Universidade do Minho
Escola de Ciências

Inês Isabel Moreira Moutinho Vieira Mendes

Systems biology approaches for the design of novel *Saccharomyces cerevisiae* winemaking strains for enhanced flavour compounds synthesis

janeiro de 2016

Systems biology approaches for the design of novel *Saccharomyces cerevisiae* winemaking strains for enhanced flavour compounds synthesis

Inês Isabel Moreira Moutinho Vieira Mendes

UMinho | 2016

Universidade do Minho
Escola de Ciências

Inês Isabel Moreira Moutinho Vieira Mendes

Systems biology approaches for the design of novel *Saccharomyces cerevisiae* winemaking strains for enhanced flavour compounds synthesis

Tese de Doutoramento em Biologia Ambiental e Molecular

Trabalho realizado sob a orientação da
Professora Doutora Maria João Sousa
e da
Doutora Sylvie Dequin

janeiro de 2016

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration. I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, _____

Full name: _____

Signature: _____

AGRADECIMENTOS
ACKNOWLEDGEMENTS

- ✓ À Professora Dorit Schuller por me ter recebido no seu grupo e me ter dado a oportunidade de realizar este doutoramento;

- ✓ To doctor Sylvie Dequin for having accepted me in her group, and for the opportunity to work in a different laboratory on a different country. Thanks for the warm welcome in Montpellier;

- ✓ À Professora Maria João Sousa que aceitou o desafio de me orientar no último ano do doutoramento e que muito contribuiu para o enriquecimento deste trabalho;

- ✓ To Carole Camarasa, Isabelle Sanchez, Jean-Roch Mouret, Jean Marie Sablayrolles, Christian Picou, Marc Perez, Audrey Bloem, Jean-Luc Legras, Pascale Brial, Magaly Angenieux, Phillipe and Evelyne for all the help in Montpellier and Pech Rouge;

- ✓ Ao CBMA e ao departamento de Biologia da Universidade do Minho por me possibilitar a realização deste trabalho;

- ✓ Aos meus colegas de laboratório pela amizade e companheirismo ao longo desta jornada o meu muito obrigado ao João, Geninha, Daniela, Nuno, Raquel, Ana, Filipa G, Filipa P., Ângela, Marlene, Gabriel, Flávio;

- ✓ To Annalisa Coi and Souhir Marsit for the friendship during my stay in Montpellier;

- ✓ A todos os técnicos e secretários porque, sem eles este trabalho não seria possível, nomeadamente: Amaro, Cristina, Isabel, Magda, Manuela C., Manuela T, Sofia, Paula e Liliana;

- ✓ À FCT pelo financiamento deste projeto de doutoramento;

- ✓ À minha família porque sempre me apoiou e sempre acreditou que eu conseguiria;

- ✓ Ao meu maridão Ricardo porque sem ele isto teria sido simplesmente impossível.

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

Abstract

Wine flavour and aroma are the result of yeast metabolism and must compounds interactions. During must fermentation thousands of volatile aroma compounds are formed, with higher alcohols, acetate esters and ethyl esters being the main aromatic compounds contributing to a floral and fruity aroma. The action of yeast, in particular of *Saccharomyces cerevisiae* strains, on the must components will build the architecture of the wine flavour and their fermentation bouquet. Only a holistic approach based on systems biology, an inter-disciplinary area combining the study of biology, chemistry, physics and mathematics, based on data-driven and model-based elucidation of complex biological systems, will allow the complete understanding on the vast and dynamical relationships between genomics, phenomics and metabolomics.

In this work a *S. cerevisiae* collection was constituded comprising 172 strains of worldwide geographical origins and different technological applications. This collection was characterized regarding 30 physiological traits that are important mostly from an oenological point of view. From the different traits studied, growth in the presence of potassium bisulphite, growth at 40 ºC, and resistance to ethanol contributed the most to strain variability, as shown by the principal component analysis. Computational methods were developed in order to assess the importance of phenotypic features to identify candidate strains to be used commercially in winemaking. In particular, the probability of a strain to be assigned to the group of commercial strains was 27% using the entire phenotypic profile and increased to 95%, when only results from the three tests suggested by the model were considered. Results show the usefulness of the mentioned approaches to simplify strain selection procedures.

In addition to the phenotypic characterization, we undertook genetic typing of the 172 *S. cerevisiae* strains, using 11 polymorphic microsatellites. We found 280 alleles, whereas microsatellite ScAAT1 contributed the most to intra-strain variability, together with the alleles 20, 9 and 16, from microsatellites ScAAT4, ScAAT5 and ScAAT6, respectively. These microsatellite allelic profiles are characteristic for both the phenotype and origin of yeast strains. Data were computationally related with the previously obtained results of the 30 phenotypic tests, and the phenotypes associated with higher number of alleles were the capacity to resist to sulphur dioxide (tested by the capacity to grow in the presence of potassium bisulphite) and the presence of galactosidase activity.

Our study demonstrated the utility of computational modelling to estimate a strain technological group and phenotype from microsatellite allelic combinations as tools for preliminary yeast strain selection.

To better understand the molecular and metabolic bases of aroma production during a fermentation process, we used comparative transcriptomic and metabolic analysis of four yeast strains from different origins and/or technological applications (cachaça, sake, wine, and laboratory), to rationally identify new targets for improving aroma production. Results showed that strains from cachaça, sake and wine presented a higher production of acetate esters, ethyl esters, acids and higher alcohols, in comparison with the laboratory strain S288c. At fermentation time T1 (5 g/L of CO_2 released), comparative transcriptomics of these three *S. cerevisiae* strains from different fermentative environments in comparison with the laboratory yeast S288c, showed an increased expression of genes related with tetracyclic and pentacyclic triterpenes metabolism, involved in sterol synthesis. Sake strain showed also an upregulation of genes *ADH7* and *AAD6*, involved in the formation of higher alcohols in the Ehrlich pathway. For fermentation time point T2 (50 g/L CO_2 released), again sake strain, but also wine strain, showed an increased expression of genes involved in formation of higher alcohols in the Ehrlich pathway, namely *ADH7*, *ADH6* and *AAD6*, which is in accordance with the higher levels of methionol, isobutanol, isoamyl alcohol and phenylethanol observed. Our approach revealed successful to integrate data from several technologies (HPLC, GC-MS and microarrays) and using different data analysis methods (PCA, MFA). The results obtained, increased our knowledge on the association of genes with the formation of metabolic compounds that contribute to the wine aroma and flavour, and showed differences in the metabolism of cachaça, sake and wine strains not yet addressed, and mainly explained by the production of fatty acids, and ethyl and acetate esters.

Resumo

O aroma e o sabor do vinho resultam das interações entre o metabolismo das leveduras e os compostos presentes no mosto. Durante a fermentação do mosto formam-se milhares de compostos voláteis, sendo que os álcoois superiores, os esteres de acetato e os esteres de etilo são os compostos que mais contribuem para o aroma floral e frutado. A ação das leveduras, em particular das estirpes de *Saccharomyces cerevisiae,* nos componentes do mosto contribui para a arquitetura do cheiro do vinho e do seu *bouquet* fermentativo. Apenas uma visão holística baseada na biologia de sistemas, uma área interdisciplinar que combina o estudo da biologia, química, física e matemática, baseada na elucidação de dados e modelos de sistemas biológicos complexos, permitirá uma completa compreensão das vastas e dinâmicas relações entre genómica, fenómica e metabolómica.

No presente trabalho uma coleção de 172 estirpes de *S. cerevisiae* foi constituída com isolados provenientes de diferentes origens geográficas e diferentes aplicações tecnológicas. O seu fenótipo foi avaliado considerando 30 testes fenotípicos, selecionados de forma a avaliar características importantes do ponto de vista enológico. O crescimento na presença de bissulfito de potássio, crescimento a 40 ˚C e resistência ao etanol foram os fenótipos que mais contribuíram para a variabilidade entre estirpes, de acordo com o revelado pela análise de componentes principais. Foram desenvolvidos métodos computacionais com o objetivo de avaliar a importância das características fenotípicas na identificação de estirpes candidatas para serem usadas comercialmente na produção de vinho. Em particular, a probabilidade de uma estirpe pertencer ao grupo das estirpes comerciais foi de 27% usando o perfil fenotípico completo e aumentou para 95% quando foram apenas considerados os 3 testes sugeridos pelo modelo. Os resultados mostraram a utilidade dos métodos mencionados para simplificar os procedimentos de seleção de estirpes.

Além da caracterização fenotípica, fez-se a tipagem genética das 172 estirpes de *S. cerevisiae,* usando para isso 11 microssatélites polimórficos. A caracterização genética identificou 280 alelos, sendo o microssatélite ScAAT1 o que mais contribuiu para a variabilidade entre estirpes, em conjunto com os alelos 20, 9 e 16 dos microssatélites ScAAT4, ScAAT5 e ScAAT6, respetivamente. Este perfil alélico de microssatélites revelou-se como característico tanto para o fenótipo como para a origem da estirpe de levedura. Estes dados foram relacionados computacionalmente com os resultados obtidos

previamente para os 30 testes fenotípicos e os fenótipos associados com o maior número de alelos foram a capacidade de resistir ao dióxido de enxofre (testado pela capacidade de crescer na presença de bissulfito de potássio) e a presença de atividade da galactosidase. Este estudo demonstrou a utilidade da modelação computacional para fazer uma estimativa do grupo tecnológico de uma estirpe e do seu fenótipo a partir de combinações alélicas de microssatélites como ferramentas para seleção preliminar de estirpes de leveduras.

De forma a contribuir para uma melhor compreensão das bases moleculares e metabólicas da produção de aroma durante o processo fermentativo, fez-se transcritómica comparativa e análise metabólica de quatro estirpes de levedura com diferentes origens e /ou aplicações tecnológicas (cachaça, saké, vinho e laboratório), para identificar novos alvos para o melhoramento da produção de aroma. Os resultados mostraram que as estirpes de cachaça, saké e vinho apresentaram uma maior produção de esteres de acetato, esteres de etilo e álcoois superiores, em comparação com a estirpe laboratorial S288c. No tempo de fermentação T1 (5 g/L de CO_2 libertado), a transcritómica comparativa das três estirpes de *S. cerevisiae* obtidas em diferentes ambientes fermentativos em comparação com a estirpe laboratorial S288c, mostrou um aumento da expressão de genes relacionados com o metabolismo dos triterpenos tetracíclicos e pentacíclicos, envolvidos na síntese de esteróis. A estirpe de saké mostrou também um aumento na regulação dos genes *ADH7* e *AAD6*, envolvidos na formação de álcoois superiores na via de Ehrlich. Para o tempo de fermentação T2 (50 g/L de CO_2 libertado), mais uma vez a estirpe de saké, mas também a estirpe vínica, mostraram um aumento da expressão de genes envolvidos na formação de álcoois superiores na via de Ehrlich, nomeadamente *ADH7*, *ADH6* e *AAD6*, o que está de acordo com elevados níveis de metionol, isobutanol, álcool isoamilico e feniletanol observados. A nossa abordagem revelou-se bem-sucedida para integrar dados de diferentes plataformas (HPLC, GC-MS e microarrays) e usando diferentes métodos de análise de dados (PCA, MFA). Os resultados obtidos aumentaram o conhecimento sobre a associação de genes com vias de formação de compostos que contribuem para o aroma e sabor do vinho, e mostraram diferenças no metabolismo das estirpes de cachaça, saké e vinho que não tinham ainda sido investigadas, e principalmente explicadas pela produção de ácidos gordos, e esteres de etilo e de acetato.

Table of contents

List of abbreviations, acronyms and initialisms

2-D	two-dimensional
2-DE	two-dimensional gel electrophoresis
3-MH	3-mercaptohexane-1-ol
3-MHA	3-mercaptohexyl acetate
4-MMP	4-mercapto-4-methyl-pentan-2-one
$A_{640\ nm}$	absorbance (optical density) measured at the wavelength of 640 nm
APCI	atmospheric pressure chemical ionization
AUC	area under the ROC curve
BCAA	branched-chain amino acid
bp	base pairs
CE	capillary electrophoresis
CoA	coenzyme A
cont.	continuation
DIMS	direct injection mass spectrometry
DMS	dimethyl sulfide
DNA	deoxyribonucleic acid
e.g.	for example (*exempli gratia)*
ESI	electrospray ionization
FAME	fatty acid methyl ester
FDR	false discovery rate
GC	gas chromatography
GEO	gene expression omnibus
GMY	genetically modified yeast
GO	gene ontology
H	hour
HPLC	high performance liquid chromatography
i.e.	that is (*is est)*

IGR	information gain ratio
KNN	*k*-nearest neighbor
LC	liquid chromatography
LM	liquid medium
MEA	malt extract agar
MAPK	mitogen-activated protein kinase
MCFA	medium-chain fatty acid
MFA	multivariate factorial analysis
min.	minutes
MLST	multilocus sequence typing
MS	mass spectroscopy
MS300	synthetic must
mtDNA	mitochondrial DNA
NAD^+	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide (reduced form)
NADPH	nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance
PC	principal component
PC-1	first principal component
PC-2	second principal component
PCA	principal component analysis
PCR	polymerase chain reaction
QTL	quantitative trait locus
RAPD	random amplified polymorphic DNA
rDNA	ribosomal DNA
Ref.	reference
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid

mRNA	messenger RNA
tRNA	transfer RNA
ROC	receiver operating characteristics
RP	reverse phase
rpm	revolutions per minute
SCX	strong cation exchange
SDS	sodium dodecyl sulphate
SM	solid medium
SNP	single nucleotide polymorphism
sPLS-DA	sparse partial least square - discriminant analysis
SSR	single sequence repeats
TA	titratable acidity
TCA	tricarboxylic acid
TOF	time of flight
Ty	transposable element of yeasts
UPGMA	unweighted pair group method with arithmetic mean
v/v	volume / volume
w/v	weight / volume
YNB	yeast nitrogen base
YPD	yeast extract-peptone-dextrose

List of figures

List of tables

Chapter I

Motivation, objectives and outline

Motivation

The market value of products derived from fermentations with *Saccharomyces cerevisiae* such as wine, is expected to increase much above the general market growth in the future. Winemaking represents today a multi-billion Euro industry that could benefit tremendously from system biology research, owing to the direct impact of yeast on several key aspect of wine quality, such as for example the production of flavour compounds mainly volatile acids, higher alcohols, esters, volatile thiols and phenols. The continuous utilization of yeasts for industrial purposes introduced an artificial pressure on the strains selected that may have also influenced genome features and novel specialization metabolic routes.

S. cerevisiae was the first sequenced eukaryotic organism, which provided a vast amount of knowledge on its molecular and cellular biology. For this reason this yeast is considered the model organism *par excellence* for genetics and genomics studies. However, a major improvement is needed regarding the understanding of the variability existing between *S. cerevisiae* strains, which will only be obtained through the integrated study of several "omic" approaches.

Objectives

In general, this thesis aims to explore the genetic and phenotypic variability that nature has created, in order to elucidate about the genetic bases that shaped inter-strain differences of flavour compound formation across wine strains of *S. cerevisiae*. An integrative approach was followed throughout the thesis, englobing genomic, transcriptomic, bioanalytical and bioinformatic approaches.

The detailed objectives of the work were:

- To constitute a genetically diverse *S. cerevisiae* strain collection, comprising isolates from differential technological applications or origins;

- To provide detailed phenotypic and genetic characterization of all strains using high-throughput methods;

- To conduct computational analysis and develop models that predict associations between a strain's phenotype and genotype;

- To validate the developed models and identify a small sub-group of *S. cerevisiae* strains with most desirable oenological phenotypic traits;

- To identify networks of genes with optimized expression patterns in selected strains, establish their relations with the aroma-relevant metabolome fraction through combined transcriptomic and metabolomic approaches and incorporate gene expression levels and aroma compound production into multivariate statistical analysis.

Thesis outline

- **Chapter I** presents the motivation that lead to the current work, together with its objectives, as well as global structure.

- In **Chapter II** an overview of the state-of-the-art is given, in particular on the use of *S. cerevisiae* in winemaking and as eukaryotic model, on the phenotypic plasticity of wine strains and the methods used for yeast genetic characterization, and about the importance of Systems Biology in modern winemaking. A special focus is given to the relation between *S. cerevisiae* and wine flavour.

- **Chapter III** comprehends the constitution of our strain collection, comprising 172 isolates of *S. cerevisiae* from different geographical and technological origins. A high-throughput method for phenotypic screen was devised to characterize all strains, and obtain a global view of the phenotypic diversity of the collection.

- **Chapter IV** focus on the genetic characterization of the *S. cerevisiae* isolates using a set of 11 highly polymorphic *S. cerevisiae* specific microsatellite loci. High genetic

variability was obtained and a set of most informative microsatellites was suggested. Computationally methods allowed associating strains phenotype with genotype in order to choose a small sub-set of 4 strains to be used in the next chapter, regarding transcriptomic and metabolic analysis.

