Mining anaerobic digestion data with Deep-omics, a new digital environmental engineering platform for omics data

Cédric Midoux, Guillaume Perréal, Marion Predhumeau, Olivier Rué, Yannick Fayolle, Mahendra Mariadassou, P. Dabert, Théodore Bouchez, Nicolas Raidelet, Ariane Bize

HAL Id: hal-04361137

https://hal.inrae.fr/hal-04361137

Submitted on 22 Dec 2023

Mining anaerobic digestion data with Deep-omics, a new digital environmental engineering platform for omics data

C Midoux, G Perréal, O Chapleur, M Predhumeau, O Rué, Y Fayolle, M Mariadassou, P Dabert, V Loux, T Bouchez, N Raidelet, A Bize

IRSTEA, Hydrosystems and Bioprocesses Research Unit, 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France; MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France, cedric.midoux@irstea.fr

IRSTEA, Direction des Systèmes d'Information et du Numérique, 5 rue de la Doua, CS 20244, 69625 VILLEURBANNE Cedex, France, guillaume.perreal@irstea.fr

IRSTEA, Hydrosystems and Bioprocesses Research Unit, 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France, olivier.chapleur@irstea.fr

IRSTEA, Direction des Systèmes d'Information et du Numérique, 5 rue de la Doua, CS 20244, 69625 VILLEURBANNE Cedex, France, manon.predumeau@irstea.fr

MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France, olivier.rue@inra.fr

IRSTEA, Hydrosystems and Bioprocesses Research Unit, 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France, yannick.fayolle@irstea.fr

MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France, mahendra.mariadassou@inra.fr

IRSTEA, Optimisation des procédés en agriculture, agroalimentaire et environnement, 17 Avenue de Cucillé, CS 64427, 35044 Rennes, France, patrick.dabert@irstea.fr

MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France, valentin.loux@inra.fr

IRSTEA, Hydrosystems and Bioprocesses Research Unit, 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France, theodore.bouchez@irstea.fr

IRSTEA, Direction des Systèmes d'Information et du Numérique, 5 rue de la Doua, CS 20244, 69625 VILLEURBANNE Cedex, France, nicolas.raidelet@irstea.fr

IRSTEA, Hydrosystems and Bioprocesses Research Unit, 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France, ariane.bize@irstea.fr

Abstract:
Taking advantage of high throughput meta-omics approaches to improve anaerobic digester design and management is a major challenge. Considering the variety of anaerobic digesters worldwide, achieving a critical mass of data is required to extract relevant and robust information thanks to meta-analyses. We present new user-friendly software tools developed to help gathering and mining such meta-omics data. Deep-omics (*Digital environmental engineering platform for omics data)* is a data warehouse dedicated to environmental biotechnologies. Its main objectives are to promote the sharing of high quality data and associated metadata, especially regarding operating conditions and process design; to offer the possibility to launch classical bioinformatics analyses on selected data; and to perform meta-analysis. Easy16S is an interactive web server tool for statistical analysis and visualisation of meta-omics data (http://genome.jouy.inra.fr/shiny/easy16S/). By conducting a meta-analysis of experimental data related to the inhibition of anaerobic digestion, we illustrate the potential of this user-friendly framework for data mining.

Keywords: data warehouse; data mining; meta-omics

Session – Microbiology of anaerobic digestion/(meta)genomic research

Introduction

The recent breakthrough of meta-omics offers new perspectives to the field of anaerobic digestion. It promises the deep understanding of microbial communities acting as catalysts and thereby the possibility to develop more cost-effective processes, through the design of biological indicators or through ecological engineering approaches. However, the variety of anaerobic digesters and operating conditions worldwide is undoubtedly very high. It is therefore difficult to extend the conclusions gained on one specific set of configurations. It highlights the need for data gathering and organization at a large scale, to take full advantage of these high throughput techniques, to perform meta-analyses and to be able to identify robust trends, in order to turn knowledge into operational outputs. We present here two new software tools developed to favour data mining of meta-omics data from environmental biotechnologies.

Material and Methods

Deep-omics is an n-tier web application: the user interface is a single page application built using the Angular framework (https://angular.io/). It accesses the data using a RESTful API. The latter is built with the Symfony framework (https://symfony.com/) and its API-platform extension (https://api-platform.com/); it connects the different server-side processes. Data are stored in a relational database (with PostgreSQL, https://www.postgresql.org/) as well as in an indexation and search engine (open-source version of Elasticsearch, https://www.elastic.co/fr/community). Heavy calculation jobs are submitted to the MIGALE computing cluster, operated by INRA (http://migale.jouy.inra.fr/) and offering a wide range of bioinformatic tools.

Easy16S is an interactive R shiny interface based on two main R packages, Shinydashboard (Chang and Ribeiro, 2017) and phyloseq (MacMurdie and Holmes, 2013). It is currently deployed on INRA-MIGALE server.

