**Fish community composition in the tropical archipelago of São Tomé and Príncipe**

Description of data files

Guillermo Porriñosa,b,c\*, Kristian Metcalfec, Ana Nunod,c, Manuel da Graçae, Katy Walkerb, Adam Dixonb,c, Márcio Guedesf, Lodney Nazaréf, Albertino dos Santosg, Liliana P. Colmanc, Jemima Dimblebyh, Marta Garcia-Docei, Annette C. Broderickc, Brendan J. Godleyc, Tiago Capela Lourençoa, Luisa Madrugab,e, Hugulay Albuquerque Maiaj, Berry Mulliganb, Philip D. Dohertyc

a cE3c - Centre for Ecology, Evolution and Environmental Changes & CHANGE - Global Change and Sustainability Institute, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

b Fauna & Flora, The David Attenborough Building, Pembroke Street, Cambridge, CB2 3QZ, UK

c Centre for Ecology and Conservation, Faculty of Environment, Science and Economy, University of Exeter, Penryn, Cornwall, TR10 9FE, UK

d Interdisciplinary Centre of Social Sciences (CICS.NOVA), School of Social Sciences and Humanities (NOVA FCSH), NOVA University Lisbon, Avenida de Berna, 26-C, 1069-061 Lisboa, Portugal

e Fundação Príncipe, Santo António, Príncipe, São Tomé and Príncipe

f Oikos – Cooperação e Desenvolvimento, Descida da Gracilda 88, Água Grande, São Tomé e Príncipe

g ONG MARAPA, Largo Bom Despacho, CP 292, São Tomé, São Tome e Príncipe

h School of Biological and Marine Sciences, University of Plymouth, Drake Circus, Plymouth, PL4 8AA

i University of La Laguna, Spain

j Universidade de São Tomé, Quinta de Santo António. C.P. 546, São Tomé e Príncipe

\* Corresponding author: gporrinos@alunos.ciencias.ulisboa.pt

Data available at <https://doi.org/10.5281/zenodo.13326274>.

Analyses available at<https://github.com/gporrinos/BRUVS_SaoTomePrincipe_2018_2020>

**See citations contained therein.**

**Table 1**: Description of files in “GEOFILES.zip” compressed folder.

|  |  |  |
| --- | --- | --- |
| **FILE** | **SOURCE** | **DESCRIPTION** |
| **STP boundaries.shp** | <https://gadm.org/> | Boundaries of São Tomé and Príncipe |
| **STP\_depth.tif** | Instituto Hidrográfico Português. (1995). Missão Hidrográfica de Angola e São Tomé – 1962 charts. | Depth data obtained from two nautical charts, digitised and interpolated into a raster using Triangular Interpolation Network in QGIS (v. 3.28.1). |

**Table 2**: Description of databases.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DATABASE** | **VARIABLE NAME** | | **DESCRIPTION** | |
| **videos.csv** | **video** | | **Unique ID for each deployment** | |
| date | | Date | |
| area | | Code of each area of sampling to aid sampling logistics. | |
| number | | Number of sampling within area of sampling. | |
| island | | Island (classified as ST, São Tomé; PC, Príncipe, and TI, Tinhosas) | |
| waypoint | | Waypoint number | |
| time | | Time of deployment | |
| depth | | Depth, measured in the field using a handheld depth sounder (EchoTest Plastimo II) | |
| camera | | Code of individual camera used in that deployment (A, B, C, D…) | |
| collection | | Time of retrieval of the camera. | |
| div\_case | | Code of individual diving case used in the deployment (useful to ID faulty equipment) | |
| battery | | Code of individual battery used in the deployment (useful to ID faulty equipment) | |
| sd\_card | | Code of individual SD card used in the deployment (useful to ID faulty equipment) | |
| valid | | Indicates whether the deployment is valid (“sim”) or not (“nao”) | |
| sampling\_round | | Which sampling round (year + season, see article) | |
| periodo | | 1st sampling period (09:00 to 14:00) or 2nd (afternoon, 15:00 onwards). Afternoon repeats were conducted to study potential differences in community composition at dawn. It was deemed unfeasible, and the data was not used in the published article due to low sample size. | |
| habitat | | Broad habitat type classified from field of view (Rock, Sand or Maerl) | |
| site | | Name of the site. Only useful for logistical reasons. | |
| lat | | Decimal latitude of the site. | |
| lon | | Decimal longitude of the site. | |
| viewer | | Who analysed the video. | |
| max\_time | | How long was the video analysed for. | |
| valid | | Was the video valid? | |
| failure | | If it was not valid, why it was not? (i.e. camera failure, loss of camera…) | |
| sampling\_type | | Random sampling or “transect”? Transects were non-random deployments. See article. | |
| sampling\_time | | Normal sampling? Or afternoon repeat? Afternoon repeats were *not* used. | |
| **maxn.csv** | | **video** | | **Unique ID for each deployment** |
| species | | Species |
| maxn | | MaxN |
| minutes | | Time in which the species appears in the video (counting from start of analysis) |
| file | | Name of the video file. |
| time\_in\_file | | Time in which the species appears *in the* *video file*. GoPro divides recording in several video files, to avoid exceedingly large video files. This column refers to the time *in the video file*. This allows to review records in the video. |
| viewer | | Who analysed this video? |
| **species.csv** | | species\_original | | Name of the species as it appears in the maxn.csv database |
| species | | Name of the species, standardised. |
| speciessimple | | Simplified species names for analyses (some species were grouped at genus level for analyses) |
| specieslist | | Only species ID’d to species level (for species list) |
| idlevel | | Taxonomic level at which the taxa was identified |
| type | | Teleost, elasmobranch, crustacean… |
| family | | Family |