diversity

MDPI

Article

Mind the Gaps: Taxonomic, Geographic and Temporal Data of Marine Invertebrate Databases from Mozambique and São Tomé and Príncipe

Marta Bento [1,2], Henrique Niza [2,3], Alexandra Cartaxana [2,4], Salomão Bandeira [5], José Paula [1,2] and Alexandra Marçal Correia [1,2,*]

1 Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
2 MARE—Marine and Environmental Sciences Centre/ARNET—Aquatic Research Network, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
3 Departamento do Mar e Recursos Marinhos, IPMA—Instituto Português do Mar e da Atmosfera, Avenida Doutor Alfredo Magalhães Ramalho 6, 1495-165 Algés, Portugal
4 Museu Nacional de História Natural e da Ciência, Universidade de Lisboa, Rua da Escola Politécnica 56–58, 1250-102 Lisboa, Portugal
5 Departamento de Ciências Biológicas, Faculdade de Ciências, Universidade Eduardo Mondlane, Maputo 1100, Mozambique
* Correspondence: amcorreia@fc.ul.pt

Abstract: One of the best ways to share and disseminate biodiversity information is through the digitization of data and making it available via online databases. The rapid growth of publicly available biodiversity data is not without problems which may decrease the utility of online databases. In this study we analyze taxonomic, geographic and temporal data gaps, and bias related to existing data on selected marine invertebrate occurrences along the coastline of two African countries, Mozambique and São Tomé and Príncipe. The final marine invertebrate dataset comprises of 19.910 occurrences, but 32% of the original dataset occurrences were excluded due to data gaps. Most marine invertebrates in Mozambique were collected in seagrasses, whereas in São Tomé and Príncipe they were mostly collected offshore. The dataset has a temporal coverage from 1816 to 2019, with most occurrences collected in the last two decades. This study provides baseline information relevant to a better understanding of marine invertebrate biodiversity data gaps and bias in these habitats along the coasts of these countries. The information can be further applied to complete marine invertebrate data gaps contributing to design informed sampling strategies and advancing refined datasets that can be used in management and conservation plans in both countries.

Keywords: coastal macroinvertebrates; knowledge gaps; natural history collections; online database; Mozambique; São Tomé and Príncipe

Citation: Bento, M.; Niza, H.; Cartaxana, A.; Bandeira, S.; Paula, J.; Correia, A.M. Mind the Gaps: Taxonomic, Geographic and Temporal Data of Marine Invertebrate Databases from Mozambique and São Tomé and Príncipe. *Diversity* **2023**, *15*, 70. https://doi.org/10.3390/d15010070

Academic Editor: Michael Wink

Received: 30 November 2022
Revised: 21 December 2022
Accepted: 2 January 2023
Published: 5 January 2023

1. Introduction

Museums and herbaria with natural history collections (NHC) are vital sources of scientific knowledge. These collections provide a vast amount of historical and current information on the world's biodiversity [1] and have proven to be a critical record of ecological change and evolution. One of the best ways to share such information is through the digitization of data and to make it available via online databases. The Global Biodiversity Information Facility (GBIF, http://www.gbif.org, accessed on 10 October 2018) aims at mobilizing biodiversity data from NHC, surveys and other sources by storing data in an online portal [2]. GBIF is not the only online biodiversity database available (e.g., OBIS, iDigBio) but it is the one that provides more information and the most widely used [3,4], with the intention to 'make the world's primary data on biodiversity freely and universally

available via the Internet' [5]. Currently, GBIF provides access to 2.2 billion occurrence records (as of August 2022).

This rapid growth of publicly available biodiversity data is not without problems, and the scientific community acknowledges taxonomic, geographical and temporal uncertainty and the existence of data gaps [6]. Differences in funding and data sharing, together with punctual geopolitical conflicts often result in spatial and species biases [2] and scarcity of data availability for some regions, requiring a thorough analysis of data quality [7]. The data available in GBIF has very diverse sources, quality, scope and accuracy [8] provided mostly by NHC, research and monitoring campaigns. The inaccuracies and uncertainties in coverage and quality of data may decrease the utility of online databases [8], and it is essential to minimize these problems through the standardization of data formats such as the Darwin Core [9] and a peer-review system for data publication [10]. Another source of inaccuracy is that the purpose of data collection may differ from the subsequent use by GBIF users [11].

Marine habitats face a wide range of pressures related to human population growth, such as industrial activities, fishing and aquaculture, maritime transportation, tourism, or urbanization of the coastline [12]. These local drivers of degradation are further strengthened with global impacts of climate change, such as sea level rise, cyclones and related floods and strong winds temperature increase and acidification. These combined pressures unbalance biotic communities and ecosystem processes, affecting biodiversity, community structure and species distributions. Marine macroinvertebrates are vital components of marine habitats due to their abundance and diversity, playing an important role in marine ecosystems. In spite of their significance there is a lack of compiled information regarding their biodiversity [13]. Developing countries are particularly vulnerable to biodiversity loss and the degradation of ecosystems. The gaps in data from developing countries are also more pronounced which is explained by limited funding and geopolitical conflicts, such as war. This means that data collection and monitoring programs, aiming to a better representation and knowledge of the natural capital, are insufficient when compared with those of developed countries, causing unbalanced online databases by underrepresentation of geographical regions [14].

The goal of this study was to assess taxonomic, geographic and temporal data gaps and bias related to existing data on selected marine invertebrate occurrences (Annelida, Arthropoda, Cnidaria, Echinodermata and Mollusca) along the coastline of two poorly studied African countries, Mozambique (MOZ) and São Tomé and Príncipe (STP), by compiling scattered data from online databases (namely GBIF and museums' NHC databases) and scientific literature. This analysis provides baseline information relevant to a better understanding of marine invertebrate biodiversity databases, their limitations and their improvement to accurately address updated biodiversity issues. Further, this dataset can be used to complete marine invertebrate data gaps and to develop informed sampling strategies leading to more refined datasets that can be applied in management and conservation plans in both countries.

2. Materials and Methods

2.1. Study Areas

Mozambique is located in the Western Indian Ocean, South-eastern Africa and has a coastline extension of 2500 km and an Economic Exclusive Zone (EEZ) of 493,000 km^2 [15]. The three most critical coastal habitats include mangrove forests, seagrass beds and coral reefs, with an approximate area of 2909, 439 and 1860 km^2, respectively [15].

São Tomé and Príncipe (STP) is an island state located at the Equator in the Gulf of Guinea at 380 km off the continental African coast [16]. It consists of one volcanic archipelago, of which the two main islands are São Tomé and Príncipe, separated by approximately 146 km, and has a coastline extension of 240 km. The EEZ covers an area of 160,000 km^2 [17]. Mangrove forests, seagrass beds and coral reefs in STP have an area of

approximately 0.8 [18], 15 [16] and 2 km^2 (estimated by comparison of satellite images of mangrove forests and seagrass beds habitats), respectively.

