

Bendima: a database for marine macro-invertebrate bycatch data designed to improve reproducibility in benthic ecology

by

Alexis MARTIN* (1), Jonathan BLETTERY (1), Agnès DETTAÏ (2),
Nicolas ROSSET (3) & Yann GOUSSEAU (3)



© SFI
Submitted: 5 Jan. 2022
Accepted: 2 Mar. 2023
Editors: F. Teletchea, A. Tilley

Key words

Benthos
Bycatch
Fisheries
Reproducibility crisis
Ecology
DNA barcoding
Database
Image

Abstract. – The difficulty of identifying marine macro-invertebrates and the lack of experts, added to the growing use of complex modeling approaches based on massive datasets, has led to a reproducibility crisis in benthic ecology. Improving the reliability of identification remains a key factor to increase the quality of raw data. We developed the database Bendima to manage benthic macro-invertebrate bycatch data from the scientific survey of the French Southern Ocean and Indian Ocean fisheries. This database is structured to store observations of macro-invertebrates in the form of images of the caught organisms associated to sampling effort data and molecular data, which allows for ongoing amendments to identifications and crossreferencing with barcode data. Once uploaded and stored as digital images, the Bendima observations data underpinning models can be fully assessed, criticized and compared. Here, we describe the Bendima system and provide an overview of the contents for teams involved in biodiversity database development, benthic ecology or fisheries monitoring.

Résumé. – Bendima : une base de données pour les captures accessoires de macro-benthos développée pour améliorer la reproductibilité des travaux de recherche en écologie benthique.

La difficulté de l'identification des macro-invertébrés marins, la pénurie de systématiciens experts de ces groupes et l'utilisation croissante de techniques de modélisation complexes basées sur des jeux de données massifs ont conduit à une crise de reproductibilité des travaux de recherche en écologie benthique. Dans ce contexte, améliorer la fiabilité des données d'identifications demeure un facteur clé de l'amélioration de la qualité des données brutes utilisées dans les modèles. Dans le cadre de notre programme de suivi scientifique des pêcheries françaises de l'océan Austral et de l'océan Indien, nous avons développé la base de données Bendima qui a pour objectif de gérer les données de captures accessoires de macro-invertébrés benthiques. Cette base de données permet de stocker les observations de macro-invertébrés sous la forme de collections d'images d'organismes, de métadonnées des protocoles d'acquisitions et de données de barcoding moléculaire, autorisant un retour aux identifications à l'origine des données brutes exploitées par les modèles. Les résultats des travaux de modélisation reposant sur les jeux de données extraits de Bendima peuvent être ainsi pleinement évalués et comparés, la bancarisation des données primaires utilisées dans les modèles n'étant pas limitée à de simples tables d'occurrences non vérifiables. Nous fournissons dans cet article une description globale de la structure de Bendima et un aperçu général de ses contenus. Notre article a pour objectif de présenter notre approche méthodologique à la communauté scientifique, notamment aux équipes travaillant sur le développement de bases de données de biodiversité, la recherche en écologie benthique ou le suivi scientifique de pêcheries.

INTRODUCTION

The concept of reproducibility consists of a series of guarantees provided by scientists that allow anyone to replicate the process they used to produce their observations (Dunlap, 1926; Popper, 1959, 1963). These guarantees ensure the possibility for other scientists to complete, confirm or refute the observations reported in a study and thus to discuss their respective results and interpretations on a rational basis (Fraser *et al.*, 2020). For experimental biology, reproducibility is the main constraint affecting the design of any experimental device (*e.g.* Desjardins *et al.*, 2021). For

field biology, reproducibility requires the full report of the biases induced by the non-controlled observation conditions (Filazzola and Cahill, 2021; Powers and Hampton, 2019).

In ecology, modern developments in bio-computing, instrumentation and web availability of datasets have resulted in a strong increase of modelling studies based on massive datasets and complex computing processes. The use of modelling also increased in response to the need of large scale biodiversity conservation studies, for which prediction techniques constitute a suitable cost-effective approach (*e.g.* Urbina-Cardona and Flores-Villela, 2010; Adhikari *et al.*, 2019; Pecchi *et al.*, 2019). The rise of modelling in ecology

(1) UMR BOREA 7208, Département Adaptations du Vivant, Muséum national d'Histoire naturelle, 57 rue Cuvier, 75231 Paris CEDEX 05, France. alexis.martin@mnhn.fr, jonathan.bletery@mnhn.fr

(2) UMR ISYEB 7205, Département Origine et Évolution, Muséum national d'Histoire naturelle, 57 rue Cuvier, 75231 Paris CEDEX 05, France. agnes.dettai@mnhn.fr

(3) UMR LTCI 5141, Laboratoire Traitement et Communication de l'Information, Télécom Paris, 19 place Marguerite Perey, 91120 Palaiseau, France. nicolas.rosset@inria.fr, yann.gousseau@telecom-paris.fr

* Corresponding author

led to what several authors name the “reproducibility crisis” (Baker, 2016; Kelly, 2019; Mondelli *et al.*, 2019; Nichols *et al.*, 2019). Not limited to ecology, this crisis affects several scientific disciplines marked by the recent huge increase of the dataset sizes, added to the strong increase of the complexity level of analytical processes (Nekrutenko and Taylor, 2012; Erdemir and Sauro, 2016; An, 2018). Indeed, the traditional way to share scientific results, in the form of reports or academic publications, does not fit anymore the requirements of the Big Data era: traditional media is not sufficient to fully report the description of complex computing processes, and tables within an article cannot be large enough to include all the massive raw data used nowadays to compute models (Fidler *et al.*, 2017). To face the reproducibility crisis in ecology, various solutions and guidelines have been proposed (Cushing *et al.*, 2018; Feng *et al.*, 2019; Powers and Hampton, 2019; Filazzola and Cahill, 2021). First, the full recording of the code and the software environment of the modelling process has been identified as an essential condition to the reproducibility of an analysis (Kasperek *et al.*, 2015; Pauliuk *et al.*, 2015; Mahajan *et al.*, 2019; Culina *et al.*, 2020). Sharing raw data and models, within supplementary materials, data papers or open-access databases, has been identified as a secondary key point of the reproducibility of the modelling studies (Reichman *et al.*, 2011; Peng, 2016; Manninen *et al.*, 2017; Radosevic *et al.*, 2020). In addition to this, providing the taxonomic references (such as taxa description, anatomical or molecular criteria, identification keys) used to produce the observations of the living organisms has been highlighted as a specific issue for ecology (Monckton *et al.*, 2020). Its purpose is to provide guarantees on the reliability of the taxa or species identifications. However, although its generalization may constitute a real improvement, this requirement is not sufficient to allow the full reproducibility of the modelling process, given the elementary unit of the analysis is still missing: the observation itself. In systematics, this issue is resolved by the conservation of reference voucher specimens on which the published measurements and observations have been performed: preserved in institutional natural history collections to ensure their long-time availability. Reference specimens provide to the scientific community a materialized proof supporting the published raw data (Remsen, 1995; Ride, 1999). In ecology, however, observations based on large number of organisms are required. For practical and ethical reasons, all the observed organisms cannot be conserved in the form of natural history specimens in order to constitute materialized proof of the field observations. This constitutes the blind spot of reproducibility in ecology: even when identification tools are mentioned, how are we to guarantee that observations made by a scientist can be compared to the observations made by another scientist? In the field, observations are only temporarily stored in the brain of each observer, and quickly

