

# SAMBA : a flexible automated workflow for the reproducible and standardized processing of eDNA metabarcoding data

L. Quintric, C. Noël, L. Leroi, A. Cormier & P. Durand

samba-sebimer@ifremer.fr

*IFREMER-IRSI-Service de Bioinformatique (SeBiMER), Centre Bretagne - 29280 PLOUZANE, FRANCE*



FAIR  
science

Optimized ASV  
approach

Sample  
decontamination

Extended  
biostatistics  
analysis

Full workflow  
HTML report

Environmental DNA (eDNA) metabarcoding → a powerful method for assessing the diversity and dynamics of microbial communities from various environmental samples (water, sediment, air).

Involves in-depth bioinformatics and biostatistics analyses → can remain complicated to use for researchers.

**Requires the implementation of automated, standardized and user-friendly solutions.**

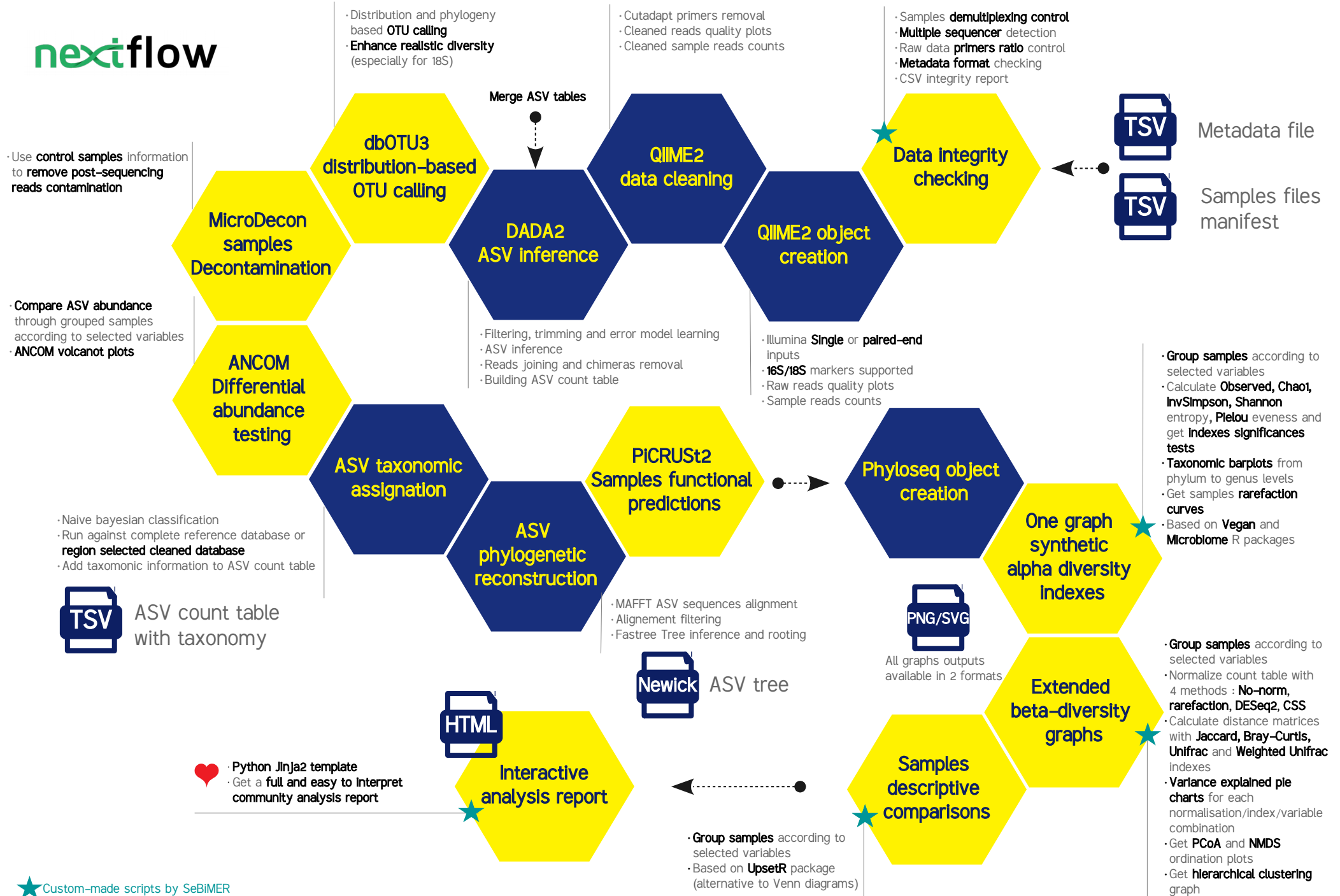
**SAMBA** → an open-source **modular workflow** to process **eDNA** metabarcoding data. It is built around three main parts: **data integrity checking**, **bioinformatics processes** and **statistical analyses**.