2.2. Compilation of Data

All available data on marine invertebrate occurrences (i.e., the presence of a species, or other taxon, at a particular place on a specified date) from MOZ and STP were downloaded from GBIF (http://www.gbif.org, accessed on 3 December 2018) (DOI's in Appendix A) into a dataset to assess taxonomic, geographic and temporal data gaps and bias of marine invertebrate data from the different coastal habitats of both countries. A total of 76,945 occurrences were assembled, of which terrestrial and freshwater species comprised more than 60% of the data (47,189 occurrences, mainly Insecta and Gastropoda occurrences). These were manually excluded from the initial dataset, as GBIF does not have an option to select marine occurrences. Data from museums' NHC databases and scientific literature (see Appendix A) were also incorporated into the dataset, following Darwin Core standards [19] (see fields used in Supporting Information Table S1). Only marine invertebrate occurrences identified at species or genus level were included in the dataset. The surveyed habitats mainly comprised mangrove forests, seagrass beds and coral reefs. Some occurrences were not collected in any of these three habitats, so they were grouped in two categories: coastal area for all other coastal habitats (e.g., rocky shores or sandy beaches) and offshore for occurrences 12 km until the limit of the Economic Exclusive Zone (EEZ).

The dataset was thoroughly reviewed and validated. A taxonomic review was conducted, and identifications were updated using the World Register of Marine Species (WoRMS, http://www.marinespecies.org, accessed on 23 February 2019). Some occurrences did not have any location data (locality and coordinates) and/or date of collection (some of this data was not accessed from the online databases sources). Occurrences without coordinates but with a good locality description were georeferenced using GEOLocate Collaborative Georeferencing tool (CoGe, https://coge.geo-locate.org, accessed on 18 March 2019) and Google Maps (https://www.google.com/maps, accessed on 18 March 2019). Missing dates on the dataset downloaded from GBIF were later retrieved by directly accessing the respective museum or institution database (e.g., Smithsonian-National Museum of Natural History or the California Academy of Sciences). The dataset was saved as an Excel spreadsheet and all basic analysis were performed using Excel tools (Microsoft Corporation, 2018, *Microsoft Excel*, available at: https://office.microsoft.com/excel).

2.3. Taxonomic Coverage

After a preliminary analysis of the dataset, the invertebrate phyla from three habitats (mangrove forests, seagrass beds and coral reefs) with more occurrences were selected for both countries. The main groups were Annelida, Arthropoda, Cnidaria, Echinodermata and Mollusca. Occurrences were analyzed by taxonomic level namely family, order, and genera.

2.4. Geographical and Temporal Coverage

Geographical marine invertebrate occurrences were analyzed by phyla and habitat along the coasts of MOZ and STP, simultaneously, allowing the comparison between both countries. Phyla occurrences in both countries were also mapped. To construct these maps, the dataset was loaded into ArcGIS ArcMap as a CSV file. Longitude and latitude of point coordinates were mapped to X and Y fields, respectively. The location of mangrove forests, seagrass beds and coral reefs in the coastal zones of MOZ and STP were retrieved as layers from the UN Environment World Conservation Monitoring Centre and the ReefBase websites. The coordinate reference system used was WGS 1984 [20].

Temporal occurrences were analyzed along the coasts of MOZ and STP by phyla and habitat. Several time intervals were considered for the temporal analysis. Occurrences before 1900 were grouped, as only 13 records were registered, while all other occurrences were grouped by decade.

3. Results

3.1. Compilation of Data

The marine invertebrate dataset had almost 30,000 occurrences in both MOZ and STP, but only 68% of these entries were used to assess taxonomic, geographic, and temporal data gaps and bias, as 32% had to be excluded. Exclusion of occurrences without location data (locality and coordinates, as occurrences that only had locality data were georeferenced) and/or date of collection (some of these data were retrieved directly from the museums online databases), represented 21% of the original dataset. The remaining 11% of exclusions were occurrences that did not have taxonomic identification at genus or species level. After data validation, a final dataset was compiled with 19,910 occurrences, of which 19,359 from MOZ and 551 from STP, and it is accessible through the GBIF portal under the Creative Commons Attribution 4.0 International license, at http://ipt.gbif.pt/ipt/resource?r=marineinvertebrate_moz_stp, accessed on 19 October 2022 [21].

3.2. Taxonomic Coverage

The total number of marine invertebrate occurrences per phyla is shown in Table 1. In both countries the phylum with more occurrences was the Arthropoda followed by the Mollusca. The phylum with the least occurrences was the Annelida.

Table 1. Marine invertebrate occurrences per phylum and corresponding frequency in Mozambique (MOZ) and São Tomé and Príncipe (STP).

Phylum	MOZ	STP	MOZ	STP
Annelida	45	21	0.23%	3.81%
Arthropoda	11,411	296	58.94%	53.72%
Cnidaria	649	24	3.36%	4.36%
Echinodermata	469	56	2.42%	10.16%
Mollusca	6785	154	35.04%	27.95%
Total	19,359	551	100.00%	100.00%

Mozambique's Arthropoda occurrences in the dataset (Figure 1) belonged mainly to the class Malacostraca (96%). Within this the order Decapoda was the most well represented (98%). Most decapod occurrences refer to the swimming crab family Portunidae (38%) and penaeid shrimps' family Penaeidae (37%).

The Mollusca occurrences in MOZ (Figure 1) were distributed mainly by two classes, Cephalopoda (52%) and Gastropoda (40%). Within the Cephalopoda, most occurrences belonged to the orders Myopsida (48%) and Sepiida (49%). All Myopsida occurrences belonged to the Myopsidae family and almost all of these belonged to the squid genus *Loligo* (96%). Sepiida occurrences were distributed in two families, Sepiidae (99%) and Sepiolidae (1%) and almost all occurrences of Sepiidae belonged to the cuttlefish genus *Sepia* (99.9%). Most of Mozambique's Gastropoda occurrences in the dataset belonged to the order Neogastropoda (55%). The families with more occurrences were the Muricidae (22%), Nassariidae (17%) and Conidae (15%).

Marine Arthropoda occurrences in STP (Figure 1) belonged mainly to the class Malacostraca (71%). Within the Malacostraca, the order Decapoda represented the majority of occurrences (56%) (and most of these belonged to the shrimp family Alpheidae (41%). As for the Mollusca in STP, most of the occurrences (Figure 1) belonged to the classes Bivalvia (41%) and Gastropoda (46%). Bivalvia occurrences belonged mainly to the order Nuculanida (24%) and most of Gastropoda occurrences belonged to the order Nudibranchia (52%).

