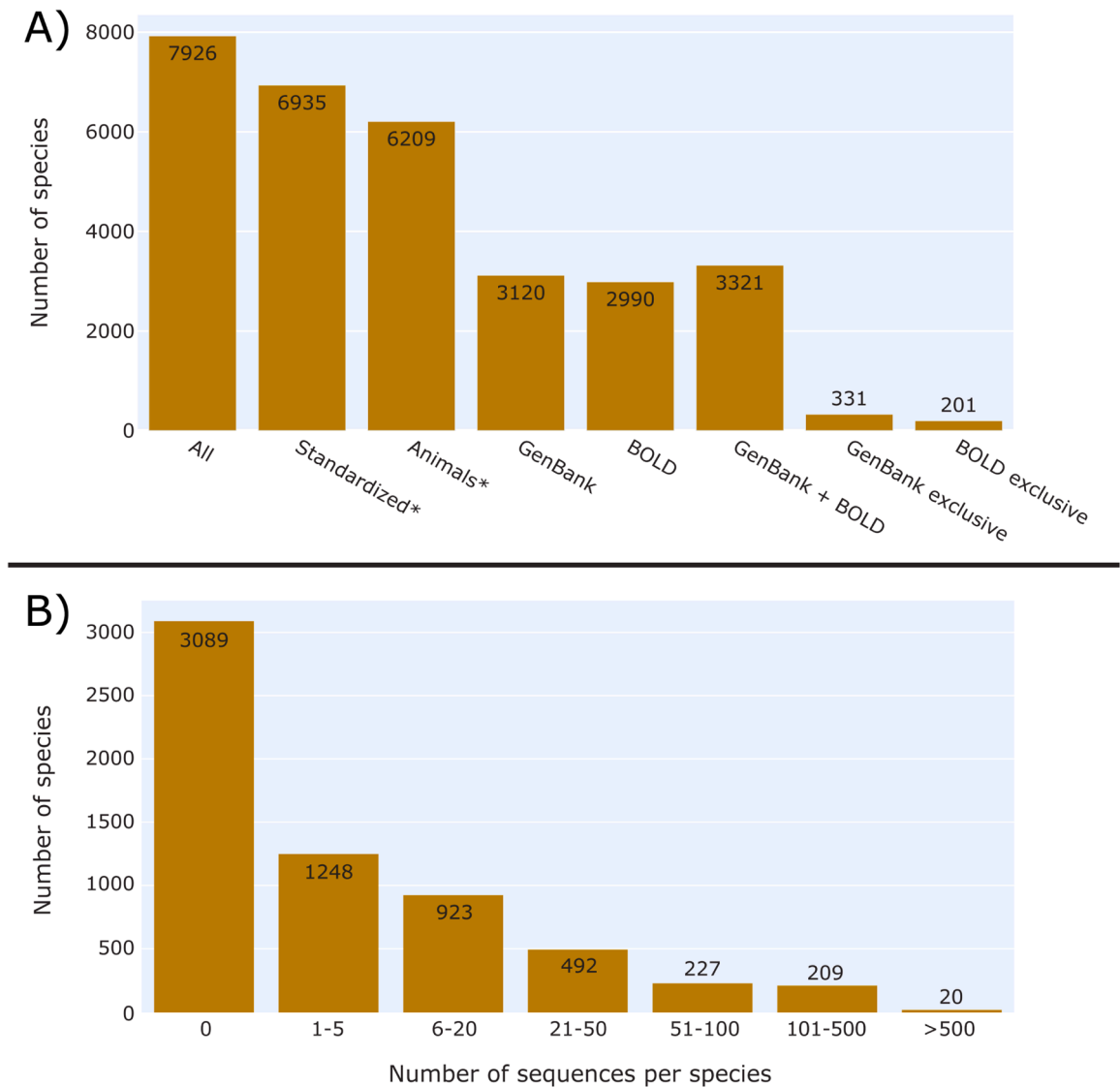

Fig. 2. (A) Summary of the number of species for every step of the study. Standardized: number of species after taxonomic standardization. Animals: number of animal species; (B) Bar plot showing the frequency of sequences for Red Sea species on GenBank. Y-axis indicates the number of species by number of sequences per species.
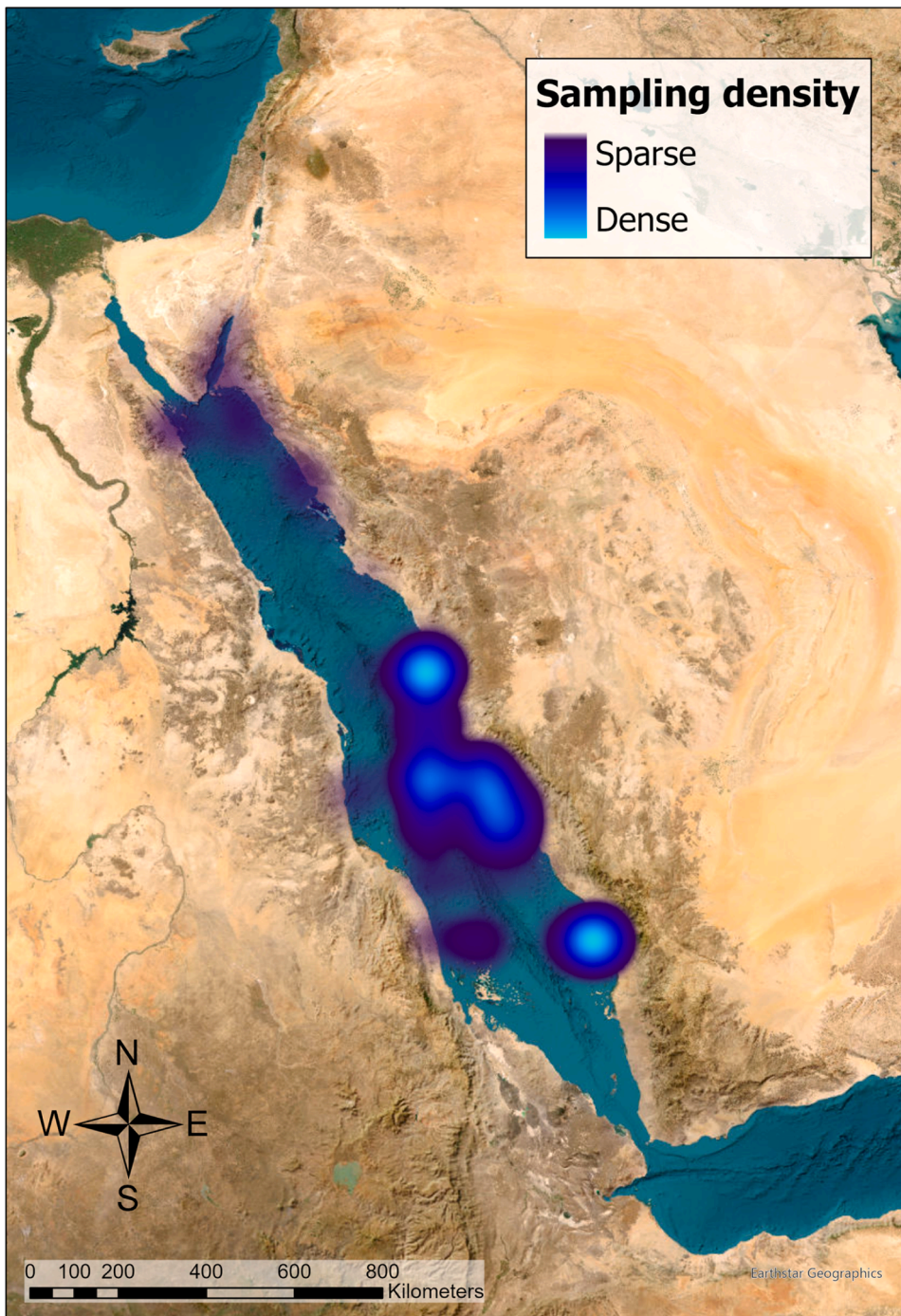
Fig. 3. Sampling density distribution map along the Red Sea. Hotspots with higher sampling densities indicate areas of increased sampling effort, resulting in a greater number of species collected from these regions. (Esri 2020. ArcGIS Pro version 2.6.0).

4. Discussion

Molecular barcoding emerges as a valuable tool for assessing biodiversity, aiding in population differentiation, speciation, phylogeographic investigations, and species identification (Grant et al., 2011). Currently, more than 245,000 marine species have been described (WoRMS), a number that grows by ~2200 new species every year (Vandepitte et al., 2018). Despite this substantial number, the Red Sea remains significantly underexplored, with only 6202 species we found across all databases. Of all the evaluated databases, WoRMS provided the most comprehensive coverage, documenting nearly 2000 unique species, yet it accounted for only 66 % of the

full species list compiled in this study for the Red Sea. The OBIS database ranked second, covering just over 50 % of the total species list and containing more than 1300 unique species (Table 1). This underrepresentation is alarming, especially when compared to other Large Marine Ecosystems (LMEs) such as the Mediterranean Sea, the North Sea, or the Caribbean Sea (Mugnai et al., 2021), and given the high amount of endemism known from the Red Sea (Dibattista et al., 2016; Berumen et al., 2019). Moreover, a large proportion of marine species remain undiscovered, with estimates suggesting that between 75 % and 90 % of marine animal species are yet to be described (May and Moths, 1988; Costello et al., 2010; Rogers et al., 2023). This lag in biodiversity cataloging underscores the need for continued efforts to accelerate the discovery and documentation of marine species, which is essential for understanding and conserving the world's oceans. Furthermore, a considerable bias exists with respect to size and commercial significance, evidenced by the predominant sampling and sequencing efforts directed towards the phylum Chordata, in particular fish. Notably absent from collections within the Red Sea are marine phyla with small body size, predominantly comprising microscopic organisms such as Placozoa, Gastrotrich, Cercozoa, and Picozoa.

