Genome-scaled phylogeny of *Saccharomyces cerevisiae* from spontaneous must fermentations

Alessia Tatti[1,2,4], Niccolò Carlino[3], Andrea Silverj[2,3], Claudia Maria Longa[4], Federico Pedrazzoli[4], Raffaele Guzzon[4], Edoardo Pasolli[5], Nicola Segata[3], Omar Rota-Stabelli[2,4]

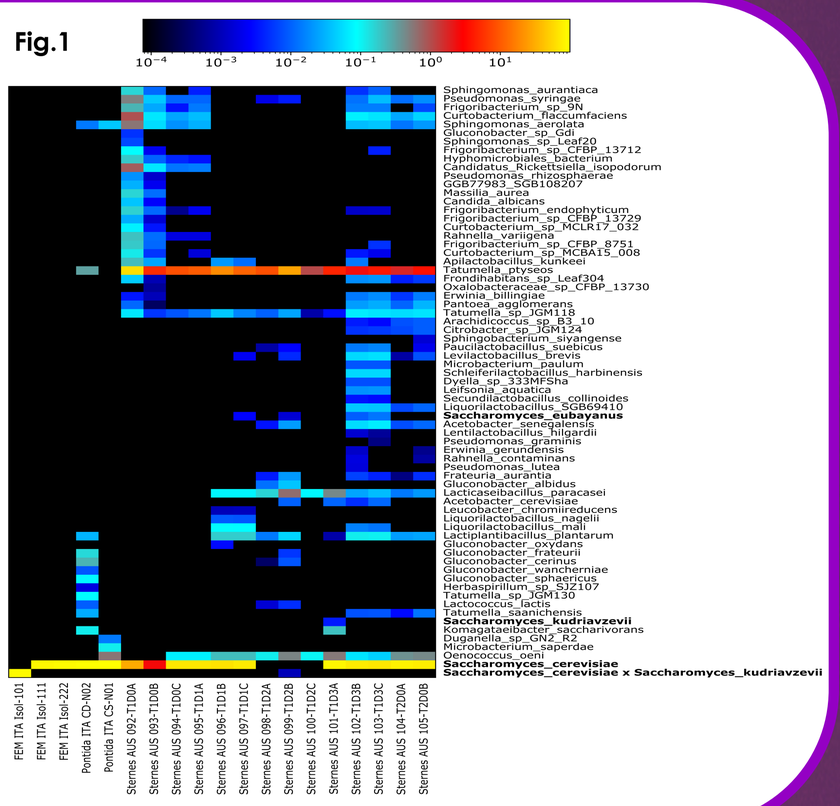
[1] Scuola Universitaria Superiore IUSS Pavia, Palazzo del Broletto, Piazza Vittoria 15, 27100 Pavia (PV);
[2] University of Trento, Centre for Agriculture Food Environment, Via. E. Mach 1, 38098, San Michele All'Adige (TN);
[3] University of Trento, CIBIO - Department of Computational, Cellular and Integrative Biology, Via Sommarinve 9, 38123, Povo (TN)
[4] Fondazione Edmund Mach , Via. E. Mach 1, 38098, San Michele All'Adige (TN);
[5] Department of Agricultural Sciences and Task Force on Microbiome Studies, University of Naples Federico II, Via Università 100, 80055, Portici (NA)

Why spontaneous yeasts? Modern winemakers commonly inoculate selected *S. cerevisiae* strains in must to obtain controlled fermentations and reproducible products. However, wine has been produced for thousands of years using spontaneous fermentations from wild strains, a practice that is experiencing a revival among small wine producers. Despite the widespread usage of such strains in the past, there is much to know about their ecology, evolution and functional potential. For example, the reciprocal affinities of these strains within the *S. cerevisiae* phylogeny have yet to be discovered, as well as the degree of their biodiversity and their impact on wine terroir. Here, we have analysed 2 must samples from North of Italy (Pontida, BG) plus 14 previously published natural must samples from Australia.

A metagenomic pipeline. Must samples were collected between the 5th and 11th days of fermentation. To test our sequencing procedure, we further processed three selected isolates of *Saccharomyces* from the Edmund Mach Foundation collection. To overcome the deleterious issue of polyphenolic compounds, we used an initial treatment of the must with sodium chloride (NaCl), polyvinylpyrrolidone (PVP), and ethylenediaminetetraacetic acid (EDTA). This enabled the successful extraction of approximately 150-400ng of DNA, which proved sufficient for shotgun sequencing paired-end 150 nucleotides (R1 and R2) with a sequencing depth of circa 5 Gbp per sample. Fourteen raw metagenomes from Australian spontaneous must fermentations (from two samples and three different time points) as described by Sternes et al. 2017 were downloaded and incorporated into the analysis.
To ensure data quality, all raw sequencing reads underwent quality control and preprocessing steps utilizing FastQC v0.12.1 and Trimmomatic v0.39. To eliminate reads originating from *Vitis vinifera* (grapevine), the host genome was filtered out using Bowtie2 v.2.5.1 and Samtools v.1.17. Subsequently, the filtered non-host paired-end and unpaired FASTQ files were employed for metagenomic profiling via MetaPhlAn version 4.0.6 using the CHOCOPhlAn SGB_202212 database. The markers from *Saccharomyces cerevisiae* EUK4932 were then extracted to perform a StrainPhlAn analysis, using as a reference the 157 yeast genomes from Maixner et al. 2021. The resulting alignment of 11480 bases was subsequently used for the computation of a maximum likelihood phylogenetic tree using RAxML v.8.2.12 with 100 bootstrap replicates and the GTR+GAMMA+I substitution model.