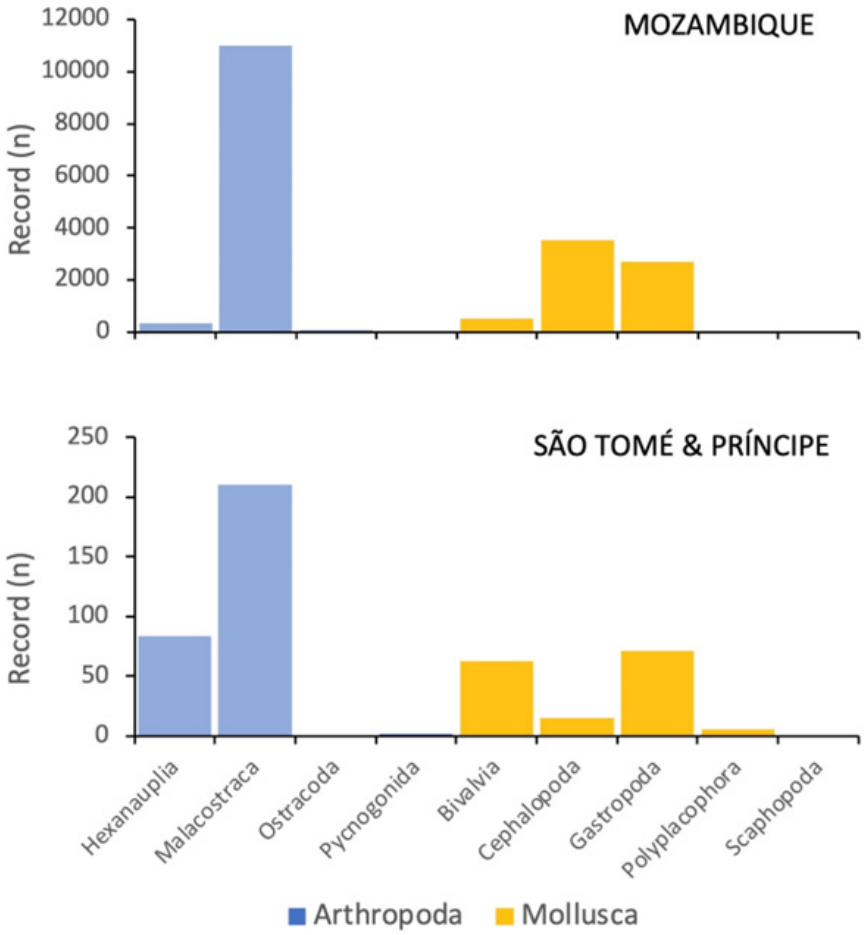

Figure 1. Represented orders under the most abundant phyla Arthropoda and Mollusca in datasets from Mozambique and São Tomé and Príncipe.

3.3. Geographical Coverage

Most marine invertebrate occurrences in MOZ (Figure 2) were collected in seagrass beds (56%) followed by offshore habitats (27%) and mangrove forests (12%). In STP (Figure 2) most marine invertebrates were collected in offshore (54%) and coastal area habitats (30%).

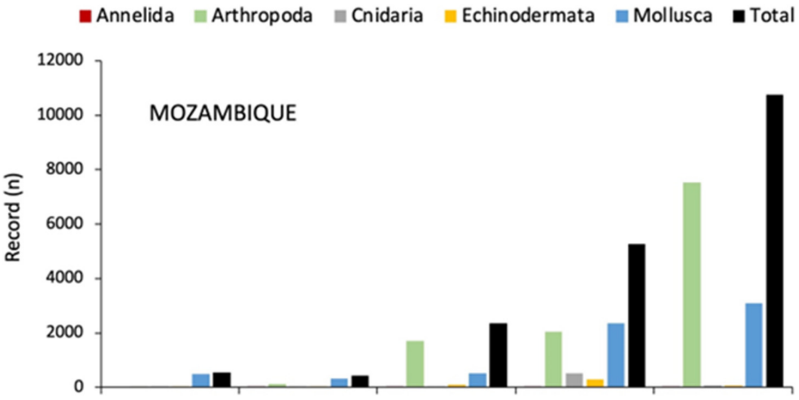

Figure 2. *Cont.*

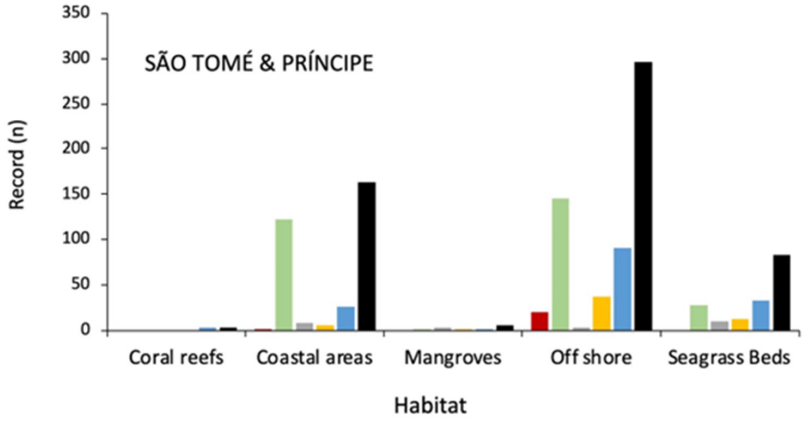

Figure 2. Most representative invertebrate phyla records within the main types of marine habitats in datasets from Mozambique and São Tomé and Príncipe.

Regarding phylum occurrences in each habitat (Figure 2), mangrove forests and seagrass beds in MOZ had a clear predominance of Arthropoda (72% and 70%, respectively) and Mollusca (22% and 29%, respectively). In coral habitats, most of the occurrences were Mollusca (89%) and only 5% of occurrences were cnidarians.

In STP (Figure 2) seagrass beds there was a prevalence of Arthropoda and Mollusca occurrences (34% and 40%, respectively). In coral reef habitats there were only Mollusca occurrences, and in mangrove forests there were only six occurrences (50% Cnidaria, 17% Arthropoda, 17% Echinodermata and 17% Mollusca).

Regarding phyla occurrences along the coastline, in MOZ (Figure 3A) Cnidaria and Mollusca were more represented in the northern provinces (Cabo Delgado and Nampula) while Arthropoda had more incidence in the central (Zambezia and Sofala) and southern provinces (Inhambane, Gaza and Maputo). In STP (Figure 3B) there were more Arthropoda occurrences in São Tomé Island and more Cnidaria and Echinodermata occurrences in the Príncipe Island.

3.4. Temporal Coverage

The dataset has a wide temporal range, the oldest occurrence was collected in 1816 and the most recent in 2019. The majority of occurrences were collected in the last two decades, representing 81% of all records (Figure 4). From 1960 to 1969 there was a slight increase in marine invertebrate occurrences comprising of 8.8% of all occurrences.

Most occurrences in MOZ (Figure 4) were registered in the decades 2000–2009 and 2010–2019. The third time interval with more occurrences was 1960–1969. In STP most occurrences (Figure 4) were registered in 1950–1959, 1970–1979 and 2000–2009.

Regarding phyla occurrences, in MOZ the Mollusca were recorded in each time interval analyzed, with 50% (or over) of occurrences in six-time intervals (1800–1899: 50%; 1900–1910: 67%; 1910–1919: 76%; 1950–1959: 51%; 1970–1979: 84%; 1990–1999: 76%). Arthropoda occurrences were most abundant in 1940–1949 (60%), 2000–2009 (66%) and 2010–2019 (65%). Other phyla occurrences in Mozambique rarely surpassed the 50% threshold, except for Cnidaria occurrences in 1930–1939 (81%) and Echinodermata occurrences in 1800–1899 (50%) and 1980–1989 (50%).