deleted from their memory after identifications, counting and measurements have been recorded in a notebook or a computer in the reduced form of a data table. With the growing use of heterogeneous datasets aggregated from various observers, this problem is compounded as the amount of data used for modelling continues to increase.

For marine invertebrates, this problem is critical. In ecology, marine invertebrates are considered as a ‘group’ in the sense of a ‘component’ of the ecosystems. Some studies are based or focused on this level of aggregation. However, marine invertebrates do not constitute a single taxonomic group. To date, the World Register of Marine Species recognizes 90,064 species of marine invertebrates in 33 phyla (not considering the Chordata phylum, which includes both invertebrates and vertebrates) (Costello *et al.*, 2013). Because this huge number of taxa requires a wide range of specialists, marine invertebrates cannot be easily identified in the field by a single observer or a limited group of observers, in particular when observers are not specialists and have to carry out multiple observation protocols (Giangrande, 2003). In addition, research in ecology of marine invertebrates faces a critical lack of specialists able to identify the taxa (Costello *et al.*, 2010). Various partial solutions have been developed. Bypassing the step of rigorous species identification by recording the presence of morphospecies is a common practice in benthic ecology (Brind’Amour *et al.*, 2014). This approach is based on the grouping of similar-looking organisms into object classes (Beattle and Oliver, 1994). The use of morphospecies assumes that the diversity of the species assemblages is signalled by the diversity of forms that organism takes (Brind’Amour *et al.*, 2014). However, the main limitation of this approach is the poor reproducibility level: morphospecies are not defined according to taxonomic hypothesis, which limits the possibility to compare, or complete, studies based on such raw data. The barcoding approach is a second solution, which has been developed to compensate for the lack of specialists in taxonomy (Telletchea, 2010). It is based on the use of molecular barcodes to identify organisms, with no need to use anatomical and/or discrete identification criteria (Stoeckle and Hebert, 2008; Grant and Linse, 2009; Adamowicz, 2015). This solution meets the requirements of reproducibility, but for marine invertebrates, this approach also induces a strong constraint: the need to extract tissues from the organisms. This prevents the use of barcoding techniques to identify all the organisms observed in the field.

Reproducibility of benthic ecology studies based on marine invertebrate data is an important issue for the scientific monitoring of the French fisheries occurring in the Southern Ocean and the south of the Indian Ocean (Duhamel and Williams, 2011; Hureau, 2011; Martin *et al.*, 2021). Since 2015, when the epibenthic macro-invertebrate component was introduced into the program, the survey has includ-

ed the characterization of the benthic habitats of the fishing areas and the mapping of the ‘Vulnerable Marine Ecosystems’ as defined by the Commission for the Conservation of Antarctic Marine Living Resources. This makes it necessary to produce raw data on the presence of the epibenthic marine invertebrate species and to use modelling approaches to characterize related benthic assemblages and species distribution patterns (Martin *et al.*, 2019). Modelling results are used to assess the impact level of the fishing activity and to design marine protected areas (MPA) (Martin *et al.*, 2019). In that context, reproducibility is a crucial requirement. The possibility to compare modelling outputs from various areas of the French territories with the results produced in other areas, is essential for global understanding of the functioning of the benthic ecosystems. This is crucial to build an integrated conservation policy for the related fisheries and to assess the efficiency of management measures. Moreover, when results are used as a basis for the design of MPAs, reproducibility also meets the requirements of transparency in a context marked by divergent stakeholders (for example fishing companies versus environmental NGOs) (Yates and Schoeman, 2015; Owusu *et al.*, 2020).

We developed the Bendima database to face the challenge of reproducibility by focusing on the improvement of the quality of the macro-invertebrate raw data, as we identified this point as a blind spot in benthic ecology. Bendima is used to manage by-catch data of macro-invertebrates collected by scientific observers during commercial fishing or scientific surveys dedicated to fish biomass assessment (Martin *et al.*, 2021). We demonstrated in a former study the possibility to use such data as inputs for modelling in order to characterize benthic ecosystems and related species assemblages (Martin *et al.*, 2019). In addition to reproducibility, a driving need for Bendima was the specific deployment context of our fishery monitoring program (Martin *et al.*, 2021). The field constraints of this program limit the availability of on-board scientific observers with sufficient expertise in marine invertebrate identification, and highlight the need for a cost-effective process for laboratory tasks (one full-time permanent benthologist with temporary participants). This study is a proof of concept of Bendima and constitutes a methodological proposal for the scientific communities involved in benthic marine invertebrate ecology, as well as for teams involved in fisheries and marine protected areas monitoring programs.

MATERIALS AND METHODS

The starting point of the creation of Bendima consisted in the choice of the nature of the raw data to be collected. We faced the following constraints:

- collection of data is performed by field observers with limited availability and no skills in benthos identification;
- rough sea conditions of the Southern Ocean and operating on commercial fishing vessels exclude the possibility of having on-board laboratory facilities;
- recording of observations in the form of a data table with non-verifiable identifications does not meet reproducibility requirements for benthos;
- all the collected organisms cannot be conserved and sent to the laboratory;
- the need for distributions of relative abundance of the taxa comprising benthic communities, not merely presence/absence data.

In order to deal with all these constraints, we decided to structure Bendima’s raw data according to four heterogeneous interrelated components:

- (1) images of benthic macro-invertebrates collected in field by scientific observers as a materialized proof of the observations, a suitable and non-time-consuming solution regarding their limited time and skills in benthos identification;
- (2) identification of the taxa and counting of the organisms performed on the images in the laboratory by benthos specialists;
- (3) sub-samples of representative specimens collected *in situ* by scientific observers and identified in the laboratory with DNA barcoding techniques (Hebert *et al.*, 2003; Ratnasingham and Hebert, 2007), providing voucher specimens and reliable taxonomic data from molecular approaches;
- (4) recording of full metadata describing sampling stations, sampling techniques and sampling effort.