- **Chapter V** embraces the transcriptomic and metabolomic analysis of four strains from different technological applications. Using different data analysis methods we successfully integrated data from different technologies (HPLC, GC-MS and microarrays) and characterized the strains aroma profile together with the parallel gene expression dynamics, which allowed to improve the knowledge of the association of some genes with the formation of metabolic compounds.

- In **Chapter VI** the global conclusions of the work are presented together with suggestions for future work.

- **Chapter VII** lists the bibliographical references cited along the thesis.

- **Chapters VIII** presents, as supporting material, supplementary data not shown in the other chapters, also as the pdf versions of the manuscripts already published.

Chapter II

General introduction

Saccharomyces cerevisiae and its use in winemaking

The production of wine is an ancestral process involving interactions between microorganisms and grape must (Pretorius 2000; Cavalieri et al. 2003). One of the first evidences of these interactions were the discovery of *Saccharomyces cerevisiae* isolates in Egypt, in the residues of the earliest known wine jars (Cavalieri et al. 2003).

Wine is a natural product that results from different biochemical reactions provided by microorganisms present on the surface of the grapes, and can be defined as multicomponent liquid solution containing water, ethanol, glycerol and organic acids as major components and other minor components such as phenolic compounds that contributes to the flavour and aroma of wine (Pizarro et al. 2007). Originally, wine was made by spontaneous fermentation taking advantage of the yeast strains present in the internal flora (Lambrechts and Pretorius 2000a).

In 1890, Müller-Thurgau introduced the concept of inoculating wine fermentations with pure yeast cultures and, as a result, the quality and quantity of wine production were vastly improved in the 20th century (Pretorius 2000). Nowadays, most wine fermentations are carried out by inoculating grape must with a pure yeast culture. Culture selection depends on the grape cultivar, must composition, general conditions of fermentation and on the final product required. From a biotechnological point of view, the use of active dry yeast has a high impact on the microbiology of the fermentation process. Unlike the natural process, the addition of pure cultures induces a clear predominance of the inoculated strain over the endogenous strains (Perez-Gonzalez et al. 1993; Beltran et al. 2002). In large-scale wine production, where rapid and reliable fermentations are essential for consistent wine flavour and predictable quality, the use of selected pure yeast inoculum of known properties is preferred (Pretorius 2000).

Industrial yeast strains of *S. cerevisiae* differ from most laboratory strains, which are either haploid or diploid. Wine industrial strains are predominantly diploid, aneuploid and occasionally polyploid (Bakalinsky and Snow 1990; Barre et al. 1993), and show a high level of chromosome length polymorphism (Bidenne et al. 1992; Rachidi et al. 1999). For several decades winemakers have been combining strains of *S. cerevisiae* on the basis of observations that mixed-cultures produce more flavour diversity and balanced wines, by introducing a greater range of flavour notes and moderating the intensity of

distinctive estery/fruity notes (Ugliano and Henschke 2009). The differences in yeast growth patterns of genetically distinct strains, that are observed when they are cultivated individually or in mixed cultures, suggest that metabolic interactions can occur between different yeasts and can be antagonistic, neutral or stimulatory (Ugliano and Henschke 2009).

The chemical composition of the wine is also dependent on the type and quality of the grapes. The primarily aim of viticultural practices is producing high quality grapes that would reflect varietal flavours and aromas and/or characters typical for a specific region or *terroir* (Styger et al. 2011b).

Some wines can undergo a secondary fermentation known as malolactic fermentation. This process is particularly desirable for high-acid wine produced in cool climate regions, as malolatic fermentation involves the deacidification of wine via the conversion of the dicarboxylic L-malic acid to the monocarboxylic L-lactic acid and carbon dioxide (Styger et al. 2011b). This process is normally carried out by lactic acid bacteria, including *Oenococcus oeni*, *Lactobacillus spp*., *Leuconostoc spp.*, and *Pediococcus spp.* (Liu 2002). In some wines from warmer regions, malolatic fermentation is also important because it changes the composition of the wine and improves its organoleptic quality (Styger et al. 2011b).

The composition of grape must offers culture conditions that are far from optimal for most microorganisms and upon inoculation yeast cells must adapt to the low pH (2.9–3.8) and high osmolarity of the new environment (sugar concentration up to 300 g/L), as well as to the high SO_2 content (40–100 mg/L) (Pizarro et al. 2007). This adaptation may change the metabolite production and as consequence will influence wine quality.

Phenotypic plasticity of *Saccharomyces cerevisiae* wine strains

Worldwide, wine producers use commercial starter yeasts to ensure the reproducibility and the predictability of wine quality. The advantages of fermentations containing *S. cerevisiae* starter cultures relies on the fact that they are rapid and produce wine with desirable organoleptic characteristics through successive processes and harvests (Fleet 1998; Schuller 2010). To select the best starter yeast one should take into account the style of the wine and/or the grape variety. To perform strain selection, certain oenological criteria are typically used, such as technological (influencing the efficiency of the fermentation process) or qualitative criteria (affecting the chemical composition and the sensorial profile of wine) (Zambonelli 1998). The most relevant physiological tests used to select *S. cerevisiae* strains were gathered by Schuller (2010) and are summarized in the **Table II-1**.

Table II-1: Oenological parameters considered for the selection of *Saccharomyces cerevisiae* wine strains.

Oenological parameters	Comments
Fermentation vigor	Maximum amount of ethanol (%, v/v) produced at the end of the fermentation Desirable: good ethanol production
Fermentation rate	Grams of CO_2 produced during the first 48 hours of fermentation Desirable: prompt fermentation initiation
Fermentation temperature	Thermotolerance and cryotolerance is related to oenological properties Optimum fermentation temperature ranges between 18 and 28°C
Mode of growth in liquid medium	Dispersed or flocculent growth, sedimentation speed Desirable: dispersed yeast growth during, but sedimentation at the end of fermentation
Foam production	Height of foam produced during fermentation Desirable: low foam production
Volatile acidity, acetic acid production	Selected strains should not release more than 100 - 400 mg/L during fermentation Desirable: low volatile acidity /acetic acid production
Malic acid degradation or production	Whether degradation or production is desirable depends on the characteristics of the must. Malic acid degradation varies between 0- 20% of the initial concentration, depending on the *S. cerevisiae* strain
Glycerol production	Desirable major fermentation by-product (5-8 g/L) contributing to wine sweetness, body and fullness
Acetaldehyde production	Desirable metabolite in sherry, dessert and port wines being an important character for selection of strains to be applied in wine ageing
Esters, higher alcohols and volatile compounds	Desirable metabolites, markedly influence wine flavour and depend on the presence of precursors related to both grape cultivar and grape maturity. Limited amounts contribute positively to global sensorial characteristics
SO_2 tolerance and production	Antioxidant and antimicrobial agent Desirable: high fermentation vigor and rate in the presence of SO_2 concentrations usually applied in winemaking; Undesirable: excessive SO_2 production
H_2S production	H_2S is detrimental to wine quality, considered as off-flavour with very low threshold value (50- 80 µg/L)
Stress resistance	Tolerance to combined acid/osmotic stress
Copper resistance	High copper concentrations may cause stuck fermentations Desirable: high copper resistance and the ability to reduce the copper content

Beside these criteria, it should also be taken into account that starter yeasts may interact with must microflora. These interactions can be positive or negative and may affect the quality of the final wine (Mannazzu et al. 2002). Certain yeasts belonging to different genera and species have the capacity to produce killer toxins (proteins and glycoproteins) that exert a microbiocidal effect on the other yeasts. The presence of this killer phenotype in *S. cerevisiae* yeasts can have high importance once it can ensure the predominance of the selected starter yeast instead of the sensitive must microflora (Mannazzu et al. 2002).

Some studies suggested the natural occurrence of *Saccharomyces* hybrid strains, such as the triple hybrid *S. cerevisiae,* x *S. bayanus* x *S. kudriavzevii* (Gonzalez et al. 2007). It was also found that *S. cerevisiae* x *S. kudriavzevii* hybrids were better adapted to fermentations carried out at lower temperatures (14-22 °C) and produced less acetic acid and more higher alcohols (Gonzalez et al. 2007). These findings have shown an evolutionary adaptation of the strains and an improvement on technological properties of the wild yeasts.

The genetic and metabolic basis responsible for the phenotype diversity among *S. cerevisiae* strains remains unclear, as they are incompletely characterized and also because some phenotypes are associated with several genes which increases the complexity of these relations. Although some efforts were made in order to complete this identification, only specific physiological parameters were characterized, such as ethanol resistance (Hu et al. 2007), temperature tolerance (McCusker et al. 1994; Steinmetz et al. 2002; Sinha et al. 2006), drug responses (Perlstein et al. 2006; Perlstein et al. 2007; Kim et al. 2009), sporulation efficiency (Primig et al. 2000; Gerke et al. 2006; Ben-Ari et al. 2006; Magwene et al. 2011) and morphology (Nogami et al. 2007).

The emergence of “omics” approaches and high-throughput methods has facilitated the analysis and comparison of the different phenotypes among strains and the establishment of relations with the genotype.

Methods for the genetic characterization of yeast strains

The study of *S. cerevisiae* genetic diversity has been exploited for many years ago, and more recently specifically for the understanding of phenotype-genotype relations. Many methods were used for intra-strain genetic characterization in different interest areas such as strain selection for their use as pure cultures (Dequin 2001; Cocolin et al. 2004), monitorization of population dynamics during fermentation of food and beverages (Granchi et al. 1999; Nadal et al. 1999; Pulvirenti et al. 2001; Granchi et al. 2003) and characterization of clinical isolates (Zerva et al. 1996; McCullough et al. 1998).

The characterization of wine yeast genotypes was initially done using traditional tools, which analysed mainly the spore segregation. With the emergence of the molecular techniques this characterization became easier and more accurate (Barnett 1992; Naumov et al. 1993). Generally, the redefinition of taxonomic grouping has become possible by analysing the similarity of DNA, RNA or cellular proteins. In order to differentiate the genera and identify species, several methods were used such as comparison of the entire yeast genome, pulsed field electrophoresis for the separation of the entire chromosomes, study of ribosomal DNA (rDNA) and mitochondrial DNA (mtDNA) by PCR and subsequent analysis of restriction fragments (Giudici and Pulvirenti 2002), which are described in the paragraphs below.

During many years chromosome pulsed-field gel electrophoresis was the method commonly used to separate DNA molecules and to analyse variations in yeast genomes. This technique was primarily applied to yeast genomes (Carle and Olson 1985) and consists in the electrophoretic separation of chromosomal DNA molecules followed by the identification of the bands by DNA-DNA hybridization. This method has shown a significant variability in the chromosomal constitution of commercial yeasts (Blondin and Vezinhet 1988) and became an advantageous method for strain identification (Degré et al. 1989; Vezinhet et al. 1990; Yamamoto et al. 1991; Querol et al. 1992; Guillamón et al. 1996; Fernández-Espinar et al. 2001; Schuller et al. 2004).

Another commonly used method to distinguish and characterize *S. cerevisiae* strains is the restriction fragment length polymorphism (RFLP) analysis of mitochondrial DNA (Dubourdieu et al. 1984; Lee and Knudsen 1985; Vezinhet et al. 1990). This method consists in the digestion of mtDNA with restriction enzymes creating high polymorphism,

due to mtDNA being very variable between species and strains, both in size and organization, and having both highly conserved species specific regions, but also other regions that evolve 10 times more rapidly than nuclear DNA (Vezinhet et al. 1990; Querol et al. 1992; Guillamón et al. 1996; Fernández-Espinar et al. 2001; López et al. 2001; Martínez et al. 2004; Schuller et al. 2004).

With the progress of polymerase chain reaction (PCR), *S. cerevisiae* strains were discriminated using quicker methods, based on the detection of polymorphisms in DNA fragment sizes or specific banding patterns, without the need of using restriction enzymes. These techniques are based on the use of oligonucleotides as primers, which bind to target sequences in each yeast DNA strand. One example is random amplified polymorphic DNA (RAPD) which is a technique characterized by the use of just one short primer (about ten nucleotides) with an arbitrary sequence and a low annealing temperature (37 ºC) during PCR (Williams et al. 1990). This allows the amplification of diverse fragments of DNA distributed along the genome and results in a pattern of amplified PCR products of different molecular weights, characteristic of each strain (Bruns et al. 1991; Paffetti et al. 1995). This method has the advantage that no information is necessary about the DNA sequence in order to design the primer. Conversely, because it depends on an intact DNA template sequence, it has some limitations when using degraded DNA samples in the amplification. RAPD has been applied with success in several projects regarding yeast strains characterization (Baleiras Couto et al. 1995; Quesada and Cenis 1995; Romano et al. 1996; Tornai-Lehoczki and Dlauchy 2000; Pérez et al. 2001; Cadez et al. 2002).

Multi-locus sequence typing (MLST) is another PCR-based method used for genetic characterization of yeast isolates using DNA sequences of internal fragments (450-500 bp) of multiple housekeeping genes. After PCR and DNA sequencing, strains are characterized by their unique allelic profile. MLST is a technique used in molecular biology for the typing of multiple loci. MLST was used in the past for identification of bacterial pathogens and then to the analysis of *S. cerevisiae* isolates (Ayoub et al. 2006). The main advantage of this method is related with a higher reliability in comparison with the electrophoretic methods allowing a high-throughput data debit and an easy sharing of results between laboratories. However, it involves high costs and the discriminatory power is affected by the fact that yeast housekeeping genes are extremely conserved.

Interdelta sequences typing is also a useful tool in genetic characterization of yeasts. Retrotransposons Ty1 and Ty2 are flanked by delta sequences (300 bp) (Cameron et al. 1979). These delta sequences are found in terminal chromosomal regions, but can occur also as single elements dispersed all over the genome. There are near 300 delta elements described in the genome of the laboratory strain S288c. In 1993, Ness and co-workers (Ness et al. 1993) proposed a protocol based on the amplification of interdelta regions by PCR, once the number and location of the delta elements have certain intraspecific variability. The primers are designed to amplify DNA regions between neighbouring delta sequences and the PCR reaction produces a mixture of differently sized fragments, specific for each strain. This technique was then optimized by designing two new primers (δ12 and δ21) that hybridize very close to the binding sites of primers δ1 and δ2, which were initially referred by Ness (Legras and Karst 2003). The use of primers δ12 and δ21 or δ12 with δ2 reveals greater polymorphism, with the appearance of a higher number of bands, resulting in a higher discriminatory power. It was shown that the combination of δ12 with δ2 was able to distinguish twice the number of strains that were discriminated by the initial primer pair δ1 and δ2 (Schuller et al. 2004). Also, Fernández-Espinar and co-workers have shown that this method requires standardization of DNA concentration. Due to the low annealing temperature (43 ˚C), "ghost bands" may be present which is another disadvantage (Fernández-Espinar et al. 2001). This can be fixed by increasing the annealing temperature to 55 ºC which reduces the number of "ghost bands", but also reduces the total number of bands obtained, and consequently the discriminatory power (Ciani et al. 2004). Analysis of PCR profiles obtained by interdelta sequences amplification were associated with a good discriminating power for the analysis of commercial strains (Lavallée et al. 1994). More recently, however, some questions have been raised regarding reproducibility between laboratories and also the influence of the DNA concentration in the electrophoretic profile obtained (Franco-Duarte et al. 2011). Even though with these limitations, this technique continues to be widely used in the present to characterize yeast strains (Pramateftaki et al. 2000; Lopes et al. 2002; Cappello et al. 2004; Ciani et al. 2004; Demuyter et al. 2004; Pulvirenti et al. 2004; Xufre et al. 2011; Bleykasten-grosshans et al. 2013).

Microsatellites or single sequence repeats (SSR) are short DNA sequences that have been shown to display a large level of size polymorphism in several eukaryotic

genomes (Richard et al. 1999), presenting also a high amount of intra-species variation. PCR amplification of these regions is a method highly discriminative for the molecular typing of indigenous *S. cerevisiae* populations (Pérez et al. 2001; Schuller et al. 2004; Schuller et al. 2005; Schuller and Casal 2005). High polymorphism, co-dominant inheritance, selective neutrality, amplification by PCR-based methods and high reproducibility are characteristics of microsatellites that make them good genetic markers. In 2001, a set of six polymorphic microsatellite loci (ScAAT1, ScAAT2, ScAAT3, ScAAT4, ScAAT5 and ScAAT6) was selected from 51 strains originated from spontaneous fermentations and that generated 44 genotypes with a total of 57 alleles (Pérez et al. 2001). It was also referred the simplicity of this technique, allowing multiplex PCR reaction in reproducible and precise way. Legras and co-workers described in 2005 another set of microsatellite loci for *S. cerevisiae* strains typing, which includes the highly polymorphic loci ScYOR267c, C4, C5, C11 and ScYPL009c (Legras et al. 2005). The use of microsatellites as a tool for identification purposes and as source of variability was extrapolated to other species, such as *Candida albicans* (Sampaio et al. 2003; Garcia-Hermoso et al. 2010), *C. parapsilosis* (Sabino et al. 2010), *C. glabrata* (Foulet et al. 2005), *S. bayanus* (Masneuf-Pomarède et al. 2007), and also with clinical applications (Correia et al. 2004; Vaz et al. 2011) and for evolutionary studies (Sampaio et al. 2005). Also, in several other fields the use of microsatellites as genetic markers has been extensively exploited, such as in paternity analyses (Jobling et al. 1997; Dow and Ashley 1998), in construction of genetic maps (Dib et al. 1996), in population genetic studies (Tautz 1989; Estoup et al. 2002) and human diseases research (Desselle et al. 2012; Manasatienkij and Rangabpai 2012; Buecher et al. 2013; Heinimann 2013; Choi et al. 2015).