In STP, several time intervals had no marine invertebrate occurrences (1900–1909; 1920–1929; 1930–1939 and 1940–1949). Mollusca occurrences were most frequent in 1910–1919 (100%), 1980–1989 (87%) and 1990–1999 (60%), and Arthropoda in 1800–1899 (67%), 1970–1979 (51%), 2000–2009 (72%) and 2010–2019 (51%). Other phyla occurrences were lower than 50%, except for Echinodermata occurrences in 1960–1969 (50%).

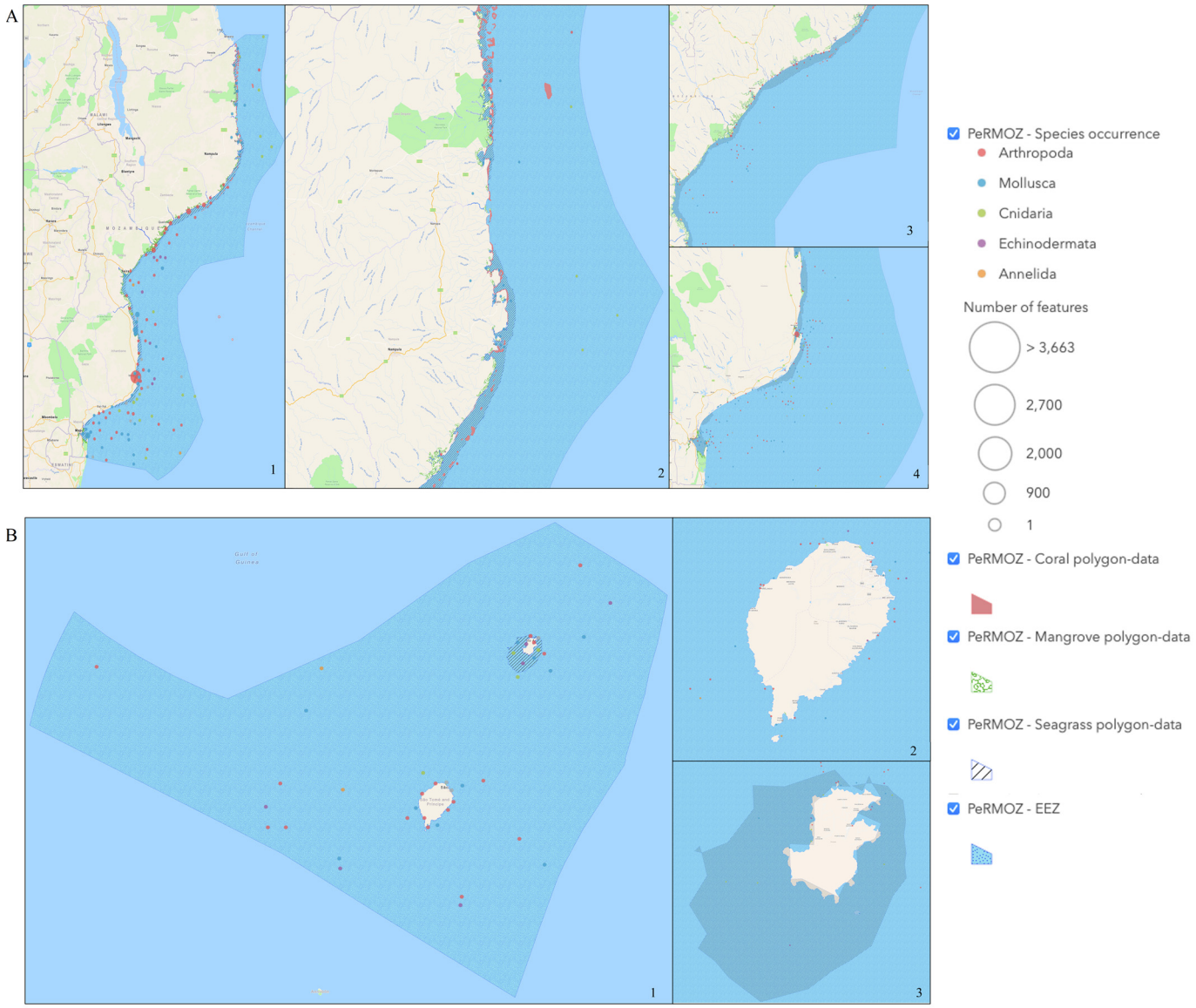

Figure 3. Phyla occurrences along the coast of (**A**) Mozambique, representing (1) a general view, (2) northern provinces, (3) central provinces and (4) southern provinces, respectively. (**B**) São Tomé and Príncipe. (1) general view, (2) São Tomé Island, (3) Príncipe Island.

The temporal distribution of marine invertebrate occurrences per studied habitat (mangrove forests, seagrass beds and coral) revealed several differences. In MOZ, occurrences were more distributed through time in mangrove forests, while in seagrass beds and coral habitats most occurrences were collected in more recent decades (2000–2009 and 2010–2019). All Annelida occurrences in mangrove forests were collected in 1950–1959. In STP, in mangrove forests and seagrass beds, most occurrences were collected in two main time intervals, 1950–1959 and 2010–2019. In coral habitats all occurrences were collected from 2010 to 2019.

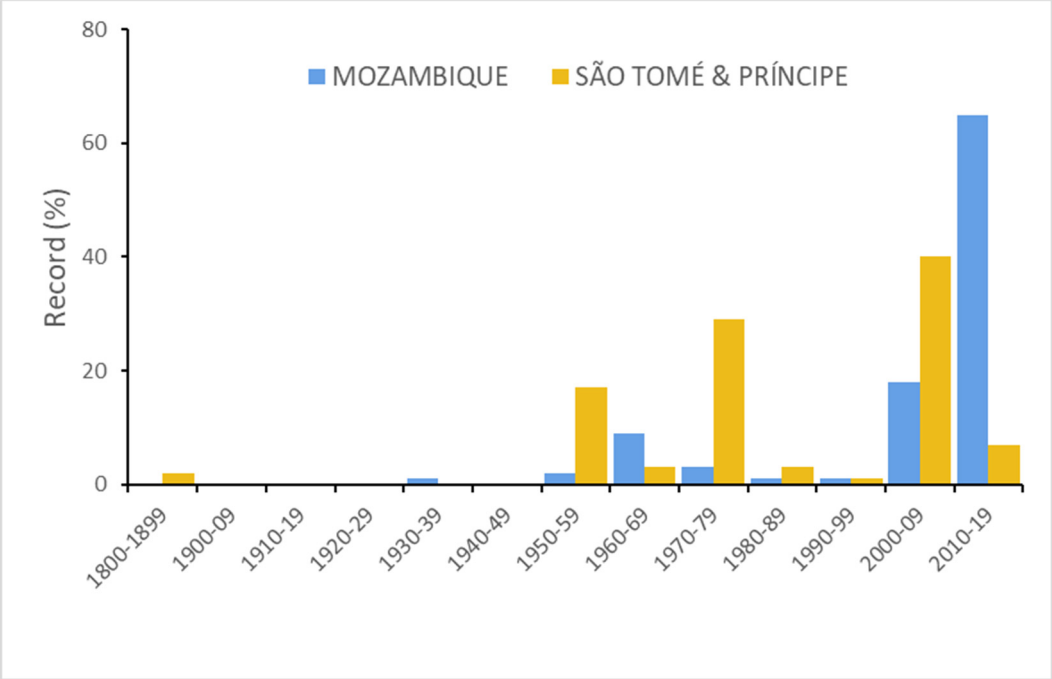

Figure 4. Pooled occurrences of invertebrate phyla of Mozambique and São Tomé and Príncipe by decade.

