



HAL
open science

Testing markers to characterize microbial communities in food ecosystems using a metagenomics approach

Claire Vincent, Serge Casaregola, Monika Coton, Céline Delbes, Hugo Devillers, Eric Dugat-Bony, Stephane Guezenec, Kate Howell, Françoise Irlinger, Jean-Luc Legras, et al.

► To cite this version:

Claire Vincent, Serge Casaregola, Monika Coton, Céline Delbes, Hugo Devillers, et al.. Testing markers to characterize microbial communities in food ecosystems using a metagenomics approach. JOBIM, Jun 2020, Montpellier, France. hal-02914960

HAL Id: hal-02914960

<https://hal.science/hal-02914960v1>

Submitted on 13 Aug 2020

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.



Testing markers to characterize microbial communities in food ecosystems using a metagenomics approach

Claire Vincent¹, Serge Casaregola², Monika Coton³, Céline Delbès⁴, Hugo Devillers⁶, Éric Dugat-Bony⁵, Stéphane Guezenc⁶, Kate Howell⁶, Françoise Irlinger⁵, Jean-Luc Legras⁶, Valentin Loux¹, Élixa Michel⁶, Jérôme Mounier³, Cécile Neuveglise⁶, Thibault Nidelet⁶, Audrey Pawtowski³, Pierre Renault², Delphine Sicard⁶, Monique Zagorec⁷ and Olivier Rué¹

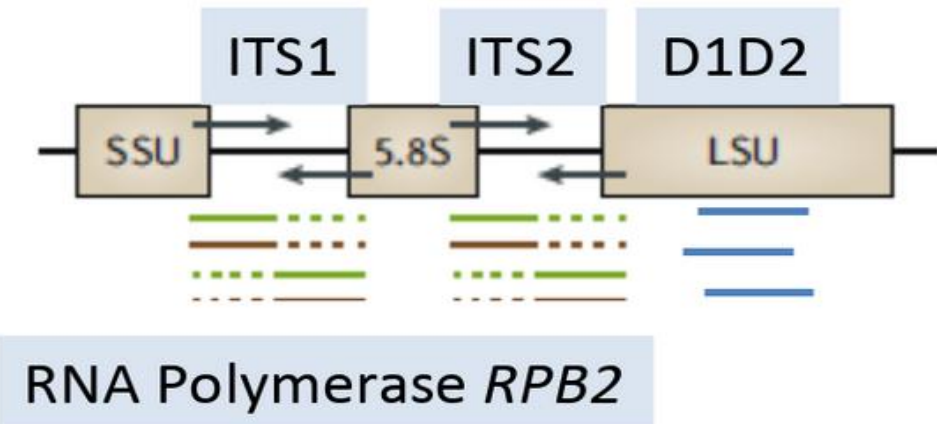
¹ INRAE - Université Paris-Saclay, MaIAGE, BioinfOmics, MIGALE bioinformatics facility, Jouy-en-Josas ² INRAE, MICALIS, Jouy-en-Josas ³ UBO, LUBEM, Plouzané ⁴ INRAE, UMR, Aurillac, ⁵ INRAE, SayFood, Thiverval-Grignon, ⁶ INRAE, SPO, Montpellier, ⁷ INRAE, SECALIM, Nantes

