



New integrative computational approaches unveil the *Saccharomyces cerevisiae* pheno-metabolomic fermentative profile and allow strain selection for winemaking



Ricardo Franco-Duarte^{a,*}, Lan Umek^b, Inês Mendes^a, Cristiana C. Castro^c, Nuno Fonseca^a, Rosa Martins^d, António C. Silva-Ferreira^d, Paula Sampaio^a, Célia Pais^a, Dorit Schuller^a

^a CBMA – Molecular and Environmental Research Centre, Department of Biology, University of Minho, Braga, Portugal

^b Faculty of Administration, University of Ljubljana, Ljubljana, Slovenia

^c IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Braga, Portugal

^d Escola Superior de Biotecnologia, Universidade Católica Portuguesa, Porto, Portugal

ARTICLE INFO

Article history:

Received 10 December 2015

Received in revised form 10 April 2016

Accepted 12 May 2016

Available online 13 May 2016

Keywords:

Saccharomyces cerevisiae

Data-fusion

Wine yeasts

Metabolomics

Matrix factorization

ABSTRACT

During must fermentation by *Saccharomyces cerevisiae* strains thousands of volatile aroma compounds are formed. The objective of the present work was to adapt computational approaches to analyze pheno-metabolomic diversity of a *S. cerevisiae* strain collection with different origins. Phenotypic and genetic characterization together with individual must fermentations were performed, and metabolites relevant to aromatic profiles were determined. Experimental results were projected onto a common coordinates system, revealing 17 statistical-relevant multi-dimensional modules, combining sets of most-correlated features of noteworthy biological importance. The present method allowed, as a breakthrough, to combine genetic, phenotypic and metabolomic data, which has not been possible so far due to difficulties in comparing different types of data. Therefore, the proposed computational approach revealed as successful to shed light into the holistic characterization of *S. cerevisiae* pheno-metabolome in must fermentative conditions. This will allow the identification of combined relevant features with application in selection of good winemaking strains.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In winemaking, the most relevant families of compounds produced by yeasts cover a large number of metabolites, including primary (e.g., sugars, organic acids, amino acids) and secondary metabolites (e.g., flavonoids and anthocyanins). These compounds play an important role in the flavor and aroma of wine (Regodón Mateos, Pérez-Navado, & Ramírez Fernández, 2006), and commercial strains are selected for their ability to contribute to its sensorial profile (Richter, Dunn, Sherlock, & Pugh, 2013; Rodríguez-Palero, Fierro-Risco, Codón, Benítez, & Valcárcel, 2013; Suárez-Lepe & Morata, 2012). *Saccharomyces cerevisiae* is one of the most versatile microorganisms and therefore the knowledge of its metabolic profiles during fermentation, together with other genetic and phenotypic characteristics may be very important to

select strains for biotechnological applications. The development of data-fusion approaches between genomics and metabolomics (qualitative versus quantitative information) is one of the major hurdles for the development of holistic characterization approaches in biotechnology (Becker & Palsson, 2008). Several methods are available currently for the inference of genomic variation between *S. cerevisiae* strains including microsatellite amplification (Legras, Ruh, Merdinoglu, & Karst, 2005), interdelta sequence typing (Franco-Duarte et al., 2011), comparative genome hybridization on array (aCGH) (Carreto et al., 2008), and single-nucleotide polymorphisms (SNPs) detection after sequencing (Liti et al., 2009; Schacherer, Shapiro, Ruderfer, & Kruglyak, 2009). Recently developed high-throughput genomic technologies, especially with the decreasing costs of sequencing, simplified significantly the characterization of biological systems at multiple levels, including detection of genomic variation (Strope et al., 2015; Via, Gignoux, & Burchard, 2010; Wilkening et al., 2013).

The study of relationships between multi-level data-types has been hampered due to lack of appropriate data resources. Within our previous work (Franco-Duarte, Mendes, Umek,

* Corresponding author at: CBMA (Centre of Molecular and Environmental Biology), Department of Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal.

E-mail address: ricardofilipeduarte@bio.uminho.pt (R. Franco-Duarte).

Drumonde-Neves, & Zupan, 2014; Franco-Duarte, Umek, Zupan, & Schuller, 2009; Mendes et al., 2013) new approaches were developed for the study of pairwise relations. In the mentioned publications, the phenotypic and genetic diversity of groups of *S. cerevisiae* strains from different geographical and technological origins was evaluated, and strain's phenotypic characteristics were estimated based on genotypic data, and using computational statistical modeling. In these works, strains were successfully identified as sharing similar genetic characteristics (microsatellite alleles) that displayed also similar phenotypes, adapting subgroup discovery techniques. Several other tools became available in the last decade relating pairwise genomic variables (Boulesteix & Strimmer, 2007; Devarajan, 2008; Hutchins, Murphy, Singh, & Graber, 2008; Kim & Park, 2007). Partial least squares regression (PLS-R – reviewed on (Boulesteix & Strimmer, 2007)) is particularly used in spectroscopy and chromatography, with successful outcomes in the discrimination of bacterial (Preisner, Lopes, Guiomar, Machado, & Menezes, 2007) and yeast strains (Kuligowski, Quintás, Herwig, & Lendl, 2012), allowing the prediction of dependent variables from a large set of independent variables. Although PLS-R is an informative method for the exploration of common features between two data sets, with this method alone, not much can be inferred about pheno-metabolomic diversity.

Advances of computational and bioinformatic tools contributed to a more powerful data analysis, incorporating integration methods that address multi-dimensional genomic, phenotypic and metabolomic data. A particular challenge was the fact that different types of genomic data (such as SNPs, microsatellite data, etc.) have different scales and units, and cannot simply be aggregated into multiple datasets. A recent breakpoint was achieved by the development of new matrix factorization methods, associated with the projection of multiple types of genomic data into a common coordinates system (Zhang et al., 2012). With this method it is possible to break down massive data sets into smaller modules that exhibit similar patterns, and, after extrapolation and adaptation of the method to other types of data, it has the potential to reveal new insights into metabolite formation pathways, which would be overlooked with only a single type of data.

In this way, our objective was to achieve a holistic characterization of a *S. cerevisiae* strain collection, using phenotypic, genetic and metabolic methods, and adapt computational approaches for the analysis of shared features between more than two data sets together. This strategy allowed, for the first time, a deep strain characterization using data from several origins, and the identification of sets of features for rapid and effective selection of good winemaking strains.

2. Methods

2.1. Strain collection and phenotypic characterization

The *S. cerevisiae* collection used in this work comprise 24 strains from different technological applications or origins (Supplemental data S1). This collection includes mainly strains used for winemaking (commercial and natural isolates that were obtained from winemaking environments – 15 strains), but also strains from natural environments (soil woodland, plants – two strains), bakery (one strain), other fermented beverages (fruit juice, palm wine, ginger beer, cachaça – four strains), and strains from unknown biological origin (two strains). All strains were stored in cryotubes containing 1 mL glycerol (30% v/v) at -80°C .

Phenotypic screening was performed using two sets of tests, the first consisting of growth tests performed using liquid medium in 96-well microplates, and the second one using agar plates, and considering a battery of tests previously established to characterize

strains from different proveniences (Mendes et al., 2013). Detailed experimental conditions can also be found in Mendes et al. (2013). In order to facilitate mathematical analysis, and considering the values of optical density, quantity of growth and change of colour in BiGGY medium, we assigned, in a growing order, the phenotypic results to classes between 0 and 3.

2.2. Genetic characterization

All 24 strains that constitute the yeast collection used in this study have been genetically characterized considering allelic combinations for 11 described microsatellites: ScaAT1, ScaAT2, ScaAT3, ScaAT4, ScaAT5, ScaAT6, YPL009, ScYOR267c, C4, C5 and C11 (Field & Wills, 1998; Franco-Duarte et al., 2014; Legras et al., 2005; Pérez, Gallego, Hidalgo, & Pe, 2001). Yeast cells cultivation, DNA isolation, multiplex primers and cycling conditions were performed as previously described (Franco-Duarte et al., 2009). Briefly, after cultivation of a frozen aliquot of yeast cells in 1 mL YPD medium (yeast extract 1% w/v, peptone 1% w/v, glucose 2% w/v) during 36 h at 28°C (160 rpm), DNA isolation was performed using a previously described method (Schuller, Valero, Dequin, & Casal, 2004) and used for microsatellite analysis. Multiplex PCR mixtures and cycling conditions were optimized and performed in 96-well PCR plates.

2.3. Individual fermentations and metabolic characterization

Individual fermentations of each strain were carried out at 18°C using white grape must (variety *Loureiro*) in Erlenmeyer flasks (100 mL) with rubber stoppers perforated with a syringe needle to allow CO_2 release. The used must had the following composition (w/v), determined by HPLC: glucose – 84.05 g/L, fructose – 54.36 g/L, tartaric acid – 1.22 g/L, glycerol – 0.19 g/L, acetic acid – 0.05 g/L, ethanol – 0.14 g/L, total acidity – 6.20 g/L, pH – 3.56. When glucose concentration was below 5 g/L and no weight variations were noted, samples were collected and frozen (-20°C) for metabolic analysis.

