



HAL
open science

Mining anaerobic digestion data with DeepOmics and Easy16S, user-friendly tools for environmental engineering meta-omics data

Ariane Bize

► To cite this version:

Ariane Bize. Mining anaerobic digestion data with DeepOmics and Easy16S, user-friendly tools for environmental engineering meta-omics data. ICBM-4 (4th International Conference on Biogas Microbiology), May 2022, Braga, Portugal. hal-04362416

HAL Id: hal-04362416

<https://hal.inrae.fr/hal-04362416>

Submitted on 22 Dec 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.



➤ Mining anaerobic digestion data with DeepOmics and Easy16S, user-friendly tools for environmental engineering meta-omics data

Ariane Bize ariane.bize@inrae.fr

4th International Conference on Biogas Microbiology (ICBM-4)

9 - 11 May 2022 | Altice Forum Braga

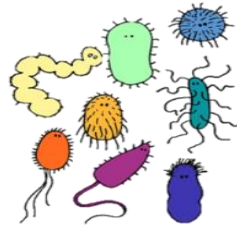
➤ Introduction

Using omics data to steer anaerobic digestion?

**Operating parameters
Process design**



**Structure, diversity, and
metabolic functions of
microbial communities**



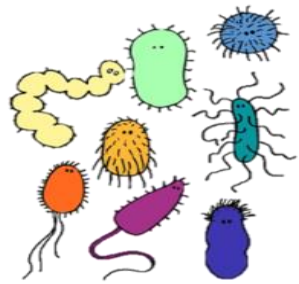
**Performance and functions
of the processes**



➤ Introduction

Challenges

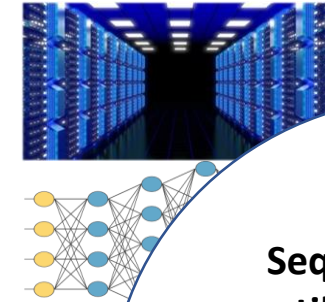
Complex microbial communities



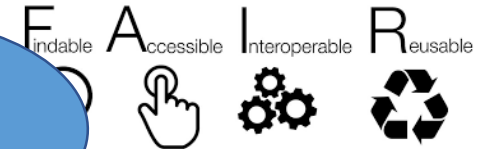
A great variety of process configuration