The compilation of the Red Sea metazoan species list from various public databases demonstrated extensive coverage at the phylum level, with 22 of the most common phyla. However, a substantial portion of these phyla are significantly underrepresented, with only a few species of each phylum. Notably, Priapulida, Hemichordata, Gnathostomulid, and Acanthocephala, are represented by only a single species, and nine phyla showed fewer than 10 species each. This dearth of representation suggests not only a deficiency in sequencing efforts for these phyla but also a scarcity of knowledge regarding these species in the Red Sea ecosystem. This represents a significant challenge, particularly for organisms with small body sizes, whose actual diversity could be ten times greater than currently recognized (Blaxter, 2004). The lack of taxonomic expertise, exacerbated by historically limited financial support for taxonomy, further complicates accurate identification and classification, thereby impeding our understanding of marine biodiversity. The result is a high number of misidentifications and erroneous records. Although recent years have seen improvements in scientific methods due to developing experimental ecological approaches, this progress has coincided with a reduction in the number of expert taxonomists for many invertebrate groups.

Nematoda, Platyhelminthes, Porifera, and Bryozoa, showed the lowest levels of COI barcoding coverage. The underrepresentation of Porifera and Bryozoa stands out as particularly concerning due to their roles as ecosystem engineers alongside corals (Rossi et al., 2017). Porifera's underrepresentation may be attributed to the standard COI marker's limitations, as it often lacks sufficient resolution to distinguish between different species within this group, and the presence of highly variable COI sequences and pseudogenes further complicates barcoding efforts. Within Porifera, the classes, Hexactinellidae and Homoscleromorpha lack any DNA barcodes, while Calcarea is represented by only one species. The majority of barcodes belong to the class Demoespongia, which is also the most abundant class in this ecosystem. Within Bryozoa, Cyclostomatida rarely constitute more than 20 % of the species recorded for regional bryozoan faunas. In the Red Sea, only five species of cyclostomes have been documented, accounting for only ~5 % of the bryozoan fauna, and lacking any sequences in the databases. They are considered to be comparatively poor competitors for living space, often being overgrown by larger animals such as sponges and ascidians, and frequently losing competition for space to cheilostome bryozoans (McKinney, 1992). So far, the barcoding of Cheilostomatida in the Red Sea has yielded fewer than ten species.

Annelida are represented exclusively by the class Polychaeta, with the most diverse order being Phyllodocida, followed by Eunicida and Sabellida. Polychaetes exhibit complex morphological and ecological characteristics, inhabiting a wide range of substrates, including sandy bottoms and rocky reefs. These diverse habitats and adaptations make polychaetes ideal candidates for the discovery of potential new species. Similarly, Echinodermata is one of the most extensively barcoded phyla, with over 180 species cataloged. Notably, ten species each are represented by more than 100 sequenced individuals. The Echinodermata, in descending order of species richness are represented by Holothuroidea, Ophiuroidea, Echinoidea, Asteroidea, and Crinoidea, all of which exhibit high barcode coverage (~60 %). These organisms thrive particularly well in warm, shallow waters and are restricted to benthic habitats. Coral reefs and coastal lagoons support a greater diversity of species compared to seagrass or mangrove environments, with rock surfaces being the preferred habitat (Campell, 1987).

Among the six classes recovered in our dataset within the phylum Cnidaria, Hexacorallia is the most represented in the occurrence databases, with nearly four times more reports compared to the second most represented class, Octocorallia. In contrast, only one species of Cubozoa and two species of Myxozoa, belonging to two separate orders, have been reported from the region, and no barcode information for these species is publicly available. Within Hexacorallia, over 90 % of the 464 reported taxa are attributed to Scleractinia. The most reported diversity within Scleractinia is retrieved in the families Acroporidae, Merulinidae, and Poritidae, which is not surrising as they are also the most speciose coral families occurring in the Indo-Pacific (Kitahara et al., 2016). However, barcoding efforts do not reflect this diversity, with two species from the family Dendrophyllidae being largerly overrepresented in both GenBank and BOLD databases. In Octocorallia, most of the recorded diversity is attributed to soft corals in the order Malacalcyonacea, with COI sequence abundance primarily driven by the families Xeniidae and Sarcophytidae.

The overrepresentation of available barcodes for certain taxa within Cnidaria could be attributed to two main reasons. First, the slow evolutionary rates of mitochondrial DNA in taxa within the subphylum Anthozoa make COI genotyping inadequate for barcoding purposes and limit its use for resolving deeper phylogenetic nodes, such as genus and family levels. This has likely influenced the sequencing effort, leading to the overrepresentation of certain taxa. For example, in Scleractinia, species-specific patterns of genetic resolution occur in both nuclear and mitochondrial DNA (Kitahara et al., 2016), which shifted the research efforts from barcoding to integrated morpho-molecular systematics and phylogenomics. Second, the limited exploration of mesophotic and deep-sea habitats in the region until recent years has led biodiversity assessments to focus more on accessible shallow-water coral reefs.