High-performance liquid chromatography with refractive index (HPLC-RI) was used to quantify ethanol and organic acids (tartaric, malic, acetic and succinic), in a EX Chrome Elite HPLC, using an Rezex[®] Ion Exclusion column. Column and refractive index detector temperatures were 60°C and 40°C , respectively, and the flow rate was 0.50 mL min^{-1} for 0–9 min, 0.25 mL min^{-1} for 10–14 min and 0.50 mL min^{-1} for 15–35 min.

Relevant metabolites (higher alcohols, esters, fatty acids) known to account for inter-strain differences and that are related to volatile compounds were determined by gas chromatography – mass spectrometry (GC-MS), in particular hexyl acetate, butanoic acid, hexanoic acid, octanoic acid, decanoic acid, dodecanoic acid, cis-3-hexenol, ethyl butanoate, ethyl hexanoate, ethyl octanoate, ethyl decanoate, ethyl dodecanoate and 2-phenylethyl acetate. Analyses were performed by solid phase microextraction (SPME), using a divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) 50/30 μm (Supelco, Sigma) fiber for 15 min. under continuous agitation and heating at 40°C . 3-Octanol (Sigma-Aldrich, 99% purity) was used as internal standard. Compounds were then desorbed from the SPME fiber directly and analyzed using a Varian CP-3800 gas chromatograph (Walnut Creek, CA, USA), equipped with a Varian Saturn 2000 mass selective detector, as previously described (Silva Ferreira et al., 2004).

2.4. Integrative data exploration from multiple experiments

Principal component analysis (PCA), available in the Unscrambler[®] X software, was used for variability analysis, using phenotypic, genetic and metabolic data.

A method of nonnegative matrix factorization (NMF) was used, as adapted from Zhang et al. (2012), to integrate the data obtained: metabolomic data obtained from GC–MS (concentrations of 13 volatile compounds) and HPLC (concentration of 5 metabolites), phenotypic results catalogued in four growth classes and microsatellite allelic presence/absence. In particular, with NMF method it was possible to integrate the results of multiple experiments within the same data set.

Results from the four experimental analyses were considered, each one with measured variables, being represented in four data matrices: X1 (GC–MS), X2 (HPLC), X3 (phenotypic characterization), X4 (microsatellite allelic presence/absence). Each of the four matrices had $n = 24$ rows, and a different number of columns – 13, 5, 30, 142 – regarding the matrix X1, X2, X3 and X4, respectively. Before NMF algorithm was applied, data was normalized using [0-1]-transformation, i.e. from each value the column minimal value was subtracted, being then the difference divided by the range. Matrices were then projected onto a common 2-dimensional system, in which each heterogeneous variable was weighted in the same projected direction forming a multi-dimensional module (md-module). To clarify, each of this md-modules contains a sub-set of strains sharing similar information regarding the experimental data, correlating in this way features of significant biological relevance. To perform this, each of the data matrices was decomposed in a common basis matrix (W) and in different coefficient matrices H_i (H_1, H_2, \dots, H_n) in a way that:

$$X_i \approx W \cdot H_i$$

Matrices W and H_i have non-negativity constraints ($W \geq 0$ and $H_i \geq 0$), and were computed in the way that X_i was as close as possible to $W \cdot H_i$, i.e., the sum over all matrices of squared differences between matrices X_i and $W \cdot H_i$ was as small as possible. In an adaptation of Zhang et al. (2012) method, generalized multiplicative update equations were used to minimize error function (sum of squared Frobenius norms). The algorithm iteratively computes the approximations of X_1, X_2, X_3 and X_4 . Then, from the factorizations, $W \cdot H_i$ subsets of strains and original variables were further defined which reflect relationships between four experimental data sets. To define which values are treated as high, Z-scores were computed for all rows in matrices H (subtracting the row mean and dividing the difference by the row standard deviation). According with the Z-scored value (greater or lower than the threshold $T = 2$) the variable was included or not in the md-module. Within this definition a particular strain can belong to different md-modules. The md-modules therefore do not form the partition since they can overlap and do not necessarily cover all strains.

With this as basis, 100 md-modules were analyzed. These modules provided an insight into the data set according with the following principles: (i) the closer the variables come, the higher is their similarity in the impact on the projection, and more related they are to each other; and (ii) the influence of a certain variable in the spatial projection is as high as their apartness from the origin. As follows, data from different matrices are projected onto a common coordinate system and correlative relationships can be inferred in the form of md-modules.

3. Results

3.1. Phenotypic and genetic characterization

A screening approach was devised, taking into consideration 30 phenotypic tests, including tests that are important for winemaking strain selection, to evaluate the extent of phenotypic variation. High-throughput testing in microplates was performed using supplemented grape must, and optical density (A_{640}) was measured

after 22 h of incubation. Growth in solid culture media was evaluated by visual scoring. All phenotypic results are summarized in Supplementary Table S2. The patterns of phenotypic variation obtained by principal component analysis can be visualized in Fig. 1, showing panel B the segregation of all 24 strains (scores) and panel A the loadings for the phenotypic variables, in the first two PCA components. 720 data points were obtained from the characterization of the 24 strains using 30 tests, and PC-1 (33%) and PC-2 (17%) explained 50% of strain variability, segregating strains by phenotypic behavior, as shown in Fig. 1B. No clear patterns were observed in relation to the strains technological group, being wine strains spread through both components. This fact shows the absence of a clear influence by any phenotypic test, which if present would separate strains clearly to a quadrant or to a direction in the PCA visualization. This demonstrates the heterogeneity of the strain collection, which is easily explained by their different locations and years of collection. The highest strain variability was associated with the growth in the presence of potassium bisulphite (KHSO_3) at both tested concentrations, at 40 °C, and regarding resistance to ethanol in liquid media (10 and 14%, v/v), as shown in Fig. 1A.

The 24 strains were also genetically characterized for allelic combinations using 11 *S. cerevisiae* specific microsatellites ScaAT1, ScaAT2, ScaAT3, ScaAT4, ScaAT5, ScaAT6, YPL009, ScYOR267c, C4, C5 and C11. A total of 142 alleles was obtained for the 24 strains (Supplementary Table S3). Microsatellites C5 and C11 were the most polymorphic ones with 18 different sized alleles obtained. Microsatellite ScaAT6 was the less polymorphic (6 different alleles). The genetic diversity of the collection is illustrated in the PCA plot of Fig. 2A, where 52% of genetic diversity is explained by the first two components (PC-1 – 32%, PC-2 – 20%). According to the PCA, microsatellites ScYOR267c, C4 and YPL009 were the most heterogeneous ones, explaining the larger part of the genetic variability observed (Fig. 2A) with some visible patterns of genetic relatedness between strains sharing the same technological group (Fig. 2B). Two of the strains obtained from fermented beverages other than wine (●) were located far on the right of the PCA, apart from all the other strains. These two strains, being collected from the fermentation of beverages cachaça and palm wine, showed higher sized alleles from microsatellite C4 than the remaining ones, together with small sized alleles for locus YPL009c (Fig. 2A). Also, two wine strains (★) were located apart in the PCA, namely one strain from France, located at the top of the PCA of Fig. 2A (higher alleles of microsatellite ScYOR267c) and one commercial wine strain located under the influence of PC-1 (right part) and the influence of PC-2 (upper part). Finally, one strain (●) obtained from nature (in detail, obtained from the plant bertram palm), was located far down on the PCA, showing one very small sized allele of locus ScYOR267c, not shared by any other strain.

3.2. Bioanalytical analysis

Bioanalytical analysis was accomplished with samples obtained at the end of fermentations from the 24 strains, to identify chemical compounds that enable the separation of wine strains from the others. A very good reproducibility was obtained between the three fermentation replicates (data not shown).

Strain-dependent differences could be observed concerning organic acids (tartaric, malic, succinic and acetic) and ethanol after HPLC quantification (Fig. 3). Tartaric acid concentration ranged between 0.9 and 1.2 g/L, whereas malic, acetic and succinic acids ranged between 5.7 and 7.3 g/L, 0.1 and 0.9 g/L, and 0.5 and 0.85 g/L, respectively (Fig. 3A). Final concentration of ethanol ranged between 103 and 121 g/L (Fig. 3B). PCA plots of HPLC data (panels C and D) explained 82% of strain variance in the first two components (PC-1 – 65%, PC-2 – 17%) and showed that strain