A critical mass of data is needed

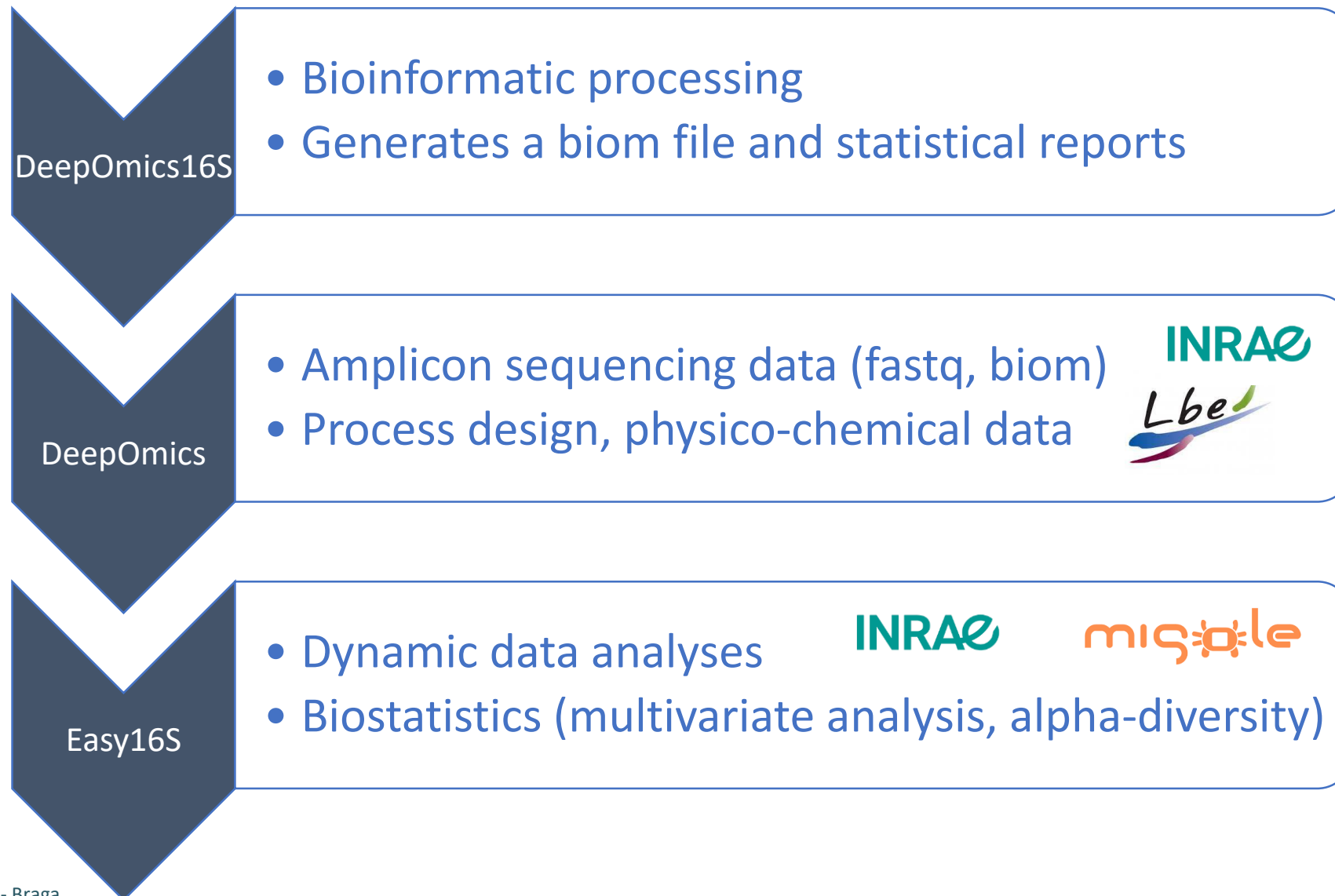


Sequencing data are still rarely reusable

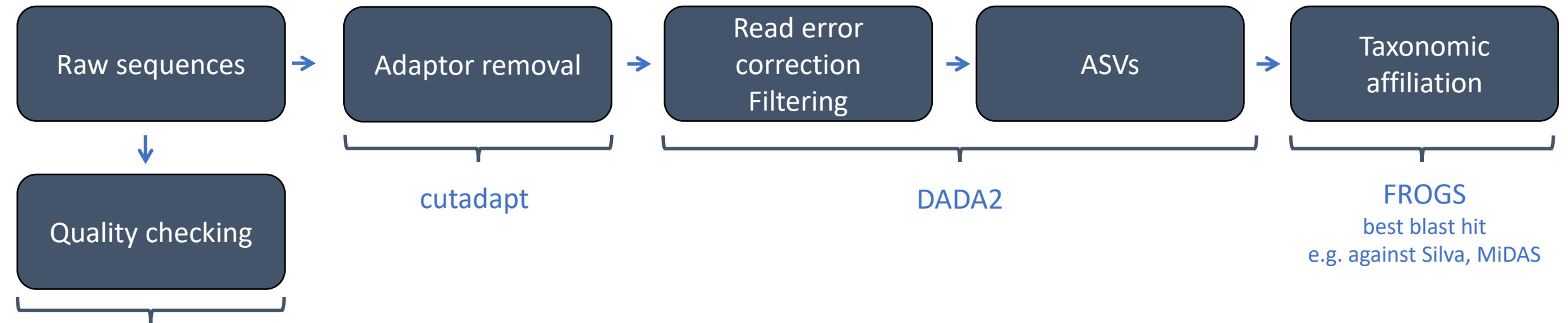


Need to promote FAIR data production
→ the **DeepOmics initiative**,
developing an Information System for
environmental biotechnology
processes