The importance of Systems Biology in modern winemaking

Systems Biology is an inter-disciplinary area combining the study of biology, chemistry, physics and mathematics, based on data-driven and model-based elucidation of complex biological systems. The term systems biology has been widely used from year 2000 onwards in the biosciences and in a variety of contexts. Systems biology approaches are mainly focused on the systematic study of entire pathways, processes, or interactions of complex biological systems, and promote a holistic view of the organism under study. The vast majority of the projects based on this subject were initiated in yeasts due to the size and the availability of the genome, and also to the easiness of manipulation and availability of tools. The objective of this area is the understanding of the genotype–phenotype relationships as well as the elucidation of the principles and mechanisms governing the behaviour of biological systems (Kitano 2002; Stelling 2004).

The large genome-scale study of relations between phenotypes and their molecular structures in genetics and protein interactions is defined by phenomics (Schork 1997; Freimer and Sabatti 2003; Fernandez-Ricaud et al. 2007; Houle et al. 2010). In earlier stage phenomics studied only a few number of phenotypes (Warringer et al. 2003; Kvitek et al. 2008; Ratnakumar et al. 2011; Chen et al. 2012). Nevertheless, with evolution in instrumentation, technology and computational approaches, phenotypic studies became possible in large sets of samples. Organisms that have to adapt to changing environmental conditions are difficult to be studied by phenomics because of their phenotypic plasticity which is characterized by the capacity of a genotype to reveal different phenotypes in different environments (Pigliucci 2001). Some studies have identified genes responsible for affecting plasticity and how they interact with each other (Mackay 2001; Remold and Lenski 2004; Kent et al. 2009). The phenotypic variation in yeast could be qualitative or quantitative. Qualitative traits obey to the Mendelian laws and are controlled by a single locus with a small effect. On the other hand, quantitative traits include a continuous distribution of a measurable character such as stress tolerance (heat and ethanol for example) (Hu et al. 2007; Parts et al. 2011). The majority of studies in yeast phenomics are based on quantitative traits controlled by multiple genetic loci and are called quantitative trait loci – QTL (Lander and Botstein 1989; Lynch and Walsh 1998). A QTL is constituted by a cluster of closely linked genes that contribute to the

quantitative trait (Mackay 2001). *S. cerevisiae* provides an ideal model for QTL analysis due to high recombination rate, richly annotated genome and the fact that genes can be directly manipulated in their genomic context. QTL mapping was used to understand several mechanisms in yeasts such as thermotolerance analysis (Steinmetz et al. 2002; Sinha et al. 2006), sporulation efficiency (Deutschbauer and Davis 2005) and drug resistance (Perlstein et al. 2007). This analysis also helped to understand genotype-phenotype relations in wine (Marullo et al. 2007; Marullo et al. 2009; Ambroset et al. 2011; Steyer et al. 2012; Brion et al. 2014) and sake (Katou et al. 2009) fermentations, and ethanol production (Hu et al. 2007).

The entire DNA nucleotide sequence content, including all genes of an organism constitutes the genome (Borneman et al. 2007). Since 1996, the *S. cerevisiae* genome is known (Goffeau et al. 1996), and it is constituted by approximately 6000 genes and 16 chromosomes. Liti and cowokers (Liti et al. 2009) sequenced several *S. cerevisiae* strains which included the reference strain S288c and also several others such as pathogenic, baking, wine, food spoilage, natural fermentations and sake strains. In this work, 235,127 high-quality single nucleotide polymorphisms (SNPs) and 14,051 nucleotide insertions or deletion were identified in the *S. cerevisiae* nuclear genome. In another work from 2008, important genomic variability was also identified between 16 yeast strains (laboratory, commercial and opportunistic human infections origins) in particular in subtelomeric regions and in Ty-element insertion sites, suggesting that this type of genome variability is the main source of genetic diversity in natural populations of yeast (Carreto et al. 2008). It was further shown that wine strains acquired large genomic regions from non-*Saccharomyces* species through horizontal gene transfer, conferring in this way molecular adaptation to conditions of high sugar, low nitrogen, and high ethanol concentrations (Novo et al. 2009). In 2008, The Australian Wine Research Institute completed the genome sequencing of the commercial yeast AWRI1631 and it was shown that this strain was substantially different from the laboratory strain (S288c), especially at the level of several single nucleotide polymorphisms (SNPs) throughout the genome (Borneman et al. 2008). Comparative genomics will become a major tool for the insightful interpretation of genomic data within the winemaking context. Until September 2011 about 74 *S. cerevisiae* strains have been sequenced, 17 of them being wine yeast strains. Several novel sequencing technologies, such as nanopore and pyrosequencing are being

optimized, as well as software tools and algorithms for automated genome annotation, together aspiring to reduce costs and time frames for genome analysis. Consequently, in the near future, many more wine yeast genomes will be sequenced and become available.

The transcriptome includes the study of the entire cellular RNA complements (mRNA, tRNA and other RNA types). The RNA population is very dynamic, being dependent both on gene regulation in response to prevailing environmental conditions, but also on variation in RNA turnover (Borneman et al. 2007). Transcriptomic approaches usually require several convergent technologies, such as DNA sequencing and amplification, synthesis of oligonucleotides, fluorescence biochemistry, and computational statistics (Rossouw and Bauer 2009b). Gene expression is variable among wild-type yeast strains and it was shown that differences in gene expression during fermentation affected co-regulated genes and distinguished yeast strains. Besides, winemaking strains deal better with stress-imposing environmental conditions and are able to manage nutrient deficiencies, such as nitrogen, in a more efficient and resourceful way (Carreto et al. 2011). The analysis of large comparative transcriptomic data of five industrial wine yeast strains was performed, and various genes/gene sets were identified that could be linked to relevant aspects of yeast performance in key areas related to flocculation, stress tolerance, and metabolism (Rossouw et al. 2009). The analysis of the transcriptomes of two phenotypically diverging wine yeast strains in two different fermentation media at three stages of wine fermentation, showed that the intersection of transcriptome datasets from fermentations using either synthetic or real grape must can help to delineate relevant changes in gene expression, mainly of genes that codify membrane transporters, in response to experimental factors such as fermentation stage and strain identity (Rossouw and Bauer 2009a).

In the same way as the transcriptome, the proteome refers to the entire protein complement of the cell. As the proteome is produced from translation of the messenger RNA portion of the transcriptome, it is also dynamic in nature. However, differences in rates of protein production from specific RNA molecules give the proteome a greater dynamic range than the transcriptome (Borneman et al. 2007). The analysis of the protein complement of the cell or its parts is possible by using two-dimensional gel electrophoresis (2DE) or chromatography coupled with various mass spectrometry methods. Because genome-scale protein quantification is not yet feasible, several

methods for determining relative levels of protein and that allow comparison between samples have been developed (Smolka et al. 2001). An important goal of functional proteomics is the identification of functional modules based on the knowledge of protein action. Protein-protein interactions methods play a crucial role in elucidating the nature of these mechanisms. Innovative methods for the cell-wide analysis of protein interactions and signalling pathways have been developed (Templin et al. 2004). Conversely, the two dimensional gel approaches have some limitations such as the low number of proteins identified, the poor gel-to-gel reproducibility, the under-representation of low-abundant and hydrophobic proteins and the poor dynamic range of detection (Fey and Larsen 2001; Rabilloud 2002). The commonly used high-performance liquid chromatographic (HPLC) approach for the separation of peptides from protein digests in complex proteomic applications is 2D nano-liquid chromatography-mass spectrometry (LC/MS). In this approach, a strong cation exchange (SCX) column is used for the first dimension and a reversed phase (RP) column for the second (Nägele et al. 2004). This 2D chromatography approach coupled with tandem mass spectrometry (MS/MS) allowed the identification of a total of 1504 yeast proteins in a single analysis (Peng et al. 2003).

The entire complement of small chemicals and metabolites present in the cell, but not including DNA, RNA or protein, constitute the metabolome. The composition of the metabolome is more dynamic than either the transcriptome or proteome as it is altered by the enzymatic action of the proteome, in addition to being directly dependent on the composition of both the intra- and extracellular environments (Borneman et al. 2007). The metabolome analysis represents the newest tool and some of its applications such as metabolic profiling, metabolic fingerprinting (Fiehn 2002; Koek et al. 2006) or metabolic footprinting (Allen et al. 2003) have been published. Metabolites are involved in systems homeostasis as key regulators. For itself, level changes of specific groups of metabolites may be descriptive of systems responses to environmental interventions. Therefore, the study of these metabolites is a powerful approach for characterizing complex phenotypes, as well as for identifying biomarkers for specific physiological responses. The major complication on metabolites study is the rapid time scales of change, or oscillations in the levels of metabolites in a pathway, even if this pathway is in a balanced, unperturbed state of equilibrium (Rossouw and Bauer 2009b). A considerable progress has been noticeable

regarding wine chemical analysis, and it is now possible to quantify a large number of chemical compounds (both volatile and non-volatile) with relative accuracy (Villas-Bôas et al. 2005). A significant correlation was shown between metabolome and transcriptome during the initial phase of growth acceleration of a *S. cerevisiae* culture, after a glucose pulse (Kresnowati et al. 2006). Several methods are available for the determination of the metabolic profile of an organism. One example is the gas-chromatography (GC) or liquid-chromatography (LC) coupled to mass-spectroscopy (MS). GC-MS has been widely used in metabolome analysis essentially in complex biological mixtures (Kind and Fiehn 2007; Lommen et al. 2007; Mas et al. 2007). This system is composed by a gas supply, an injector and a column inside an oven, which are then connected to a mass spectrometer. GC analysis can be executed using a constant pressure, a constant flow or a flow program. The combined use of GC-MS has several advantages, as for example the high capacity of mass spectra to differentiate chemically diverse metabolites, the high separation efficiency provided by the GC system, sensitivity, easiness of use, robustness, low cost and commercial and public libraries available (Villas-Bôas et al. 2005; Hollywood et al. 2006; Dettmer et al. 2007; Garcia et al. 2008). However, GC-MS requires volatile samples and the majority of metabolites are non-volatile. Also, the steps of derivatization required are time-consuming (Halket et al. 2005; Wittmann 2007; Lu et al. 2008). In order to optimize the GC-MS performance some technologies could be conjugated with it, such as GC-GC time of flight (TOF)-MS (Koek et al. 2008; Mondello et al. 2008). This method consists in the conjugation of two different GC columns improving the detection coverage and a TOF-MS that provides a very fast scanning rate and an additional sensitivity in the detection. On the other hand, this is a very expensive method and not so used regularly. Another example of metabolite analysis methods is the combination between LC and MS. This revolutionary method allows separation of non-volatile metabolites followed by electrospray ionization (ESI) or by atmospheric pressure chemical ionization (APCI) (Bakhtiar et al. 2002). In comparison with GC-MS, this technique requires lower temperatures and the volatility in the samples is not needed, so the sample preparation is easier. LC-MS is mainly used on clinical application (Bakhtiar et al. 2002), however this method was also useful in the detection of several commercial compounds, previously predicted in the *in silico* metabolomes of *Bacillus subtilis* and *Escherichia coli*, and in the determination of the complete metabolome of *S. cerevisiae* (Werf et al. 2007). Capillary

electrophoresis-mass spectrometry (CE-MS) developed in 1989, is an analytical method that combines capillary electrophoresis with mass spectrometry detection (Loo et al. 1989). In comparison with GC and LC, the advantages of CE are the higher separation efficiencies, the smaller sample injection volumes, the lower cost of the reagents used and the capacity to separate anions, cations and uncharged molecules during a unique run. This technique has been widely used to determine the metabolome of some organisms with good results in detection and quantification of different metabolites (Perrett and Ross 1992; Perrett et al. 1994; Lehmann et al. 1997; Perrett et al. 1997; Soga and Imaizumi 2001; Terabe et al. 2001; Soga et al. 2002) such as analysis of inorganic ions (Kobayashi et al. 1998), amino acids (Soga and Heiger 2000), organic acids (Shirao et al. 1994), vitamins (Schreiner et al. 2003), carbohydrates (Soga and Heiger 1998), peptides (Perrett et al. 1994), nucleotides and nucleosides (Cohen et al. 1987) and thiols (Carru et al. 2003). The main disadvantages of CE are the lower sensitivity, because of the small volumes of injection, and when coupled to MS the limited number of libraries available. NMR spectroscopy is another method frequently used for metabolomics that consists in the application of magnetic fields and radio frequency pulses to the nuclei of the atoms. NMR active nuclei, such as ^{1}H or ^{13}C when placed in a magnetic field absorb electromagnetic radiation and cause nuclear spin and the subsequent emission of radiation (Dunn and Ellis 2005). This method has the advantageous possibility of being performed in a non-invasive manner but has reduced sensitivity and lower detection limit (Pan and Raftery 2007). NMR spectroscopy has been widely used in several fields such as the analysis of cold stresses in worms, analysis of plant-cell extracts, determination of disease biomarkers and determination of biochemical action mode (Dunn and Ellis 2005; Bothwell and Griffin 2011). Some studies have shown the high applicability of this technique for example for the study of different *S. cerevisiae* strains with similar growth rate which showed different ^{1}H NMR spectra (Raamsdonk et al. 2001) and for the identification of extracellular metabolites profiles and metabolic footprints of *S. cerevisiae* (Bundy et al. 2007).

Interactome is the complement of interactions between DNA, RNA, protein and metabolites in the cell, so this is extremely complex and highly dynamic. A full study of the interactome completed with binding affinities and rate constants, would constitute the main objective of systems biology, as it would enable the creation of a complete mathematical picture of the cell (Borneman et al. 2007). Comparisons of yeast

transcriptomes and proteomes under different conditions have shown that multilevel analysis is essential for yeast systems biology (Kolkman et al. 2006). A successful systems biology study should culminate on interactomics, which aims to integrate the transfer of information between the several steps of analysis with the use of mathematical modelling and simulation tools (de Jong 2002). Various statistical methods are available to create links between large data sets and phenotypes and it is possible to extract probabilistic models that can theoretically capture cellular interactions without prior knowledge of an interaction network (Jeong et al. 2000; de Jong et al. 2003).

In summary, systems biology is an area still in development that contributes to the establishment of new tools for strain selection. Strain improvement programs and other areas of biotechnology benefits highly with the advances in systems biology studies, and, with the development of new biological tools and the improvement of computational approaches, soon it will be possible to approach organisms in an holistic manner.

Saccharomyces cerevisiae and wine flavour

Wine flavour is the result of yeast metabolism and must compounds interactions. Systems biology yeast characterization combined with all the knowledge obtained in microbiology, physiology and biochemistry of yeasts, will lead to a better understanding of wine flavour. Must is constituted by three functional groups of compounds: nutrients, flavour precursors and non-precursor flavour-active compounds. The action of yeasts on some of these compounds will build the architecture of the wine flavour and their fermentation bouquet. A description of each group of compounds contributing to the wine flavour and aroma will be detailed in the next topics. **Table II-2** presents the main chemical groups contributing to wine flavour, together with their optimal wine concentrations and formation pathways.

