# Demonstrator 2 - Notebook 2:

# Mapping the geographical distribution of plankton functional gene clusters using habitat prediction modes

**Authors:** A. Schickele, P. Debeljack, L. Guidi, S. Ayata, J.O. Irisson

**Corresponding author/maintainer:** alexandre.schickele@imev-mer.fr

# 1. Introduction

Planktonic organisms and by extension the composition and genetic potential of plankton in a given geographical area is influenced by the environmental context. In this notebook, we provide a series of tools to explore the relationship between the abundance of plankton genes and the environmental context, in order to project the biogeography of key metabolic pathways and as yet unknown plankton genes.

This notebook is divided in two services, designed for expert and non-expert users, respectively:

* **service 1 - R modeling framework:** This service describes the entire modeling pipeline and codes. It is essentially designed for expert users with a strong genomic and modeling background. The codes are accessible from the BlueCloud infrastructure and may also be duplicated and modified locally.
* **service 2 - R Shiny application:** This services provides a ready-to-use tool to explore the biogeography of plankton genetic diversity related to key metabolic pathways. It is designed for general users and easily accessible at the following link: [not yet implemented]

# 2. Service 1 installation

## 2.1. Cloning the files

The files for this service available in a zip.file at: <https://data.d4science.net/K6Ui>

To run the service within the BlueCloud infrastructure, you need to be registered at: <https://blue-cloud.d4science.org/group/planktongenomics>

* Open a Jupyter Hub **R environment, large 32Gb RAM / 8 cores**.
* Copy and unzip the bluecloud\_wp3\_d2\_s1.zip in the root. The pathway should be /home/jovyan. To unzip the file in the root, open a terminal and execute the following code:

cd  
unzip bluecloud\_wp3\_d2\_s1

## 2.2. Libraries installation

In order to run this notebook, several R packages and Python libraries are necessary. In the case they are not already installed on the server, please follow the procedure below:

* All R packages used are available on CRAN. Therefore, open a command prompt and execute the following code:

R -e "install.packages(c('raster','virtualspecies','abind','feather','reticulate','RColorBrewer','parallel','mvrsquared','tidyverse','RSQLite','RPostgreSQL','Shiny','Shinybusy'), repos='https://cran.r-project.org/', dependencies=TRUE)"

* The Python libraries are not available on public repositories. Therefore, open a command prompt and execute the following code. Please note that the path needs to correspond to the function/MBTR folder, the example corresponds to a Jupyter Hub instance on BlueCloud:

cd /home/jovyan/function/MBTR  
pip install .

### 2.3. Service description

In the root, you should find a Service\_1.ipynb file. The latter contains detailed description of the R pipeline as well as executable code. Please refer to this file for the service description and use.