The storage of all this information within a database constitutes the key element of improved reproducibility. For modelling studies based on data recorded into Bendima, it is possible to go back to the observations to check taxa identification, evaluate quality and make corrections, so modelling results can be fully assessed and compared. In addition to the storage of the computing process (the software environment, script and the model parameters), the availability of the images linked to identifications, DNA sequences, and metadata allow a complete storage of the entire modelling process, which is impossible when availability of raw data is limited to a simple table of taxa occurrences.

Various citizen science programs provided us with inspiration to design the field protocols and the pipeline for data treatment (*e.g.* Silvertown, 2009; Deguines *et al.*, 2012; Raoult *et al.*, 2016; Swanson *et al.*, 2016; Garcia-Soto and van der Meer, 2017; Hermoso *et al.*, 2021).

Field data collection protocol

The protocol to collect the data starts on-board with the spreading and the sorting of the bycatch of benthic macro-invertebrates by the scientific observers (Fig. 1A). The

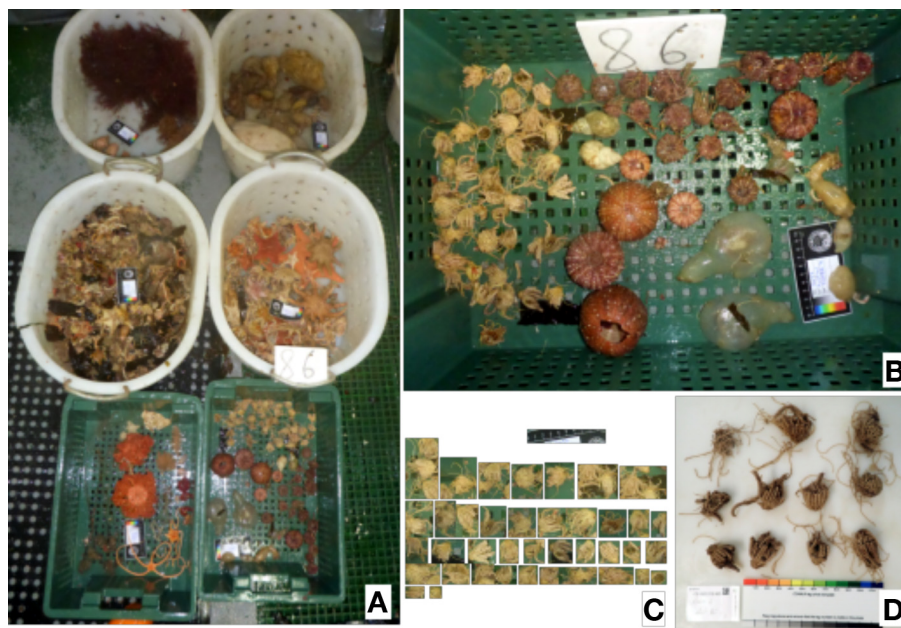


Figure 1. – Images and samples collection; sorting of the caught organisms (A), full photographing (B), taxa identification/counting/measurement and storage in the Bendima database in the form of cropped images (C), conservation of representative sub-samples for taxonomy and DNA barcoding (D).

sampling effort is defined according to the fishing gear and the geographical location. For scientific surveys based on trawling, full bycatch of benthos is sorted. For commercial activities carried out onboard industrial longliners within the French Exclusive Economic Zones, 25% of each longline is sampled during longline retrieval (for short description of the fishery monitoring program including the Bendima project and the full description of the data curation strategy see Martin *et al.*, 2021).

Organisms are grouped by morphospecies after a quick visual examination only (Fig. 1B). No detailed identification is performed at this step. Each batch of organisms is then photographed with a size reference scale (Fig. 1C). The batches of organisms are also weighed to estimate, for each station, the overall biomass impacted by the fishing activity. Whenever possible, depending on the field constraints, single organisms or colonies can be weighed and photographed separately to obtain length-weight relationship data. A subsample of representative organisms is preserved for each station. Preserved specimens are packed and labelled, and then frozen or fixated into 90% ethanol (Fig. 1D). At the end of the process, the remaining organisms are discarded.

Laboratory data process

Preserved specimens, collected images and stations metadata are sent to the laboratory at the end of each fishing cruise. Metadata include date, latitude, longitude, depth, gear type, sampling effort, fishing effort calculated in number of hooks or time of trawling, and total weight of benthos bycatch. Metadata and images raw files are uploaded to Bendima. Images are then analysed by a benthologist to identify all the taxa present and count the number of indi-

viduals. This task is performed using a tool we have created, based on the ImageJ software (Abràmoff *et al.*, 2004) with a dedicated set of macro-commands allowing partial automation of file annotation recording. All the images are examined. Every organism/colony is detected and a crop of each of them is extracted and saved as a new file into Bendima (Fig. 1C). Every organism is identified as precisely as possible according to the anatomical structures which can be observed. Identification varies from the Class level to the Species level. Identifications are recorded into Bendima in addition to the reference of the sampling station for each observed organism. Moreover, cropped images of the size scales corresponding to the pictures are also stored into Bendima, to perform measurements on the photographed organisms (Fig. 1C).

At the end of this first process, full abundance distributions for each of the taxa observed in the sampling stations can be extracted from Bendima. This allows the calculation of various indices and statistics for each taxon, such as kilometeric abundance index, complete abundance tables or impacted biomass estimations when organism counts interact with length-weight relationship data.

DNA barcoding

Preserved specimens are integrated into our natural history reference collection, to be identified by taxonomists and barcoded once DNA has been extracted. For the barcoding, a multi-marker approach is adopted. Working on multiple phyla poses problems for sequence amplifications and identification. Some markers like the protein coding mitochondrial gene Cytochrome Oxidase 1 (COI) can have limited application in amplification of some groups (Ascid-

acea for instance, Monniot *et al.*, 2011), and many phyla or even more restricted groups require custom-designed primers, considerably complicating the sequencing of molecular data. Three sequence markers were selected to cover diverse sequence divergence scales: partial COI, 18S rDNA and 16S rDNA.

COI is the reference barcoding marker for metazoans. It provides precise identifications in many benthic groups, although close species can be too similar to be differentiated in some groups for instance in Anthozoa, Porifera, but also some other groups (Neigel *et al.*, 2007). 18S rDNA sequences are largely used for interstitial fauna (Fonseca *et al.*, 2018). Closely related species are often not distinguishable, but assignation to higher taxonomic ranks is generally very good. In addition, this marker is easy to amplify for highly diverse samples as there are almost universal primers (for instance Machida and Knowlton, 2012) with PCR success rate of almost 100% on most eukaryotic groups. 16S rDNA has also been used for eukaryotic groups and has better resolution than 18SrDNA for species level. There are primers with a large taxonomic range for short and longer fragments (Palumbi and Benzie, 1991; Sarri *et al.*, 2014) with generally a good amplification success compared to COI (Leray *et al.*, 2013; Sarri *et al.*, 2014; Shokralla *et al.* 2015).