Table II-2: Aroma and flavour compounds commonly found in wine. nd - not detectable

Group	Sub-group	Example of compounds	Impact on flavour	Optimal concentration	Formation pathway	Sensorial characteristics	Ref.
Alcohols	-	Ethanol	Enhance the sensory attributes of wine	8-16 vol. %	Alcoholic fermentation	hotness; body; viscosity	1, 2, 3
	-	Glycerol	No direct impact on the aromatic characteristics	5-14 g/L	Glycolisis and redox reactions	viscosity; sweetness; odorless	4, 5, 6, 7
	Higher alcohols	Propanol	Positive and negative impact: Excessive concentrations can result in a strong, pungent smell and taste; Optimal levels impart fruity characters	9,0–68 mg/L	Ehrlich pathway	pungent, harsh	8, 9, 10
		Butanol		0,5–8,5 mg/L		fusel, spiritous	
		Isobutanol		9,0–174 mg/L		fusel, spiritous	
		Isoamyl alcohol		6,0–490 mg/L		harsh, nail polish	
		Hexanol		0,3–12,0 mg/L		green, grass	
		2-Phenylethyl alcohol		4,0–197 mg/L		floral, rose	
Non-volatile acids	-	Tartaric acid	Positive and negative impact depending on concentration; sourness; astringency	90% of tritatable acidity	Tricarboxylic acid pathway and redox reactions	sourness; astringency	11, 12, 13, 14
	-	Malic acid		2 g/L		sourness; astringency	
	-	Succinic acid		> 6g/L		salty, bitter taste	
	-	Lactic acid		nd		sourness; astringency	
	-	Citric acid				sourness; astringency	
Volatile acids	-	Acetic acid	At higher concentrations negative impact	0,4 -1,1 g/L	Redox reactions of acetaldheyde	warmth; sourness/sharpness; vinegar	15, 16
Esters	Acetate esters	Ethyl acetate	Significant effect on the fruity flavours in wine	22,5–63,5 mg/L	Lipid and acetyl-CoA metabolism	nail polish, fruity, solvent	17, 18
		Isoamyl acetate		0,1–3,4 mg/L		banana, apple, solvent	17, 19, 20
		2-Phenylethyl acetate		0–18,5 mg/L		roses, honey, apple, sweet, floral	20, 21, 22
		Hexyl acetate		0–4,8 mg/L		sweet, aromatic, fragrant	21, 23
		Isobutyl acetate		0,01–1,6 mg/L		banana, sweet, fruity	17, 21
	Ethyl fatty acid esters	Ethyl butanoate		0,01–1,8 mg/L		papaya, butter, sweet, apple, fragrant, fruity	17, 19, 20, 21, 24
		Ethyl hexanoate		0,03–3,4 mg/L		apple, fruity, sweet, aniseed-flavored	20, 21
		Ethyl octanoate		0,05–3,8 mg/L		sweet soap	8, 25
		Ethyl decanoate		0–2,1 mg/L		floral, soap	26
	Branced-chain esters	Ethyl 2-methylpropanoate		0-0,9 mg/L		strawberry like aromas	8, 26, 27
		Ethyl 2-methylbutanoate					
		Ethyl 3-methylbutanoate					

Table II-2 (cont.):

Group	Sub-group	Example of compounds	Impact on flavour	Optimal concentration	Formation pathway	Sensorial characteristics	Ref.
Carbonyl compounds	-	Acetaldehyde	Associated with oxidation off-flavours	10–75 mg/L	Decarboxylation of pyruvate	sherry, nutty, bruised apple	28, 29
	-	Diacetyl		<5 mg/L	Malolactic fermentation	buttery	30, 31
Volatile phenols	-	4-ethylphenol	Contribute positively to the aroma of some wines; higher concentrations of ethylphenols contribute to off-flavours	0,012–6,5 mg/L	Decarboxylation of hydroxycinnamic acids and phenolic acids	medicinal, barnyard	32, 33, 34
	-	4-ethyl guaiacol		0,001–0,44 mg/L		phenolic, sweet	
	-	4-vinyl phenol		0,04–0,45 mg/L		pharmaceutical	
	-	4-vinyl guaiacol		0,0014–0,71 mg/L		clove-like, phenolic	
Sulfur compounds	Sulfides	Hydrogen sulfide	Associated with off-flavours execpt the long-chain sulfur compounds	nd-370 ug/L	Sulfur metabolism	rotten egg	3, 35, 36
		Dimethyl sulfide		nd–480 ug/L		asparagus, cabbage, cooked corn	
		Diethyl sulfide		nd–10 ug/L		garlic	
		Dimethyl disulfide		nd–22 ug/L		vegetable, cabbage, onion-like	
		Diethyl disulfide		nd–80 ug/L		bad smelling, onion	
	Sulfur-containing higher alcohols	3-(methylthio)-1-propanol (methionol)		nd–4500 ug/L		potato, cauliflower, cooked cabbage	
	Thiols	Benzothiazole		nd–14 ug/L		rubber	
		Thiazole		0–34 ug/L		popcorn, peanut	
		4-methylthiazole		0–11 ug/L		green hazelnut	
		2-furanmethanethiol		0–350 ng/L		roasted coffee, burnt rubber	
		Thiophene-2-thiol		0–11 ug/L		burned, burned rubber, roasted coffee	
		4-mercapto-4-methylpentan-2-one (4MMP)		nd–0,03 ug/L		cat urine, box-tree, broom, blackcurrant	
		3-mercaptohexan-1-ol (3MH)		nd–5 ug/L		box tree, broom, passionfruit, grapefruit	
		3-mercaptohexyl acetate (3MHA)		nd–0.2 ug/L		box tree, broom, passion fruit	
Monoterpenoids	-	Linalool	Powerful odorants	0,0017–0,010 mg/L	Sterol ant terpenes biosynthetic pathway	rose	3, 37
	-	Geraniol		0,001–0,044 mg/L		rose-like	
	-	Citronellol		0,015–0,042 mg/L		citronella	
C13-norisoprenoids	-	*β*-damascenone	Powerful odorants	4-7 ug/L	Acid-catalyzed transformation of grape-derived precursors	apple, rose, honey	33, 38, 39, 40
	-	*β*-ionone		nd		violet, flower, and raspberry	
	-	1,1,6-trimethyl-1,2-dihydronaphtalene (TDN)		nd		kerosene-like	

1-(Guth and Sies 2002)**; 2-** (Gawel et al. 2007a)**; 3-** (Swiegers et al. 2005a)**; 4-** (Noble and Bursick 1984)**; 5-** (Albertyn et al. 1994)**; 6-**(Rankine and Bridson 1971)**; 7-**(Nieuwoudt et al. 2002)**; 8-** (Lambrechts and Pretorius 2000b)**; 9-** (Swiegers and Pretorius 2005)**; 10-** (Nykänen et al. 1977)**; 11-** (Whiting 1976)**; 12-** (Thoukis et al. 1965)**; 13-** (Radler 1993)**; 14-** (Coulter et al. 2004)**; 15-**(Dubois 1994)**; 16-** (Corison et al. 1979)**; 17-** (Siebert et al. 2005)**; 18-** (Boutou and Chatonnet 2007)**; 19-** (Culleré et al. 2004)**; 20-** (Escudero et al. 2004)**; 21-** (Meilgaard 1975)**; 22-** (Salo 1970)**; 23-** (Etiévant 1991)**; 24-** (Czerny et al. 2008)**; 25-** (Soles et al. 1982)**; 26-** (Boulton et al. 1998)**; 27-** (Guth 1997)**; 28-** (Schreirer 1979)**; 29-** (Berg et al. 1955)**; 30-** (Martineau et al. 1995)**; 31-** (Sponholz 1993)**; 32-** (Dubois 1983)**; 33-** (Ribéreau-Gayon et al. 2000a)**; 34-** (Chatonnet et al. 1993)**; 35-** (Mestres et al. 2000)**; 36-** (Fedrizzi et al. 2007)**; 37-** (Strauss et al. 1986)**; 38-** (Francis and Newton 2005)**; 39-** (Winterhalter et al. 1991)**; 40-** (Sabon et al. 2002)**.**

➢ **Ethanol**

The major products of alcoholic fermentation promoted by *S. cerevisiae* are ethanol and CO_2, under oenological conditions. The high sugar concentration of grape must turns the yeast metabolism into a fermentative mode (Ugliano and Henschke 2009). The production pathway of ethanol from glucose is schematized in **Figure II-1**. Ethanol content can affect the chemical, physical and sensory properties of wine such as the hotness, body and viscosity perception, and also the sweetness, acidity, aroma, flavour intensity and textural properties (Gawel et al. 2007b; Gawel et al. 2007a).

Studies have shown that ethanol can influence the relative contribution of aroma compounds in different ways. The sensorial aroma perception is also changed by the influence of ethanol; when ethanol is present at 14.5–17.2%, the odour has been described as herbaceous instead of fruity, as was perceived at low ethanol levels (Goldner et al. 2009). Reducing the alcohol levels in wine, affects the aromatic bouquet not only by strengthening the perceived interactions between woody and fruity wine odorants, but also by modifying their chemical proportions (Le Berre et al. 2007).

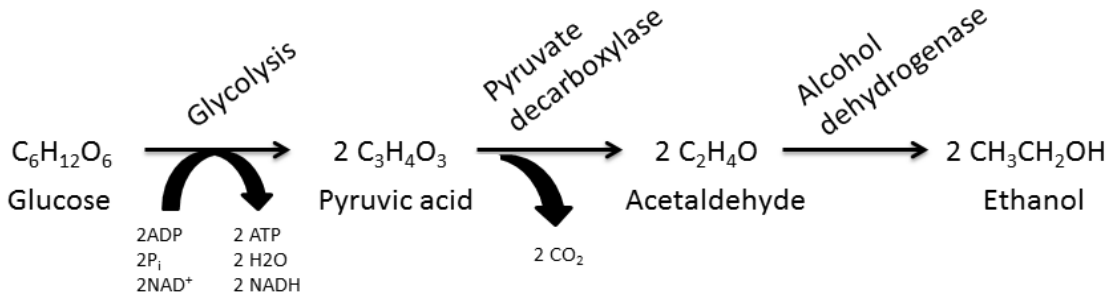


Figure II-1: Ethanol formation pathway in yeast.

➢ **Glycerol**

Glycerol is the major polyol produced during fermentation by *S. cerevisiae*. The impact of glycerol on wine flavour is not clearly understood. Early studies showed that the perceived overall flavour profile of a model wine and a white wine was not changed by the addition of glycerol, suggesting that glycerol does not play a significant role in establishing the aroma bouquet of wine (Lubbers et al. 2001). However, other studies have showed some effects of glycerol concentrations on perceivable viscosity (Noble and Bursick 1984; Nieuwoudt et al. 2002; Gawel et al. 2007b). Glycerol is formed from the

reduction of glycolytic intermediate, dihydroxyacetonephosphate, to glycerol-3-phosphate followed by dephosphorylation (**Figure II-2**). The first step is carried out by two NADH-dependent glycerol-phosphate dehydrogenase isoenzymes (Gpd1, 2p), encoded by *GPD1* and *GPD2* genes (Albertyn et al. 1994; Eriksson et al. 1995). This step is rate limiting for glycerol production since over-expression of either gene increases glycerol production.

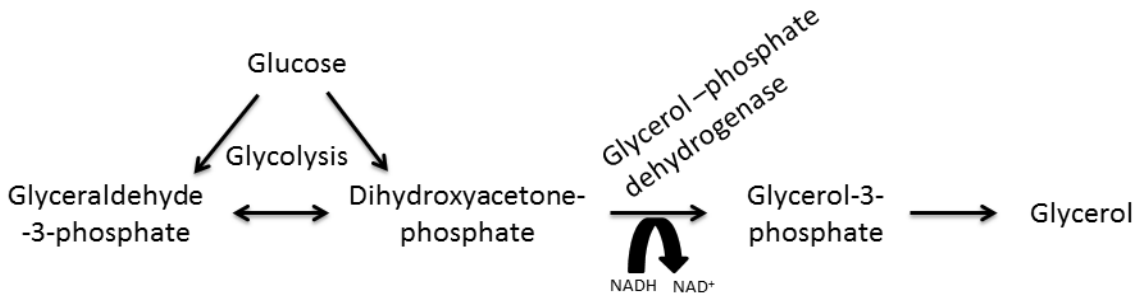


Figure II-2: Glycerol formation pathway in yeast.

- **Acetaldehyde and diacetyl**

Acetaldehyde is an important aroma compound that is directly related to alcoholic fermentation and constitutes more than 90% of the total aldehyde content of wine (Nykänen 1986). Under fermentative conditions, acetaldehyde is formed from pyruvate which is decarboxylated by pyruvate decarboxylase (**Figure II-3**). This enzyme is encoded by the genes *PDC1,5,6.* Acetaldehyde concentration is higher when the fermentation rate reaches its maximum, then decreases near the end of fermentation and gradually increases again subsequently (Lambrechts and Pretorius 2000b). High concentrations of acetaldehyde are generally associated with oxidation off-flavours in dry wines (Ugliano and Henschke 2009).

Diacetyl is formed from acetaldehyde, mainly by lactic acid bacteria during malolactic fermentation, but yeasts can also synthesize this compound during alcoholic fermentation (Styger et al. 2011b). Still, the majority of diacetyl is further metabolized to acetoin and 2,3-butanediol (Bartowsky and Henschke 2004). This compound is characterised by a 'nutty', 'toasty' or 'buttery' aroma depending on concentration (Martineau et al. 1995).

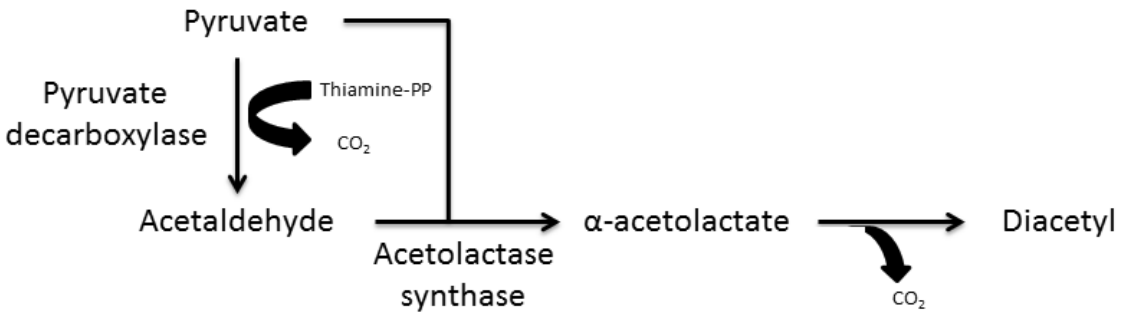


Figure II-3: Acetaldehyde and diacetyl formation pathway.

- **Non-volatile organic acids**

Wine contains a large number of organic and inorganic acids. Acidity and pH are fundamental characteristics to the sensory perception of wine, defining its structure and balance. Acids can have positive and negative impacts on aroma and flavour, depending on the concentration and type of wine (Swiegers et al. 2005b). **Figure II-4** summarizes the metabolic production of the most important organic acids.

The most important non-volatile organic acids are tartaric and malic acids, contributing with 90% for the titratable acidity (TA) of grape juice. Also, citric and lactic acids contribute to the acidity of grape juice, and succinic and keto acids are present only in residual amounts but their concentrations increase in the end of fermentation (Whiting 1976; Fowles 1992; Radler 1993; Boulton et al. 1998).

Organic acid metabolism performs an important role in wine as it produces precursors for biosynthetic pathways and it aids in the maintenance of redox balance. Tartaric acid, one of the major grape acids, is not metabolised by *S. cerevisiae* but L-malic acid can be partially degraded by most strains (Salmon 1987). However, it was shown that the commercial *S. cerevisiae* wine strain Enoferm M2 can increase the malic acid concentration of wine (Holgate 1997).

Succinic acid is the major organic acid produced by yeast metabolism and its production is highly variable among strains of *S. cerevisiae* (Radler 1993; Eglinton et al. 2000; Coulter et al. 2004). It´s formation pathway involves the reductive branch of the tricarboxylic acid (TCA) cycle during anaerobic fermentation (Roustan and Sablayrolles 2002; Camarasa et al. 2003). Oxalacetate is formed from pyruvate by a carboxylation reaction, then is reduced to L-malate and hydrated to form fumarate, being this reduced

by fumarate reductase to form succinate (Enomoto et al. 2002). Succinate can also be formed by the oxidative decarboxylation of α-ketoglutarate (Camarasa et al. 2003). Several fermentation conditions could affect succinic acid accumulation during fermentation such as yeast strain, temperature and must clarity, aeration, sugar concentartion, nutrient content, assimilable nitrogen, pH, titratable acidity and SO_2 concentration (Coulter et al. 2004).

Some other organic acids, such as pyruvic and α-ketoglutaric acids, have implications in wine stability and quality due to their ability to bind sulphur dioxide and to react with phenols (Rankine 1967; Rankine 1968a; Rankine 1968b). The keto acids are produced both during the early stages of fermentation via sugar metabolism, or from the corresponding amino acids, by the Ehrlich pathway. Strain, nitrogen type and content of the medium are the main factors affecting the keto acid production (Rankine 1968b).

