The phenotypic landscape of a Saccharomyces cerevisiae strain collection

Inês Mendes[1], Ricardo Franco-Duarte[1], Lan Umek[2], Elza Fonseca[1], João Drumonde-Neves[1,3], Sylvie Dequin[4], Blaz Zupan[2], Dorit Schuller[1]

[1]CBMA (Centre of Molecular and Environmental Biology) / Department of Biology / University of Minho, Braga, Portugal
[2] Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia
[3]Research Center for Agricultural Technology - Department of Agricultural Sciences, University of Azores
[4]INRA, UMR1083 Sciences pour l'Oenologie, F-34060 Montpellier, France

dschuller@bio.uminho.pt

Introduction

Recent research showed the vast amount of phenotypic variation among strains of *Saccharomyces cerevisiae* that originate from diverse natural habitats and that are used in distinct industrial processes. The evolution of these phenotypes is driven by biotic and abiotic environmental factors whereas a diversifying selection occurs due to unique pressures imposed after expansion into new environments.

The objective of the present work was to gain a deeper understanding of the phenotypic diversity of a strain collection comprising 172 strains from different geographical origins that are partially used for distinct technological applications. We further aimed to develop computational data mining algorithms to predict a strain´s most probable technological applications based on phenotypic results from 30 tests.

METHODS

Phenotypic characterization

Phenotypic screening was performed considering a wide range of physiological traits that are also important from an oenological point of view, using liquid (L) and solid (S) culture media. Twenty-three phenotypic tests were performed in microwell plates (in quadruplicate), using white grape must supplemented with the compounds mentioned in the results section. The final optical density (A_{640}) was determined after 22 h (30 ˚C) and classified from 0 - 3 (0: no growth; classes 1, 2 and 3: A_{640} = 0.2-0.4, A_{640} = 0.5-1, A_{640} > 1, respectively). Seven phenotypic tests were performed by inoculation of 1 μl of a cell suspension (A_{640} = 10) of each strain in Malt Extract Agar supplemented with the compounds indicated in the results section. After incubation (2- 6 days, 26 °C) growth was visually scored and assigned to a class from 0 - 3. H_2S production was evaluated by growth in BiGGY medium, whereas the results were sored (0 – 3) according to the color of the colony.

Data analysis

Principal Component Analysis (PCA), available in the "The Unscrambler X" software (Camo) was used for phenotypic variability analysis. Fisher´s exact test was applied to the phenotypic data set, with Bonferroni correction and control of false discovery rate (FDR). Yule´s coefficient of association was calculated with the objective of find relevant associations between phenotypic data and the strain's origin or technological applications. A set of standard predictive data-mining methods, such as naïve Bayesian classifier and k nearest-neighbours algorithm, as implemented in the software Orange (Curk T, *et al*. 2005), were used for the inference of prediction models. Area under the receiver operating characteristics curve (AUC) was used for prediction scoring, which estimates the probability that the predictive model would correctly differentiate between distinct origins of strains or technological applications.

