The genotypic and pheno-metabolomic landscape of a Saccharomyces cerevisiae *strain collection*

Franco-Duarte R[1], Mendes I[1], Castro CC[2], Silva JS[2], Xavier A[1], Drumonde-Neves J[3], Martins RC[1], Oliveira JM[2], Ferreira AC[4], Schuller D[1]

(1) CBMA - Molecular and Environmental Research Centre, Department of Biology, University of Minho, Braga, Portugal
(2) IBB - Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Braga, Portugal
(3) Research Center for Agricultural Technology - Department of Agricultural Sciences, University of Azores
(4) Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Porto, Portugal

ricardofilipeduarte@bio.uminho.pt

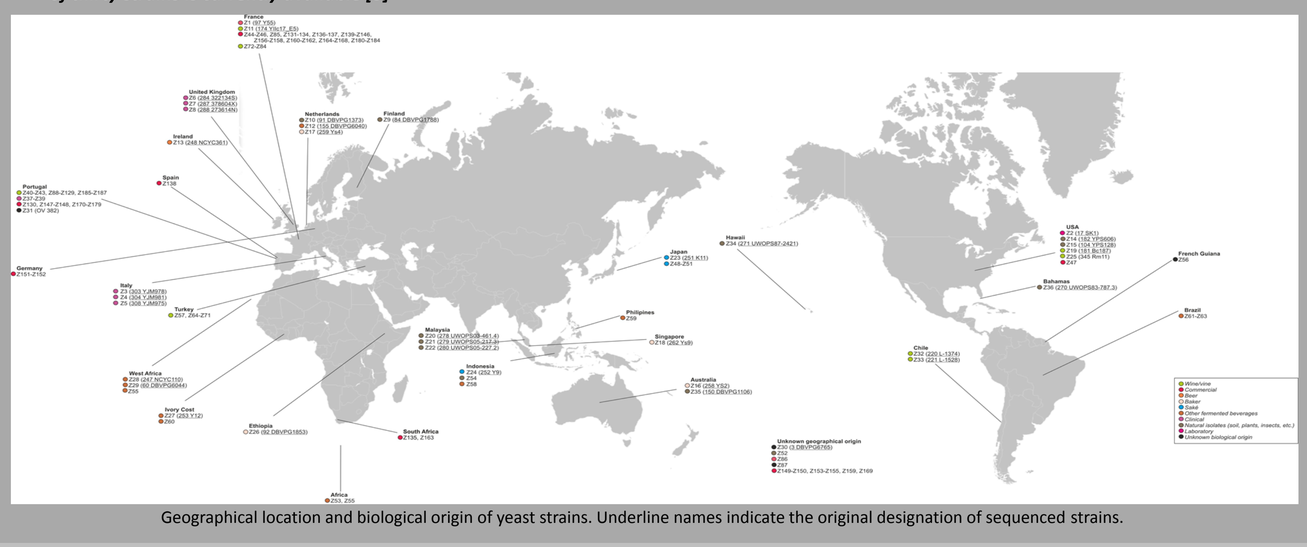
Introduction

Nowadays, most of European wine producers use commercial starter yeasts to guarantee the reproducibility and the predictability of wine quality. The advantages of fermentations containing starter cultures are related to the fact that they are rapid and produce wine with more consistent and predictable quality through successive processes and harvests [1,2]. In these fermentations the winemaker has control over the microbiology of the process, because the inoculated yeast strain predominates and suppresses the indigenous flora. The yeast diversity of many wine regions is still unexplored and it is a common practice among wineries to use commercial starter yeasts that were obtained in other winemaking regions. This arises questions regarding biodiversity preservation, one of the most important concerns of actual times, and ongoing public discussions reflect the benefits of biodiversity preservation.

The objective of the present work is to gain a deeper understanding of the phenotypic diversity of a strain collection comprising 187 strains from different geographical origins and technological applications, using high-throughput quantitative and qualitative methods in combination with bioanalytical data. Only a holistic approach between molecular biology, analytical chemistry, signal processing and bioinformatics provides detailed information on the vast and dynamical relationships between genomics, phenomics and metabolomics.

Saccharomyces cerevisiae *strain collection*

❖ ***A* Saccharomyces cerevisiae *collection was constituted, comprising 187 strains with different geographical origins and technological applications. The complete genome sequence of thirty strains is currently available [3]***



Geographical location and biological origin of yeast strains. Underline names indicate the original designation of sequenced strains.