SAMBA offers a **real alternative** to the complex use of a suite of command line bioinformatics tools while providing access to **state-of-the-art methods** and tools in the field

SAMBA is freely available at <https://github.com/ifremer-bioinformatics/samba>

# SAMBA workflow

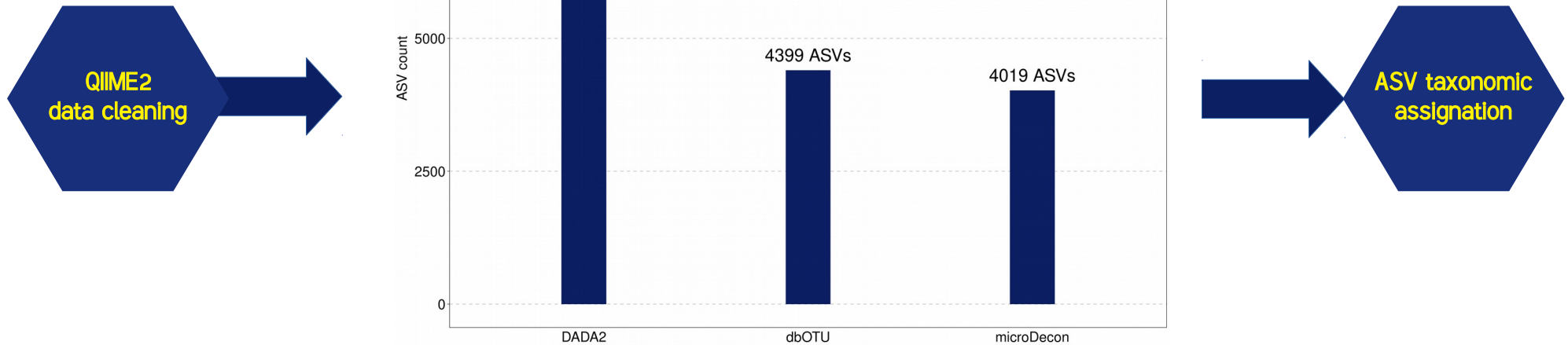
nextflow



Optimized ASV approach

Sample decontamination

Contributes to reach the real ASV diversity



→ dbOTU<sup>1</sup> : Distribution and phylogenetic based ASV clustering.

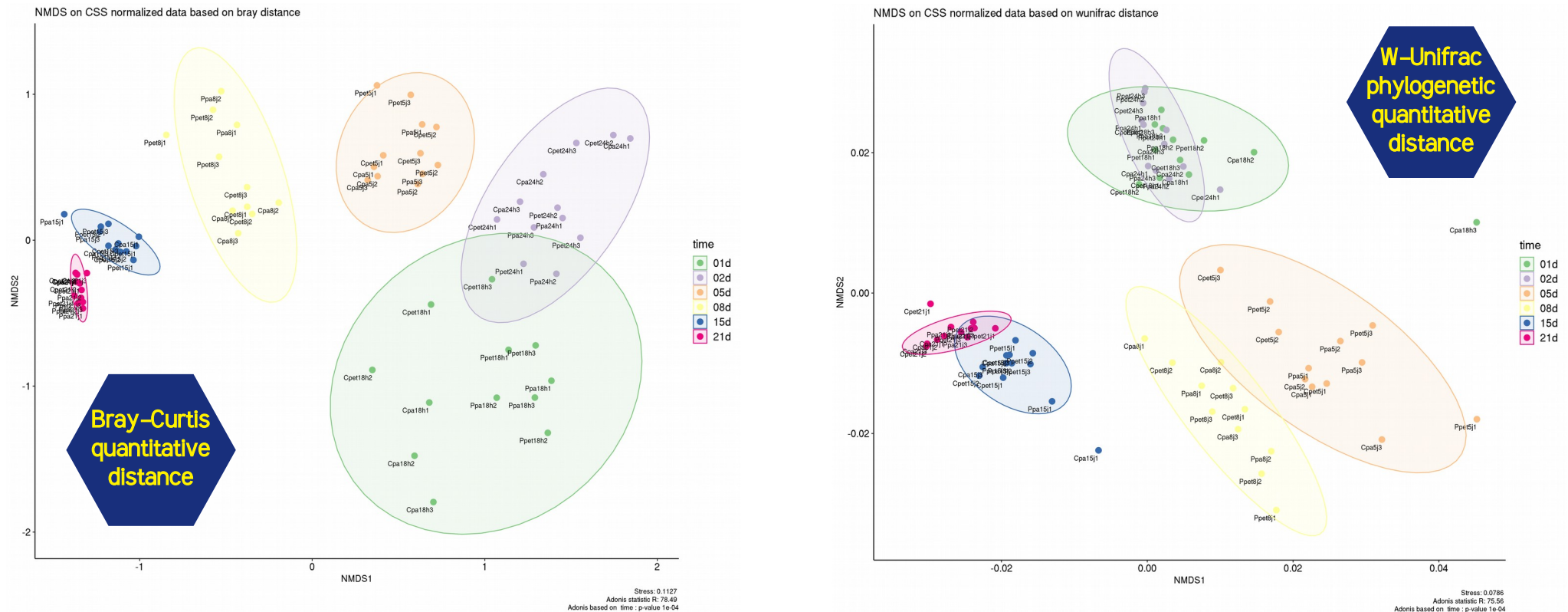
→ microDecon<sup>2</sup> : Remove contamination using experiment control samples.

