



HAL
open science

Integrating metagenetic datasets through microbial association networks to compare microbial communities from lacto-fermented vegetables

Romane Junker, Florence Valence, Michel-Yves Mistou, Stéphane Chaillou,
Hélène Chiapello

► To cite this version:

Romane Junker, Florence Valence, Michel-Yves Mistou, Stéphane Chaillou, Hélène Chiapello. Integrating metagenetic datasets through microbial association networks to compare microbial communities from lacto-fermented vegetables. ISMB ECCB, Jul 2023, Lyon, France. hal-04285493

HAL Id: hal-04285493

<https://hal.inrae.fr/hal-04285493v1>

Submitted on 14 Nov 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.

Romane Junker¹, Florence Valence², Michel-Yves Mistou¹, Stéphane Chaillou³ and Hélène Chiapello¹

romane.junker@inrae.fr

1 Université Paris-Saclay, INRAE, MaIAGE

2 INRAE, Agrocampus Ouest, STLO

3 Université Paris-Saclay, INRAE, MICALIS

Integrating metagenetic datasets through microbial association networks to compare microbial communities from lacto-fermented vegetables

Introduction

The development of low-cost sequencing technologies has generated a massive amount of microbiome datasets during the last 20 years [5]. They represent an unprecedented opportunity to address new questions and better understand microbial ecosystems. Nevertheless, it also raises many challenges related to data reusability, comparison and integration. This work explores and compares public metabarcoding datasets related to lacto-fermented vegetables using an original meta-analysis approach.

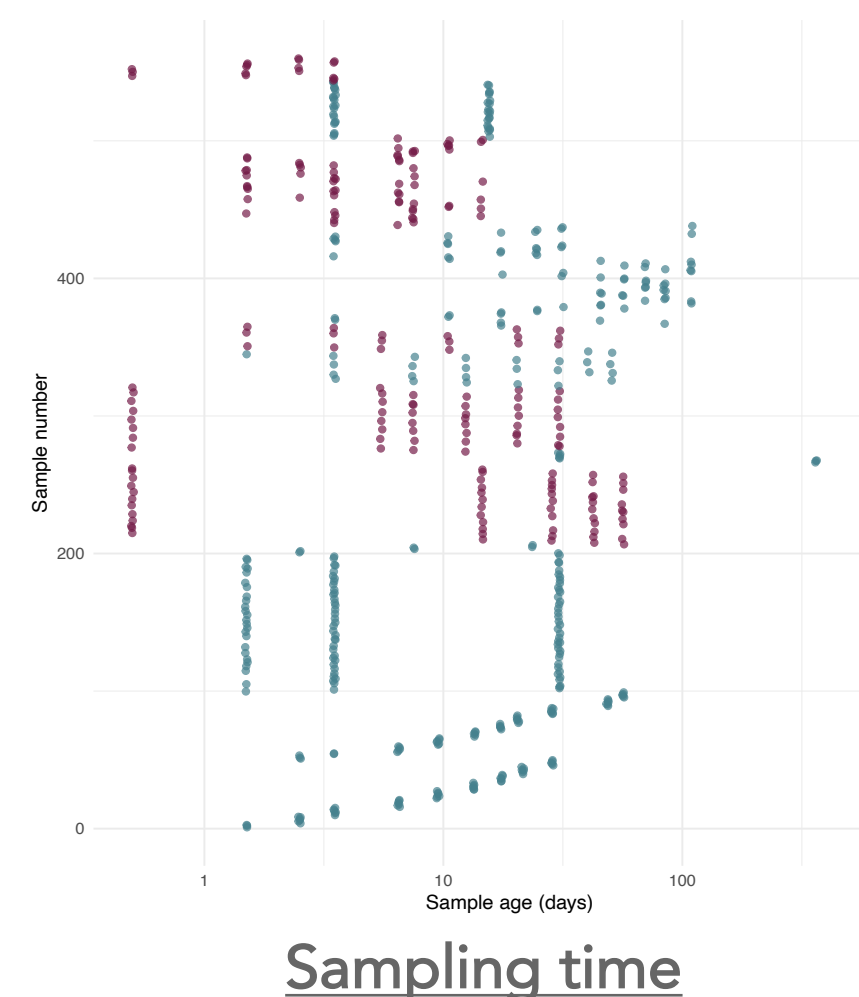
There is currently renewed interest in the bacterial communities of fermented vegetables [4], as plant-based fermented foods are at the heart of sustainability and health issues. Lacto-fermented vegetables are the product of a relatively well-described spontaneous lactic acid fermentation by lactic acid bacteria (LAB). However, community succession has received little attention in lacto-fermented vegetables.