To optimise sequencing, a double multiplex NGS sequencing approach (combining multiple indexed libraries and PCRs with tagged primers) is used: either double multiplex mixed banks with fragmentation (Hinsinger *et al.*, 2015) for the longer amplifications covering complete COI fragments and 18S or alternatively with primers tagged for each specimen (Shokralla *et al.*, 2015) for short amplifications when longer amplifications were not working. The primers and tags for COI are from Leray *et al.* (2013). The primers for 18S follow Machida and Knowlton (2012) and 16S are from Palumbi (1996), with the same tags added as described in Leray *et al.* (2013).

At the end of the barcoding process, DNA sequences and identifications obtained by comparing the sequences to reference databases (Ratnasingham and Hebert, 2007) are stored in Bendima. This allows the main set of identifications based on images only to be linked to taxonomic data from barcoding.

RESULTS AND DISCUSSION

Overview of Bendima contents and achievements

The creation of Bendima started with a first phase of development, including testing and adjustment, between 2015 and 2017. The first field protocols to collect images and specimens of benthos bycatch on board commercial vessels were provided to the fishery scientific observers during this period. Images and specimen collection protocols were also

tested in 2016 during the Pige scientific survey (Duhamel *et al.*, 2019). In the same period, images of representative subsamples of bycatch of benthos collected in 2009 and 2013 during Poker 2 and Poker 3 surveys (Duhamel *et al.*, 2019), before the creation of Bendima, have been post-processed and included into the project. Post-processing on an existing collection of available images of benthos allowed us to build Bendima and the associated tools. In 2017, we deployed the Bendima field protocols during the Poker 4 scientific survey (Duhamel *et al.*, 2019) to obtain a full observation coverage of the bycatch of benthos. This included the full collection of images of the captured benthic invertebrates to permit abundances recording, and the full DNA barcoding of the conserved representative specimens.

To date, observations of 92,447 single organisms or single colonies of marine epibenthic macro-invertebrates are stored in Bendima in the form of images with full metadata about the sampling stations. For all the datasets produced after the deployment of the database, Bendima enables to derive abundance distributions from the observations. For data based on presence/absence protocols, which were produced before the deployment of the database and post-processed to be included in Bendima, the database computes a proxy of abundances in the form of probabilities of detection.

The taxa diversity of Bendima falls into 13 phyla. The phyla present are the Echinodermata (46%), the Cnidaria (25%), the Porifera (11%) and the Chordata (8%) (Fig. 2). Organisms are identified from species up to class level. Precision of the identifications varies according to the quality of the images and the anatomical criteria observed. To date, 48% of the single organisms and colonies have been identified at the species level and 145 species with a validated name have been inventoried (not including organisms that can be grouped by morphospecies or can only be identified at the genus level or higher). Each image of an organism

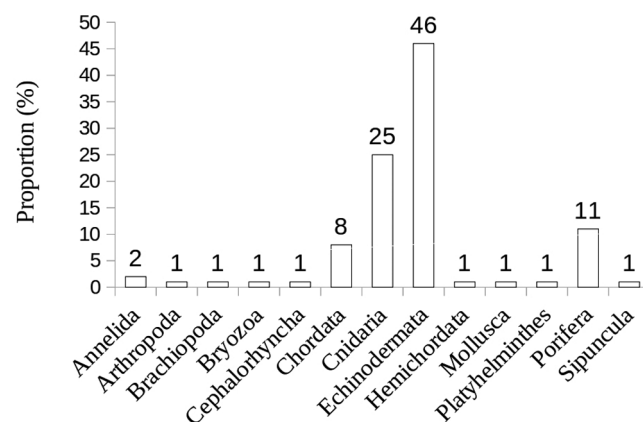
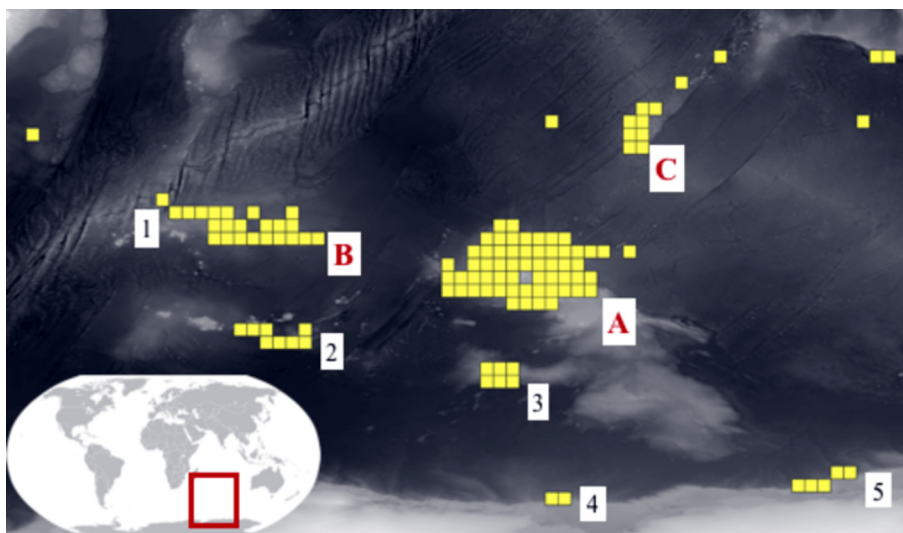


Figure 2. – The proportion of the various phyla represented in the Bendima database, based on the number of single organisms or colonies.

Figure 3. – The geographical coverage of the Bendima database; the location of the zoomed area is in provided in the inset world map; stations are aggregated into presence data according to a grid of cells of 1°; the letters and numbers relate to the name of the French Economic Exclusive Zones (EEZ) and various geomorphic structures located in international waters: A: Kerguelen EEZ, B: Crozet EEZ, C: Saint-Paul et Amsterdam EEZ, 1: Del Cano ridge, 2: Elan Bank, 3: Ob et Lena Bank, 4 and 5: Antarctica shelf.



is accompanied by a scale, allowing the measurement and grouping of organisms according to size classes. Moreover, Bendima reports the total weight of each lot, allowing biomass estimations.