Arthropoda and Mollusca stand out as highly diverse phyla within coral reef ecosystems, boasting numerous small-bodied species, yet they have remained relatively understudied in ecological research (Bouchet et al., 2002; Albano et al., 2011). Within Arthropoda, the most diverse classes are Malacostraca, predominantly represented by the order Decapoda, and Copepoda, dominated by various

groups of copepods. These classes likely inhabit different ecological niches, with decapods primarily dominating coral reef ecosystems and copepods being the principal constituents of the zooplankton. Additionally, other classes of small planktonic organisms, such as Ostracoda, Maxillopoda, and Branchiopoda, are well represented in the barcode databases, despite having significantly fewer species. In contrast, the class Pycnogonida is represented by only five species, with just one species barcoded. Mollusca, primarily composed of macrofaunal organisms, is a highly diverse group. The main classes within this group are Gastropoda and Bivalvia, with approximately 40 % barcode coverage. The class with the highest barcode coverage is Cephalopoda at 73.3 %, while the least barcoded classes are Scaphopoda, Solenogastres, and Polyplacophora. In fact the former two classes contained no barcode.

Chordata showed the highest values of species and barcoding coverage (80.4 %). Among the phylum Chordata, the class Actinopterygii is the most abundant and has been extensively studied, comprising a diverse array of fish species. This class includes 1662 species from 159 families, inhabiting environments ranging from shallow coastal waters to the deep sea. Despite their ecological significance in the Red Sea, the distribution of many fish species in this region remains poorly documented. Approximately 78 % of the fish species in the Red Sea have been barcoded to date. Coral reefs within the Red Sea support the most diverse and complex fish assemblages, with an exceptionally high rate of endemism at approximately 14.7 % (Berumen et al., 2019; Bogorodsky and Randall, 2019). In addition to Actinopterygii, other chordate classes such as Elasmobranchii, Mammalia, and Reptilia have been nearly fully barcoded, with coverage rates of 97 %, 100 %, and 100 %, respectively. Conversely, the classes Appendicularia and Thaliacea remain the least barcoded within this phylum. This disparity highlights the need for further genetic research on these lesser-studied groups. The uneven coverage is likely attributable to the size and economic significance of the organisms. Most barcodes in NCBI and BOLD databases come from commercially important species, as reflected in the FAO report on Fisheries Statistics Saudi Arabia 2021 (Alshaikhi et al., 2023), which shows that the top 10 economically important Red Sea species are well represented.

Inaccuracies in species records often arise from insufficient dataset curation, leading to erroneous entries and misidentifications within taxonomic group. To mitigate this issue, it is crucial to carefully verify and validate species lists against authoritative sources, regularly update and review datasets, and employ robust data management practices to ensure accuracy and reliability.

We found notable coverage for some select Red Sea species, with 17–23 species having more than 500 sequences, while approximately one-third of species—over 1000—are represented by merely one sequence. In genetic studies, the inclusion of multiple individuals from the same species yields multifaceted advantages. Primarily, it facilitates the comprehensive exploration of genetic diversity, thereby enhancing the statistical robustness of analyses and the reliability of research findings (Valentini et al., 2009; Hale et al., 2012). Additionally, this approach serves to ameliorate potential challenges associated with species identification, ensuring the precision of genetic analyses. By accommodating the inherent genetic variability within species, incorporating multiple individuals safeguards against misidentification arising from genetic outliers, thereby enhancing the accuracy and fidelity of genetic investigations.

Efforts to establish large-scale DNA barcoding databases, like the Barcode of Life Data System (BOLD; Ratnasingham and Hebert, 2007), have been ongoing for the past two decades. However, these databases are still incomplete. Globally, only 14 % of the known marine animal species possess COI barcodes, although certain biodiversity hotspots, such as the Large Marine Ecosystems, exhibit higher rates of coverage (Mugnai et al., 2021). For the Red Sea, the coverage stands at approximately 49 % in both GeneBank and BOLD databases, which is above the global average. However, this may reflect the limited number of known species in the region, rather than an accurate representation of its true biodiversity and COI coverage, especially considering the low number of sequenced species collected in the Red Sea. Including species from different regions in genetic databases complicates data interpretation, analysis, and application, particularly concerning the understanding of regional genetic diversity, population structure, and conservation.

The existing barcode library for the Red Sea remains insufficient for thorough biodiversity assessments. A major challenge affecting Red Sea barcoding is not only obtain more sequences but also ensuring comprehensive coverage across different geographic areas, depths, and taxonomic levels. While the Gulf of Aqaba and the Central Saudi Arabian coast have been extensively studied (Cochran et al., 2024), most of the Red Sea can be considered a barcode desert to date (Fig. 3). Expanding sampling efforts across the geographic areas is crucial for comprenhensive understanding of its genetic biodiversity. The MZG and BOLD databases contain geographic information for 2567 specimens representing 530 species, and depth information for merely 259 specimens encompassing 48 species, highlighting the low number of species effectively collected and sequenced in this region. Geographical patterns indicate a shift in the research density over the last decade from the northern Red Sea to the central east coast (Cochran et al., 2024).