This study presents an integrative bioinformatics approach to explore and compare public amplicon datasets through a meta-analysis. The workflow includes searching and selecting public time-series datasets and constructing Amplicon Sequence Variants (ASV) association networks based on co-abundance metrics. Microbial communities detection is achieved by comparison and clustering of ASV networks. We applied the workflow to ten public datasets and demonstrated its value in monitoring precisely the fermentation with the identification of the bacterial communities succession and of putative core-consortia shared by different plant fermentation types.

Results

Study	Main vegetables	Samples	Continent
1	Carrot	206	Europe
2	Cucumber	56	Europe
3	Red pepper	11	Asia
4	Kimchi cabbage, red pepper	48	Asia
5	Kimchi cabbage, red pepper	26	Asia
6	Cabbage	18	Asia
7	Red pepper	73	Asia
8	Cucumber	64	North America
9	Cabbage	41	Asia
10	Red pepper, radish, cabbage	17	North America

Study description



The ten selected datasets come from different vegetables and include between 11 and 206 samples and 32 unsynchronized timepoints.

Method

Datasets acquisition

- NCBI SRA: 1443 studies with taxonomy IDs "Food metagenome" (870726), "Fermentation metagenome" (1326787), or "Food fermentation metagenome" (1154581)
- MGNify [6]: 10 studies from the biome "Fermented vegetables"
- FoodMicrobioNet [7]: 3 studies with "Fermented" spoilage from "Vegetables and vegetable products"

Datasets selection criteria

- Several time points (to carry out a temporal monitoring)
 - More than 10 samples in the study (to compute reliable association metrics)
 - An associated publication (to ensure access to extensive metadata)
 - Same 16S region sequenced (to compare ASVs)
- 10 selected studies

Sequencing data processing

- Pipeline DADA2 [2,5]:
 - Reads quality control
 - Filtering and trimming of reads
 - Error rate learning and ASV inference
 - Paired read assembly
 - Chimera removal
 - Taxonomic assignment with Silva
 - Filter ASVs on abundance and prevalence
- 11 count tables