Context

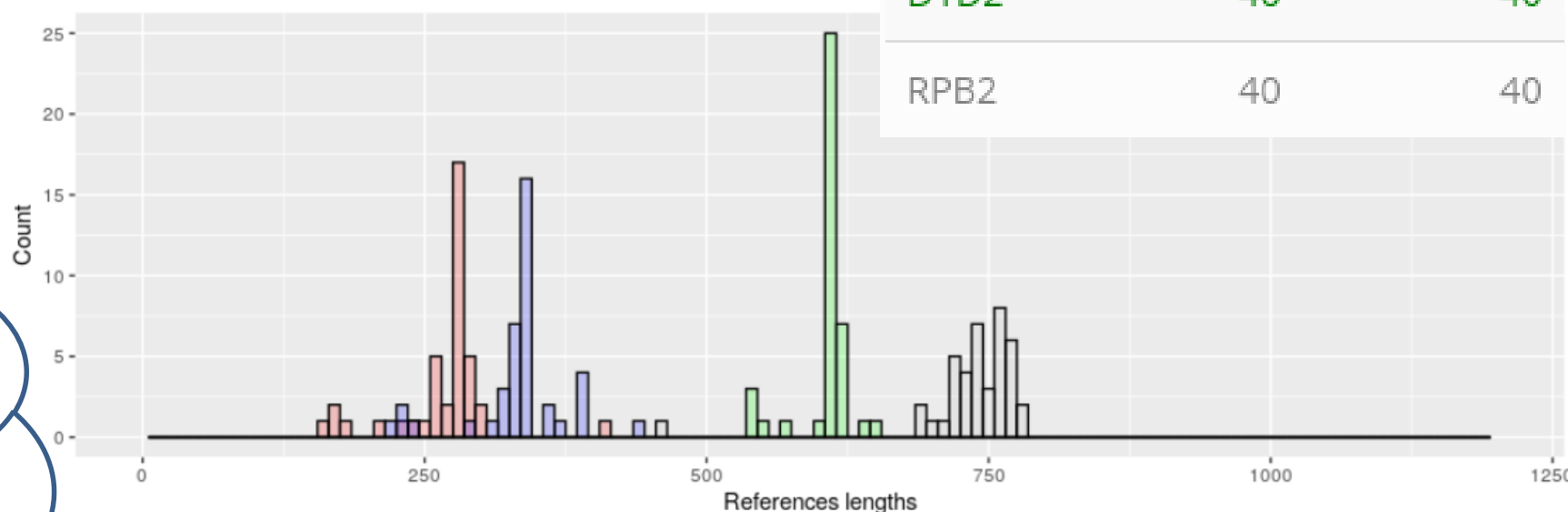
While metabarcoding is commonly used to describe prokaryotes in the microbiome of many environments, methods for describing micro-eukaryote diversity is lacking and requires better methodology and standardisation. One reason is that the universal fungal barcode, the Internal Transcribed Spacer (ITS) region, displays considerable size variation among yeasts and other micro-eukaryotes. There are also several repeats leading to sequencing errors or termination. Additionally, the ITS databases are far from complete, especially for Ascomycota that are commonly found in food. Other rDNA barcodes have been used but often do not harbor enough polymorphism to identify taxa at the Species level. In food, microbiota are usually composed of a reduced number of species compared to wild environments. Identifying micro-eukaryotes at the Species level, and potentially strain level, is therefore necessary.