Methods

Phenotypic characterization

Phenotypic screening was performed considering a wide range of physiological traits that are important from an oenological point of view. These experiments were carried out in white grape must with the addition of the mentioned compounds. After incubation (22 h, 30 ˚C, 200 rpm), optical density (A_{640}) was determined and adjusted to 1.0. Fifteen μL of this suspension were then inoculated into replicate wells of 96-well microplates containing 135 μL of culture media, so that final cellular density was 5 × 10^6 cells mL^{-1}. Quadruplicate experiments were performed for each strain in each test and final A_{640} was determined after 22 h (30 ˚C) in a microwell spectrophotometer.

Must fermentations

All 187 strains were used for fermentations with must of the grape variety Loureiro. When glucose concentration was below 5 g L^{-1}, samples were collected and used for fiber optics spectroscopy and bioanalytical analysis. Transmittance fiber optics UV-VIS-SWNIR spectroscopy of the final fermentation products was performed using a highly sensitive scientific-grade spectrometer for maximum resolution.

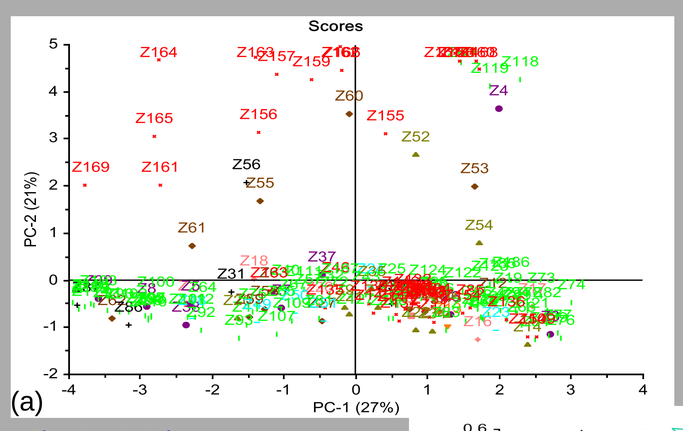
High-performance liquid chromatography with refractive index (HPLC-RI) was used to quantify primary fermentation products (fructose, glucose, ethanol, and glycerol) and organic acids (tartaric, malic, acetic and succinic).

Relevant metabolites that account for inter-strain differences and that are related to aromatic profiles (higher alcohols, esters, fatty acids) were determined by GC-MS after solid phase micro extraction (SPME).

RESULTS

1 Phenotypic characterization

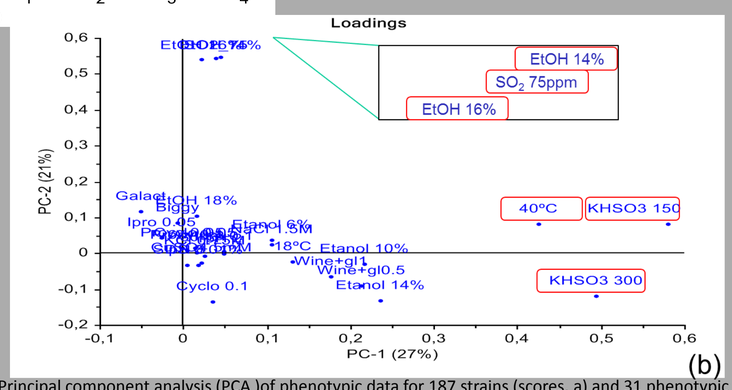
❖ ***The phenotypic diversity of 187 strains was assessed using 31 phenotypic tests with biotechnological relevance***



Laboratory
Clinical
Natural isolate
Commercial wine strains
Wine and vine
Saké
Beer
Bread
Other fermentations
Unknown biological origin

Phenotypes tested:

- Temperatures (18, 30 and 40 ˚C)
- Tolerance to stress
  - pH values (2 and 8)
  - Osmotic/saline stress (KCl and NaCl)
  - Growth in finished wines supplemented with glucose
- Growth in the presence of
  - Potassium bisulfite
  - Copper sulfate
  - Sodium dodecyl sulphate
  - Iprodion
  - Procymidon
  - Cycloheximide
- Resistance to ethanol
- Resistance to SO_2
- Galactose activity
- Biggy