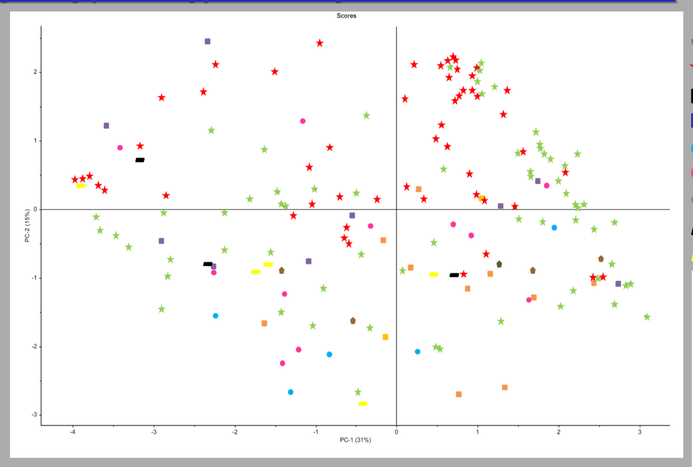
RESULTS

1 Principal Component Analysis of phenotypic data of 172 S. cerevisiae strains

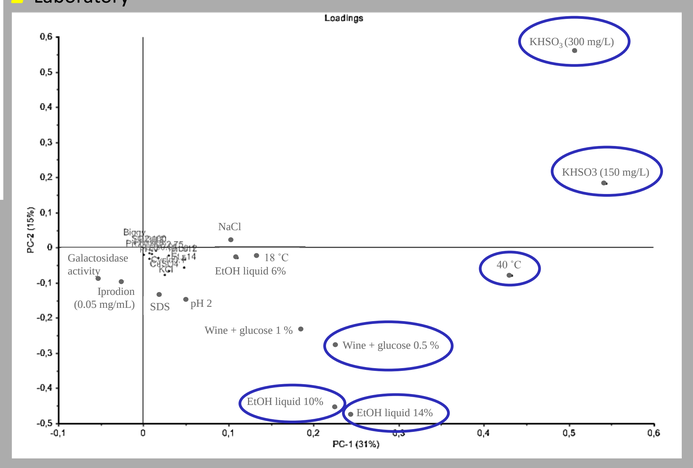
Phenotypes tested:

- Growth in :
  - Temperatures 18, 30 and 40 ˚C (L)
  - pH 2 and 8 (L)
  - KCl 0.75 M (L)
  - NaCl 1.5 M (L)
  - Wine supplemented with glucose 0.5 % (w/v) and 1% (w/v) (L)
  - $CuSO_4$ 5 mM (L)
  - SDS 0.01% (w/v)(L)
  - Iprodion 0.05 mg/mL and 0.1 mg/mL (L)
  - Procymidon 0.05 mg/mL and 0.1 mg/mL (L)
  - Cycloheximide 0.05 μg/mL and 0.1 μg/mL (L)
  - Ethanol 6% (v/v), 10% (v/v), 14% (v/v); (L)
  - Ethanol 12% (v/v), 14% (v/v), 16% (v/v), 18 % (v/v); (S)
  - $KHSO_3$ 150 mg/L and 300 mg/L (L)
  - $Na_2S_2O_5$ 75 ppm and 100 ppm (S)
- Galactosidase activity (L)
- Production of H_2S (S)

(L) - Liquid medium; (S) - Solid medium



- Natural isolate (Wine and vine)
- Natural isolate (soil woodland, plants and insects)
- Commercial wine strain
- Beer
- Baker
- Saké
- Other fermented beverages
- Clinical
- Unknown biological origin
- Laboratory



- ***The phenotypes responsible for the highest variance or strain variability (⬭) were associated with (i) differential growth in the presence of potassium bisulphite ($KHSO_3$) (ii) at 40 °C (iii) in a finished wine supplemented with glucose and (iv) in the presence of ethanol;***
- ***The majority of commercial wine strains (★) were located in the upper part of the PCA, indicative of the high resistance of strains to $KHSO_3$ in the medium and of the rather low resistance to ethanol (10 and 14%, v/v) in the liquid media;***
- ***The group of sake strains (●) were separated by the second component of the PCA, once that they were all located in the lower part of the graph, indicating that their main shared phenotype is a rather high resistance to ethanol and a low resistance of potassium bisulphite at 300 mg/L;***
- ***Natural strains of* S. cerevisiae *that were isolated from wines or from vines (★) showed a heterogeneous phenotypic behaviour since they were dispersed in the PCA plots for both components.***

2 Relationship between phenotypic results and the strain's origin or technological application

Relevant associations (adjusted $p < 0.1$) between phenotypic results and strain´s origin or technological application, using Fisher´s exact test and Yule´s coefficient of association.

Phenotypic test	Class of phenotypic result	Group (origin or technological application)	Adjusted p-value	Yule´s coefficient of association	% of strains sharing associations *
Iprodion (0.05 mg/mL)	2	Commercial wine strain	5.30×10^{-7}	0.892	82.0
Iprodion (0.05 mg/mL)	3	Natural isolate (wine and vine)	0.007	0.790	56.4
Iprodion (0.1 mg/mL)	3	Natural isolate (wine and vine)	0.071	0.848	50.6
Wine + glucose 0.5 %	0	Commercial wine strain	0.024	0.592	57.0
Wine + glucose 0.5 %	1	Sake	0.016	1.000	77.3
Wine + glucose 0.5 %	2	Natural isolate (soil woodland, plants and insects)	0.015	0.843	87.2
Wine + glucose 1 %	2	Natural isolate (soil woodland, plants and insects)	0.064	0.813	89.5
Cycloheximide (0.1 µg/mL)	2	Commercial wine strain	0.026	0.705	75.6
$KHSO_3$ (150 mg/L)	3	Commercial wine strain	0.009	0.664	59.3
Ethanol 14 % (v/v) - liquid medium	0	Commercial wine strain	0.009	0.615	64.5
Ethanol 14 % (v/v) – solid medium	0	Commercial wine strain	0.019	1.000	41.2
40 ˚C	0	Natural isolate (wine and vine)	0.057	0.613	64.0
SDS 0.01% (m/v)	0	Commercial wine strain	0.013	0.881	45.3
Procymidon (0.1 mg/mL)	2	Other fermented beverages	0.055	0.879	92.4
$CuSO_4$ (5 mM)	0	Commercial wine strain	0.024	0.708	50.6
H_2S production	2	Natural isolate (wine and vine)	0.037	0.514	61.0
Galactosidase activity	1	Natural isolate (wine and vine)	0.059	0.674	63.4
pH 2	2	Sake	0.086	0.976	97.0