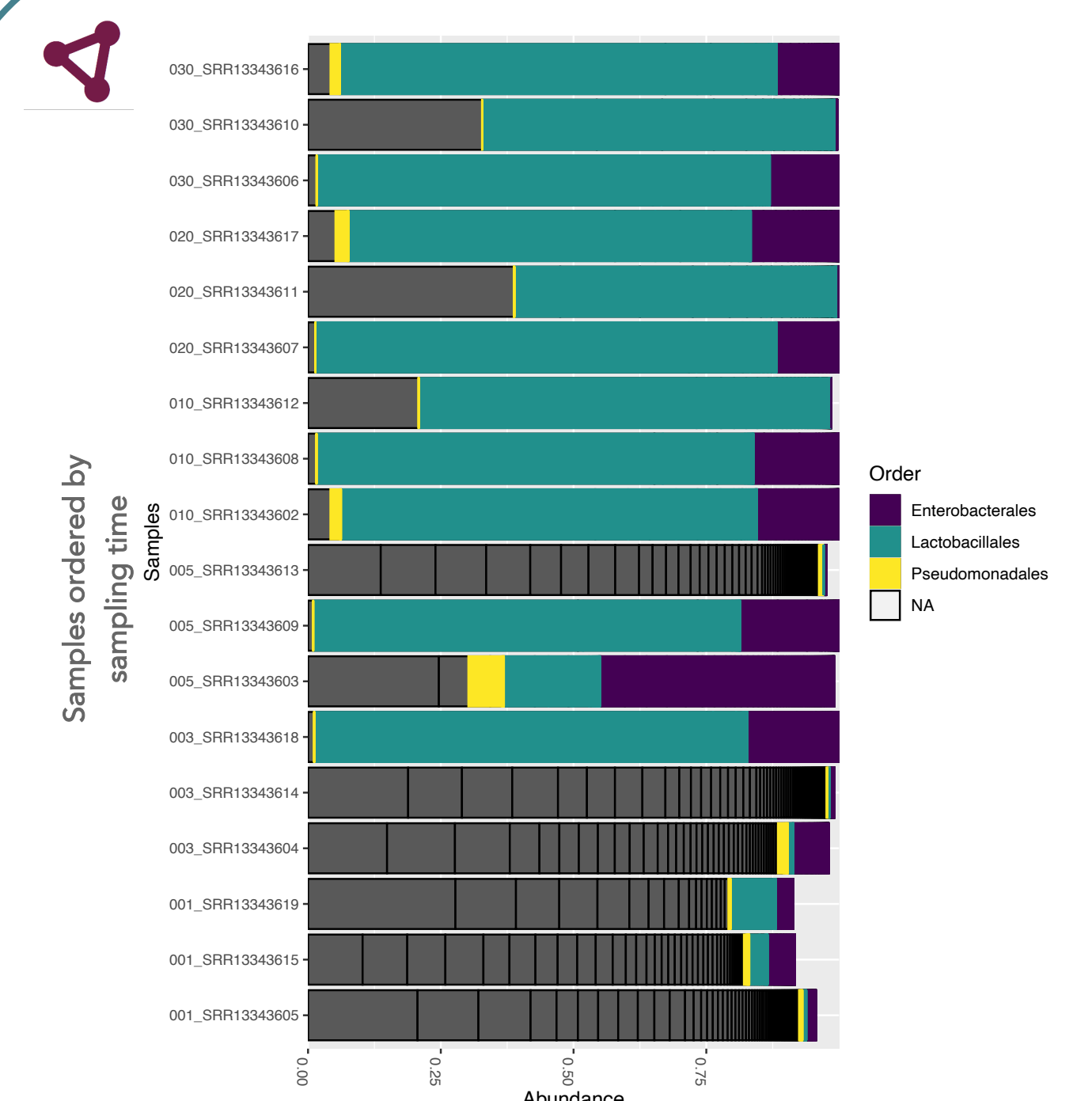
Association networks creation

- Co-presence: Jaccard distance
 - Co-abundance: Pearson and Spearman correlations, proportionality measure (on clr-transformed abundances [3])
- 11 association networks