Study Design

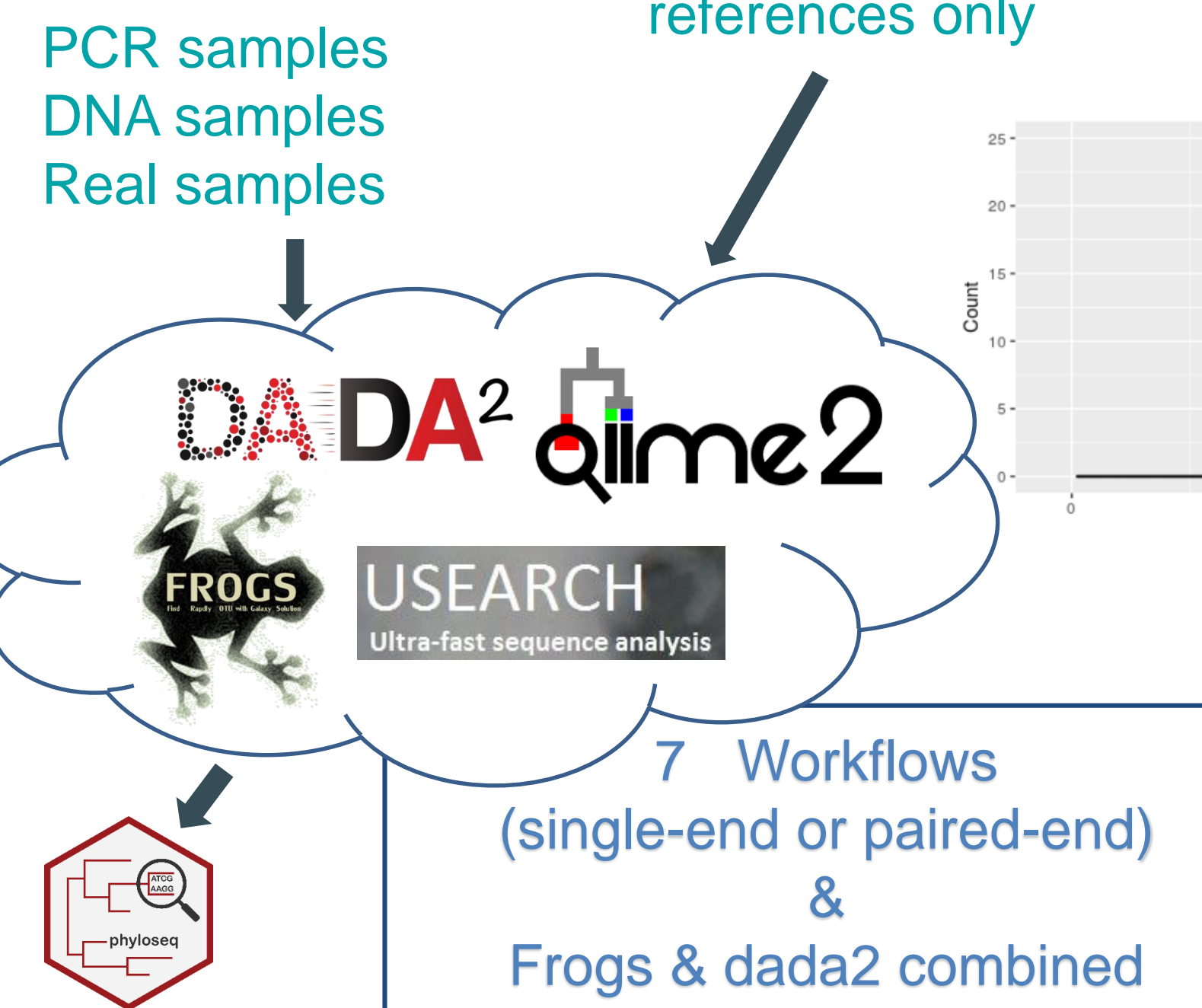
- ✓ 5 MOCKs (fermented meats, cheese, bread, wine, drosophila)
- ✓ Composed of all species reported at least twice in the ecosystem
- ✓ From 25 to 61 species per MOCK
- ✓ One strain / species
- ✓ DNA extraction of each strain by kit
- ✓ Independent PCR1 of four markers (ITS1, ITS2, D1D2, RPB2) with adaptors
- ✓ For each MOCK, DNA pooling followed by PCR1 and in parallel PCR1 products pooling
- ✓ PCR2 with labelled primers and MiSeq sequencing
- ✓ Pipeline testing