* Percentage of strains that share the phenotypic result and belong to the described group or that didn´t share the phenotypic result nor belong to that group

- ***Eighteen significant associations between the described phenotypic result and the group (origin or technological application) were obtained from a total of 910 associations, after elimination of the ones with p values above 0.1 and negative Yule´s coefficients. Between 41 and 97% of the strains share the associations described in the table above;***
- ***The most significant results were found for the resistance to Iprodion (0.05 mg/mL; lowest p-values).***
- ***Considering growth in wine supplemented with glucose, almost all values of phenotypic growth were associated with some technological group: (i) absence of growth in wine + glucose was moderately associated to commercial wine strains; (ii) class 1 of growth in wine + glucose were associated with sake strains; (iii) class 2 of growth in wine + glucose were associated to natural isolates;***
- ***The inability of commercial wine strains to growth at high percentages of ethanol was supported by the finding of two associations: class 0 of growth at ethanol 14 % (v/v) in liquid medium, and growth at ethanol 14 % (v/v) in solid medium.***

3 Prediction of technological applications based on phenotypic results

Confusion matrix indicating prediction of the technological application for 172 strains, obtained with naïve Bayesian classifier, in comparison with their real technological applications (AUC = 0,70)

Real technological application	Total number of strains	Predicted: Beer	Baker	Clinical	Commercial wine strain	Laboratory	Natural isolate (soil woodland, plants and insects)	Other fermented beverages	Sake	Unknown biological origin	Natural isolate (wine and vine)
Beer	1	**0** (0%)	0	0	0	0	1	0	0	0	0
Baker	4	0	**0** (0%)	0	0	0	3	0	0	0	1
Clinical	9	0	0	**0** (0%)	2	0	1	0	0	1	5
Commercial wine strain	47	0	0	3	**36** (77%)	0	2	1	0	0	5
Laboratory	3	0	0	1	0	**0** (0%)	0	1	0	1	0
Natural isolate (soil woodland, plants and insects)	12	0	1	2	2	0	**2** (17%)	2	0	0	3
Other fermented beverages	12	0	0	1	1	0	2	**3** (25%)	1	0	4
Sake	6	0	0	0	0	0	1	1	**2** (33%)	0	2
Unknown biological origin	4	0	0	1	0	0	0	1	0	**1** (25%)	1
Natural isolate (wine and vine)	74	0	1	3	8	1	2	3	1	1	**54** (73%)

- ***The majority of strains were correctly assigned to their technological application or origin for the groups of commercial wine strains and natural strains from wine and vine (77% and 73%, respectively);***
- ***Significantly lower associations were obtained for the remaining groups, which can be explained by the small number of strains included in these groups.***

Conclusions

We found a large phenotypic variability between strains, more associated with the technological application of the strains than their geographical origin. The extent of phenotypic variability varied with the different tests.

These results demonstrate the potential of this mathematical model to predict commercial strains based on the results of a phenotypic screen. Our approach can be used to make predictions about a strain's potential as a good commercial wine yeast strain, based on the results of a restricted number of phenotypic tests that can be performed in a high-throughput approach, using microwell plates.

References

Curk T, Demsar J, Xu Q, Leban G, Petrovic U, Bratko I, Shaulsky G, Zupan B (2005) Microarray data mining with visual programming, Bioinformatics, 21(3):396-398.

Acknowledgements

This work was funded by the fellowships SFRH/BD/74798/2010, SFRH/BD/48591/2008 and M3.1.2/F/006/2008 (DRCT). Financial support was also obtained from FEDER funds through the program COMPETE and by national funds through FCT by the projects PDTC/AGR-ALI/103392/2008 and PTDC/AGR-ALI/121062/2010.