To produce an illustrative example of meta-analysis with Deep-omics and Easy16S, we relied on datasets previously published by some of us (Poirier and Chapleur, 2018). For that purpose, reads generated by 16S rRNA gene sequencing of 90 samples were processed with FROGS (Escudié et al., 2018) on INRA-MIGALE server. Statistical analysis of the obtained BIOM table was subsequently conducted with Easy16S.

Results and Conclusions

1. Deep-omics: a digital environmental engineering platform for omics data

The objective of Deep-omics (**Figure 1.1**) is to offer the possibility to upload, request and export meta-omics data from environmental biotechnologies together with many relevant associated metadata, especially regarding operating conditions and process design.
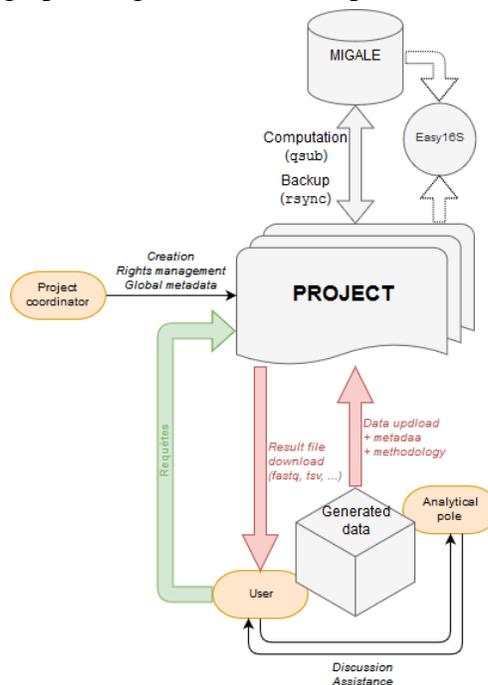

Figure 1.1 Illustrative scheme of Deep-omics design

Other functions are the graphical monitoring of analytical data and the possibility to launch bioinformatics calculations on selected data directly from Deep-omics. Finally, data collection is

organised at the project level: the project coordinator manages the rights associated to the project data (e.g. anonymised, private…). This tool is presently internal to Irstea. Several modules have been developed and are being tested and improved. Our objective is to open this tool to a wider community in the future.

2. Easy16S: a dynamic interface for statistical analysis and visualisation of 16S metabarcoding or other meta-omics data

Easy16S is a user-friendly web server tool intended for practitioners eager to explore their data and create figures rapidly and highly interactively. It is specifically focused on the mapping of covariates of interest (**Figure 2.1**). As it does not require any coding skills, it is accessible to a wide community. Easy16S accepts as entry the classical file formats associated to 16S metabarcoding (BIOM, Phyloseq Rdata, tsv) and it includes a demo dataset. Interestingly, the R code lines corresponding to each plot can be copy-pasted from the webpage. Updated on a regular basis, Easy16S is available online (http://genome.jouy.inra.fr/shiny/easy16S/) on INRA-MIGALE server and it will also be accessible from Deep-omics, to provide a united framework for data mining and meta-analysis.
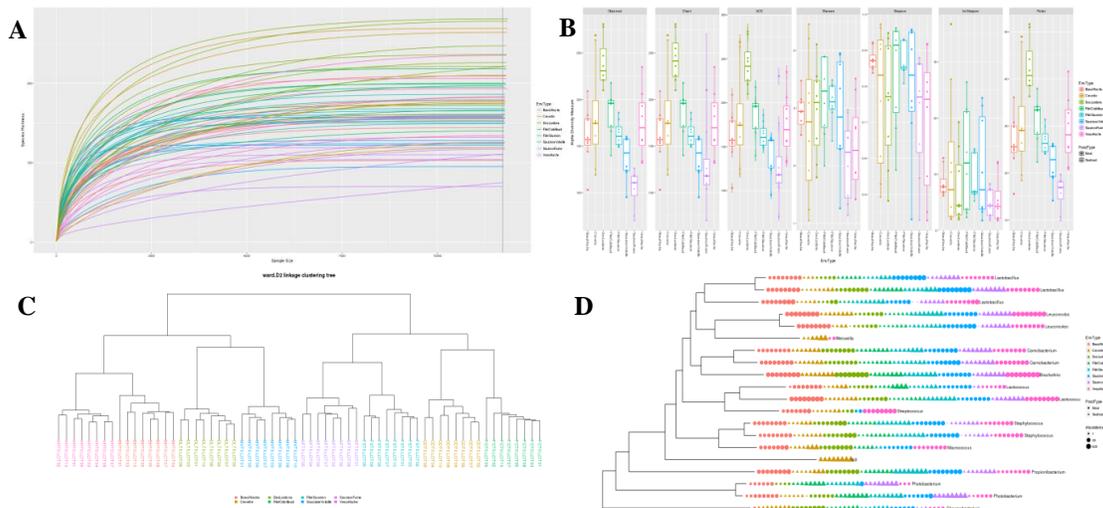
**Figure 2.1** Non exhaustive examples of Easy16S graphical outputs. A: rarefaction curves– B: α-diversity box plots – C: β-diversity, hierarchical clustering – D: phylogenetic tree

3. Case study: towards the identification of anaerobic digestion inhibition bioindicators

Performance stability is a key operational issue for anaerobic digestion (AD) and numerous compounds regularly cause digester failures. In the present example, we questioned the possibility to obtain microbial indicators revealing a perturbation due to the presence of distinct inhibitors. By jointly analyzing the data from (Poirier et al., 2016 a and b), we focused on phenol and ammonia. Indeed, in these studies, triplicate anaerobic batch toxicity assays had been conducted in parallel with ten phenol concentrations (up to 5.0 g/L) (Poirier et al., 2016 a) and ten total ammonia nitrogen concentrations (up to 50.0 g/L) (Poirier et al., 2016 b).