➤ DeepOmics and complementary tools for meta-omics data from environmental biotechnology processes



➤ Tool 1: DeepOmics16S, a bioinformatics pipe-line → on gitlab



fastQC - MultiQC

- state-of-the art **bioinformatics pipe-line for 16S metabarcoding** data analysis, to favour data homogeneity and comparison
- snakemake workflow
- in production → on gitlab
<https://forgemia.inra.fr/cedric.midoux/deepomics16S>
- by June 2022: single-end and paired-end sequences
- developped by Cédric Midoux

MultiQC: Ewels et al, *Bioinformatics*, 2016
Cutadapt: Martin M, *EMBnet. Journal*, 2011
DADA2: Callahan, et al, *Nature methods*, 2016
FROGS: Escudié, et al, *Bioinformatics*, 2017
Silva: Quast, et al, *Nucleic acids research*, 2012
MiDAS: McIlroy, et al, *Database*, 2015

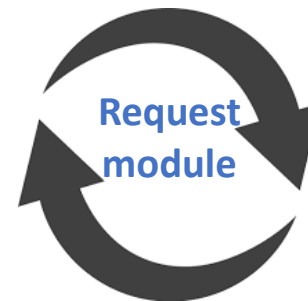
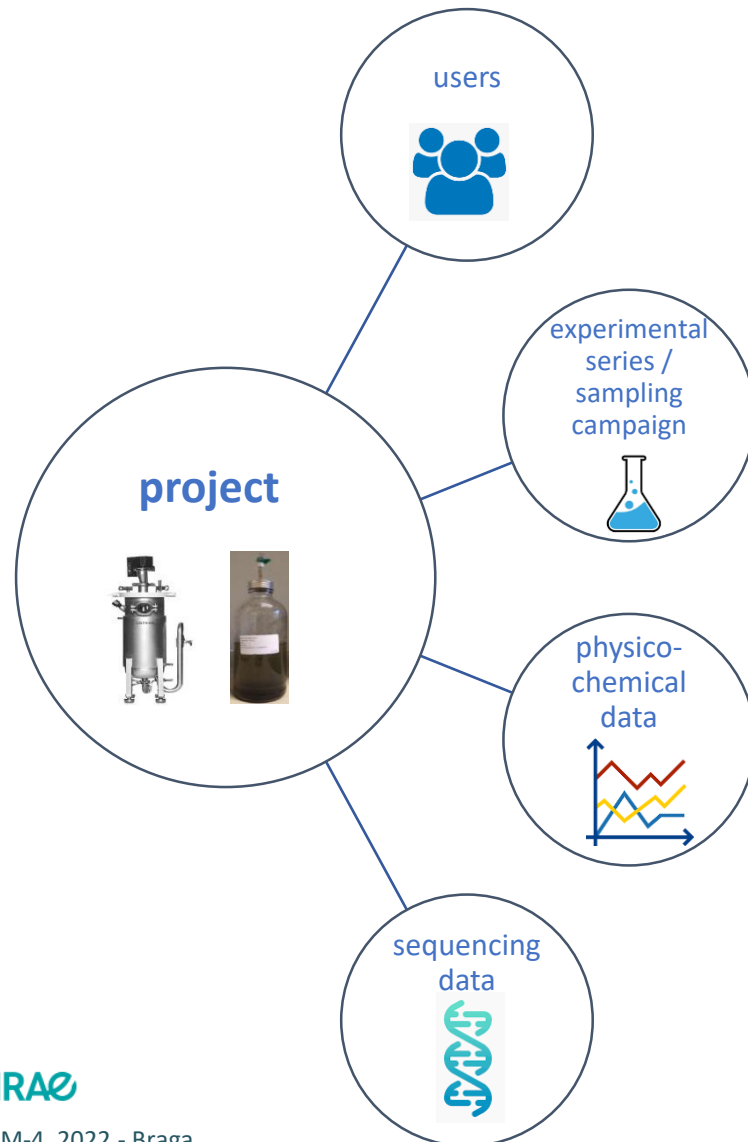