4. Discussion

Online databases such as GBIF, NHC collections, research data and published scientific papers are useful tools for gathering biodiversity data that are normally dispersed and sharing it in a structured and accessible way. We used these tools to compile a dataset with a total of 19,910 marine invertebrate occurrences from Mozambique and São Tomé and Príncipe, which enabled to assess taxonomic, geographic and temporal data gaps and bias in these two countries. Although GBIF uses Darwin Core as a standardized format for publishing data, many records did not have the complete information. For instance, 21% of the records downloaded from online databases were excluded because they had incomplete date of collection and/or location data. We were able to georeference most of the occurrences that did not have coordinates associated to the locality data, but some were too vague or just mentioned the country (e.g., West Africa, Mozambique). Another 11% of those records were also excluded because they did not have taxonomic identification at genus or species level. While this identification can be made a posteriori (and indeed this is one of the many benefits of NHC) the lack of date of collection and location data can invalidate the usefulness of an occurrence. Increasing the digitization of NHC data is essential to acquire a better insight into data gaps in both countries as many institutions are behind in this process. Furthermore, NHC managers and curators should reinforce the need to register the essential data (such as date and time of collection, coordinates, and locality—as accurate as possible, collector name, expedition name—if available, altitude or depth) when new entries are added to the collections. The data should then be added to GBIF, paving the way for FAIR data (findable, accessible, interoperable and reusable) [22].

A preliminary analysis of the dataset revealed a clear discrepancy in the volume of data between both countries. Mozambique's occurrences represent 97% of the data and São Tomé and Príncipe's only 3%. This could be due to several factors. São Tomé and Príncipe is a small country, its coastline represents 10% of that of Mozambique and its EEZ 32%. Higher education in the field of Biology is still recent in São Tomé and Príncipe. The University of São Tomé and Príncipe was established in 2014, while in Mozambique the University of Eduardo Mondlane was established in 1962 and this institution is responsible

for the functioning of the Estação de Biologia Marítima da Ilha de Inhaca (Marine Biology Station of Inhaca Island) [23].

In both countries the taxonomic groups with most occurrences represent species that are used as a food source or for decorative purposes, such as Portunidae (crabs), Penaeidae (shrimps), Myopsida (squids), Sepiidae (cuttlefish) and Neogastropoda (sea snails) in Mozambique, and Alpheidae (snapping shrimps), Nuculanida (clams) and Nudibranchia in São Tomé and Príncipe. This bias towards food resources is probably due to the fact that a large portion of the data downloaded from GBIF has been published by official institutions, such as IIP in Mozambique (National Institute of Fisheries) (see Appendix A). In the instance of Nudibranchia in São Tomé and Príncipe the majority were collected by researchers particularly interested in this taxonomic group (e.g., [24]). These gaps and limitation of taxonomic coverage can lead to biases that affect all future analyses of the data [25].

Overall, the geographical coverage in both countries has serious gaps. Considering that MOZ accounts for 60% of mangrove forests in mainland eastern Africa [15] it is surprising that only 12% of occurrences were collected in mangrove forests. In STP the three analyzed habitats were underrepresented in the data downloaded from GBIF, which is due to the sparse coverage these habitats have in the country [16,18]. In both countries, Cnidaria occurrences were underrepresented (MOZ) or completely lacking (STP) in coral habitats. This could demonstrate a bias in collecting (e.g., the main object of the studies carried out in these habitats were other taxonomic groups).

There is also a gap in historical records in both countries, more pronounced in STP than in MOZ. The decline in the number of occurrences in MOZ between the time intervals 1960–1969 and 1990–1999, can be explained by the political instability between 1964 and 1992 [26]. In STP, in the last decade (2010–2019) a decline in occurrences is noticeable (from 40% of occurrences in 2000–2009 to 7% in 2010–2019), despite the establishment of the Parque Natural Obô de São Tomé and Parque Natural Obô do Príncipe in 2006 and the Island of Príncipe Biosphere Reserve in 2012.

Marine macroinvertebrate biodiversity data gaps can only be filled by creating local databases and institutionalization of the data collection methods in an equative and sustainable way for each country. Its main justification would be the need to manage these resources, especially the endangered, unique and endemic ones. Both countries could adopt a similar approach to the Marine Strategy Framework Directive (MSFD) developed for the European Union. An inventory of the temporal occurrences, abundance and spatial distribution of marine macroinvertebrates as well as information regarding species composition, biomass and annual/seasonal variability [27] is essential.

This study provides baseline information relevant to a better understanding of marine invertebrate biodiversity data gaps and bias in mangrove forests, seagrass beds and corals along the coasts of Mozambique and São Tomé and Príncipe. This information can be further used to design informed sampling strategies leading to more refined datasets that can be used in management and conservation plans in both countries allowing future research to focus in completing data gaps.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/d15010070/s1, Table S1: Darwin Core fields used, descriptions in the Darwin Core were taken from the TDWG website (https://dwc.tdwg.org/terms/ (accessed on 29 July 2022)).

Author Contributions: Conceptualization, M.B. and A.M.C.; methodology, M.B. and H.N.; software, H.N.; validation, M.B., H.N. and A.M.C.; formal analysis, M.B.; investigation, M.B.; data curation, M.B. and H.N.; writing—original draft preparation, M.B. and A.M.C.; writing—review and editing, M.B., H.N., A.C., S.B., J.P. and A.M.C.; visualization, M.B., J.P. and A.M.C.; supervision, S.B., J.P. and A.M.C.; project administration, S.B., J.P. and A.M.C.; funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the project COBIO-NET (Coastal biodiversity and food security in peri-urban Sub-Saharan Africa: assessment, capacity building and regional networking in contrasting Indian and Atlantic Oceans) funded by Aga Khan Development Network (AKDN) and Fundação para a Ciência e a Tecnologia, I.P. (FCT). MB was funded by FCT through a PhD grant (Ref. SFRH/BD/147875/2019).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in accessible through the GBIF portal under the Creative Commons Attribution 4.0 International license, at https://www.gbif.org/dataset/3e0e4ec9-1905-4cae-9691-c0fa79361ac3 (Bento et al., 2022, DOI: 10.15468/w4s7cc) [21].