Markers	Expected	InDatabase
ITS1	40	40
ITS2	40	40
D1D2	40	40
RPB2	40	40



Methods

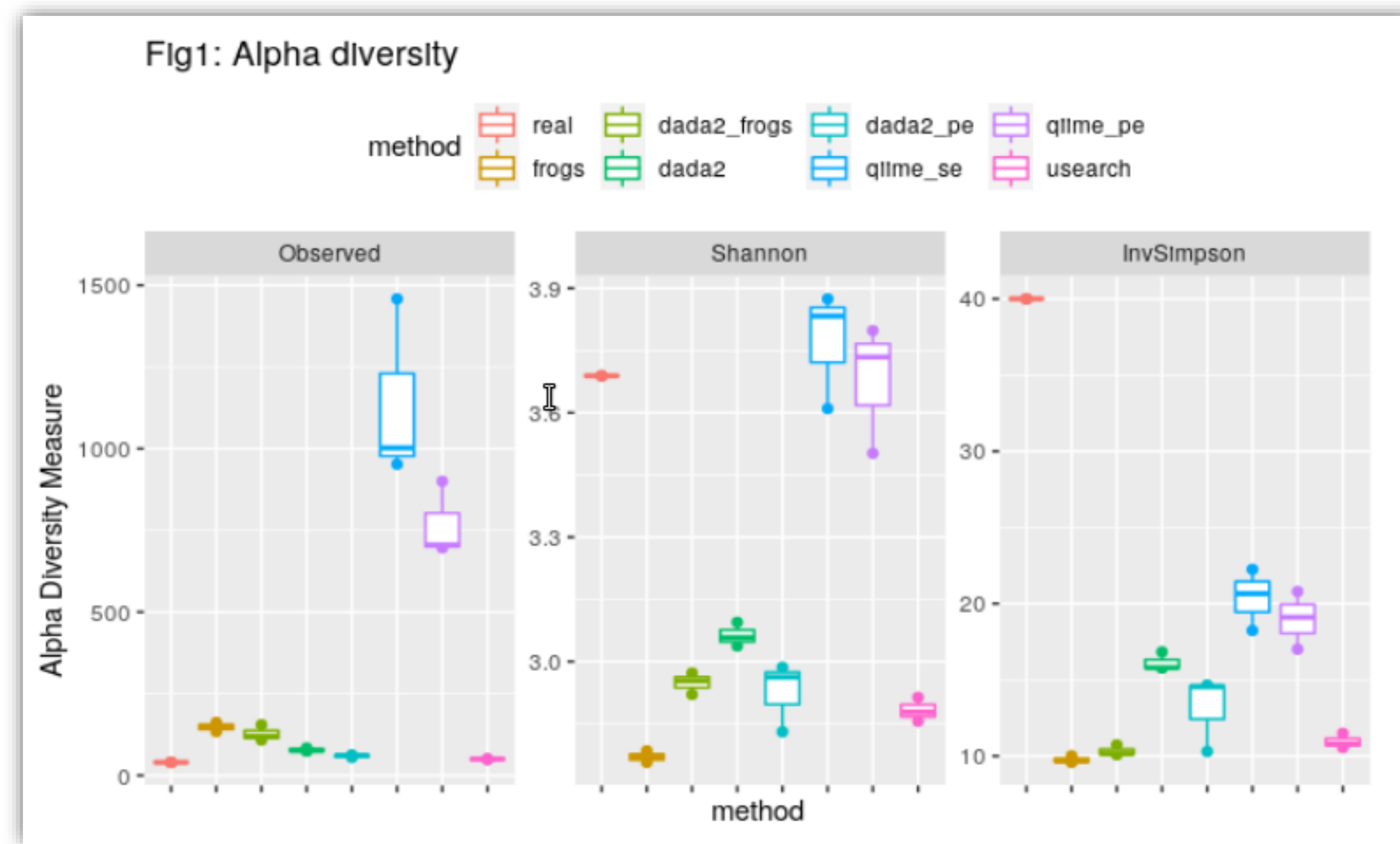


Challenge: Must deal with unmerged reads!