INRAE

ICBM-4 2022 - Braga

11th May 2022 – Ariane Bize

➤ Tool 2: DeepOmics, an information system → in pre-production stage



(to be developed)

A 3 year project (Carnot 3BCAR) is starting to pursue the development of DeepOmics and of another complementary tool, OpenSILEX EnviBIS (INRAE-LBE), driven by ontologies

INRAE



→ Better if it meets the scientific community needs

→ What are your needs?

INRAE

➤ Tool 2: DeepOmics, an information system → in pre-production stage

The screenshot displays the INRAE DeepOmics web interface. At the top, the header includes the INRAE logo, the text "DeepOmics", and the user name "ARIANE BIZE D'AVEZAC DE CASTERA". Below the header is a navigation menu with links for Home, Projects, Meta-analysis, Administration, DADA2, and Easy16S. The main content area shows a project titled "00_demo_AD_inhib_Poirier" with a "View details" link. Below this, there are three tabs: "Lab-scale process data", "Industrial and field process data", and "Meta-omics analysis", with the latter being highlighted in orange. On the left sidebar, under "Bioinformatics", the "Biosample results" link is also highlighted in orange. The main content area is titled "Biosample Results" and features a search bar and a "+ Create biomfile" button (highlighted in orange). Below these is a table with the following data:

<input type="checkbox"/>	Sequencing sample code	Metabarcoding bioinformatic run code	Processing metrics	Number of Annotation
<input type="checkbox"/>	S_00_00N2_day000 Sample : S_00_00N2_day000	NH4_bioinfo Workflow : dada2_frogs (status:private)	Number of raw reads: 53094, Post process reads: 26489, Number of ASV: 174.	174 Total annotation count : 26489
<input type="checkbox"/>	S_00_00N2_day009 Sample : S_00_00N2_day009	NH4_bioinfo Workflow : dada2_frogs (status:private)	Number of raw reads: 57022, Post process reads: 22442, Number of ASV: 183.	183 Total annotation count : 22442
<input type="checkbox"/>	S_00_00N2_day029 Sample : S_00_00N2_day029	NH4_bioinfo Workflow : dada2_frogs (status:private)	Number of raw reads: 56949, Post process reads: 25483, Number of ASV: 229.	229 Total annotation count : 25483



➤ Tool 2: DeepOmics, an information system → in pre-production stage

Project "00_demo_AD_inhib_Poirier" [View details](#)

Lab-scale process data | Industrial and field process data | Meta-omics analysis

Laboratory

- Project's input management
- Experimental series

NH4

- Operating conditions
- Biosamples

Physico-chemical and qPCR

- Data configuration
- Data import
- Data management
- Data display**

Experimental series "NH4" [View details](#)

Graph Collection "Methane cumulated production" [New graph](#) [Close](#)