Conversely, a huge gap has been detected for the west coast of the Red Sea, with the exception of only three small areas studied: the Dahlak arquipelago in Eritrea (16.600°N, 39.909°E), Port Sudan (19.606°N, 37.221°E), and Hurghada (27.263°N, 33.833°E; Fig. 3). To address all these important gaps, new barcoding initiatives should be established. It is pivotal to promote species descriptions and molecular characterization of marine species in the Red Sea to increase autochthonous coverage, similar to other areas of the world. The barcode coverage among LMEs range from 36 % to 62 % (Mugnai et al., 2021), but these values depend on the total number of species known for each region. Our review of several databases reveals that most are incomplete compared to the published literature. Furthermore, the exclusion of the Gulf of Aqaba and the Suez Canal is particularly problematic because the northern Gulf of Aqaba waters are a persistent research hotspot (Berumen et al., 2013; Cochran et al., 2024). Including geographic coordinates and depth information in genetic databases is paramount for understanding the spatial distribution and ecological context of genetic diversity within species. Such data facilitate analyses of population structure and connectivity, providing crucial insights into the environmental factors shaping genetic variation, thereby aiding conservation efforts and informing habitat-specific strategies.

Over the past few decades, there is a growing acknowledgment of the urgent need to conserve global biodiversity, especially given the alarming rates of species extinctions driven by human activities (Kerr and Currie, 1995). The lack of knowledge on the abundance of marine species has become increasingly apparent, with many species facing risks and requiring immediate conservation efforts (Savage, 1995). Coral reefs in the Red Sea serve as biodiversity hotspots and refuges for potentially temperature-tolerant species,

which could better withstand rising sea surface temperatures (Roberts et al., 2002; Fine et al., 2013). Despite its ecological significance, the overall species contribution to databases remains notably low, indicating a significant gap in research and sequencing efforts with numerous taxa inadequately represented both by DNA barcodes and the biogeographic databases. The Red Sea's coral reefs face severe threats, ranking among the world's most endangered marine ecosystems, highlighting the urgent need for conservation efforts (Gardner et al., 2003; Pandolfi et al., 2003; Knowlton and Jackson, 2008).

Meeting the goals of the Convention of Biological Diversity (CBD, https://www.cbd.int/), specifically halting biodiversity loss, will prove increasingly difficult without addressing current gaps in our understanding of marine biodiversity. We must, therefore, step up our efforts in taxonomic discovery, particularly focusing on smaller organisms, which represent a significant proportion of marine diversity and biomass, but that are often cryptic and difficult to sample and identify. Metabarcoding of environmental DNA samples offer an avenue to accelerate our inventories of local to regional biodiversity, but hinges on the existence of reference sequences in barcode repositories. Our work represents an initial resource for biodiversity assessments of the Red Sea, offering an inventory of known metazoan species including DNA barcoding information. The low number of species in the present data set highlights the need to further advance efforts in building inventories of the biodiversity of the Red Sea biodiversity to support of conservation actions.

Data Accessibility Statement

The data supporting the findings of this study are available in the supplementary material of this paper, specifically in Table S1. The code for the manual curation of the initial species list, the counting of COI sequences on NCBI GenBank, and the taxonomic standardization against GBIF are available on GitHub at https://github.com/hempelc/SIREN_project.

Ethical statement

The authors declare that the research described in this work does not involve experimentation on humans, animals, patients, or volunteers. The authors have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper. This research was funded by baseline funding from King Abdullah University of Science and Technology provided to C. M. Duarte.

CRediT authorship contribution statement

Carlos Angulo-Preckler: Conceptualization, Writing – original draft. **Christopher Hempel, Kah Kheng Lim, Sofia Frappi, Tullia Terraneo, Dirk Steinke**: Data curation. **Carlos Angulo-Preckler, Christopher Hempel**: Formal analysis, Investigation, Methodology. **Carlos M. Duarte**: Funding acquisition. **Christopher Hempel**: Software. **Francesca Benzoni, Carlos M. Duarte**: Supervision. **Carlos Angulo-Preckler, Christopher Hempel, Kah Kheng Lim, Sofia Frappi, Tullia Terraneo, Lotfi J. Rabaoui, Dirk Steinke, Francesca Benzoni, Carlos M. Duarte**: Writing – review & editing. All authors approved the final version of the manuscript.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT (version 3.5) in order to improve the readability of the manuscript and to draft some of the code utilized in this manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the published article.

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.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.gecco.2024.e03339.

Data availability

Data has been added as Supplementary material

References

Alawad, K.A., Al-Subhi, A.M., Alsaafani, M.A., Alraddadi, T.M., 2020. Decadal variability and recent summer warming amplification of the sea surface temperature in the Red Sea. PLoS ONE 15.