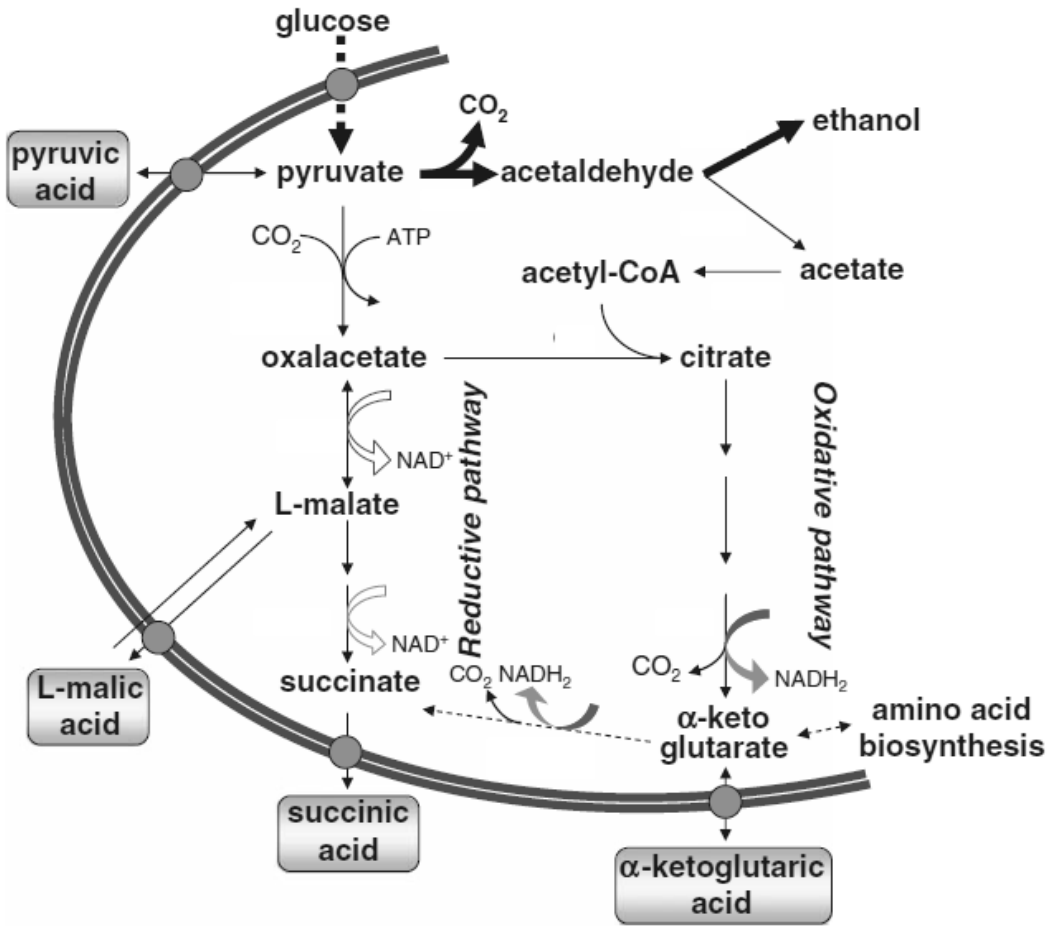


Figure II-4: Metabolism of organic acids (adapted from Ugliano & Henschke 2009).

- **Esters**

Esters can be classified as acetate esters or ethyl esters. In the group of acetate esters, the acyl group is derived from acetate (in the form of acetyl-CoA), and the alcohol group is ethanol or a complex alcohol derived from amino acid metabolism. The most significant acetate esters are ethyl acetate, isoamyl acetate and 2-phenylethyl acetate. The second group comprises the medium-chain fatty acid (MCFA) ethyl esters, where the alcohol group is ethanol, and the acyl group is derived from activated medium-chain fatty acids. Ethyl hexanoate and ethyl octanoate are the most common examples (Cordente et al. 2012).

Figure II-5 shows a simplified scheme of the biosynthetic and degradation pathway of esters. The ester formation during fermentation depends on the concentration of the co-substrates (the acyl-CoA and the alcohol) and on the activity of enzymes involved in their synthesis and hydrolysis (**Figure II-5**) (Verstrepen et al. 2003; Saerens et al. 2006; Saerens et al. 2008). Five distinct proteins were identified and characterized in *S. cerevisiae* - Atf1p, Atf2p, Eht1p, Eeb1p and Iah1p -, being the alcohol acetyltransferase Atf1p described as the protein with the greatest activity and the most studied one (Lilly et al. 2000; Lilly et al. 2006a; Saerens et al. 2010). Overexpression of *ATF1* during wine fermentation results in a significant increase in acetate ester production, whereas *ATF2* appears to play a minor role in ester formation (Lilly et al. 2000; Verstrepen et al. 2003; Lilly et al. 2006a). Overexpression of *ATF1* in yeast led to high production of ethyl acetate but this fact did not improve the fermentation bouquet and aroma of young wines, conversely, a higher hydrolysis during bottle aging was observed which caused a significant decrease in the levels of acetate esters, particularly ethyl acetate (Lilly et al. 2000). On the other hand, the deletion of both *ATF1* and *ATF2* completely stops the formation of isoamyl acetate (Verstrepen et al. 2003).

The formation of acetate ester by yeast is regulated by the *IAH1*-encoded esterase. *IAH1* regulates the accumulation of isoamyl acetate and other esters during fermentation, and consequently determines the flavour quality of wine (Lilly et al. 2006a). Yeast strains deficient in *IAH1* accumulate much higher amounts of isoamyl acetate (Fukuda et al. 1998) and its overexpression results in a significant decrease in the concentration of many esters, including isoamyl acetate, hexyl acetate, ethyl acetate, and 2-phenylethyl acetate, compared to control strains (Lilly et al. 2006a).

The availability of acetyl-CoA is also an important factor that contributes to the acetate ester formation. In a study using an *Escherichia coli* strain expressing the *ATF2* yeast gene, an increase of the levels of both CoA and acetyl-CoA was observed, which were responsible for an increase in the production of isoamyl acetate and by its turn, to an overexpression of the bacterial pantothenate kinase (*panK*) gene, which regulates CoA biosynthesis (Vadali et al. 2004).

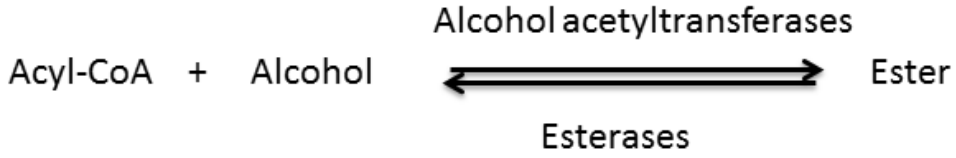


Figure II-5: The biosynthetic and degradation reactions of esters (adapted from Ugliano & Henschke 2009).

- **Higher alcohols**

Higher alcohols, also called fusel alcohols, are the most important compounds formed from amino acids produced by yeast during alcoholic fermentation (Ugliano and Henschke 2009; Styger et al. 2011b). These compounds are characterized by containing more than two carbon atoms and include the branched-chain alcohols 2-methylpropanol (isobutanol), 2-methylbutanol (amyl alcohol), and 3-methylbutanol (isoamyl alcohol), and the aromatic alcohols 2-phenylethanol and tyrosol (Ugliano and Henschke 2009). Particularly, 2- phenylethanol is considered one of the most important aromatic alcohols contributing to wine flavour (Ehrlich 1907; Hazelwood et al. 2008; Styger et al. 2011a). At high concentrations higher alcohols impart off-flavours, but lower concentrations make a crucial contribution to the flavour and aroma of wine (Nykänen 1986; Lambrechts and Pretorius 2000b).

Higher alcohols are formed by decarboxylation and subsequent reduction of α-keto-acids produced as intermediates of amino acids biosynthesis and catabolism (**Figure II-6**). The process in which the amino acids are catabolized into higher alcohols is called Ehrlich pathway (Ehrlich 1907). This pathway impacts directly or indirectly on the synthesis of other aroma compounds (Lilly et al. 2006b; Hazelwood et al. 2008).

The transamination step of Ehrlich pathway is carried out by aminotransferases: the mitochondrial and cytosolic branched-chain amino acid (BCAA) aminotransferases and the aromatic amino acid aminotransferases I and II. The mitochondrial and cytosolic branched-chain amino acid (BCAA) aminotransferases are encoded by the genes *BAT1* and *BAT2*, respectively. The aromatic amino acid aminotransferases I and II are encoded by the genes *ARO8* and *ARO9*, respectively (Eden et al. 1996; Kispal et al. 1996; Iraqui et al. 1998). In the following step, pyruvate decarboxylases, encoded by *PDC1*, *PDC5*, and *PDC6,* convert α-keto-acids into the correspondent aldehydes, which are then reduced to alcohols by alcohol dehydrogenases (Schure et al. 1998; Yoshimoto et al. 2001; Ugliano and Henschke 2009). Studies have shown that when the *BAT1* gene is overexpressed in a commercial wine yeast -VIN13-, increased concentrations of 3-methybutanol, 2-methylpropanol and 2-methylpropanoic acid are produced (Lilly et al. 2006a; Lilly et al. 2006b). On the other hand, if *BAT2* gene is deleted the formation of these compounds is decreased.

The last step of the Ehrlich pathway includes the reduction or oxidation of the fusel aldehydes to produce higher alcohols or fusel acids, respectively. The reduction of fusel aldehydes is carried out by alcohol dehydrogenases (Adh1p to Adh7p) (Larroy et al. 2002; Dickinson et al. 2003; Kondo et al. 2012), by the formaldehyde dehydrogenase Sfa1p (Hauser et al. 2007), by the NADPH-dependent aldo-keto reductase Ypr1p (Ford and Ellis 2002) and by one of the putative aryl-alcohol dehydrogenases (Aad6p) (Styger et al. 2011a). In 2011, a study has screened and characterized the ten genes with highest importance on higher alcohol formation, from which *PAD1*, *SPE1*, *OYE2* and *HOM2* were the ones that present the most significant results (Styger et al. 2011a). At industrial scale, this knowledge provides excellent tools for the improvement of the wine flavour and aroma.

The production of *n*-propanol is directly related with the initial amounts of nitrogen available and with the yeast growth, and appears not to be influenced by the structurally-related amino acids threonine and α-aminobutyric acid, therefore the production of this alcohol is not made via Ehrlich pathway (Rapp and Versini 1996).

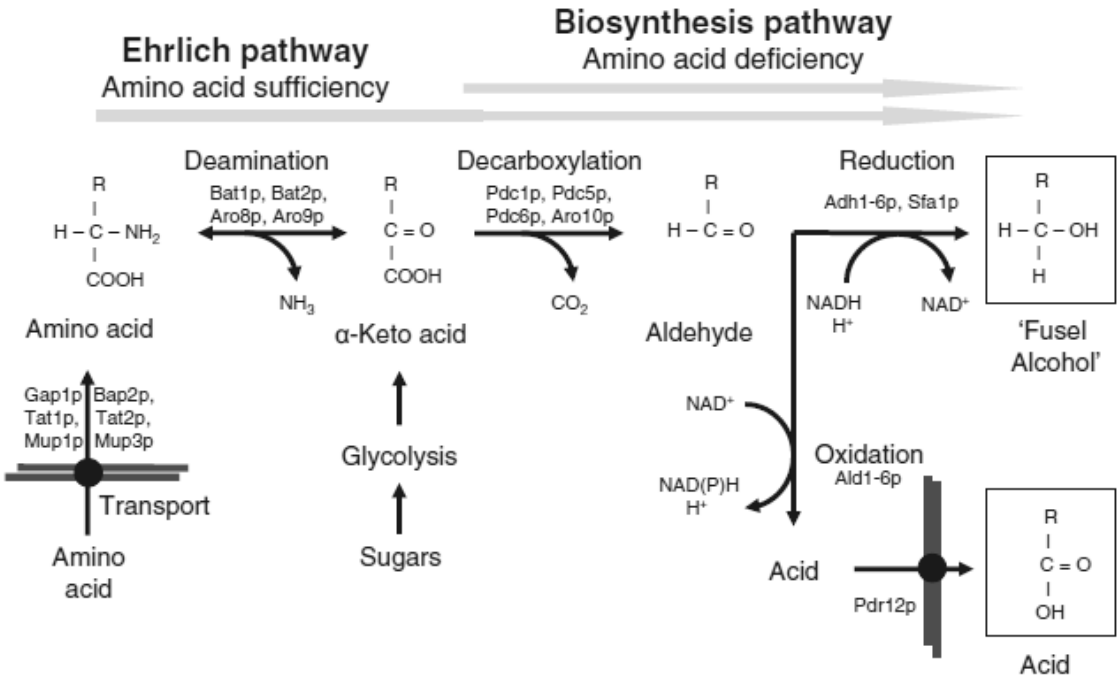


Figure II-6: Formation of higher alcohols and volatile acids from sugar and amino acids by Ehrlich pathway (adapted from Ugliano & Henschke 2009).

- **Volatile fatty acids**

Volatile fatty acids are present in wine in different forms: straight chain fatty acids usually referred to as short chain (C2–C4), medium chain (C6–C10), long chain (C12–C18), and a group of branched-chain fatty acids that include 2-methyl propanoic, 2-methyl butanoic, and 3-methyl butanoic acids (Ugliano and Henschke 2009).

Volatile fatty acids are produced via Ehrlich pathway like higher alcohols. During the last stage of Ehrlich pathway the aldehyde formed can be reduced via NADH-dependent reaction to its respective higher alcohol or it can be oxidized via a NAD^+-dependent reaction into a volatile carboxylic acid (Styger et al. 2011b). These reactions are catalysed by an alcohol dehydrogenase and by an aldehyde dehydrogenase, respectively (Dickinson et al. 2003; Vuralhan et al. 2005) – **Figure II-7**.

Acetic acid is the most important volatile fatty acid produced during alcoholic fermentation contributing with 90% of the total wine volatile acidity (Ugliano and Henschke 2009). At elevated concentrations it imparts a vinegar-like character to wine. Commonly, excessive concentrations of acetic acid in wine are largely the result of the metabolism of ethanol by aerobic acetic acid bacteria and not produced by

Saccharomyces strains. This volatile fatty acid is formed by the action of aldehyde dehydrogenases from acetaldehyde, which is derived by the decarboxylation of pyruvate (Ugliano and Henschke 2009). The cytosolic acetaldehyde dehydrogenases are encoded by *ALD6*, *ALD2* and *ALD3*, whereas the mitochondrial isoforms are encoded by *ALD4* and *ALD5* (Navarro-Avino et al. 1999; Remize and Andrieu 2000; Pigeau and Inglis 2007). The major isoforms involved in the production of acetic acid in wine are *ALD6*, *ALD4* and *ALD5* (Saint-Prix et al. 2004).

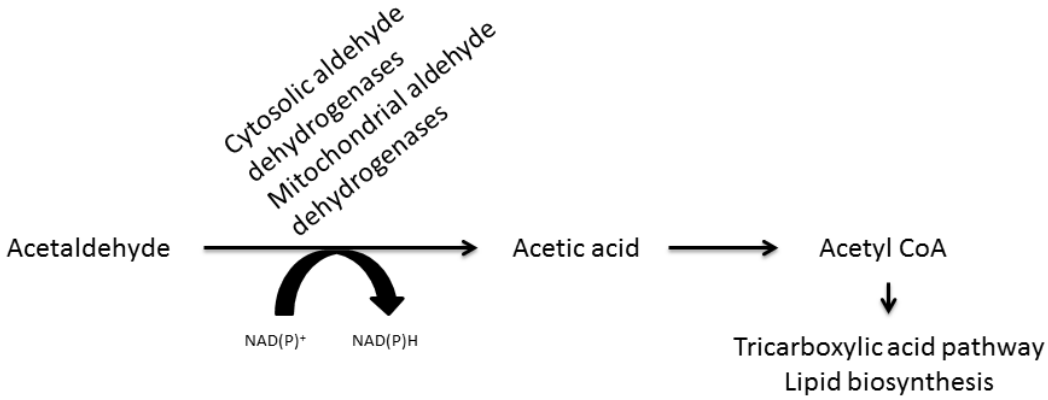


Figure II-7: Acetic acid metabolism (adapted from Ugliano & Henschke 2009).

- **Volatile sulfur compounds**

Volatile sulfur compounds are potent aroma compounds that have very low sensory threshold values, ranging from ng/L to g/L, and generally confer a negative sensory contribution to wine (Rauhut 1993; Mestres et al. 2000; Vermeulen et al. 2005). In wine, sulfur compounds appear in different categories: sulphides, polysulfides, heterocyclic compounds, thioesters and thiols (Swiegers et al. 2005b). In generally, sulfur compounds are associated with negative descriptors such as cabbage, rotten egg, sulfurous, garlic, onion and rubber (Rauhut 1993; Mestres et al. 2000; Vermeulen et al. 2005). On the other hand, there are some sulfur compounds contributing with positive aromas to wine, as for instance strawberry, passionfruit and grapefruit (Tominaga et al. 1996; Tominaga et al. 1998a; Tominaga et al. 1998b). Dimethylsulfide (DMS) is considered to be an exception since at low concentrations the asparagus, corn, molasses aromas can be considered pleasant in some type of wine (Ugliano and Henschke 2009). The development of sulfur compounds by yeasts involves different pathways such as, the metabolism of grape-

derived sulfur-containing precursors, the degradation of sulfur-containing amino acids and degradation of sulfur-containing pesticides (Mestres et al. 2000) – **Figure II-8**. Generally, the research has focused its studies on H_2S metabolism in yeast, particularly in response to changes in nutrient composition (Eschenbruch and Bonish 1976; Henschke and Jiranek 1991; Rauhut 1993; Linderholm et al. 2006; Swiegers et al. 2007).