Graph Replicate 1 | Data Replicate 1

Replicate 1

normo-mL

5000
4000
3000
2000
1000
0

Apr 2014 May 2014 Jun 2014 Jul 2014 Aug 2014

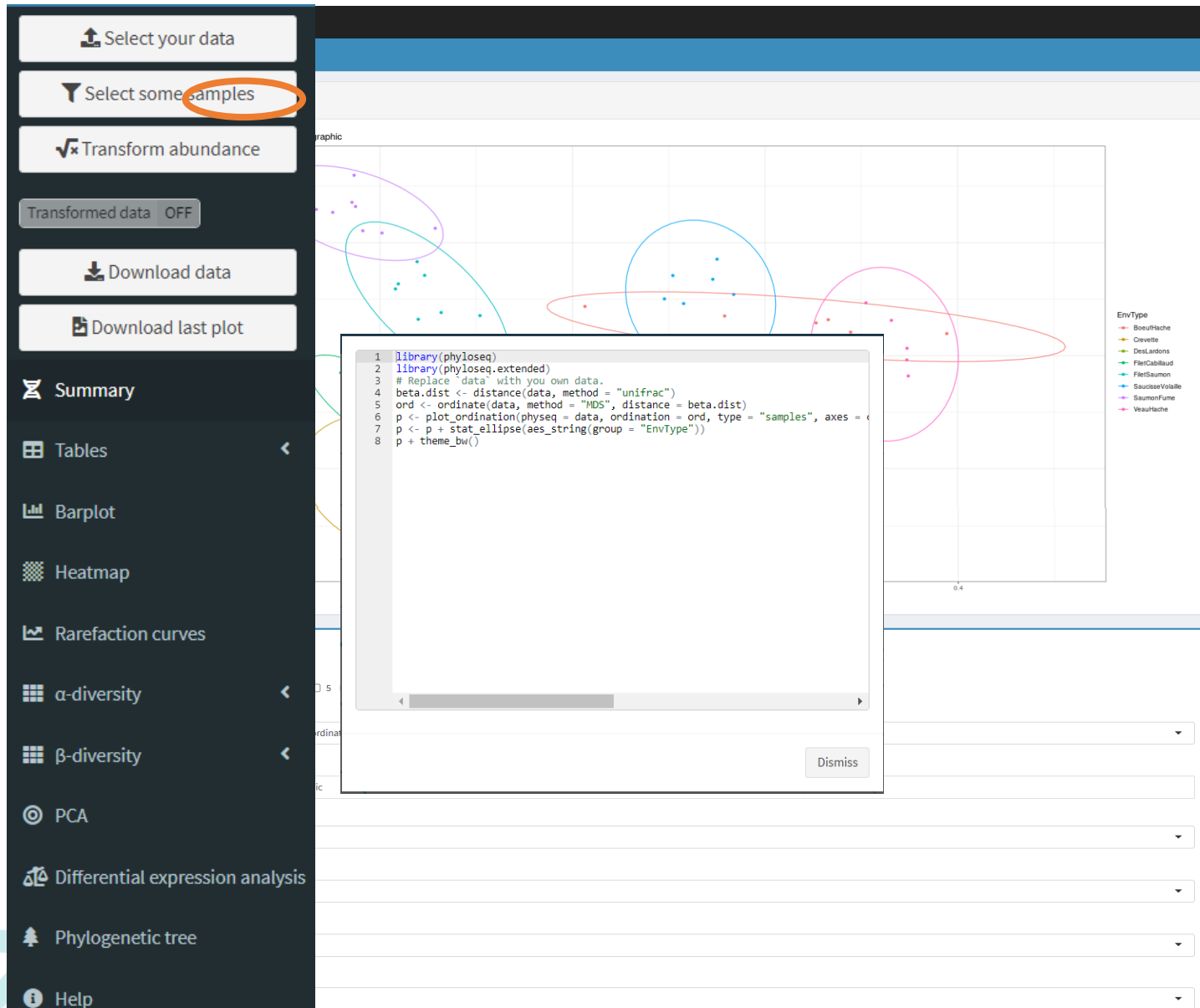
- cum_ch4 - 1 > NH4_00000_rep1
- cum_ch4 - 1 > NH4_00500_rep1
- cum_ch4 - 1 > NH4_01000_rep1
- cum_ch4 - 1 > NH4_01500_rep1
- cum_ch4 - 1 > NH4_02500_rep1
- cum_ch4 - 1 > NH4_05000_rep1
- cum_ch4 - 1 > NH4_07500_rep1
- cum_ch4 - 1 > NH4_10000_rep1
- cum_ch4 - 1 > NH4_25000_rep1
- cum_ch4 - 1 > NH4_50000_rep1

omics-test.solapp.inrae.fr/projects/6779bd4a-b938-4e48-ad18-557a7300f471/.../inputs

