

# A Dataset of Marine Macroinvertebrate Diversity from Mozambique and São Tomé and Príncipe

Marta Bento <sup>1,2,\*</sup>, Henrique Niza <sup>2,3</sup>, Alexandra Cartaxana <sup>2,4</sup>, Salomão Bandeira <sup>5</sup>, José Paula <sup>1,2</sup>  
and Alexandra Marçal Correia <sup>1,2,\*</sup>

<sup>1</sup> Department of Animal Biology, Faculty of Sciences, University of Lisbon, Campo Grande, 1749-016 Lisboa, Portugal

<sup>2</sup> MARE—Marine and Environmental Sciences Centre/ARNET—Aquatic Research Network, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal

<sup>3</sup> IPMA—Portuguese Institute for the Sea and Atmosphere, Avenida Doutor Alfredo Magalhães Ramalho 6, 1495-165 Algés, Portugal

<sup>4</sup> Museu Nacional de História Natural e da Ciência, Universidade de Lisboa, Rua da Escola Politécnica 56–58, 1250-102 Lisboa, Portugal

<sup>5</sup> Department of Biological Sciences, Eduardo Mondlane University, Maputo 1100, Mozambique

\* Correspondence: mabento@ciencias.ulisboa.pt (M.B.); amcorreia@ciencias.ulisboa.pt (A.M.C.)

**Abstract:** Marine macroinvertebrate communities play a key role in ecosystem functioning by regulating flows of energy and materials and providing numerous ecosystem services. In Mozambique and São Tomé and Príncipe marine macroinvertebrates are important for the livelihood and food security of local populations. We compiled a dataset on marine invertebrates from Mozambique and São Tomé and Príncipe through an extensive data search of digital platforms, scientific literature, and natural history collections (NHC). This dataset encompasses data from 1816 to 2023 and comprises 20,122 records, representing 617 families, 1552 genera, 2137 species, providing species occurrence in mangrove forests, seagrass beds, coral reefs, and other coastal and offshore habitats. The dataset has a Darwin Core standard format and has been fully released in the Global Biodiversity Information Facility (GBIF). It is accessible through the GBIF portal under the Creative Commons Attribution 4.0 International license. The data are standardized and validated with tools such as WoRMS, GEOLocate, and Google Maps. Therefore, they can be readily used for further studies on species richness, distribution, and functional traits. Overall, this dataset contributes baseline information on marine biodiversity for future research.

**Dataset:** <https://doi.org/10.15468/w4s7cc>.

**Dataset License:** CC-BY

**Keywords:** biodiversity; mangrove forests; seagrass beds; database; GBIF

**Citation:** 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. *Data* **2023**, *8*, 76. <https://doi.org/10.3390/data8050076>

Academic Editor: Jamal Jokar Arsanjani

Received: 7 March 2023

Revised: 10 April 2023

Accepted: 25 April 2023

Published: 25 April 2023



**Copyright:** © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Summary

Marine macroinvertebrate communities play a major role in ecosystem functioning by regulating flows of energy and materials, providing numerous ecosystem services, such as nutrient cycling and food web stability [1,2]. They are also used as a food and income source [3–5]. In Mozambique (MOZ) and São Tomé and Príncipe (STP) marine macroinvertebrates are important to the livelihood and food security of local populations.

Digitizing natural history collections (NHC) in online databases is an important source of information, both historical and current [6], that can provide an invaluable insight into how communities and ecosystems change over time. In some cases, long-term effects of global climatic changes can only be studied using NHC data [7]. Despite the numerous online databases, containing over one billion records [8], occurrence data from

marine macroinvertebrates from MOZ and STP are difficult to find. These two countries were chosen to provide comparisons between the Indian and Atlantic Ocean.

In this paper, we provide a dataset on five phyla (Annelida, Arthropoda, Cnidaria, Echinodermata and Mollusca) of marine invertebrates from Mozambique and São Tomé and Príncipe achieved through an extensive data search of digital platforms, scientific literature, and natural history collections. The taxonomic and geographical coverage were thoroughly reviewed and validated. The dataset is deposited in the Global Biodiversity Information Facility (GBIF) and can be downloaded at <https://www.gbif.org/dataset/3e0e4ec9-1905-4cae-9691-c0fa79361ac3> (accessed on 6 March 2023). All users can access the data under the Creative Commons Attribution 4.0 International license.

## 2. Data Description

### 2.1. Geographical Coverage

#### Geographical Description

Mangrove forests, seagrass beds, coral reefs, and other coastal and offshore habitats of São Tomé e Príncipe (STP) and Mozambique (MOZ).

### 2.2. Bounding Coordinates

#### 2.2.1. STP

West: 1.00  
East: 8.70  
North: 2.35  
South: −0.67

#### 2.2.2. MOZ

West: 47.43  
East: 32.49  
North: −10.18  
South: −34.12  
Geographic coordinate system is WGS84.

### 2.3. Temporal Coverage

#### 2.3.1. Begin

20 December 1816

#### 2.3.2. End

4 February 2023

### 2.4. Taxonomic Coverage

The data comprise 20,122 occurrences, representing 27 families, 42 genera, and 40 species of Annelida; 216 families, 516 genera, and 699 species of Arthropoda; 102 families, 204 genera, and 257 species of Cnidaria; 77 families, 146 genera, and 191 species of Echinodermata; and 195 families, 644 genera, and 950 species of Mollusca, totalling 617 families, 1552 genera, and 2137 species.