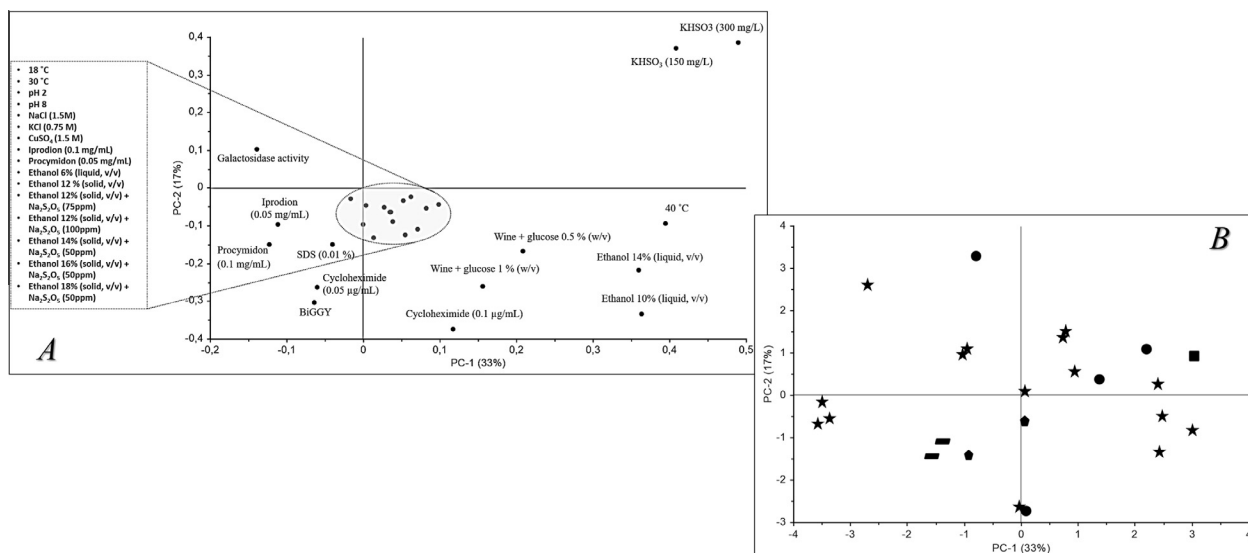


Fig. 1. Phenotypic diversity obtained for 24 *Saccharomyces cerevisiae* strains, as revealed by principal component analysis. A: loadings – 30 phenotypic tests. B: scores – 24 strains distribution. Symbols represent the strains technological applications or origin: ★ – wine; ● – other fermented beverages; ◆ – natural isolates; ■ – bread; ■ – unknown biological origin.

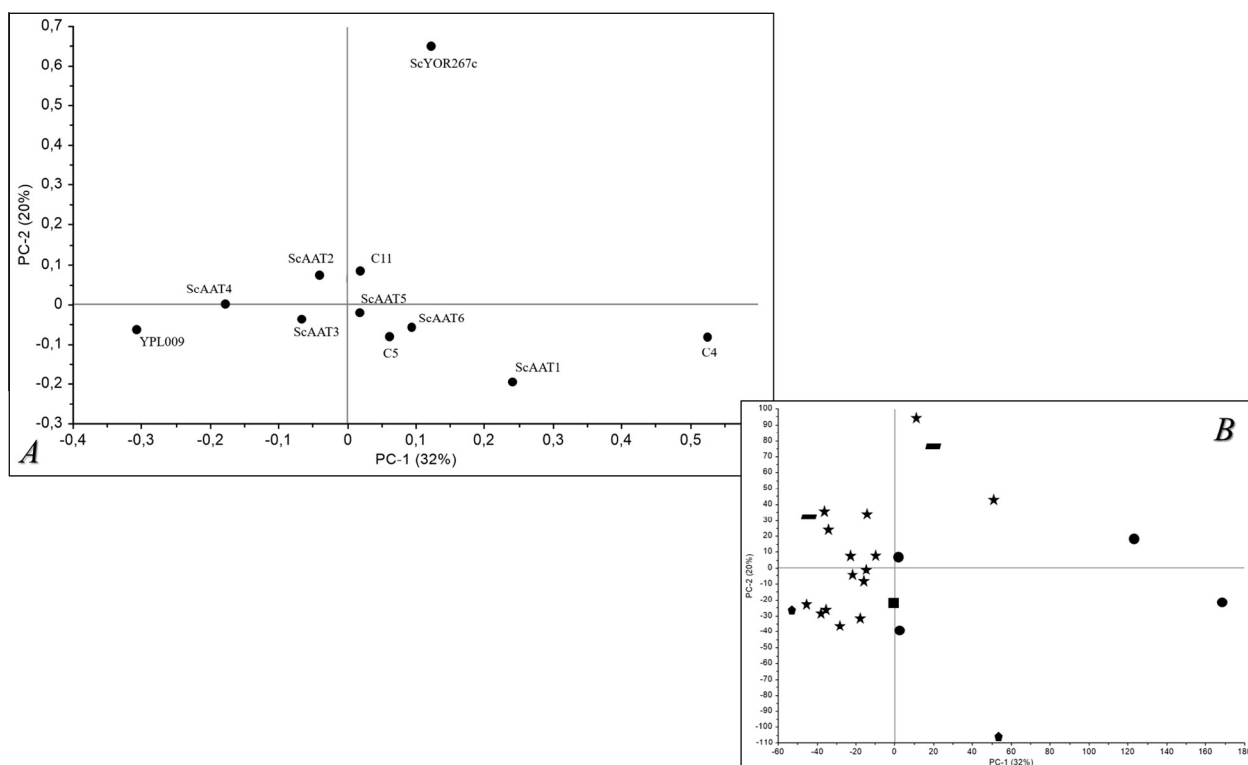


Fig. 2. Genetic diversity obtained for 24 *Saccharomyces cerevisiae* strains, as revealed by principal component analysis. A: loadings – 11 microsatellite loci. B: scores – 24 strains distribution. Symbols represent the strains technological applications or origin: ★ – wine; ● – other fermented beverages; ◆ – natural isolates; ■ – bread; ■ – unknown biological origin.

variability was mainly influenced by acetic acid production. These results, although not showing an evident separation according to the strains technological application or origin, allow to establish some patterns of distribution. In particular, acetic acid concentration discriminated strains along the first PCA component, and was highest in a natural isolate (◆, panel C), and in three strains

from fermented beverages other than wine (●), and lower in strains from unknown biological origins (■). Wine and vine strains (★) were spread throughout the PCA, showing no influence by the concentration of any compound, with the exception of a slightly lower concentration of acetic acid, as settled by their tendency to be located in the left part of the PCA.

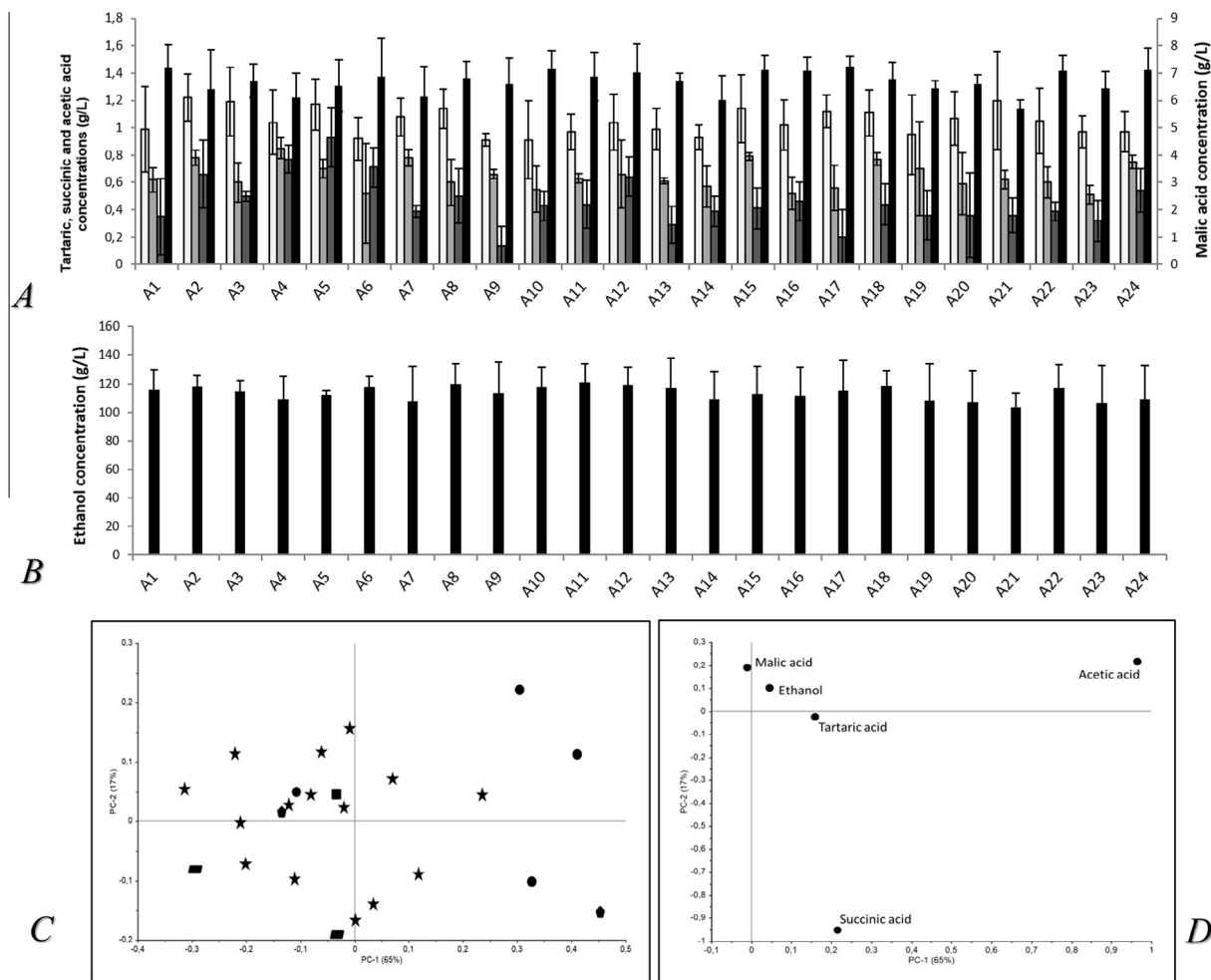


Fig. 3. HPLC analysis results obtained with 24 *Saccharomyces cerevisiae* strains. A: concentration of tartaric (□), succinic (▨), acetic (▩) and malic acids (■). B: concentration of ethanol. C: PCA plot of HPLC data showing the distribution of the 24 *S. cerevisiae* strains (scores) in the two first principal components. Symbols represent the strains technological applications or origin: ★ – wine and vine; ☆ – commercial wine strain; ■ – beer; ▨ – baker; ● – sake; ● – other fermented beverages; ● – clinical; ● – natural isolates; ■ – laboratory; ■ – unknown biological origin. D – PCA plot of HPLC data showing the distribution of the quantified compounds (loadings) in the two first principal components.

