



**HAL**  
open science

## MyGOD : Visualization and analysis interface for marine genomic observatories data

Charlotte André, Mark Hoebeke, Nicolas Henry, Cyril Noel, Patrick Durand,  
Erwan Corre

### ► To cite this version:

Charlotte André, Mark Hoebeke, Nicolas Henry, Cyril Noel, Patrick Durand, et al.. MyGOD : Visualization and analysis interface for marine genomic observatories data. JOBIM 2022, Jul 2022, Rennes, France. 10.5281/zenodo.6451438 . hal-04029627

**HAL Id: hal-04029627**

**<https://hal.science/hal-04029627v1>**

Submitted on 15 Mar 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





## Marine Genomics Observatories

Marine Genomics Observatories provide tools and protocols for long-term monitoring of marine ecosystems located in a well-defined and usually limited geographical area, in order to allow the study of their evolution over time. Hence, they yield a significant amount of data (metabarcoding data, "classical" (morphology-base) biodiversity data, supplemental physico-chemical data), which needs to be organized and stored to enable interested parties to carry out downstream analyses.



Ostreococcus lucimarinus



Cryptophyta



Dinoflagellata

## MyGOD project

The MyGOD project (Manipulate Your Genomic Observatory Data) aims to explore and interpret the data of genomic observatories projects, by providing an integrated and user-friendly web-based visualization tool.

The key component of the interface is a dashboard enabling users to assemble various graphical representations according to their needs. The user interface taps into a database for which specific integration tools, as well as (meta)data formats have been developed and/or adapted in the framework of the MyGOD project, aiming to achieve a high degree of automated ingestion.

## Development tools

MyGOD Data storage/processing

python django PostgreSQL R

Dashboard

The dashboard uses the **Vue.js** Javascript framework with the **vue-grid-layout** plugin. Graphical representations are dynamically generated with the **D3.js** Javascript library.

JS Vue.js

## An Interactive Dashboard

### How the dashboard works

The MyGOD dashboard is a multi-taxonomic search tools with save/reload functionality

#### 1 Search Panels

- High-level search criteria: dataset, date range and taxon
- Multiple independent search panels for easy comparison
- Dynamical update of dashboard graphs and plots when search criteria change

#### 2 Dashboard Controls

- Allow selection of graph types to be displayed in the dashboard
- For authenticated users, allow saving and loading of dashboards

#### 3 Plot Area

- User controllable widget sizing
- User controllable widget positioning

#### 4 Widget Parameters

- For each plot, users can modify specific parameters (taxon, physico-chemical parameter, fraction size, depth...)
- Plots can be exported as PNG or SVG, along with data and metadata files
- Plots are cloneable

Relative taxonomic composition

Representation:  Relative Abundance  Richness

Grouped by: Season

Sublevel: Genus

Taxon displayed: 5

Bar number: 20

Fraction size:  0.2-3 µm  >3 µm

## Data

### Taxonomical reference databases :

- PR2 (Eukaryota)
- SILVA (Bacteria, Archaea)
- CYANOMARKS

### Metabarcoding series :

- Astan 18S/16S** : Station of Roscoff
- Tonga 18S/16S/PetB** : Tonga archipelago
- Blanes 18S/16S** : Barcelona
- Moose 18s** : Mediterranean

## SAMBA integration

<https://github.com/ifremer-bioinformatics/samba>

Samba is a **metabarcoding analysis workflow** developed by **SeBimer**. Collaboration on MyGOD project, lead to **integration of biostatistical processing** parts of the pipeline in the visualisation tool.

- Computation of normalized data (by rarefaction)
- Computation of distance matrix for  $\beta$ -diversity

1 Toggle Taxon Search Panels

2 Dashboard

3 Summary

4 Abundance of Family/Bathycoccaceae by Season (in % on fraction 0.2 µm)

## The Web Portal

Manipulate your Genomic Observatory Data

**Search** - Use search criteria to design composite and dynamic visualisations

### Datasets

For each project, presentation of context, sampling methods is provided...

### A personal user space to:

- Access previously saved searches
- Access previously saved dashboards
- Edit dataset metadata (dataset owners)

## Challenges

### Non-Technical Challenges:

- Defining the nature of the data and metadata needed for the visualization tool
- Defining the relevant types of representations for the data
- Handling varying taxonomical frameworks (fixed or variable number of levels, explicitly named ranks or unnamed ranks)

### Technical Challenges:

- Making the best of the reactive framework : Vue/vuex/grid layout.
- Integrating third-party (R) components for analysis

## Standardized Data Format

- Definition of a lightweight, flexible format for importing tabular data:
  - One file with **sampling event descriptions (event identifiers)**
  - One file with metabarcode read data (linked through event identifiers)
  - One file with physico-chemical data (linked through event identifiers)
- Accompanying JSON metadata files specify the column structure and contents of the data tables: sample name patterns, index mapping for physico-chemical parameters.

## What's Next ?

### Upcoming Developments:

- Addition of graphical representations of normalized data :  $\beta$ -diversity (with R and Javascript)
- Enable promotion of "showcases" (dashboards with extensive comments) on portal
- Addition of "classical" (morphology-based) biodiversity data
- Addition of sequence similarity based search (Blast)