Hydrogen sulfide

Hydrogen sulfide is a highly volatile thiol that imparts a 'rotten egg' aroma and has a very low odour threshold. Because of the frequent occurrence of this compound and the low aroma threshold (50–80 µg/L), it is one of the most common problems associated with the winery (Rankine 1963; Rauhut 1993). This sulfur compound can be formed from inorganic sulfur, sulfate and sulphite, or from organic sulfur compounds, cysteine and glutathione (Rankine 1963; Eschenbruch and Bonish 1976; Rauhut 1993; Spiropoulos et al. 2000). Sulfate is accumulated by two transporters, Sul1p and Sul2p, activated with ATP and reduced to H_2S by the sulfate reductive assimilation pathway. In the last step of this pathway, sulphite is reduced to H_2S by sulphite reductase which is encoded by *MET5* and *MET10* genes (Ugliano and Henschke 2009). When present in the must, sulphite enters the cell by diffusion across the plasma membrane (Stratford and Rose 1986) and can be directly reduced to sulfide (Stratford and Rose 1985; Jiranek et al. 1996; Hallinan et al. 1999).

The formation of H_2S and other volatile compounds by degradation of the sulfur amino acids cysteine and methionine has been observed in laboratorial conditions. However, their roles under fermentative conditions are less clear (Eschenbruch and Bonish 1976; Jiranek and Langridge 1995; Moreira et al. 2002; Perpète et al. 2006). In nitrogen absence and depending on amino acid composition of the medium, cysteine is accumulated and degraded to H_2S, pyruvate and ammonia by cysteine desulfydrase (Tokuyama et al. 1973). Cysteine can also be a precursor for dimethylsulfide (DMS) and 2-mercaptoethanol (Ribéreau-Gayon et al. 2000b; Moreira et al. 2002).

Under winemaking conditions, nutrient supplementation does not eliminated the risk of H_2S production (Henschke and Jiranek 1991; Park et al. 2000; Spiropoulos et al. 2000). Some studies have shown the complexity of regulation of the sulfate reductive pathway (Spiropoulos et al. 2000; Linderholm et al. 2006; Linderholm et al. 2008). In order to

lower excess levels of H_2S into acceptable concentrations, sulfite reductase activity could be monitored. However, no commercial strains have been developed yet with this aim (Zambonelli and Mutinelli 1975; Sutherland et al. 2003). Overexpression of genes involved in H_2S sequestration (*MET17*) and *S*-amino acid biosynthesis (*CYS4* and *MET6*) suggests that metabolic flux might be a limiting factor in order to control H_2S liberation (Ugliano and Henschke 2009).

Volatile sulfur compounds

S. cerevisiae has the capacity to produce negative volatile sulfur compounds such as H_2S, as described previously. The conditions that favour H_2S production also favour the production of other volatile sulfur compounds, such as methanethiol and methanethioacetate (Rauhut et al. 1996), suggesting the relation with the methionine catabolism (Ugliano and Henschke 2009). Some sulfur-containing flavour compounds contribute positively to wine (Swiegers and Pretorius 2005), such as furfurylthiol ('roast coffee' aroma) (Tominaga et al. 2000), the 'fruity' polyfunctional thiols 3-mercaptohexan-1-ol (3MH), 4-mercapto-4-methyl-pentan-2-one (4MMP), and 3-mercaptohexyl acetate (3MHA), that impart 'passionfruit', 'grapefruit', 'gooseberry', 'guava' and 'box hedge' aromas (Swiegers and Pretorius 2005; Dubourdieu et al. 2006). However, there are some volatile organic compounds that contribute negatively to wine, including methanethiol ('cooked cabbage' aroma), dimethylsulfide, dimethyldisulfide and dimethyltrisulfide ('cabbage', 'cauliflower' and 'garlic' aromas), and methylthioesters ('cooked cauliflower', 'cheesy' and 'chives' aromas) (Cordente et al. 2012). Yeasts that have more capacity to produce H_2S also produce higher concentrations of thioacetic acid esters of methanethiol and ethanethiol (Rauhut et al. 1996).

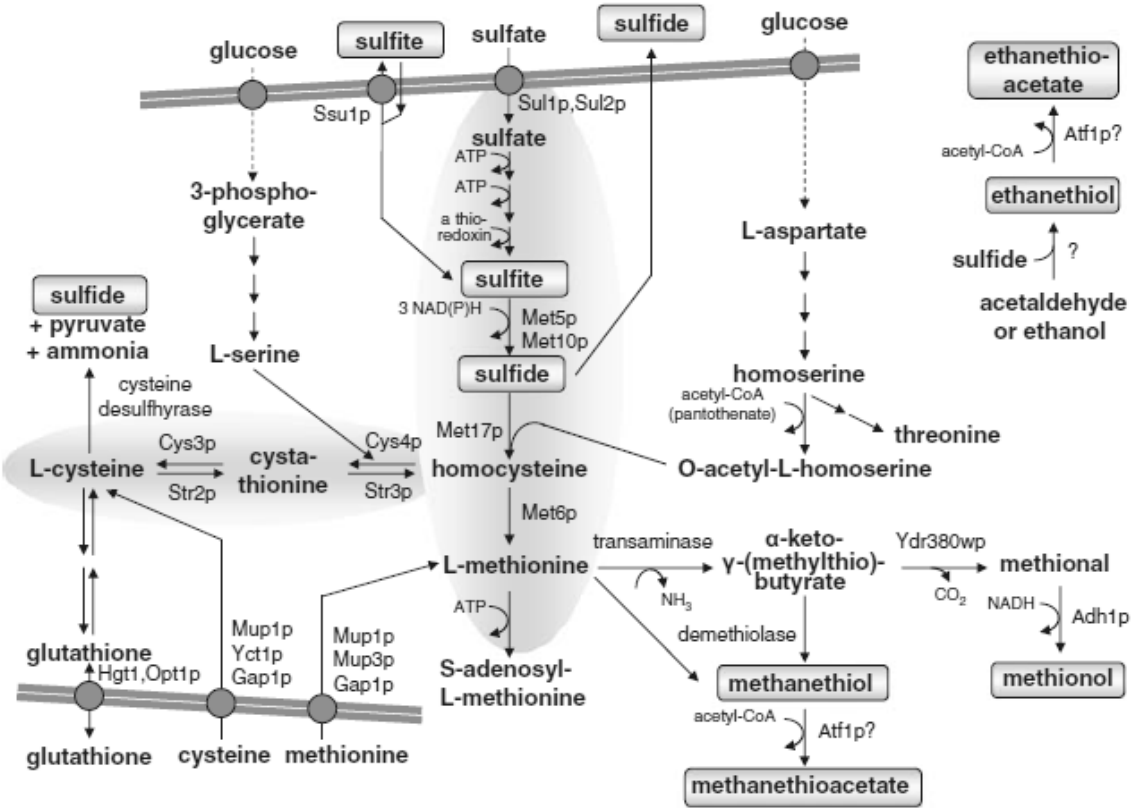


Figure II-8: Sulfur metabolism in *Saccharomyces cerevisiae* (adapted from Ugliano & Henschke 2009).

Biotechnological approaches for yeast improvement

Today, industrial biotechnology represents a well-established field with significant government, corporate and academic investment. Yeast has been extensively used for making bread, wine, beer, sake and some other fermentative products. Yeast also played an important role as a model organism in biochemistry, genetics and molecular biology. Nowadays, the need for new wine yeasts derives from both producer- and consumer-oriented requirements that aims for strains with improved technological properties and that leads to better and more diversified products.

Wine improvement can be achieved by targeting some important organoleptical and sensorial properties, such as higher ethanol tolerance, improved wholesomeness or lower ethanol content (Blondin and Dequin 1998; Dequin 2001; Verstrepen et al. 2006).

Several approaches has been evaluated with the last years with the objective of obtain strain improvement. With the emergency of molecular genetics and genomics, it became possible to develop genetically modified yeasts strains (GMY) for the biofuel, bakery and beverage industries. However, wine yeast strains obtained by these techniques are still perceived in a very controversial manner by consumers and probably will not receive approval to become commercialized (Schuller 2010). The improvement of aromatic profile of wine has been extensively exploited and several approaches have been carried out to develop *S. cerevisiae* strains able to produce higher amounts of the desirable compounds that contribute to wine aroma and flavour. Constitutive expression or overexpression of enzymes such as endoglucanases, arabinofuranosidases, endoxylanases, which catalyze the cleavage of glycosylated precursors, lead to the production of fruity wines (Perez-Gonzalez et al. 1993; Ganga et al. 1999; Manzanares et al. 2003). Swiegers and co-workers developed a *S. cerevisiae* strain which expresses tryptophanase with strong lyase activity that released 25 times more volatile thiols and displayed an intense passionfruit aroma on produced wines (Swiegers et al. 2007). Overexpression of alcohol acetyltransferase gene (*ATF1*) and ethanol hexanoyl transferase gene (*EHT1*) increased the concentrations of acetate and ethyl esters, respectively (Lilly et al. 2006a). It is known that glycerol contribute to the viscosity and sweetness of wine and its presence in wine is desirable. Overexpession of glycerol-3-phosphate dehydrogenase encoding gene *GPD1* increased glycerol production and slightly decreased ethanol formation (Michnick et al. 1997).

Another method used nowadays to obtain improved yeast strains is based on the use of quantitative traits controlled by multiple genetic loci, referred to as quantitative trait loci – QTL -, as detailed previously in this section.

Due to the controversy raised by the use of GMY, several approaches have emerged analysing the diversity that nature has created. Therefore, the study of the natural variability of yeasts is mandatory, in order to search for strains with desirable traits, and to produce wine without resort to genetic engineering. The diversifying selection that yeasts undergo after expansion into new environments and during adaptation to stressful conditions was shown to lead to strain diversity (Diezmann and Dietrich 2009; Dunn et al. 2012; Borneman et al. 2013), resulting many times in adaptive genomic changes, such as gene amplifications, chromosomal-length variations, chromosomal rearrangements

(especially amplifications and deletions) and copy-number increases (Dunham et al. 2002; Pérez-Ortín et al. 2002; Carro et al. 2003; Schacherer et al. 2007; Carreto et al. 2008; Borneman et al. 2008; Diezmann and Dietrich 2009; Liti et al. 2009; Dunn et al. 2012; Salinas et al. 2012; Bleykasten-grosshans et al. 2013; Ibáñez et al. 2014). This search for optimal natural strains and the limitations of the use of genetic modified yeasts requires the interconnection of different "omics" approaches which allow a holistic view of the different biological systems.

Chapter III

Phenotypic characterization of a S. cerevisiae *collection comprising strains from worldwide geographical origins and different technological applications*

The work presented in this chapter has been published in:
Mendes I *, Franco-Duarte R *, Umek L, Fonseca E, Drumonde-Neves J, Dequin S, Zupan B, Schuller D (2013) Computational models for prediction of yeast strain potential for winemaking from phenotypic profiles. PLoS ONE 8(7): e66523
*contributed equally

Introduction

The use of commercial starter yeasts to guarantee the control of the fermentation process and the reproducibility and predictability of wine quality is a common practice in commercial wine production. The advantages of fermentations containing *Saccharomyces cerevisiae* starter cultures relies on the fact that they are rapid and produce wine with desirable organoleptic characteristics through successive fermentations and harvests (Fleet 1998; Schuller 2010). Currently, there are about 200 commercial *S. cerevisiae* winemaking strains available, and it is a common practice among wineries to use commercial starter yeasts that were obtained in other winemaking regions.

S. cerevisiae strains from diverse natural habitats harbor a vast amount of phenotypic diversity (Camarasa et al. 2011), driven by interactions between yeast and the respective environment. In grape juice fermentations, strains are exposed to a wide array of biotic and abiotic stressors (Bisson 1999), which may lead to strain selection and generate naturally arising strain diversity. Outside the wineries, this diversifying selection occurs due to unique pressures imposed after expansion into new habitats (Frezier and Dubourdieu 1992; Sabate et al. 1998; Lopes et al. 2002; Schuller et al. 2005; Valero et al. 2007). This agrees with findings showing that wine and sake strains are phenotypically more diverse than would be expected from their genetic relatedness (Kvitek et al. 2008).

Recent phylogenetic analyses of *S. cerevisiae* strains showed that the species as a whole consists of both "domesticated" and "wild" populations. DNA sequence analysis revealed that domesticated strains derived from two independent clades, corresponding to strains from winemaking and sake. "Wild" populations are mostly associated with oak trees, nectars or insects (Greig and Leu 2009; Liti et al. 2009; Schacherer et al. 2009a). Although some *S. cerevisiae* strains are specialized for the production of alcoholic beverages, they were derived from natural populations that were not associated with industrial fermentations. This was proposed once that the oldest lineages and the majority of variation were found in strains from sources unrelated to wine production (Fay and Benavides 2005b).

The phenotypic diversity of *S. cerevisiae* strains has been explored for decades in strain selection programs to choose the ones that enhance the wine´s sensorial characteristics and confer typical attributes to specific wines. These strains are used as commercial ones by winemakers to efficiently ferment grape musts and produce desirable

metabolites, associated with reduced off-flavours (Briones et al. 1995; Ramírez et al. 1998). Strain selection approaches are mentioned in many studies aiming to characterize *S. cerevisiae* isolates obtained from winemaking regions worldwide. The most relevant physiological tests refer to fermentation rate and optimum fermentation temperature, stress resistance (ethanol, osmotic and acidic), killer phenotype, sulphur dioxide (SO_2) tolerance and production, hydrogen sulphide (H_2S) production, glycerol and acetic acid production, synthesis of higher alcohols (e.g. isoamyl alcohol, n-propanol, isobutanol), β -galactosidase and proteolytic enzyme activity, copper resistance, foam production and flocculation (Mannazzu et al. 2002).

In a previous work from our laboratory (Franco-Duarte et al. 2009) the phenotypic and genetic variability of 103 *S. cerevisiae* strains from the *Vinho Verde* wine region (Northwest Portugal) was evaluated using mainly taxonomic tests of nterest in winemaking. Several data mining procedures to estimate a strain´s phenotypic behavior based on its genotypic data were then applied. This study was, to our best knowledge, the first attempt to computationally associate genotypic and phenotypic data of *S. cerevisiae* strains. We used subgroup discovery techniques to successfully identify strains with similar genetic characteristics (microsatellite alleles) that exhibited similar phenotypes.

Within the present study we expanded the strain collection to 172 isolates from worldwide geographical origins and technological groups (wine, bread, sake, etc.) and included 30 tests with biotechnological relevance for the selection of winemaking strains. Our objective was to gain a deeper understanding of the phenotypic diversity of a global strain collection and to infer computational models that predict the biotechnological potential or geographic origin of a strain from its phenotypic profile.

Material and Methods

Strain collection

A *S. cerevisiae* strain collection was constituted, comprising 172 strains with different geographical origins and technological applications or origins (**Figure III-1** and **Supplementary data S1**). This collection includes strains used for winemaking (commercial and natural isolates that were obtained from winemaking environments), brewing, bakery, distillery (sake, cachaça) and ethanol production, laboratory strains and also strains from particular environments (e.g. pathogenic strains, isolates from fruits, soil and oak exudates). All strains were coded (Zn) and stored at -80 ˚C in cryotubes containing 1 mL glycerol (30% v/v).

Phenotypic characterization

Phenotypic screening was performed considering two sets of tests, including a wide range of physiological traits that are also important from an oenological point of view.

In a first set of phenotypic tests, strains were inoculated into replicate wells of 96-well microplates. Isolates were grown overnight in YPD medium (yeast extract 1% w/v, peptone 1% w/v, glucose 2% w/v), and the optical density (A_{640}) was then determined and adjusted to 1.0. After washing with peptone (1% w/v), 15 μL of this suspension were inoculated in quadruplicate in microplate wells containing 135 μL of white grape must of the variety *Loureiro*, to a cellular density of 5×10^6 cells/mL (A_{640} = 0.1). Final optical density was determined after 22 h (30 ˚C, 200 rpm) in a microplate spectrophotometer. All microplates were carefully sealed with parafilm, and no evaporation was observed for incubation temperatures of 30 ˚C and 40 ˚C.