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Mozambique Occurrence Data

- *Acanthocephala:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.a9e0wq, accessed on 25 October 2018
- *Annelida:* GBIF.org Occurrence Download https://doi.org/10.15468/dl.cgskgn, accessed on 25 October 2018
- *Arthropoda:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.rsvrwu, accessed on 25 October 2018
- *Decapoda MUHNAC:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.wtl0gh, accessed on 3 December 2018
- *Ascidiacea:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.ayjvmg, accessed on 10 October 2018
- *Brachiopoda:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.lwdyvc, accessed on 25 October 2018
- *Bryozoa:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.bamx4r, accessed on 25 October 2018
- *Cnidaria:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.tdo85o, accessed on 25 October 2018
- *Echinodermata:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.ob3jcd, accessed on 25 October 2018
- *Hemichordata:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.rsyyeo, accessed on 25 October 2018
- *Mollusca:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.uzkzcv, accessed on 25 October 2018
- *Nematoda:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.guvpnx, accessed on 25 October 2018
- *Nematomorpha:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.csf3en, accessed on 25 October 2018
- *Nemertea:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.egimov, accessed on 25 October 2018
- *Platyhelminthes:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.vlotks, accessed on 25 October 2018
- *Porifera:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.kkre6f, accessed on 25 October 2018
- *Rotifera:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.nb6gme, accessed on 25 October 2018
- *Sipuncula:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.vwd9ve, accessed on 25 October 2018
- *Tardigrada:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.zodhhs, accessed on 25 October 2018

- *Thaliacea:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.ihck3c, accessed on 10 October 2018

Appendix A.2. São Tomé and Príncipe Occurrence Data

- *Annelida:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.p7vvlf, accessed on 10 October 2018
- *Arthropoda:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.q4jiv0, accessed on 10 October 2018
- *Ascidiacea:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.thevng, accessed on 10 October 2018
- *Bryozoa:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.2zowxw, accessed on 10 October 2018
- *Cnidaria:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.lrjhwj, accessed on 10 October 2018
- *Echinodermata:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.isduid, accessed on 10 October 2018
- *Mollusca:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.px6i6k, accessed on 10 October 2018
- *Nematoda:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.j8lkto, accessed on 10 October 2018
- *Nematomorpha:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.e9gwz5, accessed on 10 October 2018
- *Platyhelminthes:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.f107js, accessed on 10 October 2018
- *Porifera:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.krzegd, accessed on 10 October 2018
- *Sipuncula:* GBIF.org GBIF Occurrence Download https://doi.org/10.15468/dl.dlbbow, accessed on 10 October 2018

Appendix A.3. Literature Occurrence Data

- Ahyong, S.T. Cymonomid crabs of the MAINBAZA Expedition (Decapoda: Brachyura). *Zootaxa* **2014**, *3821*, 384–390. https://10.11646/zootaxa.3821.3.7
- Baba, K.; Tirmizi, N. A New Chirostylid (Crustacea, Decapoda, Anomura) From Deeper Parts of the Japanese Waters and Off the East Coast of Africa. *Proc. Jpn. Soc. Syst. Zool.* **1979**, *17*, 52–57. https://10.19004/pjssz.17.0_52
- Berggren, M. Periclimenes Nomadophila and Tulеariocaris Sarec, Two New Species of Pontoniine Shrimps (Decapoda: Pontoniinae), from Inhaca Island, Mozambique. *J. Crustac. Biol.* 1994, 14, 782–802.
- Berggren, M. Athanopsis rubricinctuta, New Species (Decapoda: Natantia: Alpheidae), a Shrimp Associated with an Echiuroid at Inhaca Island, Moçambique. *J. Crustac. Biol.* **1991**. https://10.2307/1548555
- Berggren, M. Dasella herdmaniae (Lebour) (Decapoda: Natantia: Pontoniinae) from Moçambique and Establishment of a New Species, Dasella brucei. *J. Crustac. Biol.* **1990**, *10*, 554–559.
- Brito, A.; Pena, A. Population structure and recruitment of penaeid shrimps from the Pungué River Estuary to the Sofala Bank Fishery, Mozambique. *West. Indian Ocean. J. Mar. Sci.* **2007**, 147–158.
- Cabezas, P.; Macpherson, E.; Machordom, A. Allogalathea (Decapoda: Galatheidae): a monospecific genus of squat lobster? *Zool. J. Linn. Soc.* **2011**, *162*, 245–270. https://10.1111/j.1096-3642.2010.00681.x
- Cabezas, P.; MacPherson, E.; MacHordom, A. Taxonomic revision of the genus Paramunida Baba, 1988 (Crustacea: Decapoda: Galatheidae): A morphological and molecular approach. *Zootaxa* **2010**. https://10.11646/zootaxa.2712.1.1

- Castro, P.; Williams, A.B.; Cooper, L.L. Revision of the family Latreilliidae Stimpson, 1858 (Crustacea, Decapoda, Brachyura). *Zoosystema Paris* **2003**, *25*, 601–634.
- Clark, P.F.; Galil, B.S. A revision of the xanthid genus Pilodius Dana, 1851 (Crustacea: Brachyura: Xanthoidea). *J. Nat. Hist.* **1993**. https://10.1080/00222939300770691
- Clark, P.F.; Paula, J. Descriptions of ten xanthoidean (Crustacea: Decapoda: Brachyura) first stage zoeas from Inhaca Island, Mozambique. *Raffles Bull. Zool.* **2003**, *51*, 323–378.
- Farfante, I.P.; Ivanov, B.G. Mesopenaeus Mariae, a New Species of Shrimp (Penaeoidea: Solenoceridae), the First Record of the Genus in the Indo-west Pacific. *J. Crustac. Biol.* **1982**, *2*, 303–313.
- Flores, A.A.V.; Saraiva, J.; Paula, J. Sexual maturity, reproductive cycles, and juvenile recruitment of Perisesarma guttatum (Brachyura, Sesarmidae) at Ponta Rasa mangrove swamp, Inhaca Island, Mozambique. *J. Crustac. Biol.* **2002**, *22*, 143–156. https://10.1651/0278-0372(2002)022[0143:smrcaj]2.0.co;2
- Fratini, S.; Vannini, M.; Cannicci, S.; Schubart, C.D. Tree-climbing mangrove crabs: A case of convergent evolution. *Evol. Ecol. Res.* **2005**, *7*, 219–233.
- Freitas, A.J. The Penaeoidea of southeast Africa. III—The Family Penaeidae (excluding Genus Penaeus). *Durban* **1987**.
- Freitas, A.J. The Penaeoidea of southeast Africa. II—The Families Aristeidae and Solenoceridae. *Durban* **1985**.
- Galil, B.S. An examination of the genus Philyra Leach, 1817 (Crustacea, Decapoda, Leucosiidae) with descriptions of seven new genera and six new species. *Zoosystema* **2009**, *31*, 279–320. https://10.5252/z2009n2a4
- Galil, B.S. Crustacea Decapoda: review of the genera and species of the family Polychelidae Wood-Mason, 1874, Mémoires du Muséum national d'histoire naturelle. **2000**.
- Hayashi, K.-I. Crustacea Decapoda: revision of Pasiphaea sivado (Risso, 1816) and related species, with descriptions of one new genus and five new species (Pasiphaeidae), Mémoires du Muséum national d'Histoire naturelle. **1999**.
- Holthuis, L.B. The Indo-Pacific scyllarine lobsters (Crustacea, Decapoda, Scyllaridae). *Zoosystema* **2002**, *24*, 499–683.
- Komai, T.; Chan, T.-Y. Two new pandalid shrimps and the discovery of the second specimen of the rare hippolytid shrimp Leontocaris bulga Taylor & Poore, 1998 (Crustacea, Decapoda) from the Mozambique MAINBAZA Cruise. *Zoosystema* **2010**, *32*, 625–641. https://10.5252/z2010n4a6
- Komai, T.; Wada, K. A revision of the estuarine crab genus Ilyograpsus Barnard, 1955 (Crustacea: Decapoda: Brachyura: Macrophthalmidae), with descriptions of a new genus and one new species. *Raffles Bull. Zool.* **2008**, *56*, 357–384.
- Lai, J.C.Y.; Ng, P.K.L.; Davie, P.J.F. A revision of the Portunus pelagicus (Linnaeus, 1758) species complex (Crustacea: Brachyura: Portunidae), with the recognition of four species. *Raffles Bull. Zool.* **2010**, *58*, 199–237.
- Lemaitre, R. A worldwide review of hermit crab species of the genus Sympagurus Smith, 1883 (Crustacea: Decapoda: Parapaguridae), Tropical Deep-Sea Benthos. **2004**.
- Lemaitre, R.; Rahayu, D.L.; Komai, T. A revision of "blanket-hermit crabs" of the genus paguropsis henderson, 1888, with the description of a new genus and five new species (Crustacea, anomura, diogenidae). *Zookeys* **2018**, *2018*, 17–97. https://10.3897/zookeys.752.23712
- Litulo, C. Population and reproductive biology of the fiddler crab Uca chlorophthalmus (Brachyura: Ocypodidae) from Inhaca Island, southern Mozambique. *J. Mar. Biol. Assoc. United Kingd.* **2006**, *86*, 737–742. https://10.1017/S0025315406013646
- Litulo, C.; Abreu, D.; Tudge. Reproductive biology of Calcinus laevimanus (Randall, 1840) (Decapoda: Anomura: Diogenidae) from Inhaca Island, southern Mozambique. *Biol. Anomura II* **2006**, *6*, 57–66.
- Litulo, C.; Macia, A.; Mantelatto, F.L.M. Fecundity and sexual maturity of the crab Macrophthalmus depressus (Brachyura: Ocypodidae) from Inhaca Island, Mozambique. *Afr. J. Aquat. Sci.* **2005**, *30*, 179–183. https://10.2989/16085910509503854