Thanks to the pipelines included in Deep-omics and to Easy16S, we compared their influence on the same initial microbiota. Bacterial responses to increasing levels of inhibitors clearly revealed two distinct adaptation patterns (**Figure 3.1 A**). Increasing levels of ammonia induced a progressive shift towards bacterial families *Syntrophomonadaceae*, *Porphyromonadaceae*, *Marinilabiaceae* and *Synergystaceae;* by contrast, high levels of phenol were characterized by the emergence of other bacterial families: *Bacteroidaceae* and *Lactobacillaceae* (**Figure 3.1 B**).
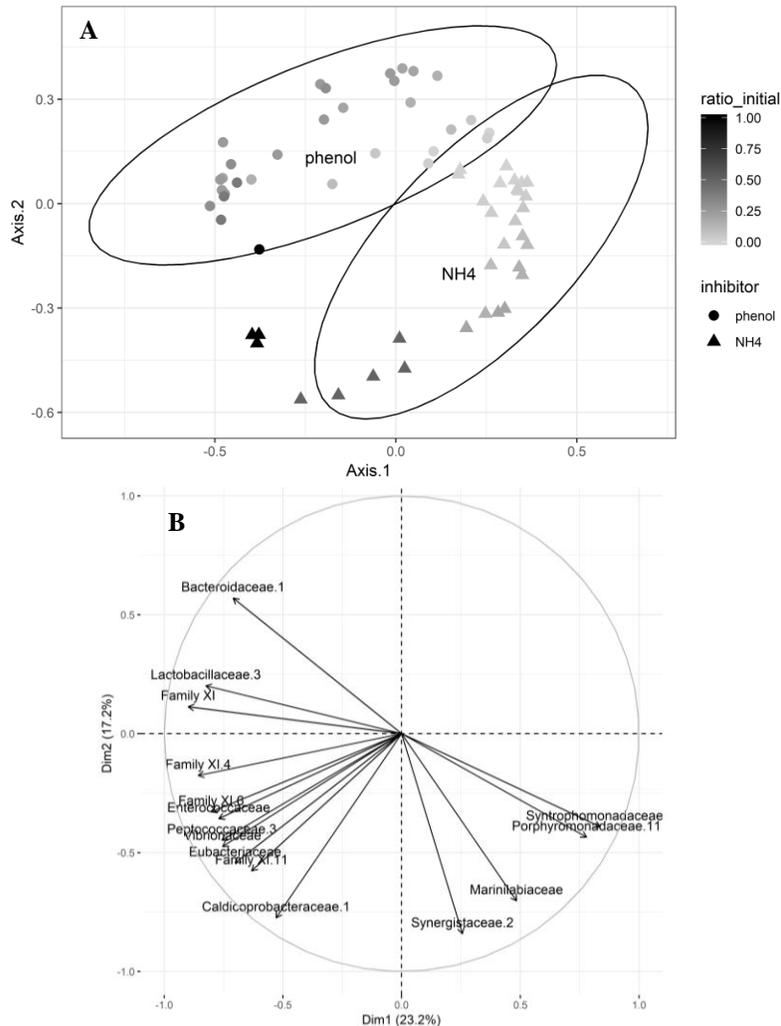
Figure 3.1 Principal Component Analysis of 16S metabarcoding data. A: samples; B: Operational Taxonomic Units

These results show the relevance of our user-friendly framework to conduct meta-analyses and search for new bioindicators.

References

Chang W & Ribeiro BB (2017) Shinydashboard: create dashboards with'Shiny'. *R package*.

Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S & Pascal G (2018) FROGS: Find, Rapidly, OTUs with Galaxy Solution. *Bioinformatics* **34**: 1287-1294.

McMurdie PJ & Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**: e61217.

Poirier S & Chapleur O (2018) Influence of support media supplementation to reduce the inhibition of anaerobic digestion by phenol and ammonia: Effect on degradation performances and microbial dynamics. *Data in Brief* **19**: 1733-1754.

Poirier S, Bize A, Bureau C, Bouchez T & Chapleur O (2016a) Community shifts within anaerobic digestion microbiota facing phenol inhibition: Towards early warning microbial indicators? *Water Research* **100**: 296-305.

Poirier S, Desmond-Le Quéméner E, Madigou C, Bouchez T & Chapleur O (2016b) Anaerobic digestion of biowaste under extreme ammonia concentration: Identification of key microbial phylotypes. *Bioresource Technology* **207**: 92-101.