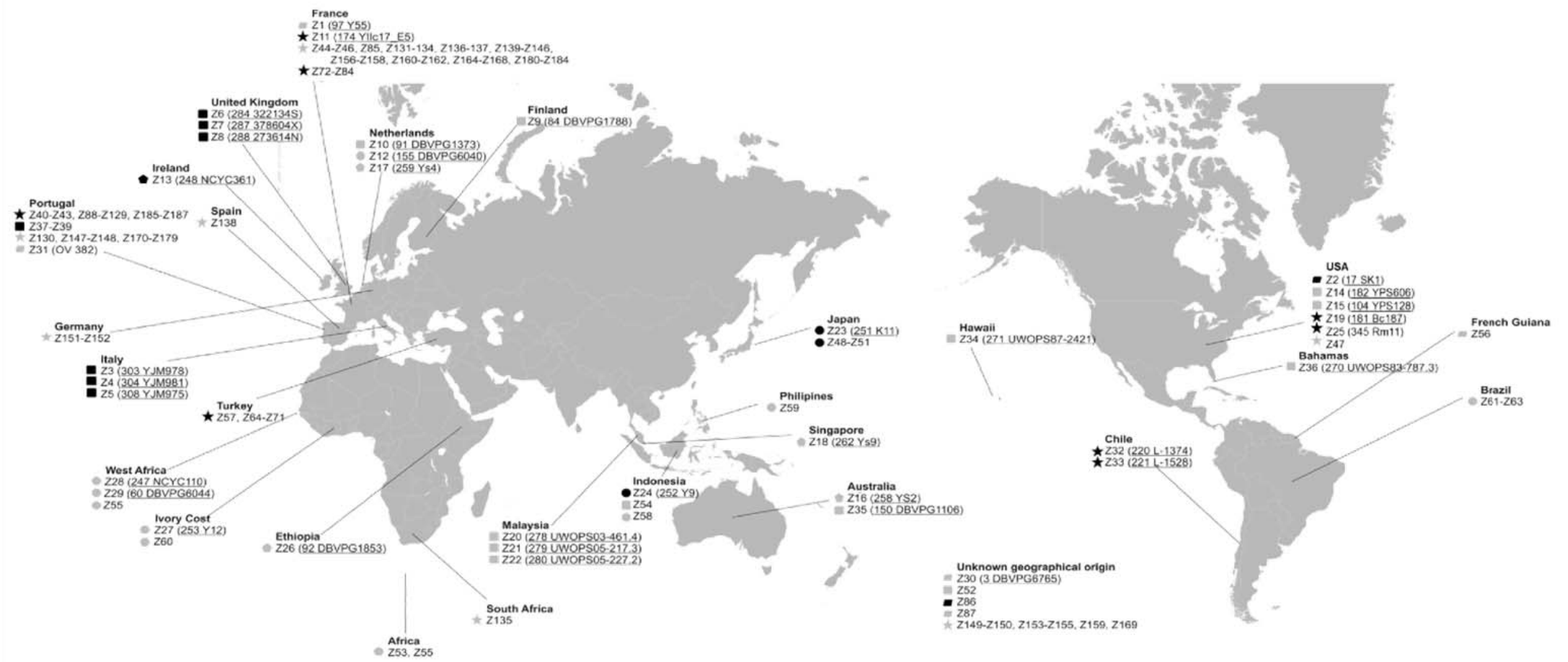


Figure III-1: Geographical location of the isolation sites of the 172 yeast strains used throughout this thesis.
Underlined identifiers indicate the original designation of sequenced strains (Liti et al. 2009).
Symbols represent strains' technological applications or origin: ★ - wine and vine; ★ - commercial wine strain; ■ - clinical; ■ - natural isolates; ● - sake; ● - other fermented beverages; ⬟ - beer; ⬟ - baker; ▬ - laboratory; ▬ - unknown biological origin.

This approach included the following tests: growth at various temperatures (18, 30 and 40 ˚C), evaluation of ethanol resistance (6, 10 and 14%, v/v), tolerance to several stress conditions caused by extreme pH values (2 and 8), osmotic/saline stress (0.75 M KCl and 1.5 M NaCl). Growth was also assessed in the presence of potassium bisulphite ($KHSO_3$, 150 and 300 mg/L), copper sulphate ($CuSO_4$, 5 mM), sodium dodecyl sulphate (SDS, 0.01%, w/v), the fungicides iprodion (0.05 and 0.1 mg/mL) and procymidon (0.05 and 0.1 mg/mL), as well as cycloheximide (0.05 and 0.1 μg/mL). These tests were carried out using *Loureiro* grape must supplemented with the mentioned compounds. The growth in finished wines was determined by adding glucose (0.5 and 1%, w/v) to a commercial white wine (12.5% v/v alcohol content). Galactosidase activity was evaluated by adding galactose (5% w/v) to Yeast Nitrogen Base (YNB, Difco™, Ref. 239210), using test tubes with 5 mL culture medium and 5×10^6 cells/mL, followed by 5 to 6 days of incubation at 26 ˚C.

Other tests were performed using solid media. Overnight cultures were prepared as previously described, adjusted to an optical density (A_{640}) of 10.0 and washed. One μL of this suspension was placed on the surface of the culture media mentioned below. Hydrogen sulphide production was evaluated using BiGGY medium (SIGMA-ALDRICH, Ref. 73608) (Jiranek *et al.* 1995), followed by incubation at 27 ˚C for 3 days. The colony colour, which represents the amount of H_2S produced was then analysed, attributing a score from 0 (no colour change) to 3 (dark brown colony). Ethanol resistance (12%, v/v) and the combined resistance to ethanol (12, 14, 16 and 18%, v/v) and sodium bisulphite ($Na_2S_2O_5$; 75 and 100 mg/L) was evaluated by adding the mentioned compounds to Malt Extract Agar (MEA, SIGMA-ALDRICH, Ref. 38954), and growth was visually scored after incubation (2 days at 27 ˚C).

All phenotypic results were assigned to a class between 0 and 3 (0: no growth (A_{640} = 0.1) or no visible growth on solid media or no colour change of the BiGGY medium; 3: at least 1.5 fold increase of A_{640}, extensive growth on solid media or a dark brown colony formed in the BiGGY medium; scores 1 and 2 corresponded to the respective intermediate values) as shown in **Table III-1**.

Data analysis

The phenotypic variability was evaluated by principal component analysis (PCA), available in the Unscrambler X software (Camo). The BioNumerics software (Applied Maths) was used for clustering, dendogram drawing and calculation of cophenetic correlation coefficients. Mann-Whitney test was applied to the phenotypic data set, including Bonferroni correction, to find relevant associations between phenotypic data and the strain's technological or geographical group. A set of standard predictive data-mining methods, such as naïve Bayesian classifier and *k* nearest-neighbors algorithm (Tan et al. 2006), as implemented in the Orange data mining suite (Demsar et al. 2004; Curk et al. 2005), were used for the inference of prediction models. For prediction scoring, area under the receiver operating characteristics (ROC) curve (AUC) was used (Hanley and McNeil 1982), which estimates the probability that the predictive model would correctly differentiate between distinct locations or distinct technological applications or origins, given the associated pairs of strains.

Results

Strain collection

A *S. cerevisiae* collection was constituted with 172 strains obtained from different geographical origins as shown in the map in **Figure III-1**. As detailed in **Supplementary data S1**, the technological applications or environments from where the strains were derived were: wine and vine (74 isolates), commercial wine strains (47 isolates), other fermented beverages (12 isolates), other natural environments – soil woodland, plants and insects (12 isolates), clinical (9 isolates), sake (6 isolates), bread (4 isolates), laboratory (3 isolates), beer (1 isolate), and four isolates with unknown origin.

The number of strains belonging to each group of technological applications or environment varies between 1 and 74. To assess a possible influence of a sample bias, due to an unequal number of representatives from each group, we determined the 95% confidence intervals for average Manhattan distance (Efron and Tibshirani 1995) between two strains in a selected group (composed by at least 5 strains). The distance was estimated based on the strains' entire phenotypic profile. The lower and upper bound of

each confidence interval were determined by percentiles of average distances for 10000 bootstraps samples. For example, with this analysis we showed that while the group of commercial strains (47 isolates) includes 31 commercial strains isolated in France, this should not bias our statistical analysis on utility of strains. Namely, the 95% confidence interval for average distances between pairwise combinations of commercial strains from France (6.37, 8.01) overlaps with the confidence interval of commercial strains from other geographical origins (4.97, 8.13). The inclusion of a higher number of strains from France does not change the limits of the confidence interval of the group of commercial strains. A similar result was observed for the group of wine and vine strains that includes numerous strains from Portugal: the 95% confidence interval for average distances between pairwise combinations of strains from Portugal (8.12, 9.83) overlaps with the same interval for wine and vine strains from other geographical locations (8.06, 9.59).

Phenotypic characterization

A phenotypic screen was devised to evaluate strain specific patterns for a set of physiological tests, including also tests that are important for winemaking strain selection. The first group of tests was performed in microplates using supplemented grape must, whereas a high reproducibility was obtained between experimental replicates. The second set of tests consisted in the evaluation of growth in solid culture media (BiGGY medium, Malt Extract Agar supplemented with ethanol and sodium metabisulphite). Galactosidase activity was evaluated by growth evaluation using Yeast Nitrogen Base supplemented with galactose, as indicated in the materials and methods section. After incubation, growth was evaluated by visual scoring (solid media) or by A_{640} determination (liquid media). **Table III-1** summarizes the number of strains belonging to each of the phenotypic classes.

Similarities between strains were evident, but each strain showed a unique phenotypic profile.

Table III-1: Number of strains belonging to different phenotypic classes, regarding values of optical density (Class 0: A_{640}=0.1; Class 1: 0.2<A_{640}<0.4; Class 2: 0.5<A_{640}<1.0; Class 3: A_{640}>1.0), growth patterns in solid media, or colour change in BiGGY medium.

Phenotypic test	Type of medium	Phenotypic class of growth			
		0	**1**	**2**	**3**
30 ˚C	liquid (must)	0	0	4	168
18 ˚C	liquid (must)	51	120	1	0
40 ˚C	liquid (must)	28	14	80	50
pH 2	liquid (must)	101	68	3	0
pH 8	liquid (must)	0	0	19	153
KCl (0.75 M)	liquid (must)	0	2	146	24
NaCl (1.5 M)	liquid (must)	84	79	9	0
$CuSO_4$ (5 mM)	liquid (must)	124	45	3	0
SDS (0.01% w/v)	liquid (must)	139	32	1	0
Ethanol 6 % (v/v)	liquid (must)	0	2	36	134
Ethanol 10 % (v/v)	liquid (must)	17	28	85	42
Ethanol 14 % (v/v)	liquid (must)	82	35	50	5
Ethanol 12 % (v/v)	solid (MEA)	150	20	1	1
Ethanol 12 % (v/v) + $Na_2S_2O_5$ (75 mg/L)	solid (MEA)	159	13	0	0
Ethanol 12 % (v/v) + $Na_2S_2O_5$ (100 mg/L)	solid (MEA)	169	3	0	0
Ethanol 14 % (v/v) + $Na_2S_2O_5$ (50 mg/L)	solid (MEA)	148	24	0	0
Ethanol 16 % (v/v) + $Na_2S_2O_5$ (50 mg/L)	solid (MEA)	163	9	0	0
Ethanol 18 % (v/v) + $Na_2S_2O_5$ (50 mg/L)	solid (MEA)	165	7	0	0
$KHSO_3$ (150 mg/L)	liquid (must)	34	11	26	101
$KHSO_3$ (300 mg/L)	liquid (must)	57	19	29	67
Wine supplemented with glucose (0.5% w/v)	liquid	103	45	24	0
Wine supplemented with glucose (1% w/v)	liquid	115	41	16	0
Iprodion (0.05 mg/mL)	liquid (must)	1	0	28	143
Iprodion (0.1 mg/mL)	liquid (must)	1	1	13	157
Procymidon (0.05 mg/mL)	liquid (must)	0	0	7	165
Procymidon (0.1 mg/mL)	liquid (must)	1	0	9	162
Cycloheximide (0.05 μg/mL)	liquid (must)	3	0	7	162
Cycloheximide (0.1 μg/mL)	liquid (must)	2	1	19	150
H_2S production	solid (BiGGY)	1	11	105	55
Galactosidase activity	liquid (YNB)	0	21	98	53

MEA: Malt Extract Agar
YNB: Yeast Nitrogen Base

A total of 5160 phenotypic data points was obtained, from 172 strains and 30 tests. The concentrations of the added compounds were chosen to obtain a wide range of tolerance patterns. As expected, all strains grew well at 30 ˚C, contrary to the growth at 40 ˚C, where a large phenotypic diversity was observed. The temperature of 18 ˚C revealed to be very limitative for the strains growth, with only one strain being capable of ferment must with final optical density above 0.4 (class higher then 2). Most strains were able to grow well at pH 8, contrarily to the pH value of 2. NaCl and CuSO4 added to the must revealed to be very limited (only 9 strains in class 2 for NaCl, three strains for CuSO4, and zero in class 3 for both cases), contrary to the KCl with 146 strains growing until class 2 of growth. Similar results were obtained when strains were exposed to SDS detergent in the medium, once that the large majority of strains (132) didn´t show any growth when this compound was present in the medium. As expected, cellular growth decreased with increasing concentrations of ethanol (6 - 14% v/v, liquid media), whereas only five isolates were able to grow well at the highest ethanol concentration of 14% (v/v). When ethanol was combined with sodium metabisulphite in solid culture media, growth was reduced with increasing concentrations of ethanol (12 to 18%, v/v) or sodium metabisulphite (50 to 100 mg/L). Resistance to sulphur dioxide, which is an antioxidant and bacteriostatic agent used in vinification, was tested by growth in the presence of wine must supplemented with potassium bisulphite ($KHSO_3$). For the concentrations of 150 and 300 mg/L, 101 and 67 strains achieved the highest class of growth, respectively. Resistance to the fungicides iprodion, procymidon and to cycloheximide was rather high at the indicated concentrations. Hydrogen sulphide production was tested using BiGGY medium. The majority of the strains were intermediate H_2S producers with the exception of one strain (from the group of wine and vine strains) that did not produce detectable levels of H_2S.

A global view of strain´s phenotypic diversity is shown in **Figures III-2** and **III-3**. Principal component analysis (PCA) of phenotypic data (**Figure III-2**) show the segregation of all 172 strains (scores) and the loadings for phenotypic variables in the first two principal components.

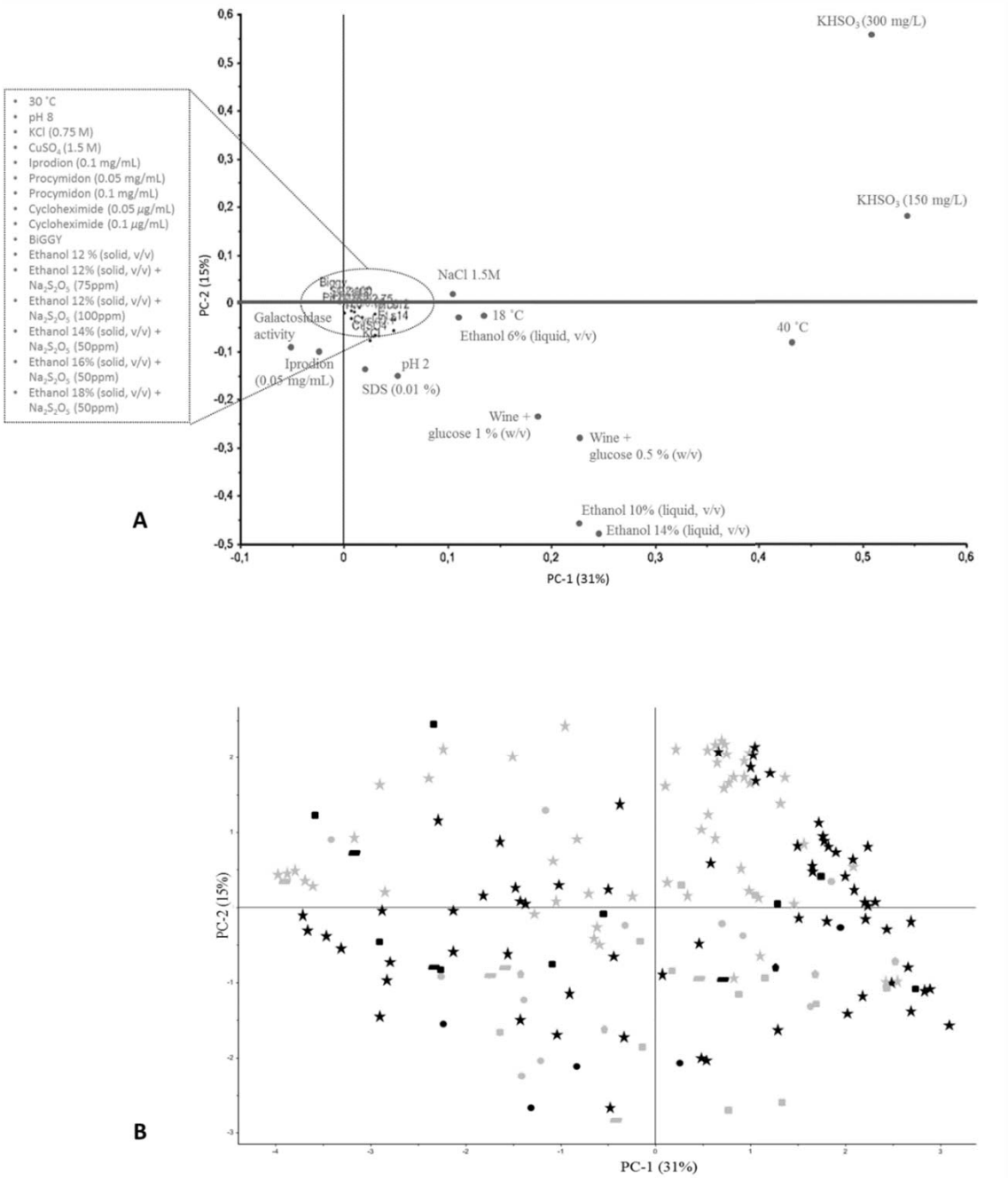


Figure III-2: Principal component analysis of phenotypic data for 172 strains:
A: 30 phenotypic tests (loadings).
B: 172 strains (scores) distribution. Symbols represent strains' technological applications or origin: ★ - wine and vine; ★ - commercial wine strain; ■ - clinical; ■ - natural isolates; ● - sake; ● - other fermented beverages; ⬟ - beer; ⬟ - bread; ▬ - laboratory; ▬ - unknown biological origin.

