



HAL
open science

Its all fun guys: a comparison of bioinformatic pipelines for metabarcoding plant and soil fungal communities

Charlie Pauvert, Marc Buée, Valerie Laval, Véronique Edel-Hermann, Laure Fauchery, Angelique Gautier, Isabelle Lesur, Jessica Vallance, Corinne Vacher

► **To cite this version:**

Charlie Pauvert, Marc Buée, Valerie Laval, Véronique Edel-Hermann, Laure Fauchery, et al.. Its all fun guys: a comparison of bioinformatic pipelines for metabarcoding plant and soil fungal communities. International Phytobiomes Conference 2018, Dec 2018, Montpellier, France. 2018. hal-01938707

HAL Id: hal-01938707

<https://hal.science/hal-01938707>

Submitted on 5 Jun 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.

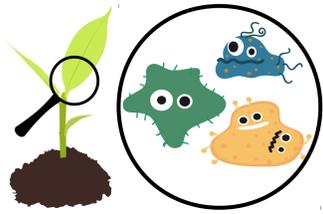


Distributed under a Creative Commons Attribution 4.0 International License

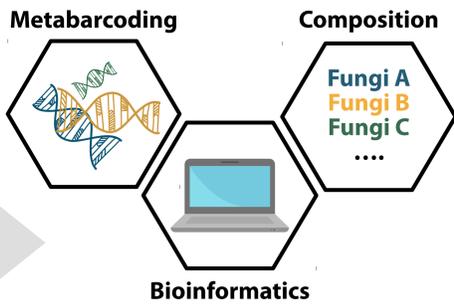
CHARLIE PAUVERT¹, MARC BUÉE², VALÉRIE LAVAL³, VÉRONIQUE EDEL-HERMANN⁴, LAURE FAUCHERY², ANGÉLIQUE GAUTIER³, ISABELLE LESUR^{1,5}, JESSICA VALLANCE⁶, CORINNE VACHER¹

Introduction

Plant and soil fungal communities influence plant fitness and ecosystem functioning.



These fungi are classically studied by **metabarcoding** approaches targeting the ribosomal internal transcribed spacer (ITS)



But there is **no clear consensus** concerning the most appropriate bioinformatic pipeline.

Objectives

1. Make and sequence a **mock community** of 189 Dikarya strains commonly found in agricultural and forest soils and in plant tissues.
2. Compare the ability of 288 **combinations of bioinformatic softwares and parameters** to recover the fungal strains present in this mock community, in the expected proportions.

Methods

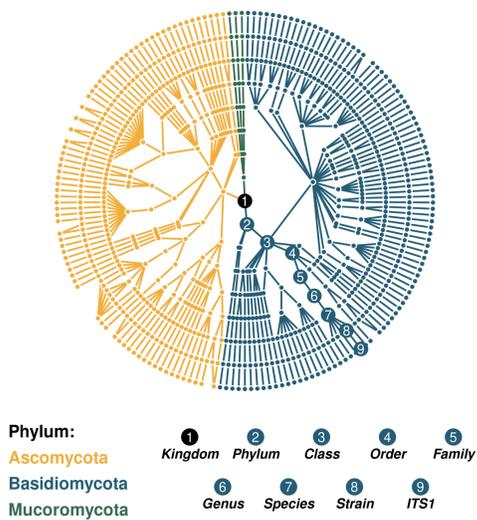


Fig 1 Taxonomic composition of the artificial fungal community. The nodes of the tree are the ITS1 sequences ($n = 175$) of the fungal strains ($n = 189$) that constitute the artificial or « mock » community.