<sup>1</sup> Olesen *et al.*, 2017. dbOTU3: A new implementation of distribution-based OTU calling

<sup>2</sup> McKnigh *et al.*, 2019. microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies

# Extended biostatistics analysis

## Contributes to the robustness and quality of biological interpretations



- Different normalization methods implemented<sup>1</sup>: rarefaction, DESeq2 and CSS.
- Four distance metrics used for  $\beta$ -diversity: Jaccard, Bray-Curtis, UniFrac and Weighted UniFrac.

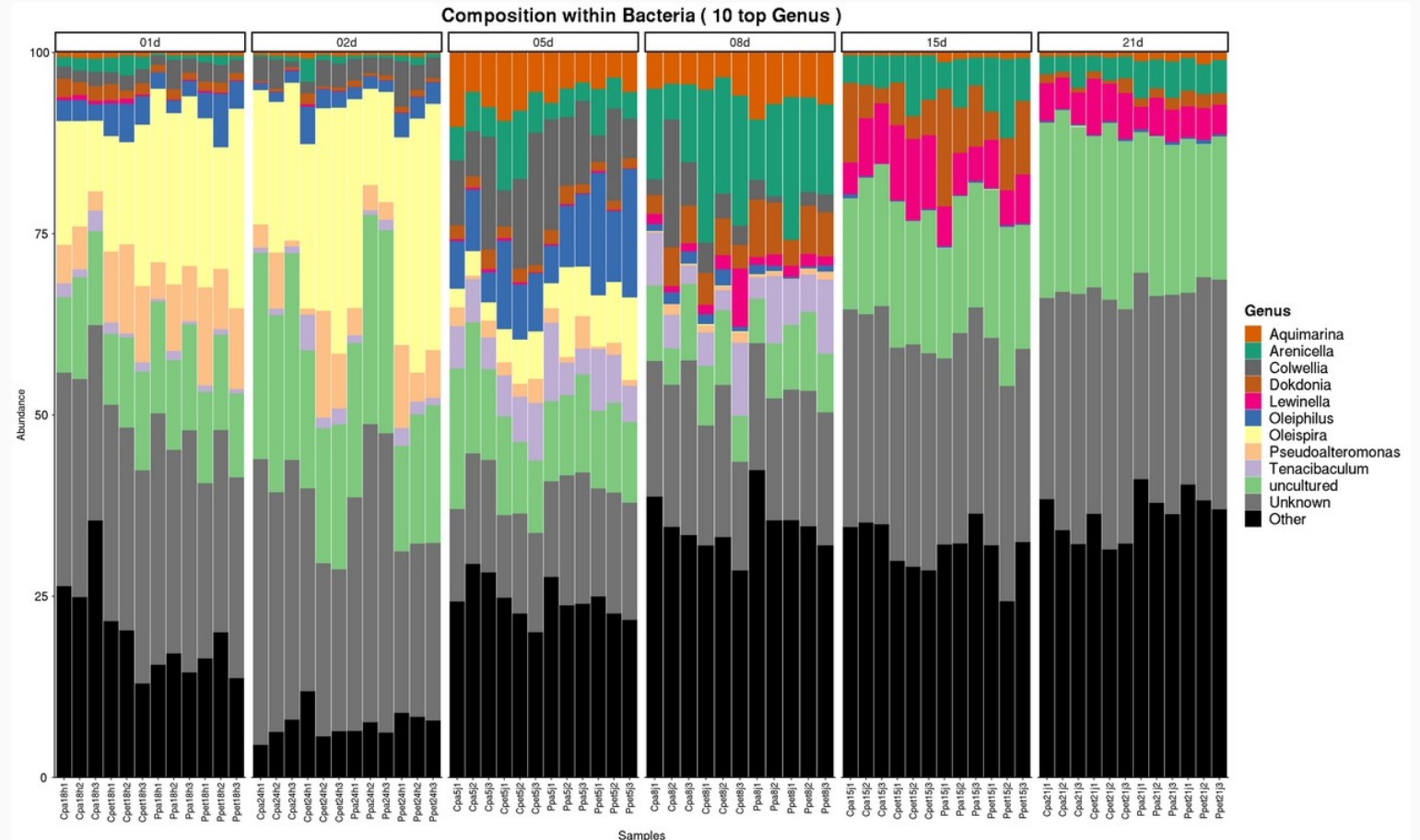
<sup>1</sup> McMurdie & Holmes, 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible

# Full workflow HTML report

<https://ifremer-bioinformatics.github.io/SAMBAExampleReport/>

- Taxonomic barplots group by **time** :

Barplot at the phylum level ; Barplot at the class level ; Barplot at the order level ; Barplot at the family level



- The results of the significance tests carried out for each variable on each diversity indexes can be viewed [here](#)

Get full HTML  
QIIME2 reports

Get full  
parameters for  
each process

Get all graphs  
from each  
process

Get results for all  
selected variables