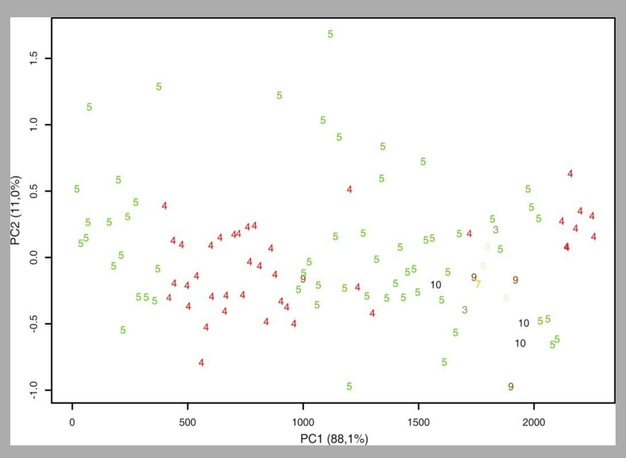
Principal component analysis (PCA)of phenotypic data for 187 strains (scores, a) and 31 phenotypic tests (loadings, b)

❖ ***Commercial strains (red) were segregated in two groups by the second component of the PCA;***

❖ ***The phenotypes responsible for the highest variance (strain variability) were associated with the growth at 40 ˚C, the presence of potassium bisulfite ($KHSO_3$), the presence of sulfur dioxide (SO_2) and with resistance to ethanol.***

2 Fiber optics spectroscopy

❖ ***Wines obtained from ninety one strains that completed fermentation were analyzed by transmittance fiber optics UV-VIS-SWNIR spectroscopy***



3 – Natural isolate (soil woodland, plants, insects)
4 – Commercial wine strains
5 – Wine and vine
7 – Beer
8 – Bread
9 – Other fermentations
10 – Unknown biological origin

PCA of fiber optics UV-VIS-SWNIR spectroscopy data for 91 strains

❖ ***Almost 100 % of strain variance is explained by PC-1 (88.1 %) and PC-2 (11.0 %);***

❖ ***Strains were segregated into groups according to their geographical origin and technological application, with the exception of strains from the wine and vine (indicated in green) that formed two groups;***

❖ ***From the spectroscopy data obtained, and in combination with microsatellite (not shown) and phenotypic data, a more restricted set of 28 most heterogeneous strains was chosen, and further bioanalytical analysis (HPLC and GC-MS) was performed.***

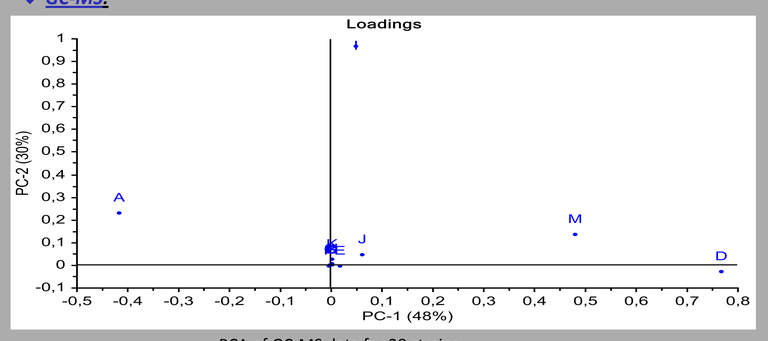
3 Bioanalytical analysis

❖ ***HPLC and GC-MS analysis were performed with samples obtained from the end of fermentation, to evaluate the chemical compounds that might be associated with the differences observed in the previous analysis.***

❖ ***HPLC:***

- ***Strain-dependent differences could be observed concerning organic acids, glycerol, fructose and ethanol;***
- ***Tartaric acid concentrations ranged between 0.5 and 1.4 g L^{-1}, whereas malic, acetic and succinic acids ranged between 4.7 – 8.2 g L^{-1}, and 0.3 – 1.3 g L^{-1}, respectively;***
- ***Final concentrations of ethanol, glycerol and fructose ranged between 80 – 138 g L^{-1}, 5 – 9.75 g L^{-1} and 0-9 g L^{-1}, respectively;***

❖ ***GC-MS:***



PCA of GC-MS data for 28 strains

A - Ethyl acetate
B - Isobutyl acetate
C - Ethyl butanoate
D - 2-methyl-1-propanol
E - Isoamyl acetate
F - 2-methyl-1-butanol
G - Ethyl hexanoate
H - Hexyl acetate
I - Ethyl lactate
J - 1-hexanol
K - cis-hex-3-en-1-ol
L - 2-phenylethyl acetate
M - 2-Phenyl ethanol