Albano, P.G., Sabelli, B., Bouchet, P., 2011. The challenge of small and rare species in marine biodiversity surveys: microgastropod diversity in a complex tropical coastal environment. Biodivers. Conserv. 20, 3223–3237.

Alshaikhi, A., Alshaye, K., Ageely, O., Samarali, R., Alharbi, B., Alhafedh, Y., Almohsen, I. & Guemes, P. (2023) Fisheries statistics: Saudi Arabia 2016–2021.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B., Ferrer, E.A., 2011. Has the Earth's sixth mass extinction already arrived? Nature 2011 471, 51–57. 7336, 471.

Berumen, M.L., Hoey, A.S., Bass, W.H., Bouwmeester, J., Catania, D., Cochran, J.E.M., Khalil, M.T., Miyake, S., Mughal, M.R., Spaet, J.L.Y., Saenz-Agudelo, P., 2013. The status of coral reef ecology research in the Red Sea. Coral Reefs 32, 737–748.

Berumen, M.L., Roberts, M.B., Sinclair-Taylor, T.H., Dibattista, J.D., Saenz-Agudelo, P., Isari, S., He, S., Khalil, M.T., Hardenstine, R.S., Tietbohl, M.D., Priest, M.A., Kattan, A., Coker, D.J., Berumen, M.L., Sinclair-Taylor, T.H., Isari, · S., Hardenstine, R.S., Tietbohl, · M.D., Kattan, · A., Coker, · D.J., Roberts, M.B., Dibattista, J. D., Saenz-Agudelo, P., He, S., Khalil, M.T. & Priest, M.A. (2019) Fishes and Connectivity of Red Sea Coral Reefs. 157–179.

Blaxter, M.L., 2004. The promise of a DNA taxonomy. Philos. Trans. R. Soc. Lond. Ser. B: Biol. Sci. 359, 669–679.

Bogorodsky, S.V., Randall, J.E., 2019. Endem. Fishes Red. Sea 239–265.

Bosworth, W., Huchon, P., McClay, K., 2005. The Red Sea and Gulf of Aden Basins. J. Afr. Earth Sci. 43, 334–378.

Bouchet, P., Duarte, C.M., 2006. Explor. Mar. Biodivers.: Sci. Technol. Chall.

Bouchet, P., Lozouet, P., Maestrati, P., Heros, V., 2002. Assessing the magnitude of species richness in tropical marine environments: exceptionally high numbers of molluscs at a New Caledonia site. Biol. J. Linn. Soc. 75, 421–436.

Bucklin, A., Steinke, D., Blanco-Bercial, L., 2011. DNA barcoding of marine metazoa. Annu. Rev. Mar. Sci. 3, 471–508.

Campell, A.C., 1987. Echinoderms of the Red. Sea. Red. Sea 215–232.

Chaidez, V., Dreano, D., Agusti, S., Duarte, C.M., Hoteit, I., 2017. Decadal trends in Red Sea maximum surface temperature. Sci. Rep. 2017 7 (1), 1–8. 7.

Cochran, J.E.M., Kattan, A., Langner, U., Knochel, A.M., Carvalho, S., Coker, D.J., Fitzgerald, L., Ford, K., Justo, M.S.S., Hardenstine, R.S., McIvor, A.J., Peinemann, V. N., Pombo-Ayora, L., Predragovic, M., Rich, W.A., Scott, K., Shchepanik, H., Tanabe, L.K., Tietbohl, M.D., Williams, C.T., Berumen, M.L., 2024. Fine-scale spatial and temporal trends in Red Sea coral reef research. Reg. Stud. Mar. Sci. 71.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., De Hoon, M.J.L., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422.

Costello, M.J., Coll, M., Danovaro, R., Halpin, P., Ojaveer, H., Miloslavich, P., 2010. A Census of Marine Biodiversity Knowledge, Resources, and Future Challenges. PLOS ONE 5, e12110.

Dibattista, J.D., Berumen, M.L., Gaither, M.R., Rocha, L.A., Eble, J.A., Choat, J.H., Craig, M.T., Skillings, D.J., Bowen, B.W., 2013. After continents divide: comparative phylogeography of reef fishes from the Red Sea and Indian Ocean. J. Biogeogr. 40, 1170–1181.

Dibattista, J.D., Howard Choat, J., Gaither, M.R., Hobbs, J.P.A., Lozano-Cortés, D.F., Myers, R.F., Paulay, G., Rocha, L.A., Toonen, R.J., Westneat, M.W., Berumen, M. L., 2016. On the origin of endemic species in the Red Sea. J. Biogeogr. 43, 13–30.

Dirzo, R., Raven, P.H., 2003. Global state of biodiversity and loss. Annu. Rev. Environ. Resour. 28, 137–167.