The recent scientific surveys (Pige survey and Poker program) represent a collection of 67,856 identified organisms over 777 trawling stations located from 90 to 1000 m depth. A collection of voucher specimens has been constituted during Poker 4, including 2,389 samples and DNA sequences of 16S rDNA, 18S rDNA and CO1 markers, over 219 stations in the Kerguelen Exclusive Economical Zone (EEZ) ranging from 90 to 1000 m depth. Moreover, all the French commercial activity is covered since 2015. With a collection of 24,960 identified organisms, this represents to date 27% of the observations recorded into Bendima. In addition to this, a collection of approximately 1,700 samples and conserved specimens from the commercial fishing operations, still to be barcoded, have been constituted. The theoretical sampling effort is linked to the level of the commercial activity, which reaches approximately 8,000 to 9,000 organisms/colonies photographed and identified every year over 3,000 to 6,000 commercial fishing operations (longlines/traps), from 50 to 2900 m depth.

To date, a large region is covered by Bendima, including three French EEZ and various continuous or isolated spots of research/prospecting fishing in the international seas (Fig. 3). In accordance with its importance regarding both commercial fishing and scientific activity, the main area appears to be the Kerguelen EEZ. The database covers a large part of the EEZ since 2015, with an important level of observation coverage for the benthic ecosystems of the northern part of the Kerguelen Plateau. The second region is the Crozet EEZ. Observation coverage is mainly limited to the western part of the EEZ and the archipelago. This is due to the limited extension of the commercial fishing activity and the absence

of recent scientific surveys. In the EEZ of Saint-Paul and Amsterdam, the coverage of Bendima is also mainly restricted to the proximity of the archipelago, where the commercial fishing activity is concentrated. The Del Cano ridge, the Elan bank and Ob and Lena banks constitute secondary areas in terms of covered surface. However, they present an important interest for the monitoring of the subantarctic benthic ecosystems, with a gradient according to both latitudinal and longitudinal axes. Bendima includes supplementary data from isolated spots of prospecting fishing in the Indian Ocean and the Southern Ocean. Two spots are located on the Antarctica eastern continental shelf (Fig. 3), corresponding to the most southern stations recorded in Bendima.

Bendima allowed us to build an efficient information system for marine benthic macro-invertebrates in a relatively short time. This large dataset, in terms of the number of occurrences, sampling effort control and covering a wide range of taxa, has been constituted over just a few years. Raw data extracted from Bendima is based on reviewable taxa identifications, partly validated by specialists who examined the images and by the DNA barcoding of the specimens collected during the Poker 4 survey. In addition to this reliable taxa-based information, Bendima also improves reproducibility when using morphospecies. In most of the studies based on such approach, morphospecies are not described and are just mentioned as well as taxa (*e.g.* Ramalho *et al.*, 2017; Kuhnz *et al.*, 2020; Salinas-de-León *et al.*, 2020) occasionally with images of a few organisms provided as an example (*e.g.* Hibberd, 2016). The full storage of the observations in the form of images induces two improvements. First, if a series of morphospecies is defined within a dataset stored in Bendima, the full observations of the organisms belonging to these morphospecies can be provided. This is of major interest to compare results from various studies, differing in space and/or time. For such studies, the access

to the images of organisms may allow the construction of correspondence tables between heterogeneous series of morphospecies, or to reclassify all the observations in a common series of morphospecies to be used in a common modelling process. Secondary, objective criteria drawn from the image analysis can be provided to define rigorously each morphospecies, such as size range, colour scheme, shape, and anatomical structures.

Future improvements

The next development stage of Bendima concerns the contributor community management, which has been identified as a key factor for the success of citizen science projects (Conrad and Hilchey, 2011; Pandya, 2012; Pecl *et al.*, 2019). Most of our efforts consisted of the creation of the tools and the production of the datasets presented here. Community management of contributors was limited to the training of the fishery observers working onboard commercial vessels and to the real-time follow up of their field activities. Training focused on the application of the various field protocols (detailed in a complete handbook), practical exercises in the sorting of the organisms and a series of presentations of the scientific studies we performed on the ecology of the various species. When they are working in the field, fishery observers can contact us to obtain, for instance, complementary information about the application of the protocols or identifications of organisms they have observed. They can also provide feedback and suggestions for the improvement of the protocols. The recent publication of a newsletter dedicated to the presentation of the studies performed by the team involved in the scientific monitoring of the fisheries allowed us to start sending our own feedback about the use of the benthos data. The effort with the contributor community management should be significantly increased in the future. More scientific feedback from researchers and more training should be organized for fishery observer empowerment, to structure among them a community of field benthologists. Moreover, a second community of contributors should be organized for the identification and analysis of the database contents. The creation of a dedicated online tool to make Bendima available through the web could be a solution to allow the involvement of citizens and worldwide specialists in the identification of the taxa. In order to start this part of the project, a prototype of web site for now called “Halieutique” (<http://halieutique.mnhn.fr/sector/kerguelen>), designed to share the contents of Bendima, is under development.

Beyond the scientific monitoring of the French fisheries, we also aim to develop tools and procedures to make Bendima more interoperable with other international databases. Secretariats of the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) and Southern Indian Ocean Fisheries Agreement (SIOFA) maintain global

databases, including observations of macro-invertebrate bio-indicators of “Vulnerable Marine Ecosystems”. Direct transmission of data from Bendima to these two systems could be a first step. Moreover, a significant goal for Bendima development should be the connection to SeaLifeBase (Palomares and Pauly, 2021), to provide benthic fauna data of the Southern Ocean and the Indian Ocean.

Acknowledgements. – We would like to acknowledge the Direction des Pêches Maritimes et de l’Aquaculture du Ministère de l’Agriculture et de l’Alimentation and the Muséum national d’Histoire naturelle for support and funding. We also would like to acknowledge all the fishery observers from the Terres Australes et Antarctiques Françaises who provided raw data to be integrated into Bendima. The authors thank two anonymous reviewers for their useful comments.