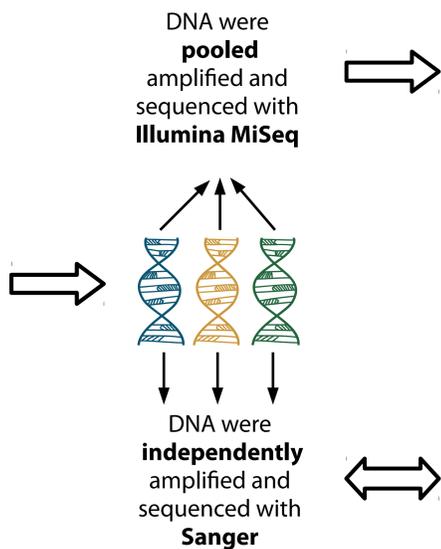
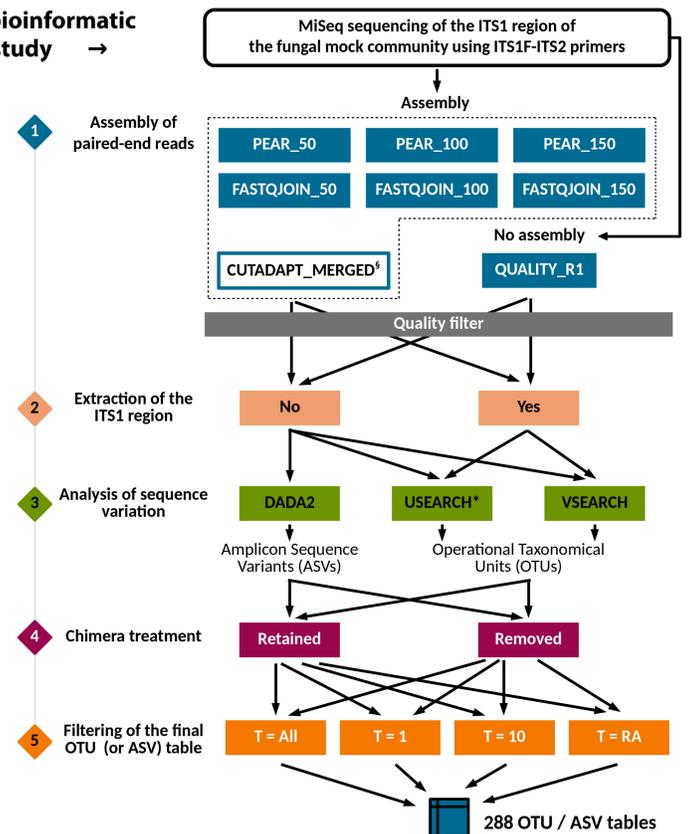
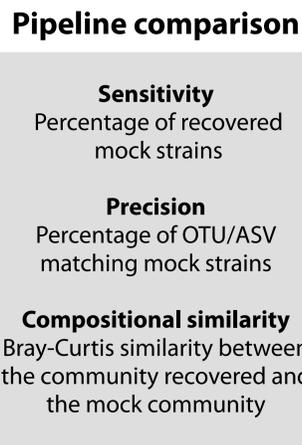


Fig 2 Overview of the bioinformatic pipelines compared in this study

The MiSeq sequences were analyzed with 288 unique combinations of bioinformatic softwares and parameters (referred to hereafter as pipelines)



Results

	Highest Sensitivity	Highest Precision	Highest Similarity	Best Trade-off
1. Assembly	QUALITY_R1	FASTQJOIN_150	QUALITY_R1	QUALITY_R1
2. Extraction	No	No	No	No
3. Variation	VSEARCH	DADA2	DADA2	DADA2
4. Chimeras	Retained	Removed	Retained	Removed
5. Filtering	All	10	All	All

Table 1 Pipelines with the best performances for each three criteria and the pipeline with the best trade-off. Pipelines are labeled below the table and the steps are numbered as in Fig 2.

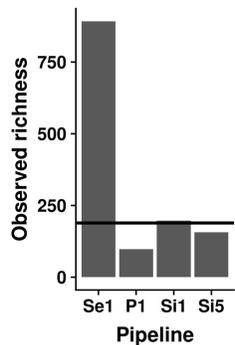


Fig 3 Total number of OTUs (or ASVs) retrieved by the top four pipelines (Table 1). The black horizontal line indicates the expected richness.

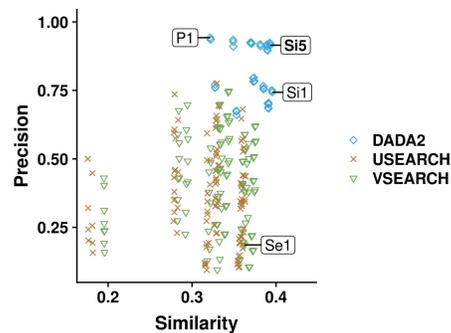


Fig 4 Values of precision and compositional similarity to the mock fungal community, for all 288 bioinformatic pipelines. Only the top four pipelines are labeled

None of the pipelines scored highest in all three criteria (Table 1).

Observed fungal richness depended strongly on the bioinformatic pipeline (Fig 3).

Community richness and composition were well recovered by DADA2 (Fig 3 & 4).

Conclusion

The **Si5 pipeline** used single forward (R1) sequences with DADA2 and no filter other than the removal of low-quality and chimeric sequences.



No overestimation of the number of fungal strains. One of the best to recover the composition of the mock community. Good performance on precision and sensitivity as well.

Acknowledgments Attendance at this congress was possible thanks to the award for the best presentation of the Pathobiome 2018 congress. We thank Matthieu Barret, Martial Briand, Lucas Auer, Gregory Gambetta, Guilherme Martins, Frédéric Barraquand and Tania Fort for useful comments and discussions. We also thank the INRA MEM metaprogramme (Meta-Omics of Microbial Ecosystems) for financial and scientific support. The mock community sequencing was funded by the INRA MEM MetaBAR project (PI: MB) and the bioinformatic analyses were performed as part of the INRA MEM Learn-biocontrol project (PI: CV). Additional funding was received from the LABEX COTE (ANR-10-LABX-45) and the LABEX CEBA (ANR-10-LABX-25-01). CP's PhD grant was funded by the INRA and Bordeaux Sciences Agro (BSA). We thank the Genotoul sequencing facility (Get-PlaGe) for sequencing the mock community and the Genotoul bioinformatics facility (Bioinfo Genotoul) for providing computing and storage resources.