GC–MS analysis after solid phase microextraction (SPME) was used to determine aromatic compounds from the final fermentation stage. Table 1 shows the concentration of the 13 quantified volatile compounds, including the respective sensorial thresholds and odor descriptors. Concentrations above the sensorial detection threshold described for wines were detected for 8 of the 13 compounds: ethyl butanoate, ethyl hexanoate, ethyl octanoate, ethyl decanoate, 2-phenylethyl acetate, hexanoic acid, octanoic acid and decanoic acid. Ethyl hexanoate and ethyl octanoate were observed for all the 24 strains. A large variance among strains was observed for other compounds, being some of them produced in concentrations above the sensorial threshold by a small number of strains, such as hexanoic acid (4 strains) and decanoic acid (6 strains). Hexyl acetate, ethyl dodecanoate, butanoic acid, dodecanoic acid and cis-3-hexenol were produced in concentrations below the detected threshold by all strains.

The PCA plotted in Fig. 4 segregated the strains (panel A – loadings; panel B – scores) according to the aromatic profiles, and the first two components explained 70% of the observed variability between isolates (PC-1 – 53%, PC-2 – 17%). A clear separation of strains according to the type of compound produced was revealed by PCA (Fig. 4A): esters were located in the upper-right part of the PCA, whereas acids were predominant in the lower-left part, under

influence of both principal components. This division was not clearly related with the strains technological origin, but particular groups of strains showed a different behavior regarding these compounds (Fig. 4B): (i) wine strains (★) showed intermediate concentrations of both esters and acids; (ii) strains from unknown biological origin (■) showed a high production of esters, with a particularly higher production of hexyl acetate by one of the isolates; (iii) some strains from fermented beverages other than wine (●) positioned in the right part of the PCA plot mainly due to a higher production of decanoic acid and ethyl decanoate, among others; (iv) natural isolates (●) and isolates from bread (■) were positioned near the plot origin, showing no significant influence by any particular compounds. The position of wine strains as intermediate producers of both esters and volatile acids, in opposition for example to strains from other fermented beverages, is in agreement with the importance of both families of compounds in the aromatic profiles.

3.3. Pheno-metabolome portrayal by identification of multi-dimensional modules

From the initial data set (30 phenotypic tests with results catalogued in classes from 0 to 3, 142 microsatellite allelic sizes,

Table 1
Concentration (mg/L) of aromatic compounds determined by GC–MS in the sub-group of 24 *Saccharomyces cerevisiae* strains. Concentrations above the sensorial threshold are underlined.

Compounds	Hexyl acetate	Butanoic acid	Hexanoic acid	Octanoic acid	Decanoic acid	Dodecanoic acid	Cis-3-hexenol	Ethyl butanoate	Ethyl hexanoate	Ethyl octanoate	Ethyl decanoate	Ethyl dodecanoate	2-Phenylethyl acetate
Sensorial threshold	0.640	2.200	8.000	8.800	6.000	0.610	0.400	0.200	0.005	0.002	0.200	Not available	0.250
Odor description	Sweet, aromatic, fragrant	Cheese, rancid	Cheese, sweaty	Rancid, harsh	Fatty	Soapy, waxy	Green leefs, banana, sweet; herb	Acid fruit	Green apple	Sweet, soap	Pleasant, soap	Soapy, estery	Fruity, flowery with a honey note
References	Etiévant and Etiévant (1991) and Meilgaard (1975)	Meilgaard (1975)	Amerine and Roessler (1976)	Salo (1970)	Amerine and Roessler (1976)	Vilanova, Genisheva, Masa, and Oliveira (2010)	Cullere, Escudero, Cacho, Ferreira, and Culleré (2004), Escudero et al. (2004), Ferreira, López, Cacho, Lo, and Lopez (2000) and Gewu and Guth (1997)	Gewu and Guth (1997) and Meilgaard (1975)	Gewu and Guth (1997) and Meilgaard (1975)	Gewu and Guth (1997) and Meilgaard (1975)	Ferreira et al. (2000) and Meilgaard (1975)	Siebert et al. (2005)	Lilly et al. (2000)
A1	0.247	0.461	4.201	<u>11.496</u>	1.464	0.151	0.153	0.181	<u>0.666</u>	<u>1.807</u>	0.140	0.008	<u>0.285</u>
A2	0.234	0.786	<u>8.487</u>	<u>14.131</u>	<u>6.258</u>	0.255	0.185	<u>0.211</u>	<u>0.889</u>	<u>1.880</u>	<u>1.126</u>	0.039	<u>0.279</u>
A3	0.197	0.509	5.387	<u>14.869</u>	5.459	0.110	0.152	0.109	<u>0.760</u>	<u>0.987</u>	<u>0.245</u>	0.007	<u>0.266</u>
A4	0.165	0.641	7.043	<u>19.455</u>	<u>8.499</u>	0.157	0.161	0.191	<u>0.720</u>	<u>1.555</u>	<u>0.984</u>	0.014	<u>0.287</u>
A5	0.214	0.298	5.139	<u>11.503</u>	4.928	0.155	0.125	0.098	<u>0.552</u>	<u>1.427</u>	<u>0.840</u>	0.015	<u>0.229</u>
A6	0.282	0.455	<u>11.194</u>	<u>20.128</u>	<u>10.413</u>	0.323	0.173	<u>0.202</u>	<u>0.950</u>	<u>3.101</u>	<u>1.774</u>	0.057	<u>0.419</u>
A7	0.206	0.605	<u>8.465</u>	<u>16.930</u>	<u>7.269</u>	0.243	0.162	<u>0.266</u>	<u>0.805</u>	<u>1.364</u>	<u>0.797</u>	0.031	<u>0.610</u>
A8	0.207	0.469	5.369	<u>11.774</u>	4.547	0.179	0.144	0.080	<u>0.777</u>	<u>1.490</u>	<u>0.520</u>	0.012	<u>0.255</u>
A9	0.517	0.625	5.136	<u>12.785</u>	2.696	0.116	0.133	<u>0.308</u>	<u>1.090</u>	<u>2.284</u>	<u>0.253</u>	0.013	<u>0.462</u>
A10	0.223	0.498	4.521	<u>13.656</u>	1.209	0.058	0.165	<u>0.247</u>	<u>0.712</u>	<u>1.213</u>	0.107	0.003	0.212
A11	0.414	0.474	6.142	<u>17.507</u>	4.543	0.126	0.144	0.079	<u>0.991</u>	<u>2.623</u>	<u>0.652</u>	0.014	<u>0.300</u>
A12	0.220	0.479	5.257	<u>14.320</u>	4.213	0.304	0.139	<u>0.250</u>	<u>0.736</u>	<u>0.997</u>	<u>0.283</u>	0.014	<u>0.285</u>
A13	0.203	0.377	4.070	<u>10.849</u>	1.506	0.141	0.135	0.161	<u>0.717</u>	<u>1.054</u>	0.055	0.009	0.210
A14	0.253	0.475	4.473	<u>13.909</u>	2.483	0.075	0.122	<u>0.247</u>	<u>0.590</u>	<u>1.381</u>	<u>0.262</u>	0.009	0.201
A15	0.249	0.653	6.451	<u>11.805</u>	3.490	0.190	0.181	<u>0.238</u>	<u>1.000</u>	<u>2.291</u>	<u>0.330</u>	0.020	<u>0.349</u>
A16	0.229	0.582	<u>8.812</u>	<u>30.719</u>	<u>11.497</u>	0.511	0.141	0.123	<u>0.947</u>	<u>2.510</u>	<u>0.854</u>	0.013	<u>0.253</u>
A17	0.242	0.352	3.960	<u>11.553</u>	2.311	0.052	0.115	<u>0.202</u>	<u>0.596</u>	<u>1.219</u>	<u>0.219</u>	0.005	<u>0.254</u>
A18	0.413	0.699	5.244	<u>12.057</u>	2.816	0.120	0.193	<u>0.218</u>	<u>0.873</u>	<u>2.031</u>	<u>0.458</u>	0.017	<u>0.364</u>
A19	0.351	0.560	7.341	<u>18.477</u>	<u>7.735</u>	0.278	0.132	<u>0.255</u>	<u>1.084</u>	<u>2.175</u>	<u>0.873</u>	0.019	<u>0.313</u>
A20	0.230	0.455	4.385	<u>9.449</u>	1.854	0.078	0.164	<u>0.204</u>	<u>0.555</u>	<u>1.342</u>	<u>0.334</u>	0.009	0.231
A21	0.085	0.492	5.411	<u>18.661</u>	3.782	0.105	0.117	<u>0.265</u>	<u>0.536</u>	<u>0.764</u>	<u>0.207</u>	0.008	0.094
A22	0.093	0.484	3.395	8.065	1.322	0.086	0.117	0.183	<u>0.386</u>	<u>0.681</u>	0.098	0.005	0.184
A23	0.253	0.543	5.453	<u>16.448</u>	2.482	0.151	0.123	<u>0.228</u>	<u>0.786</u>	<u>1.894</u>	0.195	0.015	<u>0.303</u>
A24	0.326	0.370	3.401	<u>9.668</u>	1.867	0.090	0.148	0.188	<u>0.581</u>	<u>1.319</u>	<u>0.231</u>	0.006	<u>0.442</u>