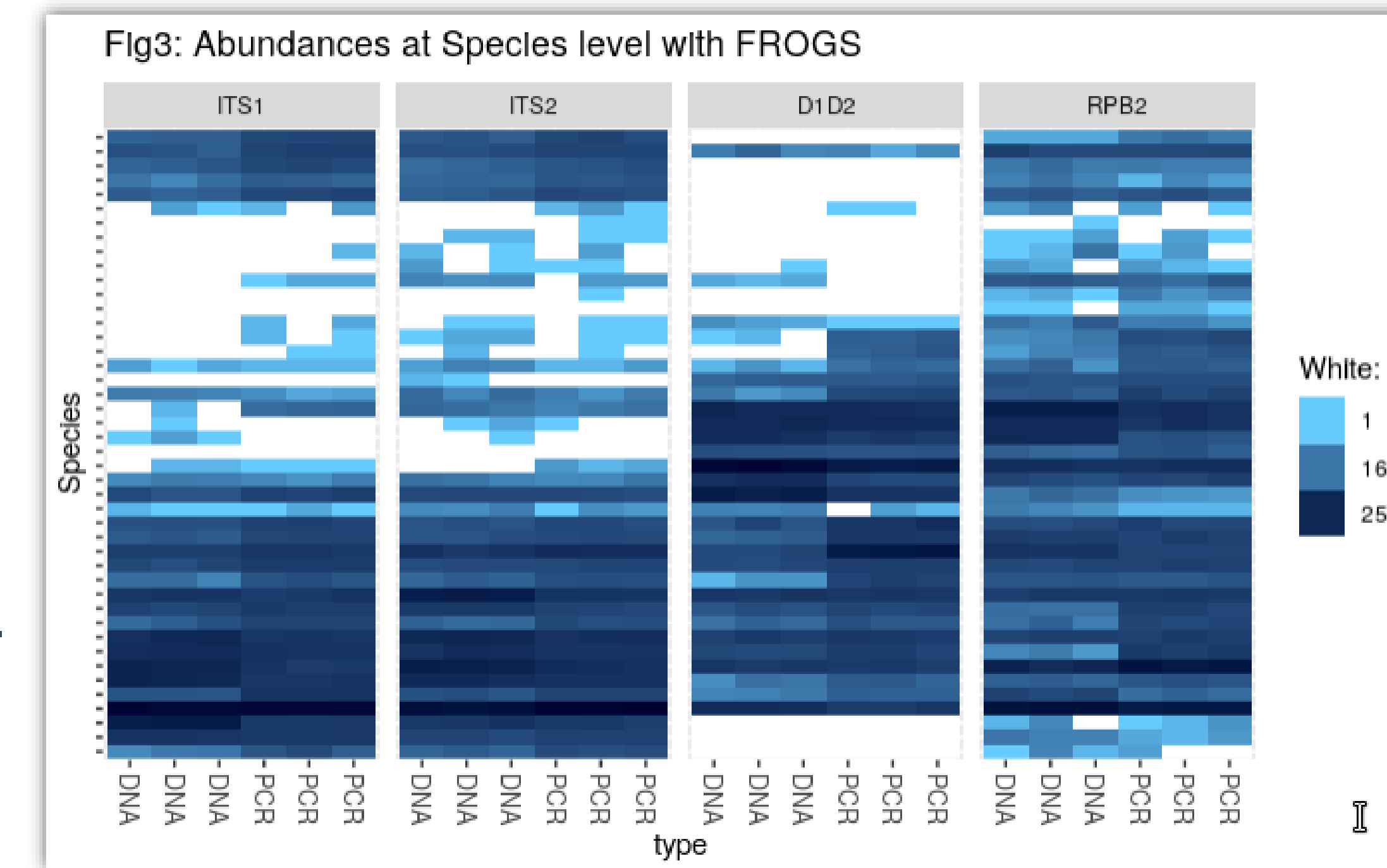
Reference expertise: Each of the 572 sequences present in MOCKs was curated. A complete reference database, without ambiguities was therefore available for taxonomic affiliation.

Results in fermented meats (ITS1 / DNA)