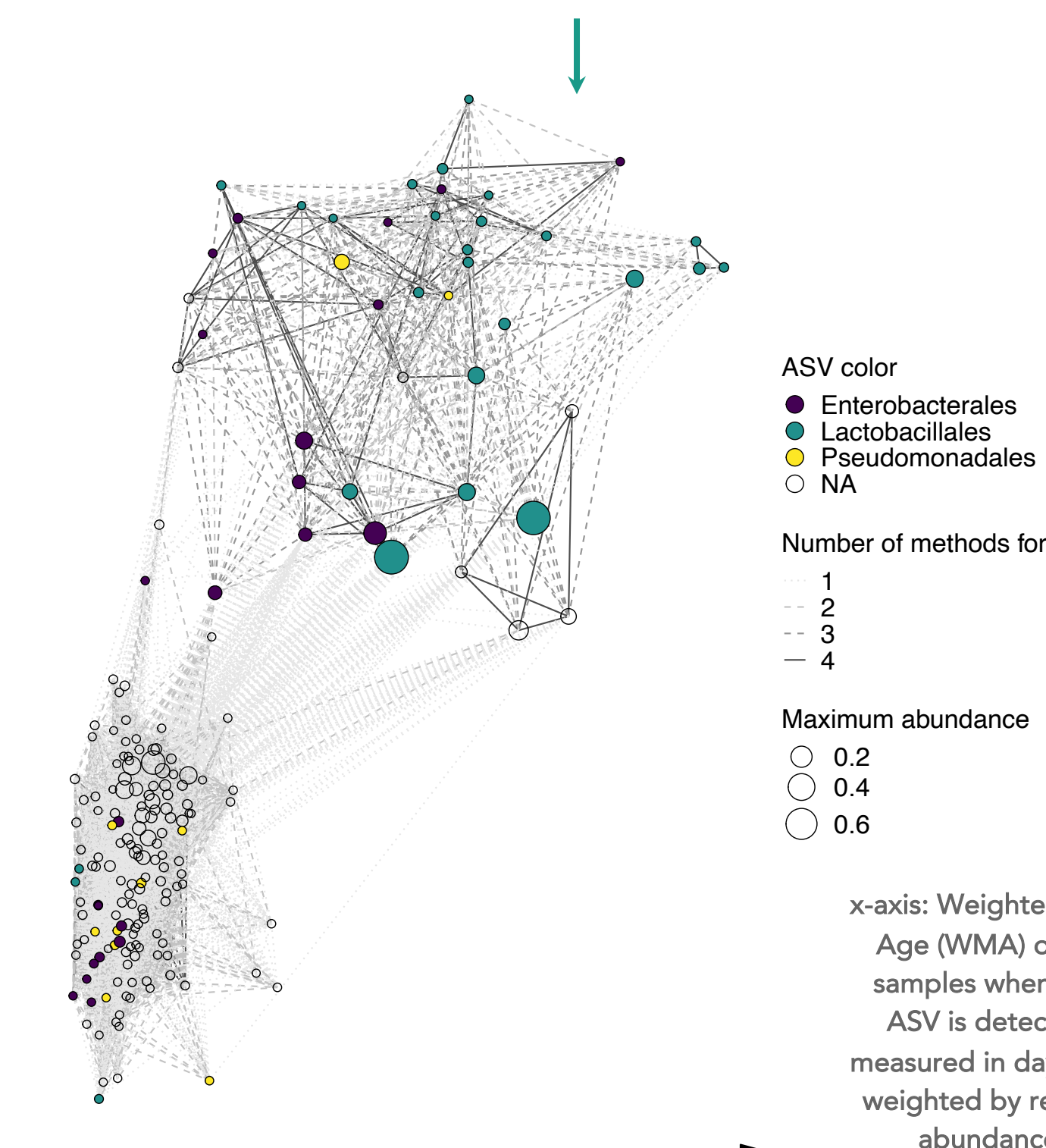
Networks comparison

- Core network construction (with edges found in at least 3 networks)
 - Clustering multiplex "stochastic block model" [1]
- 10 ASVs clusters