- ***PCA revealed ethyl lactate as the one with highest weight in strains variability regarding PC-2, and 2-methyl-1-propanol, 2-Phenyl ethanol, and ethyl acetate concerning PC-1.***

4 Computational analysis

❖ ***Prediction of geographical location and technological application***

- ***Phenotypic data was used to predict strains geographical location and technological application, using knn (k-nearest neighbor) algorithm (the one that performed the best among the ones tested);***

REAL versus PREDICTED DATA		REAL DATA	
		wine and vine strains	commercial wine strains
PREDICTED DATA	wine and vine strains	67	8
	commercial wine strains	8	53

AUC = 0.935

- ***Table shows an example of a prediction of technological group, based on phenotypic data. A correct assignment of a strains to their technological application was obtained for 89.3 % and 86.8 % for wine and commercial strains respectively;***
- ***Are under receiver operating characteristics score (AUC) is very high proving the accuracy of the algorithm used.***

❖ ***Prediction of phenotypic features based on the HPLC data set for the 28 strains***

Phenotypic test	Statistic Coefficient	Standard Error	*p* value
Ethanol 14%	0.137	0.025	0
$KHSO_3$ (300 mg L^{-1})	-0.087	0.022	0
Iprodion (0.1mg mL^{-1})	-0.193	0.060	0.004
SO_2 0ppm	0.148	0.054	0.011
SO_2 75ppm	-0.124	0.050	0.021
Ethanol 6%	-0.093	0.044	0.045
KHSO3 (150 mg L^{-1})	0.050	0.028	0.078
Ethanol 10%	-0.040	0.023	0.098

- ***HPLC data set was used by Orange software [4] to predict each one of the phenotypic features;***
- ***Table shows the phenotypes presenting a good correlation with HPLC data;***
- ***Results are sorted according to the level of significance being the most relevant ones the resistance to ethanol at 14% and the growth in the presence of KHSO3 (300 mg L-1) with a p-value of 0;***

➢ ***Additional data mining is currently underway to explore relations between (i) the genotype and phenotype of the strains, (ii) the genotype of the strains and the chemical composition of final fermentations, (iii) the strains phenotypic features and the chemical composition of final fermentations which may contribute to the establishment of new tools for strains selection and improvement approaches.***

Conclusions

- ***The present work contributes to a better understanding of intra-strain differences regarding the pheno-metabolomic characterization of* S. cerevisiae *isolates***
- ***A set of 31 phenotypic characteristics explained large part of strain diversity***
- ***Fiber-optics spectral analysis of wines obtained with 91 strains separated them according to their geographical origins and technological use***
- ***Aromatic profiles of final fermentations obtained by GC-MS analysis, and quantification of primary fermentation products by HPLC, revealed that organic acids, glycerol, fructose, ethanol, ethyl lactate, 2-methyl-1-propanol, 2-Phenyl ethanol, and ethyl acetate mostly account for inter-strain variability***
- ***Computational approaches showed the potential to predict the strains technological application from phenotypic data, and to predict the phenotypes responsible for the highest variance (ethanol resistance, growth in the presence of $KHSO_3$, SO_2 and iprodion)***

References

[1] Fleet, G. H. (2008). Wine yeasts for the future. FEMS Yeast Res 8, 979-95.

[2] Schuller, D. (2010). Better yeast for better wine - Genetic improvement of *Saccharomyces cerevisiae* wine strains in Rai, M. and Koevics, G. (Eds), Progress in Mycology, Scientific Publishers (India), pp. 1-51.

[3] Liti, G., *et al.* (2009). Population genomics of domestic and wild yeasts. Nature 458, 337-41.

[4] Demsar, J., *et al.* (2004). Orange: from experimental machine learning to interactive data mining. White Paper, Faculty of Computer and Information Science, University of Ljubljana.

Acknowledgements

R. Franco-Duarte, C. Castro, J. Silva and I. Mendes are recipients of a fellowship from the Portuguese Science Foundation (FCT, SFRH/BD/48591/2008, SFRH/BD/46737/2008, SFRH/BD/48616/2008 and SFRH/BD/74798/2010, respectively). This work was financially supported by the Portuguese Science Foundation (Fundação para a Ciência e Tecnologia, FCT), within the project AGR-ALI/103392/2008 and the European Community´s Seventh Framework Program (FP7/2007-2013) under grant agreement nº 232454.