Duarte, C.M., Agusti, S., Barbier, E., Britten, G.L., Castilla, J.C., Gattuso, J.P., Fulweiler, R.W., Hughes, T.P., Knowlton, N., Lovelock, C.E., Lotze, H.K., Predragovic, M., Poloczanska, E., Roberts, C., Worm, B., 2020. Rebuilding marine life. Nature 2020 580, 39–51. 7801, 580.

Ellis, J.I., Jamil, T., Anlauf, H., Coker, D.J., Curdia, J., Hewitt, J., Jones, B.H., Krokos, G., Kürten, B., Hariprasad, D., Roth, F., Carvalho, S., Hoteit, I., 2019. Multiple stressor effects on coral reef ecosystems. Glob. Change Biol. 25, 4131–4146.

Elshanawany, R., Zonneveld, K.A.F., 2016. Dinoflagellate cyst distribution in the oligotrophic environments of the Gulf of Aqaba and northern Red Sea. Mar. Micropaleontol. 124, 29–44.

Fine, M., Gildor, H., Genin, A., 2013. A coral reef refuge in the Red Sea. Glob. Change Biol. 19, 3640–3647.

Gardner, T.A., Côté, I.M., Gill, J.A., Grant, A., Watkinson, A.R., 2003. Long-term region-wide declines in Caribbean corals. Science 301, 958–960.

Grant, R.A., Griffiths, H.J., Steinke, D., Wadley, V., Linse, K., 2011. Antarctic DNA barcoding; a drop in the ocean? Polar Biol. 34, 775–780.

Hale, M.L., Burg, T.M., Steeves, T.E., 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLOS ONE 7, e45170.

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B: Biol. Sci. 270, 313–321.

Kerr, J.T., Currie, D.J., 1995. Effects of Human Activity on Global Extinction RiskEfectos de la actividad humana sobre el riesgo de extinción global. Conserv. Biol. 9, 1528–1538.

Kitahara, M.V., Fukami, H., Benzoni, F. & Huang, D., 2016, The new systematics of scleractinia: Integrating molecular and morphological evidence. The Cnidaria, past, present and Future: The World of Medusa and her Sisters, 41–59..

Knowlton, N., Jackson, J.B.C., 2008. Shifting baselines, local impacts, and global change on coral reefs. PLOS Biol. 6, e54.

Martin, D., Aguado, M.T., Álamo, M.A.F., Britayev, T.A., Böggemann, M., Capa, M., Faulwetter, S., Fukuda, M.V., Helm, C., Petti, M.A.V., Ravara, A., Teixeira, M.A.L., 2021. On the Diversity of Phyllodocida (Annelida: Errantia), with a Focus on Glyceridae, Goniadidae, Nephtyidae, Polynoidae, Sphaerodoridae, Syllidae, and the Holoplanktonic Families. Diversity 2021 13, 131.

May, R.M., Moths, B.I., 1988. How many species are there on earth? Science 241, 1441–1449.

Mckinney, W., 2010. Data Struct. Stat. Comput. Python.

McKinney, F.K., 1992. Competitive interactions between related clades: evolutionary implications of overgrowth interactions between encrustin cyclostome and cheilostome bryozoans. Mar. Biol. 114, 645–652.

Miller, S.E., 2007. DNA barcoding and the renaissance of taxonomy. Proc. Natl. Acad. Sci. USA 104, 4775–4776.

Mugnai, F., Meglécz, E., Costantini, F., Abbiati, M., Bavestrello, G., Bertasi, F., Bo, M., Capa, M., Chenuil, A., Colangelo, M.A., De Clerck, O., Gutiérrez, J.M., Lattanzi, L., Leduc, M., Martin, D., Matterson, K.O., Mikac, B., Plaisance, L., Ponti, M., Riesgo, A., Rossi, V., Turicchia, E., Waeschenbach, A., Wangensteen, O.S., 2021. Are well-studied marine biodiversity hotspots still blackspots for animal barcoding? Glob. Ecol. Conserv. 32.

Ngugi, D.K., Antunes, A., Brune, A., Stingl, U., 2012. Biogeography of pelagic bacterioplankton across an antagonistic temperature–salinity gradient in the Red Sea. Mol. Ecol. 21, 388–405.

O'Brien, T.D., Blanco-Bercial, L., Questel, J.M., Batta-Lona, P.G., Bucklin, A., 2024. MetaZooGene Atlas and database: reference sequences for marine ecosystems. Methods Mol. Biol. 2744, 475–489.

Pandolfi, J.M., Bradbury, R.H., Sala, E., Hughes, T.P., Bjorndal, K.A., Cooke, R.G., McArdle, D., McClenachan, L., Newman, M.J.H., Paredes, G., Warner, R.R., Jackson, J.B.C., 2003. Global trajectories of the long-term decline of coral reef ecosystems. Science 301, 955–958.

Pievani, T., 2014. The sixth mass extinction: Anthropocene and the human impact on biodiversity. Rend. Lince-.-. 25, 85–93.