Results



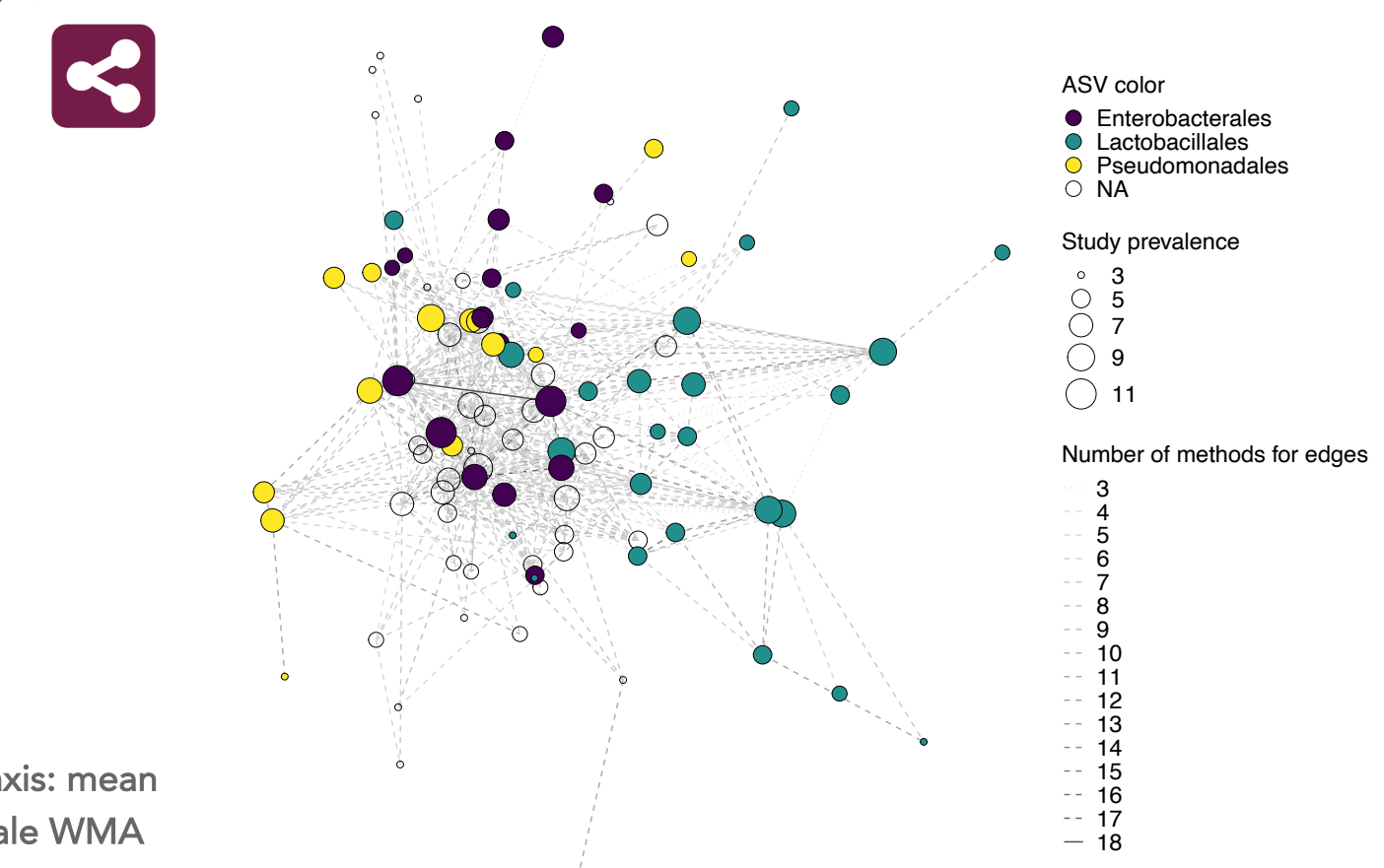
Barplot featuring relative abundance for each sample of PRJNA689239 study on pao cai (18 samples during 6 timepoints)



ASV association network built from samples of PRJNA689239 study

The association network highlights the dynamic evolution of microbial communities during pao cai fermentation. The network is composed of two distinct patterns: a high diversity of ASV (belonging to the initial vegetable microbiota) and, on the other hand, Enterobacteriales and a majority of Lactobacillales.

Results



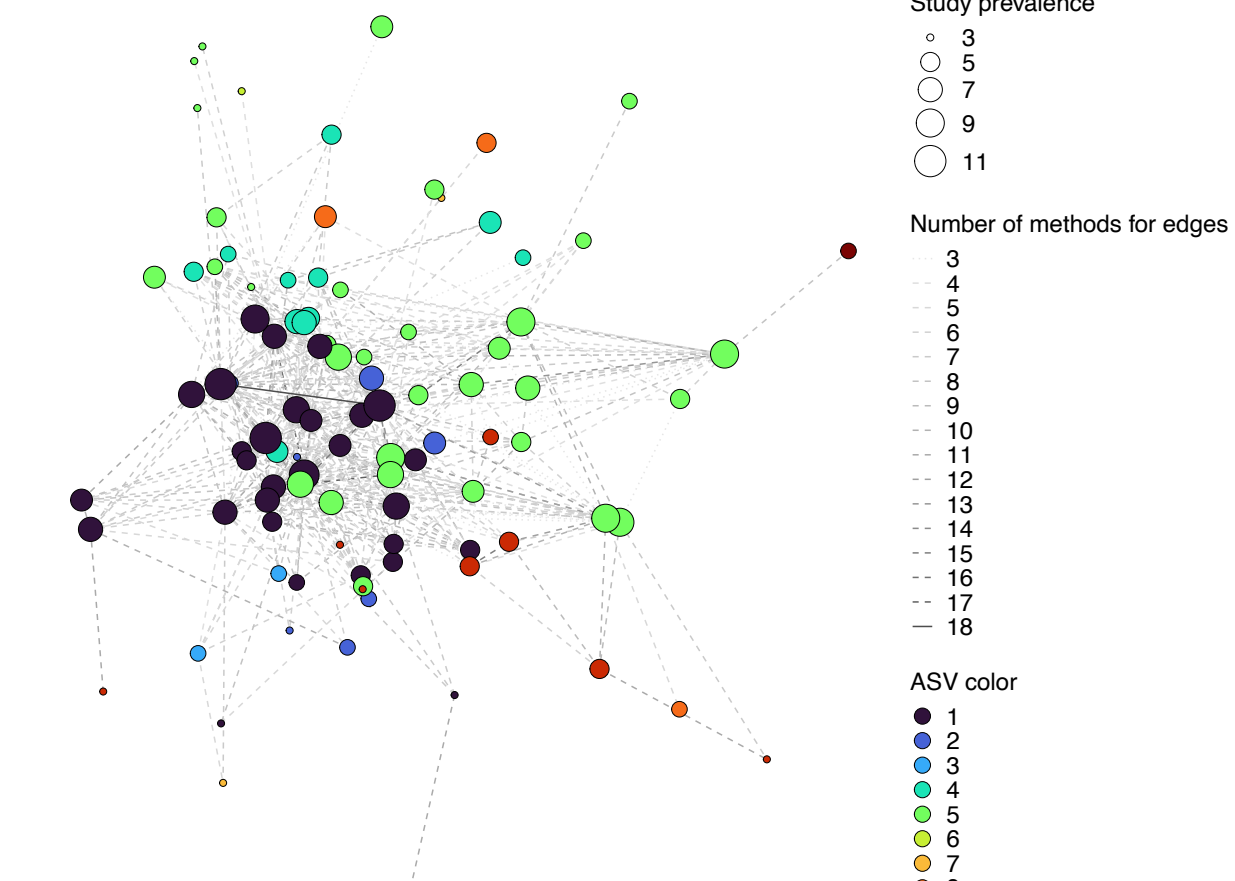
Core network built from 560 samples of 11 studies

