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, Francoise Irlinger, Jean-Luc Legras, et al.

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

[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, UMRF, Aurillac, [5] INRAE, SayFood, Thiverval-Grignon, [6] INRAE, SPO, Montpellier, [7] INRAE, SECALIM, Nantes

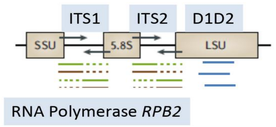
Claire Vincent[1], Serge Casaregola[2], Monika Coton[3], Céline Delbès[4], Hugo Devillers[6], Éric Dugat-Bony[5], Stéphane Guezenec[6], Kate Howell[6], Francoise Irlinger[5], Jean-Luc Legras[6], Valentin Loux[1], Élisa Michel[6], Jérôme Mounier[3], Cécile Neuveglise[6], Thibault Nidelet[6], Audrey Pawtowski[3], Pierre Renault[2], Delphine Sicard[6], Monique Zagorec[7] and Olivier Rué[1]

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

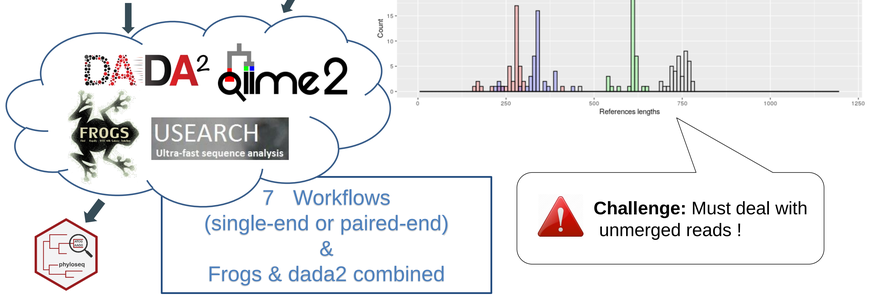
- ✓ 5 MOCKs (fermented meats, cheese, bread, wine, drosophila)
- ✓ Composed of all species reported at least twice in the ecosystem
- ✓ From 25 to 61 species per MOCK
- ✓ One strain / species
- ✓ DNA extraction of each strain by kit
- ✓ Independent PCR1 of four markers (**ITS1**, **ITS2**, **D1D2**,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



Methods

PCR samples
DNA samples
Real samples

Database with MOCK references only



7 Workflows
(single-end or paired-end)
&
Frogs & dada2 combined

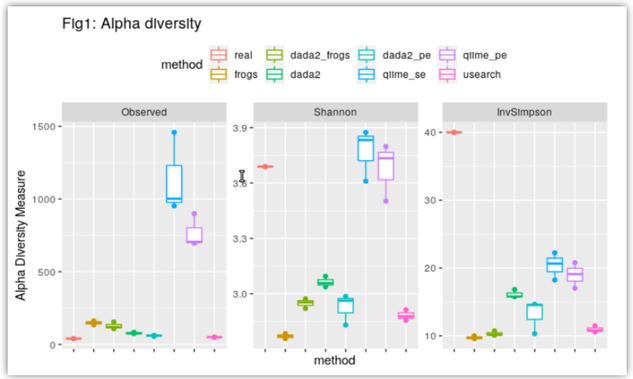
Markers	Expected	InDatabase
ITS1	40	40
ITS2	40	40
D1D2	40	40
RPB2	40	40

Challenge: Must deal with unmerged reads !

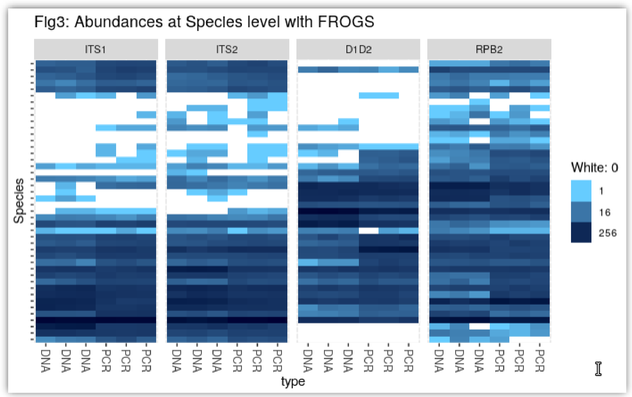
Reference expertise: Each of the 572 sequences present in MOCKs was curated. A complete reference database, without ambiguities was therefore available for taxonomic affiliation.

Results in fermented meats (ITS1 / DNA)

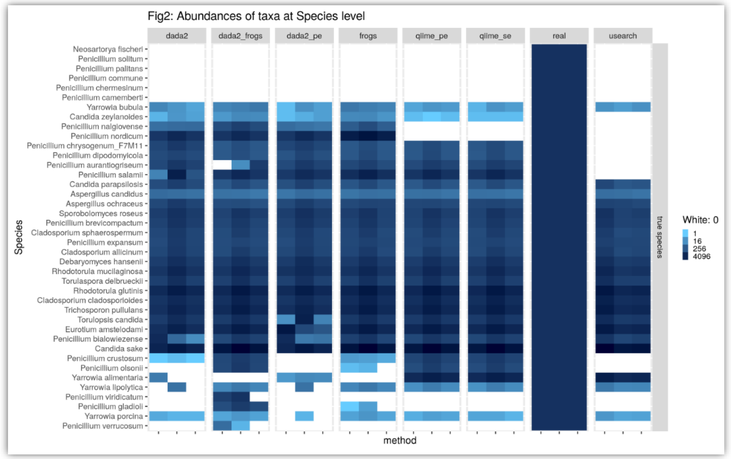
- ❑ α-diversity varies considerably according to bioinformatic workflows (Fig1)



- ❑ 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 worflows. Results may reflect only a small part of the reality.
- ❖ Some species are not targeted with some markers.
- ❖ This knowledge will help scientists choice 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.

Escudié, 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).