Metagenomic profiling of must. The metagenomic profiling (Fig 1, alongside) exhibited pronounced *Saccharomyces cerevisiae* prevalence across most samples, while isolate FEM-101, a vintage Edmund Mach collection strain, revealed a potential *S. cerevisiae* and *S. kudriavzevii* hybrid. Compared to the original analysis of Sternes et al. (2017) we recovered a smaller taxonomic diversity (and more unassigned reads). This is because our profiling was hindered by the CHOCOPhlan database which is currently focused on the human metagenome. We aim to enlarge the CHOCOPhlan database to employ all *S. cerevisiae* reference genomes plus most genomes associated with wine fermentations.

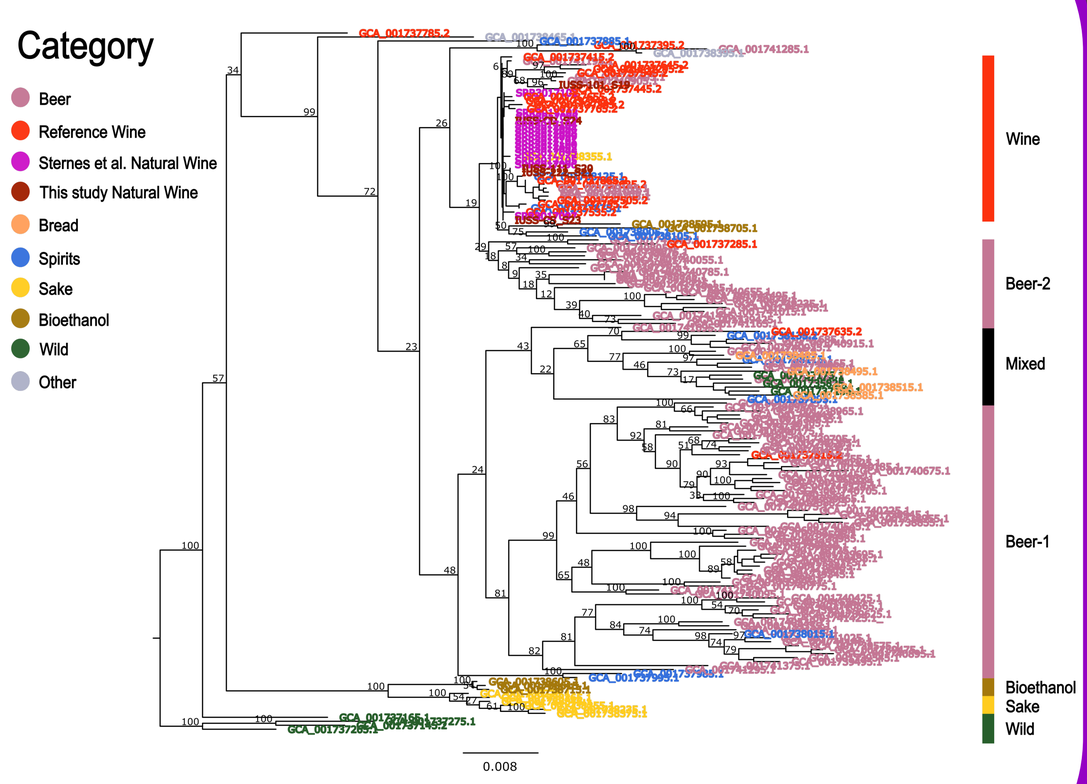
Fig.1



Genome assemblies. We used SPAdes to assemble genomes. Results were excellent for Saccharomyces isolates (e.g., Genome Size 11481095, N50 115761). However, challenges emerged when assembling from must metagenomes (e.g., Genome Size 31273219, N50 38448). This is likely due to the co-presence of different *Saccharomyces cerevisiae* strains in the must and/or natural variability in the population of the same strain. Future analyses aim to address this by employing alternative tools capable of discerning distinct strains within the sample.

Phylogeny of *S. cerevisiae*. The obtained maximum likelihood phylogenetic tree (Fig. 2, alongside) derived from the StrainPhlAn markers alignment was rooted with the wild-African lineage. It exhibited a consistent topology mirroring existing literature, displaying distinctive clustering based on food type, including well-recognized groups like Sake, Mixed, Beer 1, and Beer 2. It permitted the assignment of the extracted *S. cerevisiae* reads to the Wine clade. Yet, the Wine clade's resolution proved to be insufficient to differentiate selected and natural *S. cerevisiae* strains, possibly due to taxa and marker biases.

Fig. 2



What's next? To evaluate the biodiversity of *S. cerevisiae* in Italian wines, must samples from spontaneous fermentations will be gathered across various regions and grape varieties, all obtained at the same stage of fermentation. The phylogenetic tree obtained from the StrainPhlAn marker alignment will be improved by expanding taxon sampling with over 1000 reference genomes. This strategic enhancement aims to achieve a more precise and accurate phylogenetic resolution. We will employ a dedicated assembly pipeline for *S. cerevisiae*, experimenting with various approaches involving diverse settings and software programs. Furthermore, standard ortholog extraction will bolster a >100kb alignment, augmenting future research prospects.
This approach will enable us to uncover the biodiversity of *S. cerevisiae* and its influence on wine production.