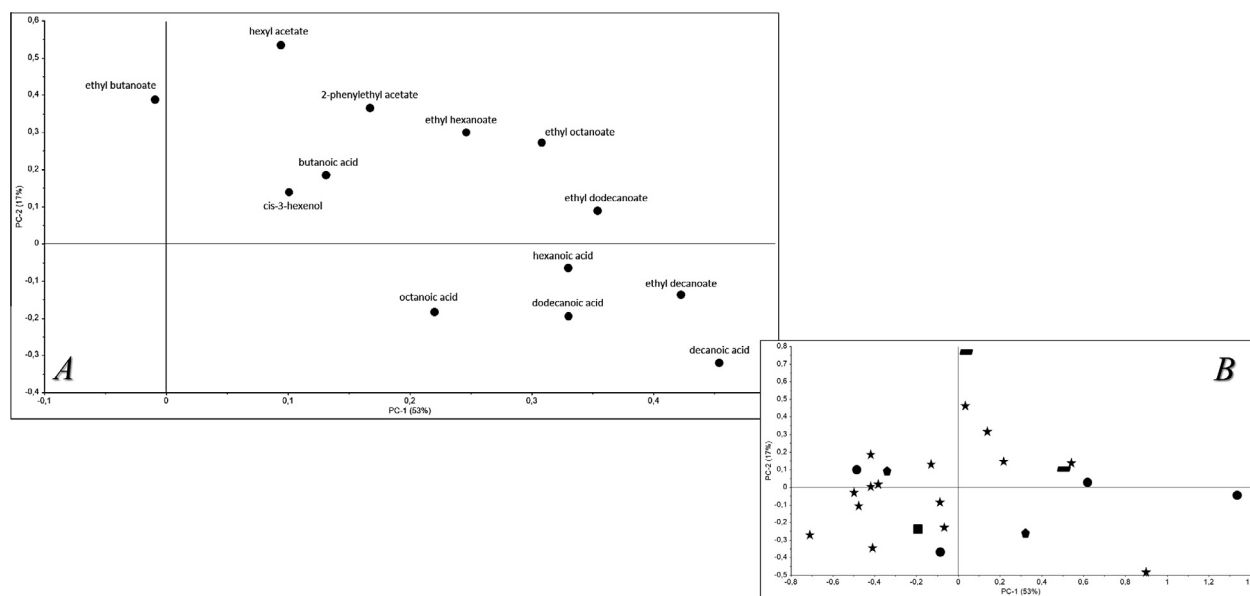


Fig. 4. Principal component analysis of GC-MS data for 24 *Saccharomyces cerevisiae* strains. A: loadings: 13 metabolic compounds. B: scores: 24 strains distribution. Symbols represent the strains technological applications or origin: ★ – wine; ● – other fermented beverages; ■ – natural isolates; ■ – bread; ■ – unknown biological origin.

concentration of 13 compounds obtained by GC-MS analysis and concentration of 5 metabolites determined by HPLC) across 24 strains, features present in less than 5% of the strains were removed. From this new data set (26 phenotypic tests, 66 microsatellite alleles and 13 + 5 metabolic compounds) a common basis matrix (W) was composed, as described in Section 2. Matrix W projection is presented in [Supplementary data S4](#), showing also how variables correlate between each other. As closer to each other, more similar is their impact on the projection and when they are more apart from the origin, the correlation coefficient increases. With this approach an attempt was explored to search how variables correlate in a way that we can group them in terms of similar behavior in certain conditions. After parameter optimizations, the 4 large matrices were broken down into basic building blocks, from which 100 multi-dimensional correlated modules (md-modules) were obtained. For clarification, these md-modules consists of sub-sets of most related data obtained from the projection presented in [Supplementary data S4](#), and are composed by a certain number of strains that have a similar behavior for the variables tested. In [Table 2](#) and [Supplementary data S5](#), the 17 statistical most relevant md-modules are presented, combining for the first time, sets of most-correlated features of significant biological relevance. Three of the 17 modules contain only strains from wine environments: modules number 29, 34 and 47. Good capacity to grow in cycloheximide and at 18 °C was a transversal feature to the three modules, which was already shown in our previous work to be a phenotypic trait associated with wine strains ([Mendes et al., 2013](#)). Cycloheximide is an inhibitor of protein synthesis, and it was shown that spontaneous mutants of *S. cerevisiae* that are resistant to this compound can be isolated from industrial fermentations ([Perez, Regodon, Valdes, De Miguel, & Ramirez, 2000](#)). In md-module number 29 it was possible to associate the phenotypic characteristics of growth in cycloheximide, iprodion, 18 °C and ethanol 6% (w/v) of the four mentioned wine strains, with the results obtained in the GC-MS quantification for 2-phenylethyl acetate. This compound contributes to the fruity and flowery aroma of wines ([Lilly, Lambrechts, & Pretorius, 2000](#)), but may mask some varietal aromas if present in high concentrations. The formation of this ester is especially promoted in slow fermenta-

tions, at low temperatures and in the absence of oxygen ([Ribéreau-Gayon, 2000](#)). These facts are in agreement with the relations found with phenotypic characteristics of module 29, especially the temperature of 18 °C (strains having the highest growth at this temperature were integrated in this module) and the presence of ethanol (strains obtaining the highest growth class in the presence of 6% (w/v) ethanol). In this way, results show that some features have high potential to be used in strain selection, especially directed for winemaking application. Particularly, some features were present in the md-modules in a higher proportion than others, having also a particular predominance in modules with only winemaking strains: good capacity to grow (highest phenotypic class) in cycloheximide (both at 0.05 or 0.1% w/v) – 19 occurrences; good capacity to grow (highest phenotypic class) in iprodion (0.05 or 0.1% w/v) – 13 occurrences; capacity to grow at 18 °C (phenotypic class 1) – 9 occurrences; presence of homozygous alleles ScaAT6-256 and ScaAT5-256 – 7 and 4 occurrences respectively; good production of the compounds 2-phenylethyl acetate and ethyl butanoate (4 occurrences each), and also of the compounds ethyl hexanoate and ethyl octanoate (3 occurrences each).

4. Discussion

In recent years, research has focused on the investigation of relationships between metabolic pathways and phenotypic and genetic fingerprints. However, systematic analysis of such multi-dimensional data to discover biological patterns is still a difficult task. A great number of tools were developed for 1- or, at most, 2-dimensional data, with satisfactory results. In our previous work we developed computational methods to establish associations between phenotypes and genotypes of *S. cerevisiae* yeasts from different origins ([Franco-Duarte et al., 2009](#); [Franco-Duarte et al., 2014](#); [Mendes et al., 2013](#)). In the present work computational approaches were improved to be applied to multi-dimensional data, obtained through metabolic, phenotypic and genetic characterization of the yeast collection, which was not yet routinely possible. Computational biology can benefit from this knowledge,

Table 2 (continued)