➤ Tool 3: Easy16S → a webserver

<https://shiny.migale.inrae.fr/app/easy16S/>

INRAE migale

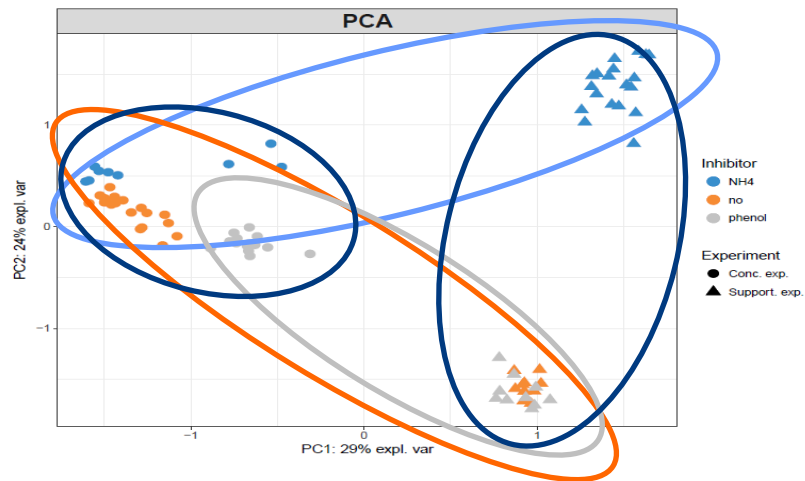


A Shiny application,
developped by Cédric Midoux

➤ Application potential, example 1: AD inhibition

2 experimental series
144 samples analyzed by 16S metabarcoding

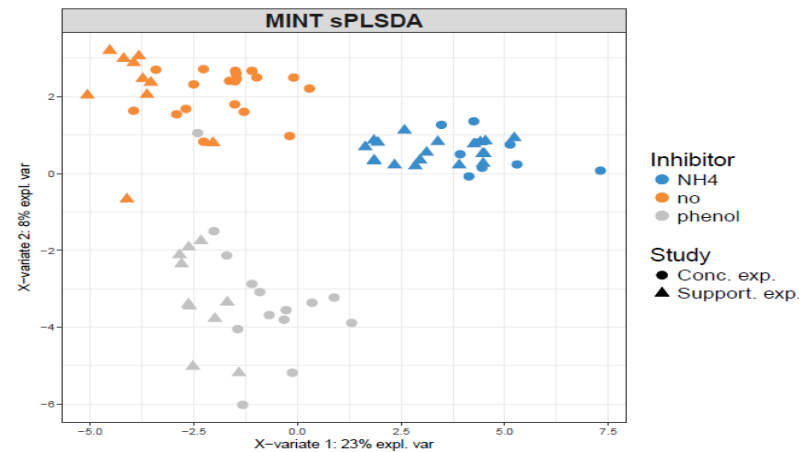
Sample ordination with PCA



Important “study” effect
No possibility to generalize?