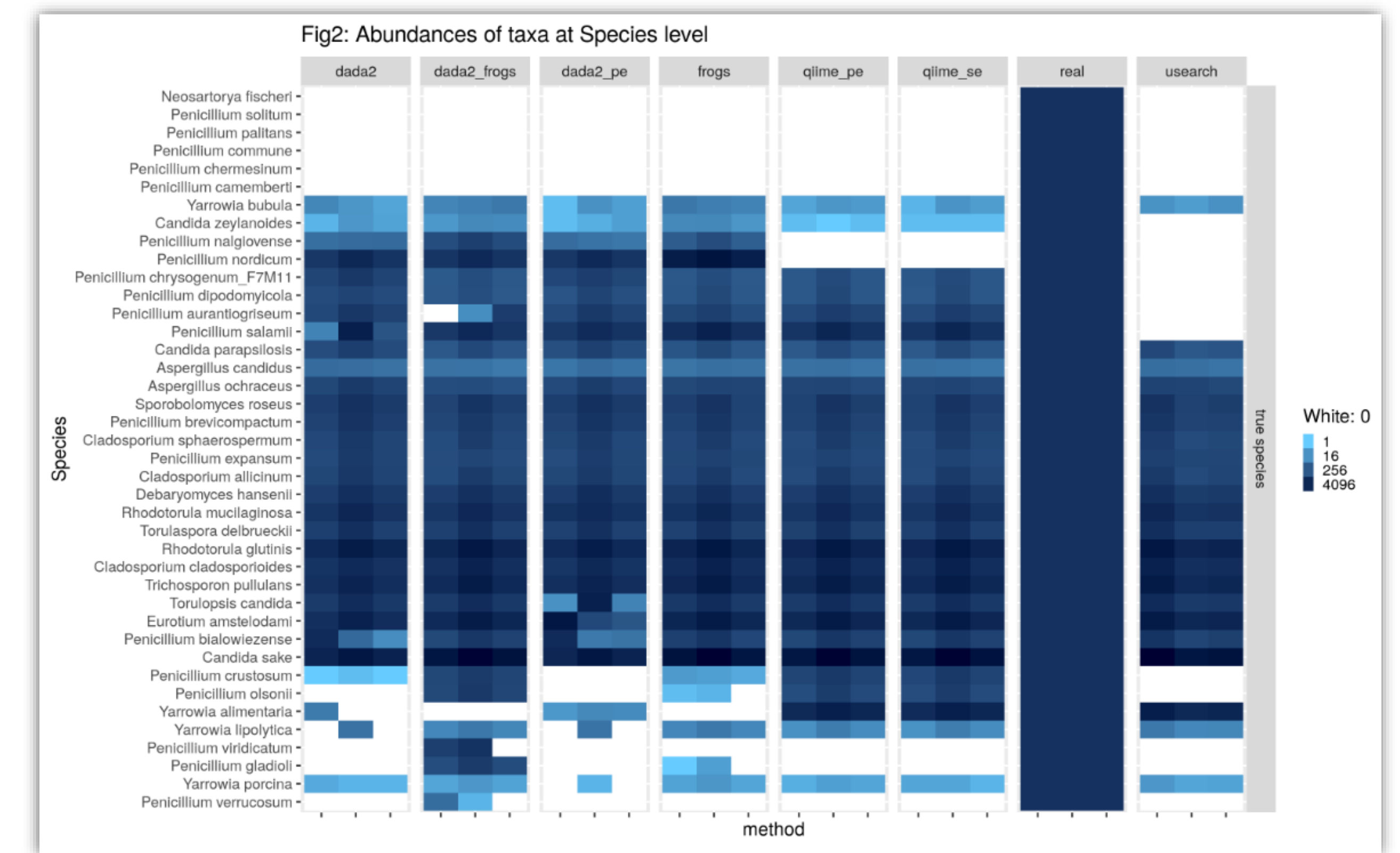
- α-diversity varies considerably according to bioinformatic workflows (Fig1)



- Replicates are essential to avoid missing species (Fig3)
- The choice of marker is crucial to detect particular species (Fig3)



- Qiime2 and Usearch miss more species than others (Fig2)
- Multi-affiliations indicate the impossibility to choose between several references



Conclusions

- ❖ Real MOCKs are important to compare bioinformatic workflows. Results may reflect only a small part of the reality.
- ❖ Some species are not targeted with some markers.
- ❖ This knowledge will help scientists choose the most appropriate marker for their ecosystems.

Perspectives

- ❖ Are the MOCKs species detectable in real samples?
- ❖ What is the effect of a less-curated databank on taxonomic resolution?
- ❖ Fermented meat results identical to other ecosystems?

This work was financially supported by the INRAE-funded MEM metaprogram

References

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., ... & Bai, Y. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature biotechnology*, 37(8), 852-857.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods*, 13(7), 581.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460-2461.

Escudé, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., ... & Pascal, G. (2018). FROGS: find, rapidly, OTUs with galaxy solution. *Bioinformatics*, 34(8), 1287-1294.

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS one*, 8(4).