A temporal structure is shared, with a shift from the initial vegetable microbiota to Lactobacillales via Enterobacteriales



ASV prevalence computed from 560 samples of 11 studies

The initial microbiota, and the LAB present mainly at the end of fermentation, are more specific to each experiment



Core network with clustering results

Cluster 5 (in bright green) highlights an association between Enterobacteriales and a number of Lactobacillales: this community may represent an early fermentation core-consortium.

Conclusion

- The comparison of microbial ASV association networks is valuable for exploring bacterial community successions and guiding the exploration of potential microbial consortia responsible for different fermentation phases in lacto-fermented vegetables.
- This work validates the recurring pattern of the Enterobacteriales presence at the beginning of fermentation, thus raising questions about their ecological role.
- However, the biological interpretation of this work is limited by its scale and the metabarcoding technology.
- Ultimately, this study highlights the potential of re-using data and the need for complementary functional studies (shotgun metagenomics, metatranscriptomics) with sampling over time to improve our understanding of vegetable fermentation.

References

- Aubert, J., Barbillon, P., Donnet, S. and Miele, V. (2022). Using Latent Block Models to Detect Structure in Ecological Networks. In Statistical Approaches for Hidden Variables in Ecology (eds N. Peyrard and O. Gimenez).
- Callahan, Benjamin J., et al. "DADA2: High-resolution sample inference from Illumina amplicon data." Nature methods 13.7 (2016): 581-583.
- Gloor, Greg. "ALDEX2: ANOVA-Like Differential Expression tool for compositional data." ALDEX manual modular 20 (2015): 1-11.
- Leech, et al. (2020). Fermented-Food Metagenomics Reveals Substrate-Associated Differences in Taxonomy and Health-Associated and Antibiotic Resistance Determinants. MSystems, 5(6), e00522-20.
- McMurdie, Paul J., and Susan Holmes. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." PLoS one 8.4 (2013): e61217.
- Mitchell, Alex L., et al. "MGNify: the microbiome analysis resource in 2020." Nucleic acids research 48.D1 (2020): D570-D578.
- Parente, et al. "FoodMicrobioNet v4: A large, integrated, open and transparent database for food bacterial communities." International Journal of Food Microbiology 372 (2022): 109696.

Acknowledgements

We are grateful to the SFBI for the funding of our participation in this conference, to MICA Division and the ENS Paris-Saclay for the funding of this PhD and to INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12) for providing computing and storage resources.