



Katsuhiko Tanaka and Kazuo Inaba

## Abstract

Increasing scientific information promote scientists to collect scattered data, put them together and organize them, and make them accessible with the best usability. These data include identification, characterization, habitats and biogeographical distribution, as well as their genomic information. Much of such data are now integrated so as to be accessible for scientists as databases, which are enormously useful to study marine biology. Here we describe some biographic databases and genome/transcriptome databases that are commonly accessed by scientists.

## 29.1 Biogeographic Database

The identification of marine species from the field, with information on the morphology, phylogeny, and habitat, is indispensable to elucidate marine geological and biological environments. A picture guidebook is generally useful for these purposes; now, however, several web-based databases are well-developed for the collection and storage of a large quantity of information and searching of users' queries. In particular, accumulating information on marine species with genomic data enables one to obtain various data types, such as the gene expressions of a species in several tissues and their phylogenetic relationships.

Biogeographic databases provide information on the distribution of an organism, as well as their biodiversity at a regional to global scale. For marine organisms, the

---

K. Tanaka  
School of Marine Science and Technology, Tokai University, Shizuoka, Japan

K. Inaba (✉)  
Shimoda Marine Research Center, University of Tsukuba, Shimoda, Shizuoka, Japan  
e-mail: [kinaba@shimoda.tsukuba.ac.jp](mailto:kinaba@shimoda.tsukuba.ac.jp)

Ocean Biogeographic Information System (OBIS) has collected global data from about 2500 regional/local databases and integrated over 52 million occurrence records of over 120,000 marine species. Each record includes the locality of a species in a collection/observation. OBIS was constructed by a global research/monitoring project (Census of Marine Life [CoML]; Decker and O’Dor 2002; Halpin et al. 2006) and was opened to the public in 2000. It is currently operated as shared property under the Intergovernmental Oceanographic Data and Information Exchange in the United Nations Educational, Scientific and Cultural Organization (UNESCO). The Autonomous Reef Monitoring Structures (ARMS) Program is aimed toward collecting sessile organisms for CoML projects, and the database is managed by the Smithsonian Institute (Plaisance et al. 2009). Using OBIS, users can freely search records by organism name, geographic region, etc., and can then visualize their distribution on a scalable map. OBIS data are also downloadable for secondary use or further analyses. The information on OBIS contributes to understanding the distributions and diversity of marine organisms and has been used for the prediction of marine species distributions, such as future marine biodiversity patterns under climate change, in combination with environmental datasets/models.

In Japan, the Biological Information System for Marine Life (BISMaL) constructed by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC) is a useful data source to learn marine biodiversity around Japan (Fig. 29.1). BISMaL is an integrated database dealing with different kinds of information. It provides taxonomic information (scientific names of species and their higher classification) on marine organisms around Japan as well as their occurrence records from several datasets. An organism is searchable by name/hierarchical taxonomic tree. Descriptions of the morphology and ecology, as well as photographs, are available for some organisms, and movies taken during deep-sea surveys by JAMSTEC are also provided for deep-sea species. Users can search and visualize occurrence records on a map, filter them by taxonomic group, dataset, and depth, and download them for further analyses.

BISMaL has collected over 450,000 occurrence records for over 6200 species. The data is still too poor to completely cover the marine biodiversity around Japan; thus, the Japan Ocean Biogeographic Information System Center is calling for research communities in Japan to provide data to BISMaL to establish a robust data source for related fields. Such data collections may also be beneficial for data providers because BISMaL, as an OBIS node providing regional information to OBIS, send data to OBIS, which are then integrated with other data from various regions and opened to the public. A regional database such as BISMaL can be not only a data source, but also a useful tool for data publication online, and provides a way for every scientist/scientific community to contribute to the global dataset. Other than BISMaL, several excellent public and private databases are available to access information on marine organisms online. For example, the Japanese Association for Marine Biology (JAMBIO) is continuously conducting a joint survey and summarizes the marine organisms collected as the database “RINKAI” (Inaba 2015).

The image displays the BISMAL (Biological Information System for Marine Life) web interface. At the top left is the BISMAL logo. The main content area includes a search bar with the text "Please search name." and a magnifying glass icon. Below the search bar, the taxonomic classification for the species is listed: "Classification : Eukarya - Opisthokonta - Animalia - Arthropoda - Crustacea - Malacostraca - Eumalacostraca - Eucarida - Decapoda - Pleocyemata - Anomura - Galatheoidea - Galatheidae - Shinkaia". The species name "***Shinkaia crosnieri*** Baba & Williams, 1998" is prominently displayed. Below the name are three interactive buttons: "Show literature" with a document icon, "Taxonomy" with a tree icon, and "Occurrence records" with an information icon. To the right of the main content is a vertical navigation menu with three options: "Occurrence records" (selected), "Map", and "Summary".

At the bottom of the interface, there are two charts. The left chart, titled "[Environment data]", is a bar chart showing the distribution of records across different depths. The x-axis is labeled "Depth (m)" and ranges from 0 to 500. The y-axis represents the number of records. The data points are as follows:
 

Depth (m)	Records
662 - 1092	1
692 - 1092	1
1092 - 1292	1
1292 - 1492	2
1492 - 1892	3
1892 - 1892	1
1892 - 2092	1
2092 - 2292	1

The right chart, titled "Temperature", is a bar chart showing the distribution of records across different temperature ranges. The x-axis is labeled "Temperature (degrees)" and ranges from 3.0 to 8.0. The y-axis represents the number of records. The data points are as follows:
 

Temperature (degrees)	Records
3.0 - 4.0	3
4.0 - 5.0	4
5.0 - 6.0	1
6.0 - 7.0	1
7.0 - 8.0	1
8.0 - 9.0	1

Fig. 29.1 An integrated database for marine organisms, the Biological Information System for Marine Life (BISMAL).