MN	S	TG	W	Phenotypic test	PC	W	Microsatellite allele	H	W	HPLC quantified compound	NC/QC	W	GC–MS quantified compound	NC/QC	W
18	A1	Natural isolate	3.84	Wine + glucose (0.5% w/v)	1	1.00	ScAAT6-256	2	0.02						
	A23	Wine and vine	2.48	40 °C	2	0.08									
	A13	Wine and vine	2.23	Ethanol 6% (v/v) – Im Iprodion (0.1 mg/mL) 18 °C	3	0.04									
20	A11	Wine and vine	4.06	Ethanol 6% (v/v) – Im	3	0.08	ScAAT4-329	2	0.14				Hexyl acetate	0.75/0.39	0.04
	A9	Unknown	3.40	Iprodion (0.1 mg/mL)	3	0.03							Ethyl hexanoate	0.93/1.01	0.03
	A16	Wine and vine	2.02												
29	A15	Wine and vine	2.29	Cycloheximide (0.1 µg/mL)	3	0.08							2-Phenylethyl acetate	0.52/0.32	0.03
	A12	Wine and vine	1.19	Ethanol 6% (v/v) – Im	3	0.06									
	A22	Wine and vine	1.01	18 °C	1	0.02									
	A24	Wine and vine	1.00	Iprodion (0.05 mg/mL)	3	0.02									
				Iprodion (0.1 mg/mL)	3	0.02									
				Cycloheximide (0.05 µg/mL)	3	0.02									
34	A16	Wine and vine	3.65	SDS (0.01% w/v)	1	0.13	YPL009c-307	2	0.05						
	A12	Wine and vine	3.46	Iprodion (0.1 mg/mL)	3	0.04	ScAAT5-219	2	0.05						
	A13	Wine and vine	1.86	Iprodion (0.05 mg/mL)	3	0.03									
				NaCl (1.5 M) 18 °C	1	0.03									
					1	0.02									
				Cycloheximide (0.05 µg/mL)	3	0.02									
47	A20	Wine and vine	5.25	KHSO ₃ (300 mg/L)	3	0.13	ScAAT2-378	2	0.15						
	A19	Wine and vine	5.07	18 °C	1	0.06	ScAAT5-256	2	0.05						
	A15	Wine and vine	3.22	H ₂ S production	2	0.06	ScAAT6-256	2	0.03						
				Cycloheximide (0.1 µg/mL)	3	0.04									
61	A22	Wine and vine	1.84	Iprodion (0.05 mg/mL)	3	0.08	ScAAT6-256	2	0.04						
	A13	Wine and vine	1.46	CuSO ₄ (5 mM)	1	0.05									
	A1	Natural isolate	1.24	Cycloheximide (0.1 µg/mL)	3	0.04									
	A15	Wine and vine	0.94												
	A12	Wine and vine	0.90												
71	A19	Wine and vine	2.53	Cycloheximide (0.05 µg/mL)	3	0.05	ScAAT5-256	2	0.06				Ethyl hexanoate	0.96/1.04	0.03
	A9	Unknown	1.62	Cycloheximide (0.1 µg/mL)	3	0.03	ScAAT6-256	2	0.04						
	A16	Wine and vine	1.61	Iprodion (0.1 mg/mL) 18 °C	3	0.02									
					1	0.02									
78	A13	Wine and vine	3.86	Wine + glucose (1% w/v)	1	0.24	ScAAT5-256	2	0.04	Acetic acid	0.64/0.44	0.04			
	A3	Bread	2.24	Galactosidase activity	2	0.06	ScAAT5-219	2	0.04						
	A12	Wine and vine	1.82												

(continued on next page)

Table 2 (continued)

MN	S	TG	W	Phenotypic test	PC	W	Microsatellite allele	H	W	HPLC quantified compound	NC/QC	W	GC-MS quantified compound	NC/QC	W
80	A6	Other fb	3.30	KCl (0.75 M)	2	0.05	ScAAT3-241	2	0.11				Hexyl acetate	0.47/0.24	0.04
	A2	Other fb	2.24	Cycloheximide (0.1 µg/mL)	3	0.05							Ethyl octanoate	0.69/2.14	0.03
	A5	Other fb	2.03	Cycloheximide (0.05 µg/mL)	3	0.04							Ethyl decanoate	0.70/1.25	0.02
				18 °C	1	0.04									
				Ethanol 14% (v/v) – lm	2	0.02									
85	A4	Natural isolate	3.02	H ₂ S production	2	0.09							Ethyl decanoate	0.32/0.57	0.02
	A18	Wine and vine	2.17	Procymidon (0.1 mg/mL)	3	0.08							Ethyl butanoate	0.71/0.22	0.02
	A14	Wine and vine	1.65	Cycloheximide (0.05 µg/mL)	3	0.06									

MN – module number; S – strains characterizing the module; TG – technological group; W – weight of the feature in the module; PC – phenotypic classes (0–3) according to the amount of growth (see Section 2); H – heterozygous alleles (1) or homozygous allele (2); NC/QC – normalized concentration (g/L) and quantified concentration according to the mentioned method; lm – liquid must; fb – fermented beverages.

since the development of effective methods that are able to combine numeric and alpha-numeric data types in an easily-extendible way is an important objective of nowadays research.

Phenotypic characterization of the 24 *S. cerevisiae* strains was performed based on approaches that are generally applied for the selection of yeast winemaking strains (Mannazzu, Clementi, & Ciani, 2002). This assessment revealed a high phenotypic diversity, which is in agreement with other studies using natural yeast populations (Camarasa, Sanchez, Brial, Bigey, & Dequin, 2011; Franco-Duarte et al., 2015; Goddard, Anfang, Tang, Gardner, & Jun, 2010; Liti et al., 2009; Schacherer et al., 2009; Warringer et al., 2011). This high intra-strain diversity can be explained by genetic rearrangements which are characteristic for this species (Dunn, Levine, & Sherlock, 2005; Schuller et al., 2007). Also, Camarasa (Camarasa et al., 2011) provided evidence for phenotypic evolution driven by environmental adaptation, namely some phenotypes (resistance to high sugar concentrations, ability to complete fermentation and low acetate production) that were capable to distinguish groups of strains according to their ecological niches. In parallel with the phenotypic characterization, a high genetic diversity (Fig. 2, Table S3), was obtained for the collection of 24 isolates, with a total of 142 alleles obtained with 11 polymorphic microsatellites. Contrarily to our previous studies (Franco-Duarte et al., 2009) using microsatellites to characterize yeast strains, the most polymorphic microsatellites were C5 and C11, followed by ScAAT1 and ScAAT5 (Table S3). This seems to indicate that when considering a strain collection more biased to winemaking, the higher genetic variation occurs in the allele size of these loci, pointing to their importance in the choice of winemaking candidate strains.

After phenotypic and genetic characterization, strains were also analyzed regarding their metabolic profiles by HPLC and GC-MS. These methods revealed to be accurate to determine aromatic and volatile compounds from the final fermentation stage. HPLC analysis revealed an opposite contribution of acetic acid and the remaining acids and ethanol regarding the first PCA component (PCA, Fig. 3), which is in agreement with reported effects of acetic acid on the fermentation yield and yeast growth (Maiorella, Blanch, & Charles, 1983; Taherzadeh, Niklasson, & Lidn, 1997; Thomas, Hynes, & Ingledew, 2002). Acetic acid is an important end-product of energy metabolism (Tielens, van Grinsven, Henze, van Hellemond, & Martin, 2010), and is used as an antimicrobial agent in the food and beverage industries (Luck & Jager, 1997) due to the enhanced production of its precursor acetyl-CoA. Results obtained in several organisms showed the association of acetic acid with the

capacity to survive to unfavorable conditions (Tielens, Rotte, van Hellemond, & Martin, 2002). In this way, the significant presence of acetic acid in the end of fermentation, mainly in natural isolates in opposition to wine strains, is in agreement with the survival of these strains in environmental conditions.

GC-MS was able to detect concentrations above the sensorial detection thresholds in 8 compounds, from the 13 quantified (Table 1). The main limitation of using GC-MS approaches is the difficult identification of compounds in an unsupervised way, due to the inexistence of a universal spectral library. PCA of these results revealed a clear separation between acids and esters, in terms of concentrations produced by the strains (Fig. 4). Esters, produced by yeasts during alcoholic fermentation, are known, both in the case of ethyl acetate esters and fatty acid esters, to have a significant influence on the fruity aromas of the final product as documented in Table 1. In the case of volatile fatty acids, their concentration influenced also the PCA position of wine strains. Concentration of these compounds in wine were reported as being usually between 500 and 1000 mg/L (Swiegers, Bartowsky, Henschke, & Pretorius, 2005). The concentration of volatile acids is of particular relevance, being associated with unpleasant odors and tastes in concentrations above 300 mg/L, such as a pungent smell and taste. In concentrations below that level, volatile acids can have a positive impact with fruity and floral aromas (González Álvarez, González-Barreiro, Cancho-Grande, & Simal-Gándara, 2011), mainly due to the obstruction of their esters hydrolysis.

An holistic matrix factorization approach was assessed and adapted from Zhang et al. (2012) to project data onto a common system of coordinates, in which the most related variables were weighted together and placed apart from the axis origin. The used NMF approach provided two major advantages as being applied to our data: (a) it reduced the complexity across multiple data sets contrarily to other available algorithms, once that the discovered md-modules (modules composed with strains sharing identical experimental results) identified vertical associations between multiple data sets which would be ignored if only one (PCA) or two (PLS) data sets were used; (b) provided a global overview of the inherent structure of our data, because the modular approach captured the associations among sets of different types of variables (GC-MS, phenotypic results, microsatellite allelic presence/absence) and md-modules identification stratified strains into distinct groups, which helped to identify the complex associations between the three analyzed data sets. From NMF analysis, a subset of 17 statistical significant multi-dimensional modules

(md-modules) were revealed (Table 2), combining for the first time, sets of most-correlated features of significant biological relevance.