Multivariate INTEgrative method, MINT



Correcting for the “study effect”
Results can be generalized

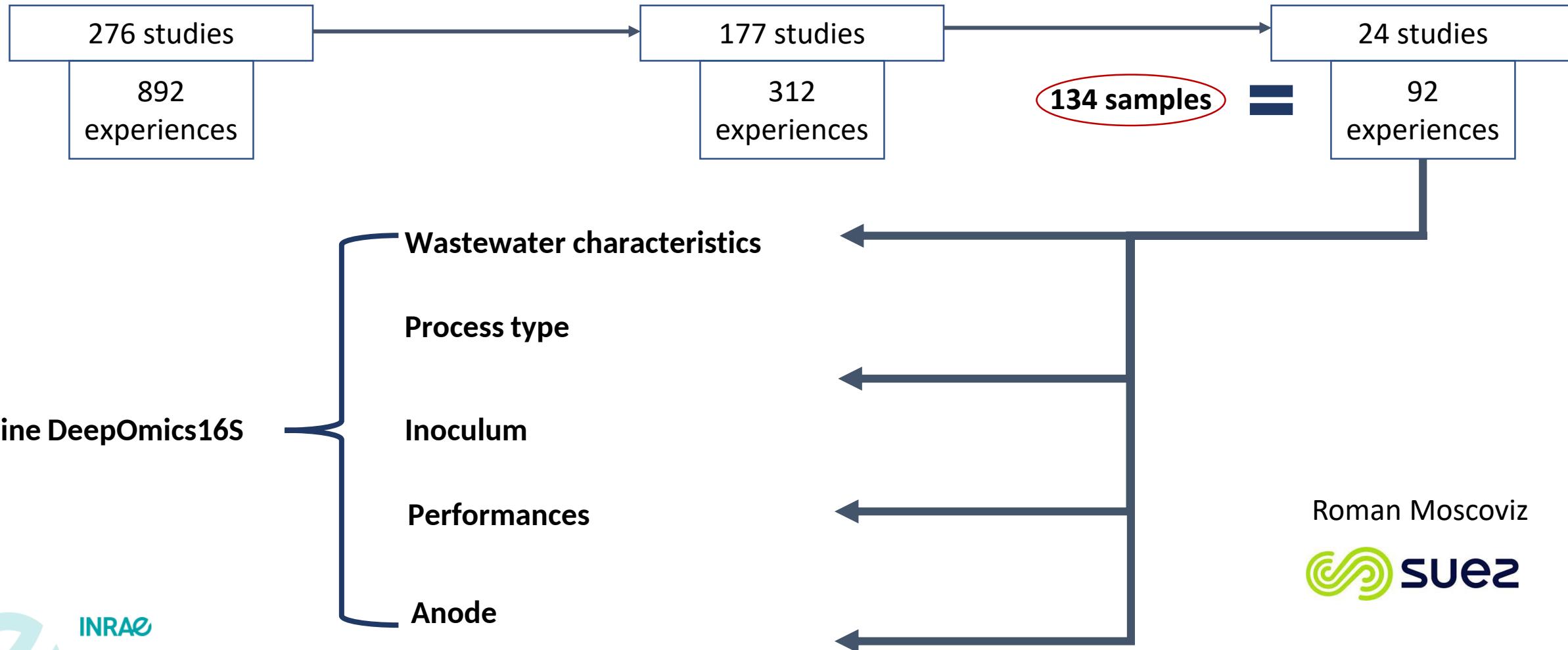


Increase of the **Clostridiales** relative abundance is an early warning of AD inhibition.

Presence of **Cloacimonetes** is associated with good performance of methane production.