- Litulo, C.; Tudge, C. Population structure and breeding season of the hermit crab Diogenes brevirostris Stimpson, 1858 (Decapoda, Anomura, Diogenidae) from southern Mozambique. *J. Nat. Hist.* **2005**, *39*, 2887–2899. https://10.1080/00222930500206107
- Macia, A.; Quincardete, I.; Paula, J. A comparison of alternative methods for estimating population density of the fiddler crab Uca annulipes at Saco Mangrove, Inhaca Island (Mozambique). *Hydrobiologia* **2001**, *449*, 213–219. https://10.1023/A:1017502817087
- Macpherson, E.; Rodríguez-Flores, P.C.; Machordom, A. New sibling species and new occurrences of squat lobsters (Crustacea, Decapoda) from the western Indian ocean. *Eur. J. Taxon.* **2007**, *2017*, 1–61. https://10.5852/ejt.2017.343
- Mkare, T.K.; Groeneveld, J.C.; Teske, P.R.; Matthee, C.A. Comparative genetic structure in two high-dispersal prawn species from the south-west Indian Ocean. *Afr. J. Mar. Sci.* **2017**, *39*, 467–474. https://10.2989/1814232X.2017.1402089
- Penha-Lopes, G.; Torres, P.; Macia, A.; Paula, J. Population structure, fecundity and embryo loss of the sea grass shrimp Latreutes pymoeus (Decapoda: Hippolytidae) at Inhaca Island, Mozambique. *J. Mar. Biol. Assoc. United Kingd.* **2007**, *87*, 879–884. https://10.1017/S0025315407056767
- Saito, T.; Komai, T. A review of species of the genera Spongicola de Haan, 1844 and Paraspongicola de Saint Laurent & Cleva, 1981 (Crustacea, Decapoda, Stenopodidea, Spongicolidae). *Zoosystema* **2008**, *30*, 87–147.
- Sakai, K. Upogebiidae of the World (Decapoda, Thalassinidea). *Crustac. Monogr.* **2006**, *6*, 186.
- Sakai, K. Dr. R. Plante's collection of the families Callianassidae and Gourretiidae (Decapoda, Thalassinidea) from Madagascar, with the description of two new genera and one new species of the Gourretiidae sakai, 1999 (new status) and two new species of the Callia. *Crustaceana* **2004**, *77*, 553–601. https://10.1163/1568540041718019
- Serène, R. Crustacés Décapodes Brachyoures de l'océan Indien Occidental et de la Mer Rouge, Xanthoidea: Xanthidae et Trapeziidae. Faune Trop. **1984**.
- Simbine, L.; Marques, C.G.; Freitas, P.D.; Samucidine, K.E.; Gusmão, J.; Tavares, C.; Junior, P.G. Metapenaeus dobsoni (Miers, 1878), an alien Penaeidae in Mozambican coastal waters-confirmation by mtDNA and morphology analyses. *West. Indian Ocean J. Mar. Sci.* **2018**, *17*, 1–12.
- Spiridonov, V.A.; Apel, M. A new species and new records of deep-water Calappidae (Crustacea: Decapoda) from the Indian Ocean with a key to the Mursia Desmarest, 1823 species of the region. *J. Nat. Hist.* **2007**, *41*, 2851–2890. https://10.1080/00222930701770786
- Torres, P.; Penha-lopes, G.; Macia, A.; Paula, J. Population structure and egg production of the seagrass shrimp hippolyte kraussiana stimpson, 1860 (decapoda: Hippolytidae) at inhaca island, mozambique. *Invertebr. Reprod. Dev.* **2007**, *50*, 145–153. https://10.1080/07924259.2007.9652239
- Torres, P.; Penha-Lopes, G.; Narciso, L.; Macia, A.; Paula, J. Fecundity and brood loss in four species of fiddler crabs, genus Uca (Brachyura: Ocypodidae), in the mangroves of Inhaca Island, Mozambique. *J. Mar. Biol. Assoc. United Kingd.* **2009**, *89*, 371–378. https://10.1017/S0025315408002701
- Williams, A.B. Revision of the genus Latreilia Roux. *Quad. del Lab. di Tecnol. della Pesca* **1982**, *3*, 227–255.