The method presented revealed to be a successful way to reduce the dimensionality of the data and/or to uncover hidden patterns, to combine significant multi-scale information from different analytical origins, being its wider application to facilitate the interpretation of the data. It has to be mentioned that some of the associations found relate to phenotypes with small interest for the winemakers. However, it was important for the objectives of the present work not to focus only on tests relevant for winemaking and/or strain selection. In this way, the associations found were important to obtain the exploratory method that can now be easily extendible to other phenotypes or other sets of numeric or alpha-numeric data types.

By identifying md-modules it was possible to break down data sets into smaller blocks, and search for correlated patterns. This method has the potential to be easily applied to any sets of data, allowing the identification of good strains in selection programs without the need of extensive and laborious characterizations.

5. Conclusions

In the present work powerful data analysis techniques were adapted to the results obtained with the selected *S. cerevisiae* strain collection, in order to contribute to two current challenges in today's science: *i*) analytical methods allow the debit of several gigabytes of data in just a few minutes, but data analysis is not capable to scrutinize them in a proper way, ignoring a large part of its potential; *ii*) although several methods have been suggested and improved to find associations between two data sets, no method allowed yet the comparison between three or more sets of data.

The focus of this work was to develop and adapt already existing strategies to combine multi-scale data from different origins (phenotypes, microsatellites and metabolic data). We consider our approach to be successful and innovative, by the use of new approaches of matrix factorization that allow the identification of multi-dimensional correlated modules with significant biological relevance. Our work shed light into the possibility of obtaining a holistic view of the *S. cerevisiae* pheno-metabolome, which was not yet routinely possible with the current state of the art methods. These findings may be of great importance for several biotechnological applications, namely identification of strains to be used in winemaking by the replacement of extensive and laborious experiments. In particular the presented method allowed to focus in specific features that were shown to be associated with good winemaking strains: good capacity to grow in cycloheximide, in iprodion and at 18 °C, presence of homozygous alleles ScAAT6-256 and ScAAT5-256, good production of the compounds 2-phenylethyl acetate, ethyl butanoate, ethyl hexanoate and ethyl octanoate.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RFD and DS designed the experiments; RFD, LU, IM, CCC, NF and RM performed the experiments; RFD, PS, CP and DS wrote the manuscript; ACSF, PS, CP and DS supervised the work. All authors contributed to the discussion of the research and read and approved the final manuscript.

Acknowledgements

Inês Mendes was recipient of a fellowship from the Portuguese Science Foundation, FCT (SFRH/BD/74798/2010). This work was supported by FCT I.P. through the strategic funding UID/BIA/04050/2013, and the project PTDC/AGR-ALI/121062/2010.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.foodchem.2016.05.080>.

References

- Amerine, M., & Roessler, E. (1976). *Wines, their sensory evaluation*. N. Y.: WH Freeman.
- Becker, S. A., & Palsson, B. O. (2008). Three factors underlying incorrect in silico predictions of essential metabolic genes. *BMC Systems Biology*, 2, 14. <http://dx.doi.org/10.1186/1752-0509-2-14>.
- Boulesteix, A., & Strimmer, K. (2007). Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, 8(1), 32–44. <http://dx.doi.org/10.1093/bib/bbl016>.
- Camarasa, C., Sanchez, I., Brial, P., Bigey, F., & Dequin, S. (2011). Phenotypic landscape of *Saccharomyces cerevisiae* during wine fermentation: Evidence for origin-dependent metabolic traits. *PLoS ONE*, 6(9), e25147. <http://dx.doi.org/10.1371/journal.pone.0025147>.
- Carreto, L., Eiriz, M. F., Gomes, A. C., Pereira, P. M., Schuller, D., & Santos, M. A. S. (2008). Comparative genomics of wild type yeast strains unveils important genome diversity. *BMC Genomics*, 9, 524. <http://dx.doi.org/10.1186/1471-2164-9-524>.
- Cullere, L., Escudero, A., Cacho, J., Ferreira, V., & Culleré, L. (2004). Gas chromatography-olfactometry and chemical quantitative study of the aroma of six premium quality Spanish aged red wines. *Journal of Agricultural and Food Chemistry*, 52(6), 1653–1660. <http://dx.doi.org/10.1021/jf0350820>.
- Devarajan, K. (2008). Nonnegative matrix factorization: An analytical and interpretive tool in computational biology. *PLoS Computational Biology*, 4(7), e1000029. <http://dx.doi.org/10.1371/journal.pcbi.1000029>.
- Dunn, B., Levine, R. P., & Sherlock, G. (2005). Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. *BMC Genomics*, 6, 53. <http://dx.doi.org/10.1186/1471-2164-6-53>.
- Escudero, A., Gogorza, B., Melús, M. A., Ortín, N., Cacho, J., Ferreira, V., ... Ortín, N. (2004). Characterization of the aroma of a wine from maccabeo. Key role played by compounds with low odor activity values. *Journal of Agricultural and Food Chemistry*, 52(11), 3516–3524. <http://dx.doi.org/10.1021/jf0353411>.
- Etiévant, P. X., & Etiévant, P. (1991). Wine. In H. Maa (Ed.), *Volatile compounds in food and beverages* (pp. 483–546).
- Ferreira, V., López, R., Cacho, J. F. J., Lo, R., & Lopez, R. (2000). Quantitative determination of the odorants of young red wines from different grape varieties. *Journal of the Science of Food and Agriculture*, 80, 1659–1667.
- Field, D., & Wills, C. (1998). Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proceedings of the National Academy of Sciences of the United States of America*, 95(4), 1647–1652.
- Franco-Duarte, R., Bigey, F., Carreto, L., Mendes, I., Dequin, S., Santos, M. A. S., & Schuller, D. (2015). Intra-strain genomic and phenotypic variability of the commercial *Saccharomyces cerevisiae* strain Zymaflore VL1 reveals micro-evolutionary adaptation to vineyard environments. *FEMS Yeast Research*, 15 (6). <http://dx.doi.org/10.1093/femsyr/fov063>.
- Franco-Duarte, R., Mendes, I., Gomes, A. C., Santos, M. A. S., de Sousa, B., & Schuller, D. (2011). Genotyping of *Saccharomyces cerevisiae* strains by interdelta sequence typing using automated microfluidics. *Electrophoresis*, 32(12), 1447–1455. <http://dx.doi.org/10.1002/elps.201000640>.
- Franco-Duarte, R., Mendes, I., Umek, L., Drumonde-Neves, J., & Zupan, B. (2014). Computational models reveal genotype – phenotype associations in *Saccharomyces cerevisiae*. *Yeast*, 31(May), 265–277. <http://dx.doi.org/10.1002/yea.3016>.
- Franco-Duarte, R., Umek, L., Zupan, B., & Schuller, D. (2009). Computational approaches for the genetic and phenotypic characterization of a *Saccharomyces cerevisiae* wine yeast collection. *Yeast*, 26, 675–692. <http://dx.doi.org/10.1002/yea.1728>.
- Gewu, W., & Guth, H. (1997). Quantitation and sensory studies of character impact odorants of different white wine varieties. *Journal of Agriculture and Food Chemistry*, 45, 3027–3032 (Table 1).
- Goddard, M. R., Anfang, N., Tang, R., Gardner, R. C., & Jun, C. (2010). A distinct population of *Saccharomyces cerevisiae* in New Zealand: Evidence for local dispersal by insects and human-aided global dispersal in oak barrels. *Environmental Microbiology*, 12(1), 63–73. <http://dx.doi.org/10.1111/j.1462-2920.2009.02035.x>.