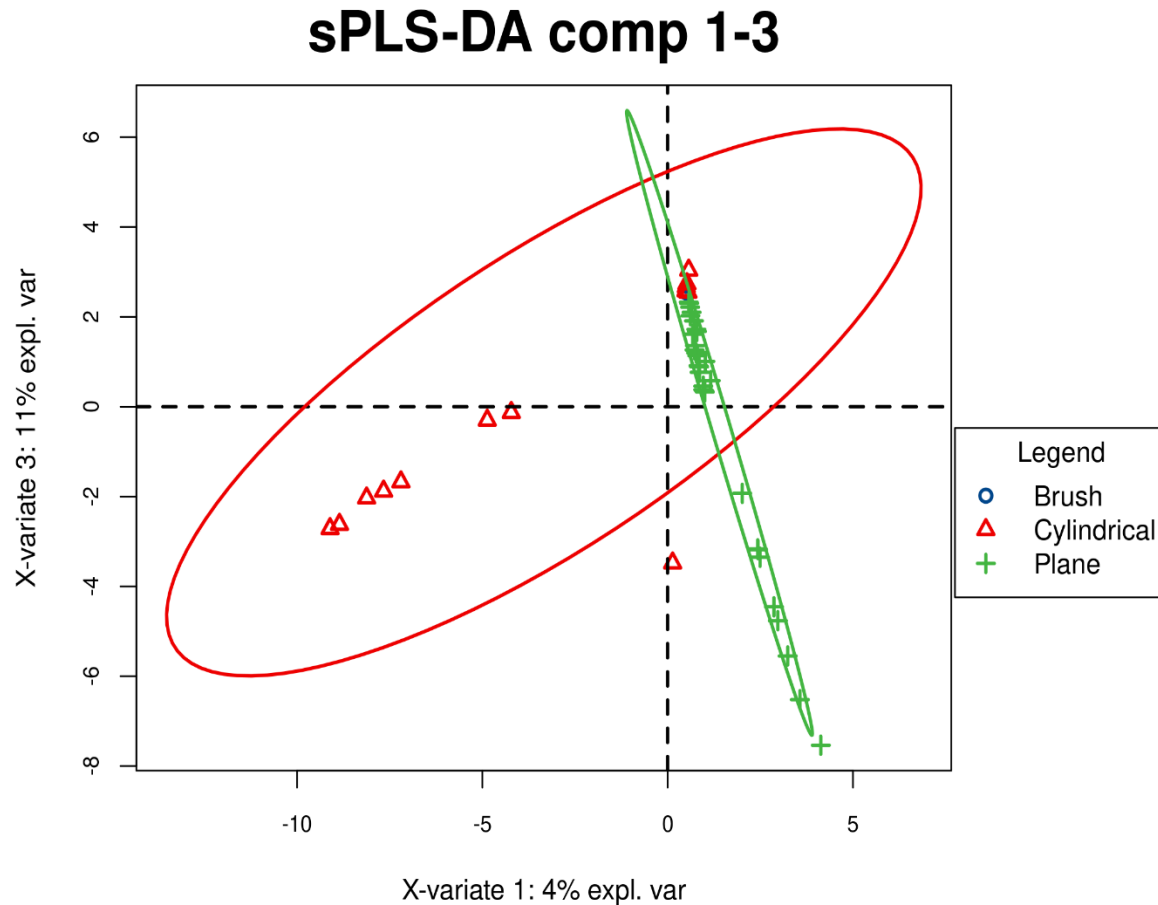
➤ Application potential, example 2: bioelectrochemical systems



Pipeline DeepOmics16S

Roman Moscoviz
 **suez**

➤ Application potential, example 2: bioelectrochemical systems



Main outcome

Geobacter members, known to be the main electroactive microorganisms at bioanodes, were more abundant in the planar geometry which would suggest that this geometry would have an effect on the current density of an MFC

Other microorganisms were identified such as *Nitrosomonas* and *Desulfuromonas acetexigens* bacteria

Roman Moscoviz



INRAE

ICBM-4 2022 - Braga

11th May 2022 – Ariane Bize

➤ Conclusions and perspectives

- **DeepOmics16S / DeepOmics / Easy16S**

- complementary tools
- to capitalize on data
- and help practitioners in the production of FAIR* data
- to facilitate AI analyses
- in the field of anaerobic digestion

- **Perspectives**

- consult potential users to better understand their needs
- ensure the interoperability (or even integration) of these different tools
- develop a userfriendly request module
- extend DeepOmics to shotgun sequencing data



➤ Thank you for your attention!



Cédric Midoux
Véronique Jamilloux
Sleheddine Kastali
Olivier Chapleur
Yannick Fayolle
Théodore Bouchez



Patrick Dabert



Elie Le Quémener
Virginie Rossard
Eric Latrille
Emilie Fernandez



Guillaume Perréal
Nicolas Raidelet
Christophe Biasani
Aurélie Gramusset
Marion Predhumeau



Valentin Loux



Project 3BCAR MEMOS (21-A3)



INRAE