## 29.2 Genome and Transcriptome Database

For the DNA-based identification of a specific organism or a mass organism survey in marine environments, DNA barcoding and environmental DNA (eDNA) are powerful tools. The former is a method for the rapid identification of known or unknown species by using a short fragment of mitochondrial cytochrome oxidase I (COI or COX1) or the nuclear ribosomal RNA internal transcribed spacer (ITS). The latter is a mass metagenomic analysis of single and multi species from a variety of environmental samples, such as sea water, containing diverse DNA sources from tissue debris and feces. It provides information on species distributions, population, and ecosystems and is useful for environmental biomonitoring.

After complete determination of genome sequences in the pufferfish *Takifugu rubripes* and the ascidian *Ciona intestinalis* in 2002, genomic information has been rapidly accumulated for many other marine organisms, due to the development of next generation sequencing. The marine organisms in which complete genome sequences are available include the diatom (*Thalassiosira pseudonana*), brown algae (*Ectocarpus siliculosus*), sponge (*Amphimedon queenslandica*), cnidarians (*Nematostella vectensis*, *Clytia hemisphaerica*, and *Acropora digitifera*), ctenophores (*Mnemiopsis leidyi* and *Pleurobrachia bachei*), sea urchins (*Strongylocentrotus purpuratus*, *Hemicentrotus pulcherrimus* [the latter is a Japanese species]), lancelet (*Branchiostoma floridae*), bivalve (*Spisula solidissima*), shrimp (*Litopenaeus vannamei*), and many others. In some marine organisms without full genome sequences, many sequence data are available from RNAseq analysis of expressed genes (transcriptomes).

The sequences determined independently by researchers with annotations from a single gene or those from mass sequencing are, as ever, submitted to a fundamental initiative for sequence database collaboration that operates the DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) (ENA; European Nucleotide Archive), and National Center for Biotechnology Information (NCBI) (GenBank). The genome information can be accessed from their genome browsers, where online search tools, including BLAST searches, can be used against the sequence data of a specific organism or mass sequence data. It is often necessary to access the genome information for multiple marine organisms. In these cases, it is useful to obtain information through portal sites. For prokaryotes, there are well-organized databases for marine microbial genomes: MarRef, MarDB, and MarCat. MarBEF is one of the two networks for European marine ecosystem research, along with Marine Genomics Europe (MGE) (Klemetsen et al. 2018). These are easily accessed through the Marine Metagenomics Portal (MMP) (<https://mmp.sfb.uit.no/>). Other useful metagenomic databases include the Genomes OnLine Database (GOLD), Viral Informatics Resource for Metagenome Exploration (VIROME), MGnify (formerly EBI Metagenomics), Integrated Microbial Genomes and Microbiomes (IMG/M), and Marine Life Genome Database (MLGD).

A few mass collections of marine organisms are operated, such as the Global Ocean Sampling (GOS) Expedition on Sorcerer II (J. Craig Venter Institute) and the

---

Tara Expedition (Tara Foundation). The latter implemented an expedition named "Tara Pacific" from 2016 to 2018, which included the investigation of Japanese coasts (Carradec et al. 2018). Samples from the previous expedition, "Tara Oceans," are mapped to the external Tara Oceans Sample Registry at PANGAEA (<http://www.pangaea.de/>). Their information and the primary sequence data can be accessed through EMBL-EBI ([www.ebi.ac.uk/ena/data/view/PRJEB402](http://www.ebi.ac.uk/ena/data/view/PRJEB402); [www.ebi.ac.uk/metagenomics/](http://www.ebi.ac.uk/metagenomics/); [www.ebi.ac.uk/metagenomics/projects/ERP001736/](http://www.ebi.ac.uk/metagenomics/projects/ERP001736/)).

---

## References

- Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al. (2018). A global ocean atlas of eukaryotic genes. *Nature Communications*, 9, 373. <https://doi.org/10.1038/s41467-017-02342-1>.
- Decker, C. J., & O'Dor, R. (2002). A census of marine life: Unknowable or just unknown? *Oceanologica Acta*, 25, 179–186.
- Halpin, P. N., Read, A. J., Best, B. D., Hyrenbach, K. D., Fujioka, E., et al. (2006). OBIS-SEAMAP: Developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds, and sea turtles. *Marine Ecology Progress Series*, 316, 239–246.
- Inaba, K. (2015). Japanese marine biological stations: Preface to the special issue. *Regional Studies in Marine Sciences*, 2, 154–157.
- Klemetsen, T., Raknes, I. A., Fu, J., Agafonov, A., Balasundaram, S. V., Tartari, G., et al. (2018). The MAR databases: Development and implementation of databases specific for marine metagenomics. *Nucleic Acids Research*, 46(D1), D692–D699.
- Plaisance, L., Knowlton, N., Paulay, G., & Meyer, C. (2009). Reef-associated crustacean fauna: Biodiversity estimates using semi-quantitative sampling and DNA barcoding. *Coral Reefs*, 28, 977–986.