Raitsos, D.E., Hoteit, I., Prihartato, P.K., Chronis, T., Triantafyllou, G., Abualnaja, Y., 2011. Abrupt warming of the Red Sea. Geophys. Res. Lett. 38, 14601.

Ratnasingham, S., Hebert, P.D.N., 2007. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Mol. Ecol. Notes 7, 355–364.

Ribeiro, B.R., Velazco, S.J.E., Guidoni-Martins, K., Tessarolo, G., Jardim, L., Bachman, S.P., Loyola, R., 2022. bdc: A toolkit for standardizing, integrating and cleaning biodiversity data. Methods Ecol. Evol. 13, 1421–1428.

Roberts, C.M., McClean, C.J., Veron, J.E.N., Hawkins, J.P., Allen, G.R., McAllister, D.E., Mittermeier, C.G., Schueler, F.W., Spalding, M., Wells, F., Vynne, C., Werner, T.B., 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science 295, 1280–1284.

Rogers, A.D., Appeltans, W., Assis, J., Ballance, L.T., Cury, P., Duarte, C., Favoretto, F., Hynes, L.A., Kumagai, J.A., Lovelock, C.E., Miloslavich, P., Niamir, A., Obura, D., O'Leary, B.C., Ramirez-Llodra, E., Reygondeau, G., Roberts, C., Sadovy, Y., Steeds, O., Sutton, T., Tittensor, D.P., Velarde, E., Woodall, L., Aburto-Oropeza, O., 2022. Discovering marine biodiversity in the 21st century. Adv. Mar. Biol. 93, 23–115.

Rogers, A.D., Appiah-Madson, H., Ardron, J.A., Bax, N.J., Bhadury, P., Brandt, A., Buttigieg, P.L., De Clerck, O., Delgado, C., Distel, D.L., Glover, A., Gobin, J., Guilhon, M., Hampton, S., Harden-Davies, H., Hebert, P., Hynes, L., Lowe, M., MacIntyre, S., Madduppa, H., Mazzuco, A.C., de, A., McCallum, A., McOwen, C., Nattkemper, T.W., Odido, M., O'Hara, T., Osborn, K., Pouponneau, A., Provoost, P., Rabone, M., Ramirez-Llodra, E., Scott, L., Sink, K.J., Turk, D., Watanabe, H.K.,

Weatherdon, L.V., Wernberg, T., Williams, S., Woodall, L., Wright, D.J., Zeppilli, D., Steeds, O., 2023. Accelerating ocean species discovery and laying the foundations for the future of marine biodiversity research and monitoring. Front. Mar. Sci. 10, 1224471.

Rossi, S., Bramanti, L., Gori, A., Orejas, C., 2017. An overview of the animal forests of the world an overview of the animal forests of the world (ed. by). Marine Animal Forest. Springer, pp. 1–25 (ed. by).

Savage, J.M., 1995. Systematics and the biodiversity crisis. BioScience 45, 673–679.

Steinke, D., Prosser, S.W.J., Hebert, P.D.N., 2016. DNA barcoding of marine metazoans. Methods Mol. Biol. 1452, 155–168.

Valentini, A., Pompanon, F., Taberlet, P., 2009. DNA barcoding for ecologists. Trends Ecol. Evol. 24, 110–117.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., Gaboriaud, C., Jean, P., Poulet, N., Roset, N., Copp, G.H., Geniez, P., Pont, D., Argillier, C., Baudoin, J.M., Peroux, T., Crivelli, A.J., Olivier, A., Acqueberge, M., Le Brun, M., Møller, P.R., Willerslev, E., Dejean, T., 2016. Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol. Ecol. 25, 929–942.

Vandepitte, L., Vanhoorne, B., Decock, W., Vranken, S., Lanssens, T., Dekeyzer, S., Verfaille, K., Horton, T., Kroh, A., Hernandez, F., Mees, J., 2018. A decade of the World Register of Marine Species – General insights and experiences from the Data Management Team: Where are we, what have we learned and how can we continue? PLOS ONE 13, e0194599.

Villalobos, R., Aylagas, E., Pearman, J.K., Curdia, J., Coker, D., Bell, A.C., Brown, S.D., Rowe, K., Lozano-Cortés, D., Rabaoui, L.J., Marshell, A., Qurban, M., Jones, B., Berumen, M.L., Carvalho, S., 2024b. Biodiversity patterns of the coral reef cryptobiota around the Arabian Peninsula. Sci. Rep. 2024 14 (1), 1–13. **14**.

Villalobos, R., Aylagas, E., Ellis, J.I., Pearman, J.K., Anlauf, H., Curdia, J., Lozano-Cortes, D., Mejia, A., Roth, F., Berumen, M.L., Carvalho, S., 2024a. Responses of the coral reef cryptobiome to environmental gradients in the Red Sea. PLOS ONE 19, e0301837.

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.D.B., Eisendle, U., Freyhof, J., Gadawski, P., Graf, W., Haegerbaeumer, A., van der Hoorn, B.B., Japoshvili, B., Keresztes, L., Keskin, E., Leese, 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.