REFERENCES

- ABRÀMOFF M., MAGALHÃES P. & RAM S., 2004. – Image Processing with ImageJ. *Biophotonics*, July, 7.
- ADAMOWICZ S.J., 2015. – International Barcode of Life: evolution of a global research community. *Genome*, 58: 151–162. <https://doi.org/10.1139/gen-2015-0094>
- ADHIKARI D., TIWARY R., SINGH P.P., UPADHAYA K., SINGH B., HARIDASAN K.E., BHATT B.B., CHETTRI A. & BARIK S.K., 2019. – Ecological niche modeling as a cumulative environmental impact assessment tool for biodiversity assessment and conservation planning: a case study of critically endangered plant *Lagerstroemia minuticarpa* in the Indian Eastern Himalaya. *J. Environ. Manage.*, 243: 299–307. <https://doi.org/10.1016/j.jenvman.2019.05.036>
- AN G., 2018. – The crisis of reproducibility, the denominator problem and the scientific role of multi-scale modeling. *Bull. Math. Biol.*, 80: 3071–3080. <https://doi.org/10.1007/s11538-018-0497-0>
- BAKER M., 2016. – 1,500 scientists lift the lid on reproducibility. *Nat. News*, 533: 452–454. <https://doi.org/10.1038/533452a>
- BEATTLE A.J. & OLIVER I., 1994. – Taxonomic minimalism. *Trends Ecol. Evol.*, 9: 488–490. [https://doi.org/10.1016/0169-5347\(94\)90320-4](https://doi.org/10.1016/0169-5347(94)90320-4)
- BRIND’AMOUR A., LAFFARGUE P., MORIN J., VAZ S., FOVEAU A. & LE BRIS H., 2014. – Morphospecies and taxonomic sufficiency of benthic megafauna in scientific bottom trawl surveys. *Cont. Shelf Res.*, 72: 1–9. <https://doi.org/10.1016/j.csr.2013.10.015>
- CONRAD C.C. & HILCHEY K.G., 2011. – A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environ. Monit. Assess.*, 176: 273–291. <https://doi.org/10.1007/s10661-010-1582-5>
- COSTELLO M.J., COL M., DANOVARO R., HALPIN P., OJAVEER H. & MILOSLAVICH P., 2010. – A census of marine biodiversity knowledge, resources, and future challenges. *PLOS ONE*, 5: e12110. <https://doi.org/10.1371/journal.pone.0012110>

- COSTELLO M.J., BOUCHET P., BOXSHALL G., FAUCHALD K., GORDON D., HOEKSEMA B.W., POORE G.C.B., VAN SOEST R.W.M., STÖHR S., WALTER T.C., VANHOORNE B., DECOCK W. & APPELTANS W., 2013. – Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. *PLOS ONE*, 8: e51629. <https://doi.org/10.1371/journal.pone.0051629>
- CULINA A., VAN DEN BERG I., EVANS S. & SÁNCHEZ-TÓJAR A., 2020. – Low availability of code in ecology: a call for urgent action. *PLOS Biol.*, 18: e3000763. <https://doi.org/10.1371/journal.pbio.3000763>
- CUSHING J.B., LACH D., ZANOCCO C. & HALAMA J., 2018. – Scientific visualization and reproducibility for “open” environmental science. In: 2018 IEEE International Conference on Big Data (Big Data). Presented at the 2018 IEEE International Conference on Big Data (Big Data), IEEE, Seattle, WA, USA: 3211-3216. <https://doi.org/10.1109/BigData.2018.8622039>
- DEGUINES N., JULLIARD R., DE FLORES M. & FONTAINE C., 2012. – The whereabouts of flower visitors: contrasting land-use preferences revealed by a country-wide survey based on citizen science. *PLOS ONE*, 7: e45822. <https://doi.org/10.1371/journal.pone.0045822>
- DESJARDINS E., KURTZ J., KRANKE N., LINDEZA A. & RICHTER S.H., 2021. – Beyond standardization: improving external validity and reproducibility in experimental evolution. *BioScience*, 71: 543–552. <https://doi.org/10.1093/biosci/biab008>
- DUHAMEL G. & WILLIAMS R., 2011. – History of whaling, sealing, fishery and aquaculture trials in the area of the Kerguelen Plateau. In: The Kerguelen Plateau: Marine Ecosystem and Fisheries. Presented at the Symposium on Kerguelen Plateau Marine Ecosystem and Fisheries. Société Française d’Ichtyologie, Concarneau, France: 15-28.
- DUHAMEL G., PÉRON C., SINÈGRE R., CHAZEAU C., GASCO N., HAUTECŒUR M., MARTIN A., DURAND I. & CAUSSE R., 2019. – Important readjustments in the biomass and distribution of groundfish species in the northern part of the Kerguelen Plateau and Skiff Bank. In: The Kerguelen Plateau: Marine Ecosystem + Fisheries. Presented at the Second symposium on Kerguelen Plateau Marine Ecosystem and Fisheries, Australian Antarctic Division, Hobart, Australia: 135-184.
- DUNLAP K., 1926. – The experimental methods of psychology. In: Powell Lecture in Psychological Theory, Apr, 1925, Clark University, Worcester, MA, US; Portions of This Research Were Presented at the Powell Lecture in Psychological Theory at Clark University, April 21, 1925. Clark University Press, Worcester, USA. <https://doi.org/10.1037/11020-022>
- ERDEMIR A. & SAURO H.M., 2016. – Editorial reproducibility of computational models. *IEEE Trans. Biomed. Eng.*, 63: 1995-1996. <https://doi.org/10.1109/TBME.2016.2594702>
- FENG X., PARK D.S., WALKER C., PETERSON A.T., MEROW C. & PAPEŞ M., 2019. – A checklist for maximizing reproducibility of ecological niche models. *Nat. Ecol. Evol.*, 3: 1382-1395. <https://doi.org/10.1038/s41559-019-0972-5>
- FIDLER F., CHEE Y.E., WINTLE B.C., BURGMAN M.A., MCCARTHY M.A. & GORDON A., 2017. – Metaresearch for evaluating reproducibility in ecology and evolution. *BioScience*, 67: 282-289. <https://doi.org/10.1093/biosci/biw159>
- FILAZZOLA A. & CAHILL J.F., 2021. – Replication in field ecology: identifying challenges and proposing solutions. *Methods Ecol. Evol.*, 12: 1780-1792. <https://doi.org/10.1111/2041-210X.13657>
- FONSECA G., FONTANETO D. & DI DOMENICO M., 2018. – Addressing biodiversity shortfalls in meiofauna. *J. Exp. Mar. Biol. Ecol.*, 502: 26-38. <https://doi.org/10.1016/j.jembe.2017.05.007>
- FRASER H., BARNETT A., PARKER T.H. & FIDLER F., 2020. – The role of replication studies in ecology. *Ecol. Evol.*, 10: 5197-5207. <https://doi.org/10.1002/ece3.6330>
- GARCIA-SOTO C. & VAN DER MEEREN G. (Eds), 2017. – Advancing Citizen Science for Coastal and Ocean Research. European Marine Board, Ostend, Belgium.
- GIANGRANDE A., 2003. – Biodiversity, conservation, and the taxonomic impediment? *Aquat. Conserv. Mar. Freshw. Ecosyst.*, 13: 451-459. <https://doi.org/10.1002/aqc.584>
- GRANT R.A. & LINSE K., 2009. – Barcoding Antarctic biodiversity: current status and the CAML initiative, a case study of marine invertebrates. *Polar Biol.*, 32: 1629-1637. <https://doi.org/10.1007/s00300-009-0662-x>
- HEBERT P.D.N., RATNASINGHAM S. & DE WAARD J.R., 2003. – Barcoding animal life: cytochrome c oxidase subunit I divergences among closely related species. *Proc. R. Soc. Lond., B*: 270. <https://doi.org/10.1098/rsbl.2003.0025>
- HERMOSO M., NARVÁEZ S. & THIEL M., 2021. – Engaging recreational scuba divers in marine citizen science: Differences according to popularity of the diving area. *Aquat. Conserv. Mar. Freshw. Ecosyst.*, 31: 441-455. <https://doi.org/10.1002/aqc.3466>
- HIBBERD T., 2016. – Describing and predicting the spatial distribution of benthic biodiversity in the sub-Antarctic and Antarctic. PhD Thesis. University of Tasmania, Hobart, Australia.
- HINSINGER D.D., DEBRUYNE R., THOMAS M., DENYS G.P.J., MENNESSON M., UTAGE J. & DETTAI A., 2015. – Fishing for barcodes in the Torrent: from COI to complete mitogenomes on NGS platforms. *DNA Barcodes*, 3. <https://doi.org/10.1515/dna-2015-0019>
- HUREAU J.C., 2011. – Marine research on the Kerguelen Plateau: from early scientific expeditions to current surveys under the CCAMLR objectives. In: The Kerguelen Plateau: Marine Ecosystem and Fisheries. Presented at the Symposium on Kerguelen Plateau Marine Ecosystem and Fisheries, Société Française d’Ichtyologie, Concarneau, France: 5-13.
- KASPEREK D., MAISENBACHER S., KOHN A., LINDEMANN U. & MAURER M., 2015. – Increasing the reproducibility of structural modelling. *J. Eng. Design*, 26: 259-281. <https://doi.org/10.1080/09544828.2015.1026883>
- KELLY C.D., 2019. – Rate and success of study replication in ecology and evolution. *PeerJ*, 7: e7654. <https://doi.org/10.7717/peerj.7654>
- KUHNZ L.A., RUHL H.A., HUFFARD C.L. & SMITH K.L., 2020. – Benthic megafauna assemblage change over three decades in the abyss: Variations from species to functional groups. *Deep Sea Res. Part II: Top. Stud. Oceanogr.*, 173: 104761. <https://doi.org/10.1016/j.dsr2.2020.104761>
- LERAY M., YANG J.Y., MEYER C.P., MILLS S.C., AGUDELO N., RANWEZ V., BOEHM J.T. & MACHIDA R.J., 2013. – A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Front. Zool.*, 10: 34. <https://doi.org/10.1186/1742-9994-10-34>
- MACHIDA R.J. & KNOWLTON N., 2012. – PCR primers for metazoan nuclear 18S and 28S ribosomal DNA sequences. *PLOS ONE*, 10: e0134314. <https://doi.org/10.1371/journal.pone.0134314>