- González Álvarez, M., González-Barreiro, C., Cancho-Grande, B., & Simal-Gándara, J. (2011). Relationships between Godello white wine sensory properties and its aromatic fingerprinting obtained by GC-MS. *Food Chemistry*, 129(3), 890–898. <http://dx.doi.org/10.1016/j.foodchem.2011.05.040>.
- Hutchins, L. N., Murphy, S. M., Singh, P., & Graber, J. H. (2008). Position-dependent motif characterization using non-negative matrix factorization. *Bioinformatics*, 24(23), 2684–2690. <http://dx.doi.org/10.1093/bioinformatics/btn526>.
- Kim, H., & Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics*, 23(12), 1495–1502. <http://dx.doi.org/10.1093/bioinformatics/btm134>.
- Kuligowski, J., Quintás, G., Herwig, C., & Lendl, B. (2012). A rapid method for the differentiation of yeast cells grown under carbon and nitrogen-limited conditions by means of partial least squares discriminant analysis employing infrared micro-spectroscopic data of entire yeast cells. *Talanta*, 99, 566–573. <http://dx.doi.org/10.1016/j.talanta.2012.06.036>.
- Legras, J.-L., Ruh, O., Merdinoglu, D., & Karst, F. (2005). Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. *International Journal of Food Microbiology*, 102(1), 73–83. <http://dx.doi.org/10.1016/j.ijfoodmicro.2004.12.007>.
- Lilly, M., Lambrechts, M. G., & Pretorius, I. S. (2000). Effect of increased yeast alcohol acetyltransferase activity on flavor profiles of wine and distillates. *Applied and Environment Microbiology*, 66(2), 744–753. <http://dx.doi.org/10.1128/AEM.66.2.744-753.2000>.
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., ... Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, 458(7236), 337–341. <http://dx.doi.org/10.1038/nature07743>.
- Luck, M., & Jager, M. (1997). Acetic acid. *Antimicrobial food additives: Characteristics, uses, effects* (pp. 137–143). Verlag: Springer.
- Maiorella, B., Blanch, H. W., & Charles, R. (1983). By-product inhibition effects on ethanolic fermentation by *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, 25(1), 103–121. <http://dx.doi.org/10.1002/bit.260250109>.
- Mannazzu, I., Clementi, F., & Ciani, M. (2002). Strategies and criteria for the isolation and selection of autochthonous starter. In M. Ciani (Ed.), *Biodiversity and biotechnology of wine yeasts* (pp. 19–35). Trivandrum: Research Signpost.
- Meilgaard, M. C. (1975). Flavor chemistry of beer: Part II: Flavor and threshold of 239 aroma volatiles. *MBAA Technical Quarterly*, 12, 151–168.
- Mendes, I., Franco-Duarte, R., Umek, L., Fonseca, E., Drumonde-Neves, J., Dequin, S., ... Schuller, D. (2013). Computational models for prediction of yeast strain potential for winemaking from phenotypic profiles. *PLoS ONE*, 8(7), e66523. <http://dx.doi.org/10.1371/journal.pone.0066523>.
- Pérez, M. A., Gallego, F. J., Hidalgo, P., & Pe, M. A. (2001). Evaluation of molecular techniques for the genetic characterization of *Saccharomyces cerevisiae* strains. *FEMS Microbiology Letters*, 205(2), 375–378. [1111/j.1574-6968.2001.tb10975.x](http://dx.doi.org/10.1111/j.1574-6968.2001.tb10975.x).
- Perez, F., Regodon, J. A., Valdes, M. E., De Miguel, C., & Ramirez, M. (2000). Cycloheximide resistance as marker for monitoring yeasts in wine fermentations. *Food Microbiology*, 17(2), 119–128. <http://dx.doi.org/10.1006/fmic.1999.0271>.
- Preisner, O., Lopes, J. A., Guimar, R., Machado, J., & Menezes, J. C. (2007). Fourier transform infrared (FT-IR) spectroscopy in bacteriology: Towards a reference method for bacteria discrimination. *Analytical and Bioanalytical Chemistry*, 387(5), 1739–1748. <http://dx.doi.org/10.1007/s00216-006-0851-1>.
- Regodón Mateos, J. A., Pérez-Navado, F., & Ramírez Fernández, M. (2006). Influence of *Saccharomyces cerevisiae* yeast strain on the major volatile compounds of wine. *Enzyme and Microbial Technology*, 40(1), 151–157. <http://dx.doi.org/10.1016/j.enzmictec.2005.10.048>.
- Ribéreau-Gayon, P. (2000). *Handbook of enology. The chemistry of wine stabilization and treatments* (Vol. 2) Chichester [etc.]: Wiley.
- Richter, C. L., Dunn, B., Sherlock, G., & Pugh, T. (2013). Comparative metabolic footprinting of a large number of commercial wine yeast strains in Chardonnay fermentations. *FEMS Yeast Research*, 13(4), 394–410. <http://dx.doi.org/10.1111/1567-1364.12046>.
- Rodríguez-Palero, M. J., Fierro-Risco, J., Codón, A. C., Benítez, T., & Valcárcel, M. J. (2013). Selection of an autochthonous *Saccharomyces* strain starter for alcoholic fermentation of Sherry base wines. *Journal of Industrial Microbiology and Biotechnology*, 40(6), 613–623. <http://dx.doi.org/10.1007/s10295-013-1251-0>.
- Salo, P. (1970). Determining the odor thresholds for some compounds in alcoholic beverages. *Journal of Food Science*, 35, 95–99.
- Schacherer, J., Shapiro, J. A., Ruderfer, D. M., & Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature*, 458(7236), 342–345. <http://dx.doi.org/10.1038/nature07670>.
- Schuller, D., Pereira, L., Alves, H., Cambon, B., Dequin, S., & Casal, M. (2007). Genetic characterization of commercial *Saccharomyces cerevisiae* isolates recovered from vineyard environments. *Yeast*, 24(May), 625–636. <http://dx.doi.org/10.1002/yea.1496>.
- Schuller, D., Valero, E., Dequin, S., & Casal, M. (2004). Survey of molecular methods for the typing of wine yeast strains. *FEMS Microbiology Letters*, 231(1), 19–26. [http://dx.doi.org/10.1016/S0378-1097\(03\)00928-5](http://dx.doi.org/10.1016/S0378-1097(03)00928-5).
- Siebert, T. E., Smyth, H. E., Capone, D. L., Neuwöhner, C., Pardon, K. H., Skouroumounis, G. K., ... Neuwöhner, C. (2005). Stable isotope dilution analysis of wine fermentation products by HS-SPME-GC-MS. *Analytical and Bioanalytical Chemistry*, 381(4), 937–947. <http://dx.doi.org/10.1007/s00216-004-2992-4>.
- Silva Ferreira, A. C., Guedes de Pinho, P., César, A., Ferreira, S., Pinho, P. G., & Ferreira, A. C. S. (2004). Nor-isoprenoids profile during port wine ageing—influence of some technological parameters. *Analytica Chimica Acta*, 513(1), 169–176. <http://dx.doi.org/10.1016/j.aca.2003.12.027>.
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., ... McCusker, J. H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Research*, 1–13. <http://dx.doi.org/10.1101/gr.185538.114>.
- Suárez-Lepe, J. A., & Morata, A. (2012). New trends in yeast selection for winemaking. *Trends in Food Science & Technology*, 23(1), 39–50. <http://dx.doi.org/10.1016/j.tifs.2011.08.005>.
- Swiegers, J. H. H., Bartowsky, E. J. J., Henschke, P. A. A., & Pretorius, I. S. S. (2005). Yeast and bacterial modulation of wine aroma and flavour. *Australian Journal of Grape and Wine Research*, 11(2), 139–173. <http://dx.doi.org/10.1111/j.1755-0238.2005.tb00285.x>.
- Taherzadeh, M. J., Niklasson, C., & Lidn, G. (1997). Acetic acid friend or foe in anaerobic batch conversion of glucose to ethanol by *Saccharomyces cerevisiae*? *Chemical Engineering Science*, 52(15), 2653–2659. [http://dx.doi.org/10.1016/S0009-2509\(97\)00080-8](http://dx.doi.org/10.1016/S0009-2509(97)00080-8).
- Thomas, K. C., Hynes, S. H., & Ingledew, W. M. (2002). Influence of medium buffering capacity on inhibition of *Saccharomyces cerevisiae* growth by acetic and lactic acids. *Applied and Environment Microbiology*, 68(4), 1616–1623. <http://dx.doi.org/10.1128/AEM.68.4.1616>.
- Tielens, A. G., Rotte, C., van Hellemond, J. J., & Martin, W. (2002). Mitochondria as we don't know them. *Trends in Biochemical Sciences*, 27(11), 564–572. [http://dx.doi.org/10.1016/S0968-0004\(02\)02193-X](http://dx.doi.org/10.1016/S0968-0004(02)02193-X).
- Tielens, A. G. M., van Grinsven, K. W. A., Henze, K., van Hellemond, J. J., & Martin, W. (2010). Acetate formation in the energy metabolism of parasitic helminths and protists. *International Journal for Parasitology*, 40(4), 387–397. <http://dx.doi.org/10.1016/j.ijpara.2009.12.006>.
- Via, M., Gignoux, C., & Burchard, E. G. (2010). The 1000 genomes project: New opportunities for research and social challenges. *Genome Medicine*, 2(1), 3. <http://dx.doi.org/10.1186/gm124>.
- Vilanova, M., Genisheva, Z., Masa, A., & Oliveira, J. M. (2010). Correlation between volatile composition and sensory properties in Spanish Albariño wines. *Microchemical Journal*, 95(2), 240–246. <http://dx.doi.org/10.1016/j.microc.2009.12.007>.
- Warringer, J., Zörgö, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., ... Blomberg, A. (2011). Trait variation in yeast is defined by population history. *PLoS Genetics*, 7(6), e1002111. <http://dx.doi.org/10.1371/journal.pgen.1002111>.
- Wilkening, S., Tekkedil, M. M., Lin, G., Fritsch, E. S., Wei, W., Gagneur, J., ... Steinmetz, L. M. (2013). Genotyping 1000 yeast strains by next-generation sequencing. *BMC Genomics*, 14(1), 90. <http://dx.doi.org/10.1186/1471-2164-14-90>.
- Zhang, S., Liu, C., Li, W., Shen, H., Laird, P. W., & Zhou, J. (2012). Discovery of multi-dimensional modules by integrative analysis of cancer genomic data. *Nucleic Acids Research*, 40(19), 9379–9391. <http://dx.doi.org/10.1093/nar/gkn000>.