## 2.5. Data Structure

### 2.5.1. Data File

The original dataset of MarineInvertebrateMozSTP.csv is a text file according to Darwin Core terms defined by TDWG (2022)

### 2.5.2. File Format

The data files are text based in the UTF-8 encoding.

### 2.5.3. Variable Definitions

Variable definitions are shown in Supporting Information Table S1. Descriptions in the Darwin Core were taken from the TDWG website (<https://dwc.tdwg.org/terms/>; accessed on 29 July 2022).

## 2.6. Accessibility

### 2.6.1. License

This dataset is provided under a Creative Commons Attribution 4.0 International License (CC BY 4.0 international; <https://creativecommons.org/licenses/by/4.0/legalcode>, accessed on 30 August 2022).

### 2.6.2. Location of Storage

<https://www.gbif.org/dataset/3e0e4ec9-1905-4cae-9691-c0fa79361ac3>, accessed on 19 October 2022 (Bento et al., 2022, DOI: 10.15468/w4s7cc, accessed on 19 October 2022) [9].

## 3. Methods

### 3.1. Compilation of Data

Data on marine invertebrate occurrences from MOZ and STP were downloaded from GBIF (<http://www.gbif.org>, accessed on 10 to 27 December 2018 and on 1 to 5 March 2023) into a dataset. A total of 78,259 occurrences were assembled, of which 60% were terrestrial and freshwater species. These were manually excluded, as GBIF does not have an option to only select marine occurrences. Data from museums' collection databases and scientific literature were also incorporated into the dataset, following Darwin Core standards [10]. Missing dates on the dataset downloaded from GBIF were later retrieved by directly accessing the respective museum or the institution's own database.

### 3.2. Taxonomy and Systematics

A thorough taxonomic review was performed, and unaccepted identifications were updated using the World Register of Marine Species (WoRMS, <http://www.marinespecies.org>, accessed on 5 to 28 February 2019 and on 2 to 6 March 2023). Only occurrences identified at species or genus level were considered.

After a preliminary analysis, the phyla with most occurrences were selected: Annelida, Arthropoda, Cnidaria, Echinodermata, and Mollusca.

### 3.3. Georeference

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) were used to georeference occurrences without coordinates but with a good locality description.

### 3.4. Updates and Maintenance

Following the original methodology, a yearly update to include newly digitized data deposited in GBIF will be carried out. The most recent update increased the temporal

range from 2018 to 2023 and added 212 additional occurrences (three Annelida, 50 Arthropoda, 58 Cnidaria, 36 Echinodermata, and 65 Mollusca).

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/data8050076/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). M.B. 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 and 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> (accessed on 6 March 2023) [9].

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

## References

1. Worm, B.; Barbier, E.B.; Beaumont, N.; Duffy, E.; Folke, C.; Halpern, B.S.; Jackson, J.B.C.; Lotze, H.K.; Micheli, F.; Palumbi, S.R.; et al. Impacts of Biodiversity Loss on Ocean Ecosystem Services. *Science* **2006**, *314*, 787–790.
2. Stachowicz, J.J.; Bruno, J.F.; Duffy, J.E. Understanding the Effects of Marine Biodiversity on Communities and Ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **2007**, *38*, 739–766. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095659>.
3. Bandeira, S.; Gullström, M.; Balidy, H.; Samussone, D.; Cossa, D. Seagrass Meadows in Maputo Bay. In *The Maputo Bay Ecosystem*; Paula, J., Bandeira, S., Eds.; WIOMSA: Zanzibar Town, Tanzania, 2014; pp. 147–170, ISBN 9789987955930.
4. Paula, J.; Macamo, C.; Bandeira, S. Mangroves of Maputo Bay. In *The Maputo Bay Ecosystem*; Paula, J., Bandeira, S., Eds.; WIOMSA: Zanzibar Town, Tanzania, 2014; pp. 109–126.
5. Paula, J. Overall Assessment of the State of the Coast in the Western Indian Ocean. In *The Regional State of the Coast Report: Western Indian Ocean*; Paula, J., Ed.; UNEP-Nairobi Convention and WIOMSA: Nairobi, Kenya, 2015; p. 546.
6. Smith, V.S.; Blagoderov, V. Bringing Collections out of the Dark. *Zookeys* **2012**, *209*, 1–6. <https://doi.org/10.3897/zookeys.209.3699>.
7. Davis, C.L.; Guralnick, R.P.; Zipkin, E.F. Challenges and Opportunities for Using Natural History Collections to Estimate Insect Population Trends. *J. Anim. Ecol.* **2023**, *92*(2), 237–249. <https://doi.org/10.1111/1365-2656.13763>.
8. Ball-Damerow, J.E.; Brenskelle, L.; Barve, N.; Soltis, P.S.; Sierwald, P.; Bieler, R.; LaFrance, R.; Ariño, A.H.; Guralnick, R.P. Research Applications of Primary Biodiversity Databases in the Digital Age. *PLoS ONE* **2019**, *14*, e0215794. <https://doi.org/10.1371/journal.pone.0215794>.
9. 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*; Version 1.10; Occurrence Dataset; Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa: Lisboa, Portugal, 2022; <https://doi.org/10.15468/w4s7cc>.
10. Wiczorek, 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. <https://doi.org/10.1371/journal.pone.0029715>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.