- MAHAJAN S., EVANS K.J., KENNEDY J.H., XU M., NORMAN M.R. & BRANSTETTER M.L., 2019. – Ongoing solution reproducibility of earth system models as they progress toward exascale computing. *Int. J. High Perform. Comput. Appl.*, 33: 784-790. <https://doi.org/10.1177/1094342019837341>
- MANNINEN T., HAVELA R. & LINNE M.L., 2017. – Reproducibility and comparability of computational models for astrocyte calcium excitability. *Front. Neuroinform.*, 11. <https://doi.org/10.3389/fninf.2017.00011>
- MARTIN A., TROUSLARD E., HAUTECOEUR M., BLETTERY J., MOREAU C., SAUCÈDE T., AMÉZIANE N., DUHAMEL G. & ELÉAUME M., 2019. – Ecoregionalisation and conservation of benthic communities in the French exclusive economic zone of Kerguelen. In: *The Kerguelen Plateau: Marine Ecosystem + Fisheries*. Presented at the Second Symposium on Kerguelen Plateau Marine Ecosystems and Fisheries, Australian Antarctic Division, Hobart, Australia: 279-303.
- MARTIN A., CHAZEAU C., GASCO N., DUHAMEL G. & PRUVOST P., 2021. – Data curation, fisheries and ecosystem-based management: the case study of the Pecheker database. *Int. J. Digit. Curation*, 16: 31 p. <https://doi.org/10.2218/ijdc.v16i1.674>
- MONCKTON S.K., JOHAL S. & PACKER L., 2020. – Inadequate treatment of taxonomic information prevents replicability of most zoological research. *Can. J. Zool.*, 98: 633-642. <https://doi.org/10.1139/cjz-2020-0027>
- MONDELLI M.L., PETERSON A.T. & GADELHA Jr L.M.R., 2019. – Exploring Reproducibility and FAIR Principles in Data Science Using Ecological Niche Modeling as a Case Study (Preprint No. arXiv:1909.00271 [cs]). National Laboratory for Scientific Computing and Biodiversity Institute, University of Kansas, Petropolis, Brazil and Lawrence, USA.
- MONNIOT F., DETTAI A., ELEAUME M., CRUAUD C. & AMEZIANE N., 2011. – Antarctic Ascidians (Tunicata) of the French-Australian survey CEAMARC in Terre Adélie. *Zootaxa*, 2817: 1. <https://doi.org/10.11646/zootaxa.2817.1.1>
- NEIGEL J., DOMINGO A. & STAKE J., 2007. – DNA barcoding as a tool for coral reef conservation. *Coral Reefs*, 26: 487-499. <https://doi.org/10.1007/s00338-007-0248-4>
- NEKRUTENKO A. & TAYLOR J., 2012. – Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat. Rev. Genet.*, 13: 667-672. <https://doi.org/10.1038/nrg3305>
- NICHOLS J.D., KENDALL W.L. & BOOMER G.S., 2019. – Accumulating evidence in ecology: once is not enough. *Ecol. Evol.*, 9: 13991-14004. <https://doi.org/10.1002/ece3.5836>
- OWUSU K.A., ACEVEDO-TREJOS E., FALL M.M. & MERICO A., 2020. – Effects of cooperation and different characteristics of Marine Protected Areas in a simulated small-scale fishery. *Ecol. Complex.*, 44: 100876. <https://doi.org/10.1016/j.ecocom.2020.100876>
- PALOMARES M.L.D. & PAULY D., 2021. – SeaLifeBase [WWW Document]. <https://www.sealifebase.ca/> (accessed 1.3.22).
- PALUMBI S.R., 1996. – Nucleic acids II: the polymerase chain reaction. In: *Molecular Systematics*. Sinauer & Associates Inc, Sunderland, USA: 205-247.
- PALUMBI S.R. & BENZIE J., 1991. – Large mitochondrial DNA differences between morphologically similar penaeid shrimp. *Mol. Mar. Biol. Biotechnol.*, 1: 27-34.
- PANDYA R.E., 2012. – A framework for engaging diverse communities in citizen science in the US. *Front. Ecol. Environ.*, 10: 314-317. <https://doi.org/10.1890/120007>
- PAULIUK S., MAJEAU-BETTES G., MUTEL C.L., STEUBING B. & STADLER K., 2015. – Lifting industrial ecology modeling to a new level of quality and transparency: a call for more transparent publications and a collaborative open source software framework. *J. Ind. Ecol.*, 19: 937-949. <https://doi.org/10.1111/jieec.12316>
- PECCHI M., MARCHI M., BURTON V., GIANNETTI F., MORIONDO M., BERNETTI I., BINDI M. & CHIRICI G., 2019. – Species distribution modelling to support forest management. A literature review. *Ecol. Model.*, 411: 108817. <https://doi.org/10.1016/j.ecolmodel.2019.108817>
- PECL G.T., STUART-SMITH J., WALSH P., BRAY D.J., KUSSETIC M., BURGESS M., FRUSHER S.D., GLEDHILL D.C., GEORGE O., JACKSON G., KEANE J., MARTIN V.Y., NURSEY-BRAY M., PENDER A., ROBINSON L.M., ROWLING K., SHEAVES M. & MOLTSCHANIWSKYJ N., 2019. – Redmap Australia: challenges and successes with a large-scale citizen science-based approach to ecological monitoring and community engagement on climate change. *Front. Mar. Sci.*, 6: 349. <https://doi.org/10.3389/fmars.2019.00349>
- PENG G., 2016. – Moving toward model reproducibility and reusability. *IEEE Trans. Biomed. Eng.*, 63: 1997-1998. <https://doi.org/10.1109/TBME.2016.2603418>
- POPPER K., 1959. – The propensity interpretation of probability. *Br. J. Philos. Sci.*, 10: 25-42. <https://doi.org/10.1093/bjps/X.37.25>
- POPPER K.R., 1963. – Science as falsification. *Conject. Refutations*, 1: 33-39.
- POWERS S.M. & HAMPTON S.E., 2019. – Open science, reproducibility, and transparency in ecology. *Ecol. Appl.*, 29. <https://doi.org/10.1002/eap.1822>
- RADOSEVIC N., DUCKHAM M., LIU G.J. & SUN Q., 2020. – Solar radiation modeling with KNIME and Solar Analyst: Increasing environmental model reproducibility using scientific workflows. *Environ. Model. Softw.*, 132: 104780. <https://doi.org/10.1016/j.envsoft.2020.104780>
- RAMALHO S.P., LINS L., BUENO-PARDO J., CORDOVA E.A., AMISI J.M., LAMPADARIOU N., VANREUSEL A. & CUNHAM R., 2017. – Deep-sea mega-epibenthic assemblages from the SW Portuguese margin (NE Atlantic) subjected to bottom-trawling fisheries. *Front. Mar. Sci.*, 4: 350. <https://doi.org/10.3389/fmars.2017.00350>
- RAOULT V., DAVID P.A., DUPONT S.F., MATHEWSON C.P., O'NEILL S.J., POWELL N.N. & WILLIAMSON J.E., 2016. – GoPros™ as an underwater photogrammetry tool for citizen science. *PeerJ*, 4: e1960. <https://doi.org/10.7717/peerj.1960>
- RATNASINGHAM S. & HEBERT P.D.N., 2007. – Bold: the barcode of life data system. *Mol. Ecol. Notes*, 7: 355-364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x>
- REICHMAN O.J., JONES M.B. & SCHILDHAUER M.P., 2011. – Challenges and opportunities of open data in ecology. *Science*, 331: 703-705. <https://doi.org/10.1126/science.1197962>
- REMSEN J.V., 1995. – The importance of continued collecting of bird specimens to ornithology and bird conservation. *Bird Conserv. Int.*, 5: 146-180. <https://doi.org/10.1017/S095927090000099X>
- RIDE W. (Ed.), 1999. – International Code of Zoological Nomenclature. International Trust for Zoological Nomenclature, London, UK.