References

1. Smith, V.S.; Blagoderov, V. Bringing Collections out of the Dark. *Zookeys* **2012**, *209*, 1–6. [CrossRef]
2. Beck, J.; Böller, M.; Erhardt, A.; Schwanghart, W. Spatial Bias in the GBIF Database and Its Effect on Modeling Species' Geographic Distributions. *Ecol. Inform.* **2014**, *19*, 10–15. [CrossRef]
3. Beck, J.; Ballesteros-Mejia, L.; Buchmann, C.M.; Dengler, J.; Fritz, S.A.; Gruber, B.; Hof, C.; Jansen, F.; Knapp, S.; Kreft, H.; et al. What's on the Horizon for Macroecology? *Ecography Cop.* **2012**, *35*, 673–683. [CrossRef]
4. Jetz, W.; McPherson, J.M.; Guralnick, R.P. Integrating Biodiversity Distribution Knowledge: Toward a Global Map of Life. *Trends Ecol. Evol.* **2012**, *27*, 151–159. [CrossRef] [PubMed]

5. Chapman, A.D. *Principles and Methods of Data Cleaning: Primary Species and Species-Occurrence Data*; Version 1; GBIF: Copenhagen, Denmark, 2005.
6. Maldonado, C.; Molina, C.I.; Zizka, A.; Persson, C.; Taylor, C.M.; Albán, J.; Chilquillo, E.; Rønsted, N.; Antonelli, A. Estimating Species Diversity and Distribution in the Era of Big Data: To What Extent Can We Trust Public Databases? *Glob. Ecol. Biogeogr.* **2015**, *24*, 973–984. [CrossRef] [PubMed]
7. Vandepitte, L.; Bosch, S.; Tyberghein, L.; Waumans, F.; Vanhoorne, B.; Hernandez, F.; De Clerck, O.; Mees, J. Fishing for Data and Sorting the Catch: Assessing the Data Quality, Completeness and Fitness for Use of Data in Marine Biogeographic Databases. *Database* **2015**, *2015*, bau125. [CrossRef] [PubMed]
8. Faith, D.; Collen, B.; Ariño, A.; Patricia Koleff, P.K.; Guinotte, J.; Kerr, J.; Chavan, V. Bridging the Biodiversity Data Gaps: Recommendations to Meet Users' Data Needs. *Biodivers. Inform.* **2013**, *8*, 41–58. [CrossRef]
9. Chapman, A.D. *Principles of Data Quality*; GBIF: Copenhagen, Denmark, 2005.
10. Costello, M.J.; Michener, W.K.; Gahegan, M.; Zhang, Z.Q.; Bourne, P.E. Biodiversity Data Should Be Published, Cited, and Peer Reviewed. *Trends Ecol. Evol.* **2013**, *28*, 454–461. [CrossRef] [PubMed]
11. Chapman, A.D. *Uses of Primay Species-Occurrence Data*; GBIF: Copenhagen, Denmark, 2005.
12. Garcia, D.A.; Amori, M.; Giovanardi, F.; Piras, G.; Groppi, D.; Cumo, F.; de Santoli, L. An Identification and a Prioritisation of Geographic and Temporal Data Gaps of Mediterranean Marine Databases. *Sci. Total Environ.* **2019**, *668*, 531–546. [CrossRef] [PubMed]
13. Lam-Gordillo, O.; Baring, R.; Dittmann, S. Ecosystem Functioning and Functional Approaches on Marine Macrobenthic Fauna: A Research Synthesis towards a Global Consensus. *Ecol. Indic.* **2020**, *115*, 106379. [CrossRef]
14. Soberón, J.; Peterson, A.T. Monitoring Biodiversity Loss with Primary Species-Occurrence Data: Toward National-Level Indicators for the 2010 Target of the Convention on Biological Diversity. *AMBIO A J. Hum. Environ.* **2009**, *38*, 29–34. [CrossRef] [PubMed]
15. Samoilys, M.; Pabari, M.; Andrew, T.; Maina, G.W.; Church, J.; Momanyi, A.; Mibei, B.; Monjane, M.; Shah, A.; Menomussanga, M.; et al. *Resilience of Coastal Systems and Their Human Partners in the Western Indian Ocean*; Fitzgerald, M.A., Ed.; IUCN ESARO, WIOMSA, CORDIO and the UNEP Naiorbi Convention: Nairobi, Kenya, 2015; ISBN 978-2-8317-1724-1.
16. Alexandre, A.; Silva, J.; Ferreira, R.; Paulo, D.; Serrão, E.A.; Santos, R. First Description of Seagrass Distribution and Abundance in São Tomé and Príncipe. *Aquat. Bot.* **2017**, *142*, 48–52. [CrossRef]
17. Ministério dos Recursos Naturais e Meio Ambiente Primeiro Relatório Nacional Da Biodiversidade. 2003. Available online: https://www.terrabrasilis.org.br/ecotecadigital/pdf/serie-biodiversidade--03-legislacao-ambiental-brasileira-grau-de-adequacao-a-convencao-sobre-diversidade-biologica.pdf (accessed on 25 November 2018).
18. Haroun, R.; Herrero Barrencua, A.; Abreu, A.D. Mangrove Habitats in São Tomé and Príncipe (Gulf of Guinea, Africa): Conservation and Management Status. *Coast. Res. Libr.* **2018**, *25*, 589–605. [CrossRef]
19. Wieczorek, J.; Bloom, D.; Guralnick, R.; Blum, S.; Döring, M.; Giovanni, R.; Robertson, T.; Vieglais, D. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* **2012**, *7*, e29715. [CrossRef] [PubMed]
20. Niza, H.; Bento, M.; Lopes, L.F.; Cartaxana, A.; Correia, A.M. A Picture Is Worth a Thousand Words: Using Digital Tools to Visualise Marine Invertebrate Diversity Data along the Coasts of Mozambique and Sao Tomé & Príncipe. *Biodivers. Data J.* **2021**, *9*, e68817. [CrossRef] [PubMed]
21. Bento, M.; Niza, H.; Cartaxana, A.; Bandeira, S.; Paula, J.; Correia, A.M. A Dataset of Marine Macroinvertebrate Diversity from Mozambique and São Tomé and Príncipe. Available online: https://www.gbif.org/dataset/3e0e4ec9-1905-4cae-9691-c0fa79361ac3 (accessed on 19 October 2022).
22. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Arie, B.; Blomberg, N.; Boiten, J.-W.; Bonino da Silva Santos, L.; Bourne, P.E.; et al. Comment: The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
23. Paula, J.; Bandeira, S. An Introduction to the Maputo Bay. In *The Maputo Bay Ecosystem*; Paula, J., Bandeira, S., Eds.; WIOMSA: Zanzibar Town, Tanzania, 2014; pp. 3–10. ISBN 978-9987-9559-3-0.
24. Pola, M.; Gosliner, T.M. The First Molecular Phylogeny of Cladobranchian Opisthobranchs (Mollusca, Gastropoda, Nudibranchia). *Mol. Phylogenet. Evol.* **2010**, *56*, 931–941. [CrossRef] [PubMed]
25. Mora, C.; Tittensor, D.P.; Myers, R.A. The Completeness of Taxonomic Inventories for Describing the Global Diversity and Distribution of Marine Fishes. *Proc. R. Soc. B Biol. Sci.* **2008**, *275*, 149–155. [CrossRef] [PubMed]
26. Neves, I.Q.; da Luz Mathias, M.; Bastos-Silveira, C. Mapping Knowledge Gaps of Mozambique's Terrestrial Mammals. *Sci. Rep.* **2019**, *9*, 18184. [CrossRef] [PubMed]
27. Howarth, W. The Marine Strategy Framework Directive. *J. Water Law* **2008**, *19*, 95–97.