- SALINAS-DE-LEÓN P., MARTÍ-PUIG P., BUGLASS S., ARNÉS-URGELLÉS C., RASTOIN-LAPLANE E., CREEMERS M., CAIRNS S., FISHER C., O'HARA T., OTT B., RAINEAULT N.A., REISWIG H., ROUSE G., ROWLEY S., SHANK T.M., SUAREZ J., WATLING L., WICKSTEN M.K. & MARSH L., 2020. – Characterization of deep-sea benthic invertebrate megafauna of the Galapagos Islands. *Sci. Rep.*, 10: 13894. <https://doi.org/10.1038/s41598-020-70744-1>
- SARRI C., STAMATIS C., SARAFIDOU T., GALARA I., GODOSOPOULOS V., KOLOVOS M., LIAKOU C., TAST-SOGLU S. & MAMURIS Z., 2014. – A new set of 16S rRNA universal primers for identification of animal species. *Food Control*, 43: 35-41. <https://doi.org/10.1016/j.foodcont.2014.02.036>
- SHOKRALLA S., PORTER T.M., GIBSON J.F., DOBOSZ R., JANZEN D.H., HALLWACHS W., GOLDING G.B. & HAJIBABAEI M., 2015. – Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci. Rep.*, 5: 9687. <https://doi.org/10.1038/srep09687>
- SILVERTOWN J., 2009. – A new dawn for citizen science. *Trends Ecol. Evol.* 24: 467-471. <https://doi.org/10.1016/j.tree.2009.03.017>
- STOECKLE M. & HEBERT P., 2008. – Barcode of Life. *Sci. Am.*, 299: 82-88.
- SWANSON A., KOSMALA M., LINTOTT & C. PACKER C., 2016. – A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. *Conserv. Biol.*, 30: 520-531. <https://doi.org/10.1111/cobi.12695>
- TELETCHEA F., 2010. – After 7 years and 1000 citations: comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA*, 21: 206-226
- URBINA-CARDONA N. & FLORES-VILLELA O., 2010. – Ecological-niche modeling and prioritization of conservation-area for Mexican herpetofauna. *Conserv. Biol.*, 24: 1031-1041. <https://doi.org/10.1111/j.1523-1739.2009.01432.x>
- YATES K.L. & SCHOEMAN D.S., 2015. – Incorporating the spatial access priorities of fishers into strategic conservation planning and marine protected area design: reducing cost and increasing transparency. *ICES J. Mar. Sci.*, 72: 587-594. <https://doi.org/10.1093/icesjms/fsu122>