The phenotypes responsible for the highest strain variability (**Figure III-2A**) were associated with growth patterns in the presence of potassium bisulphite ($KHSO_3$), at 40 ˚C, in a finished wine supplemented with glucose (0.5%, w/v), and resistance to ethanol in liquid media (10 and 14%, v/v). PC-1 (31%) and PC-2 (15%) explained 46% of strain variability and segregated strains by phenotypic behavior into some patterns, as shown in **Figure III-2B**. The group of sake strains (●) and the group of natural strains (■) tended to be separated by the second principal component, accumulating in the lower part of the PCA, indicating that they were influenced by the presence of ethanol in the medium (higher resistance), and by the growth in the presence of potassium bisulphite (300 mg/L, lower resistance).

Strains isolated from vines or wine (★) showed a heterogeneous phenotypic behaviour since they were dispersed throughout the PCA plot for both components. A similar tendency was observed for commercial strains (★); however, the majority of strains tended to concentrate in the upper part of the PCA, indicative of a trend to higher $KHSO_3$ resistance and lower ethanol resistance. The nine clinical strains (■) were distributed in both PCA components, showing no discriminant results in any of the phenotypic tests.

UPGMA (Unweighted Pair Group Method with Arithmetic Mean) algorithm was used to hierarchical cluster the 172 strains. The dissimilarity between two strains was measured using Euclidean distance (**Figure III-3**). The combined phenotypes of wine strains did not separate this group of strains that were rather scattered throughout all the clusters. Commercial strains (★) tended to be more predominant in the clusters shown in the lower part of the dendogram, where some of the clusters are constituted only by commercial strains. Remaining groups show no individual grouping, which confirmed the results of the PCA.

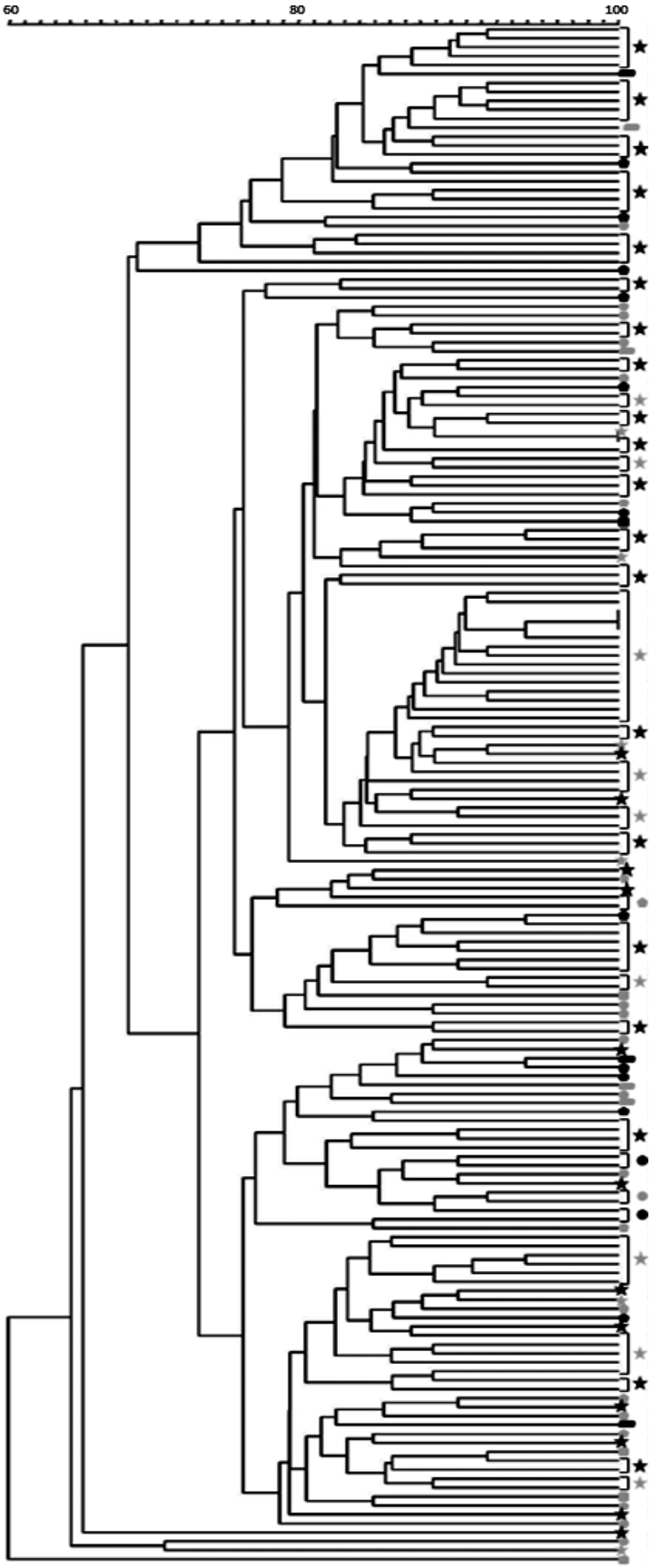

60
80
100

Figure III-3: Dendogram showing phenotypic variation of 172 strains under 30 growth conditions.

Strains are organized according to UPGMA-based hierarchical clustering (cophenetic correlation factor = 0.75), using Euclidean distance correlation to estimate phenotypic profile similarities. Symbols represent strains' technological applications or origin: ★ – wine and vine; ★ – commercial wine strain; ■ – clinical; ■ – natural isolates; ● – sake; ● – other fermented beverages; ⬟ – beer; ⬟ – baker; ▬ – laboratory; ▬ – unknown biological origin.

We further analysed phenotypic diversity through *k*-means clustering algorithm. Using silhouette score (Rousseeuw 1987) we identified 3 distinct clusters (**Table III-2**), composed of 38, 90 and 44 strains, respectively. The phenotypes that most distinguished the strains, as indicated by high values of information gain to classify strains into clusters, were growth at the highest and lowest temperatures tested (18 and 40 ˚C). Cluster 2 was constituted of strains that didn´t grow at both 18 and 40 ˚C, whereas cluster 1 and 3 included strains that grew at both temperatures, but with more pronounced growth at 40 ˚C, in particular for strains of cluster 3. Other tests that were also relevant for the cluster separation included growth in the presence of NaCl (1.5M), $KHSO_3$ (150 and 300 mg/L), ethanol 6% (v/v) and at pH 2. The strain cluster membership is displayed in the phenotypic data PCA visualization (**Supplementary data S3**).

Table III-2: Phenotypic tests mostly contributing for the division of strains into three clusters, in terms of information gain, obtained with *k*-means clustering algorithm.
Numbers in the last three columns represent the most characteristic value in terms of phenotypic class of strains included in the clusters, for the mentioned phenotypic tests.

Phenotypic test	Information gain	Cluster		
		1	2	3
18 ˚C	0.33	1	0	1
40 ˚C	0.33	2	0	3
NaCl (1.5M)	0.26	0	0	1
$KHSO_3$ (300 mg/L)	0.23	3	0	3
Ethanol 6% (v/v) – liquid medium	0.23	3	2	3
pH 2	0.21	0	0	1
$KHSO_3$ (150 mg/L)	0.21	3	0	3
Total number of strains		38	90	44

Computational analysis

Phenotypic results of the 172 *S. cerevisiae* strains was further used in statistical tests in order to search for relationships with the strains origin. Mann-Whitney test is mostly used to identify statistically significant associations between two data sets in which data instances in each group are measured on ordinal level and when there are an unequal number of members in the classes to be compared. This test was used to search for relationships between phenotypic results for the 172 strains and their shared geographical origin or technological application group. This test was used as described in (Mendes *et al.* 2013) and revealed 300 associations between phenotypes and technological application or origin of strains, whereas statistical significance was found for 11 associations (Bonferroni adjusted *p*-value lower than 0.1). In this analysis the most significant associations between a phenotypic class and a technological group were reported, together with the results of the computation of the probability of each phenotypic class (0-3) according to its contribution to the observed association (Mendes

et al. 2013 – Table 3). The reported results identified two associations for the resistance to iprodion, whereas class 3 and 2 were associated with strains collected from wine/vineyards and with commercial strains, respectively. Capacity to grow in the presence of potassium bisulphite (150 mg/L, classes 2 and 3) was associated with commercial wine strains. Natural isolates (87% – 89%) were associated with class 2 of growth in wine supplemented with glucose, both at 0.5 and 1% (w/v), contrarily to 57% of commercial strains that were unable to grow in wine supplemented with glucose (0.5%, w/v). The lower ability of commercial strains to grow at higher ethanol concentrations was also supported by the finding of one significant association for absent growth (class 0) in liquid medium containing ethanol (14%, v/v). About half of the strains included in this group shared the inability to grow in must containing SDS (0.01%, w/v) and $CuSO_4$ (5 mM), but grew well in cycloheximide supplemented must (76% of strains, class 2).

The results present in this work were also used in the same publication to construct a model that would predict a strain´s technological group from its phenotypic profile. *K*-nearest neighbor algorithm (*k*NN) and naïve Bayesian classifiers (Tan et al. 2006), as implemented in the Orange data mining suite were used for modelling, and the predictive performance of both classifiers was evaluated in terms of area under the Receiver-Operating-Characteristics (ROC) curve, using 5-fold cross validation (Hanley and McNeil 1982). In Table 4 of (Mendes *et al.* 2013) a confusion matrix is shown considering naïve Bayesian classifications in test data sets of cross-validation. Cross validated AUC score was 0.70. Correct assignments were found for the larger groups of commercial wine strains and strains obtained from wine and vineyards, where 36 (77%) and 54 (73%) strains, respectively, were accurately allocated. The same computational technique was also used to explore which phenotypes mostly contributed to the assignment of a strain to the commercial wine group. In Figure 3 of the same publication a nomogram is represented that shows naïve Bayesian classifier results (Mozina et al. 2004). Three phenotypes were considered by the classifier as the ones contributing more positively to build the model, having the remaining ones a smaller impact. To predict the commercial potential of a strain, the contribution of each phenotype was scored in the scale from -100 to 100, and the individual scores were summed-up to readout the probability of the predicted class. For the present data set, growth in must containing the fungicide iprodion (0.05 mg/mL), in cycloheximide (0.1 μg/mL) and in the presence of potassium bisulphite

(150 mg/L) were the three features with the most relevant contribution for the mathematical assignment of a strain to the commercial group. The probability of a strain to be assigned to the group of commercial strains is 0.27 (27%) when considering the strains entire phenotypic profile and increases to 0.95 (95%) when only the three phenotypic results mentioned in panel A are taken into consideration, as shown in the probability scale present in panel B.

Discussion

Franco-Duarte et al. (2009) performed a genotypic and phenotypic characterization of 103 *S. cerevisiae* strains from a Vinho Verde Portuguese winemaking region and developed computational techniques to relate both set of data. Subgroups were found for strains sharing allelic combinations and specific phenotypes such as low ethanol resistance, growth at 30 ˚C and growth in media containing galactose, raffinose or urea. In the present work, we extended the work to a phenotypically more heterogeneous strain collection comprising 172 *S. cerevisiae* isolates from worldwide origins, and used computational methods to relate the phenotype with the strain´s origins and to make predictions about a strain´s biotechnological potential based on phenotypic data. The group of phenotypic tests used herein was based on approaches that are generally applied for the selection of *S. cerevisiae* winemaking strains (Mannazzu et al. 2002).

The collection of 172 strains from worldwide geographical origins revealed a high phenotypic diversity (**Figures III-2** and **S2**, and **Table III-2**), which is in agreement with previous studies (Brandolini et al. 2002; Agnolucci et al. 2007; Kvitek et al. 2008; Franco-Duarte et al. 2009; Salinas et al. 2010; Camarasa et al. 2011; Warringer et al. 2011). A significantly higher phenotypic diversity was observed in the present study compared to our results from 2009 using 103 Portuguese wine yeast strains (Franco-Duarte et al. 2009). In particular, the inclusion of new tests compared to our previous study allowed a more detailed analysis of the phenotypic variability of strains associated with winemaking environments. Recent studies aimed to describe the elements that shaped the genomes of *S. cerevisiae* strains*,* suggesting that populations comprise distinct domesticated and natural groups, as well as mosaics within these groups, based on strain´s origin and application (Schacherer et al. 2007; Liti et al. 2009; Goddard et al. 2010). Clinical isolates for example, do not derive from a common ancestor, but rather represent multiple events in which environmental strains opportunistically colonize humans (Schacherer et al. 2007; Muller and McCusker 2009).

Genetic rearrangements and intra-strain variation are characteristic for this species (Dunn and Ellis 2005; Valero et al. 2007), which might explain the rather high phenotypic variability that was described in recent studies. Camarasa et al. (2011) showed that some phenotypes (resistance to high sugar concentrations, ability to complete fermentation and low acetate production) were able to distinguish groups of strains according to their

ecological niches, providing evidence for phenotypic evolution driven by environmental adaptation. This high phenotypic variation in stressful conditions was also revealed by Kvitek et al. (2008) showing the existence of unique features shared by strains from similar habitats. Our data are in agreement with the previously mentioned studies regarding the high phenotypic diversity. They also confirm the findings of Legras and co-workers (Legras et al. 2007a), that found population substructures of *S. cerevisiae* strains according to their technological application or origin, using multilocus microsatellite typing. In the work of Legras, only 28% of the diversity was associated with geographical origins, which suggests local domestication events. We herein investigated the utility of data mining to improve our understanding of relations between phenotypes and the strains' technological application or origin. The developed models can also be useful to optimize screening tests and to find commercial wine yeast candidates from strain collections.

The present work produced results that were analysed using statiscal methods, being the main outcomes already published (Mendes et al. 2013). Using Mann-Whitney test, 11 significant associations were found between a particular phenotypic result and a technological group. The most significant results were found for the resistance to iprodion, growth in potassium bisulphite and in wine supplemented with glucose. Iprodion is a dicarboximide contact fungicide used to control a wide variety of fungal pests on vegetables, ornamentals, pome and stone fruit, root crops, cotton and sunflowers. *S. cerevisiae* shows higher resistance to this fungicide than other yeast species such as *Candida albicans.* In this species, iprodion stimulates glycerol synthesis and inhibits the cell growth for several days, contrarily to *S. cerevisiae* where a low toxicity was observed (Ochiai et al. 2002; Čadež et al. 2010). Our results showed that iprodion resistance (0.05 mg/mL) was higher in strains from wine and vineyards in comparison to commercial wine strains. The higher iprodion resistance among strains obtained from wineries and vineyards might be explained by the evolution of this trait upon recurrent exposure, which does not apply for commercial wine strains that are added to clarified musts that should not contain this fungicide. The low ethanol resistance of commercial wine strains in liquid media containing 14% (v/v) ethanol was somehow unexpected, because these strains are usually selected for high ethanol resistance. This could be explained by the fact that the mathematical relations were observed for ethanol concentrations above the values that

usually occur in wines (10-13%, v/v). Results showed also that commercial strains tended to a better growth in media containing potassium bisulphite, a compound used as wine antiseptic and antioxidant, reflecting also an adaptive mechanism among this group of strains.

We found that the large phenotypic variability between strains could be associated with the technological application or origin of the strains rather than their geographical origin, once that no relevant relations were found for the last analysis (data not shown). The naïve Bayesian classifier was used to assign a strain to their technological group, based on their phenotypic profile (Mendes *et al.* 2013). This association was achieved for the majority of strains belonging to the commercial and wine and vine groups (77% and 73%, respectively). The cross-validated performance of this method yielded an AUC score of 0.70, that is considered as moderate (Hanley and McNeil 1982) and lies in between the values of an arbitrary and perfect classification (AUC=0.5 and 1.0, respectively). Poor results were obtained for the remaining groups, which is due to the corresponding small number of isolates. These results demonstrate the potential of the predictive models to classify strains based on results of phenotypic screens. Bayesian classifier used the strains phenotypic profiles for prediction of commercial strains, and identified 3 of the 30 phenotypic tests (growth in musts containing iprodion (0.05 mg/mL), cycloheximide (0.1 μg/mL) or potassium bisulphite (150 mg/L)) as the ones providing more information for the assignment of strains to the commercial group. When using only 3 tests, rather than the entire phenotypic profile, the probability of a strain to be classified as commercial increases significantly (from 27% to 95%).

In conclusion, our results demonstrate the usefulness of computational approaches to describe phenotypic variability among groups of *S. cerevisiae* strains that also might occur as adaptive mechanisms in specific environments. The herein developed models can make predictions about the biotechnological potential of strains and simplify the selection of candidate strains to be used as commercial wine strains.

Chapter IV

Genetic characterization of a S. cerevisiae *collection and establishement of phenotype-genotype associations*

The work presented in this chapter has been published in:

Franco-Duarte R*, Mendes I*, Umek A, Drumonde-Neves J, Zupan B, Schuller D (2014) *Computational models reveal genotype-phenotype associations in* Saccharomyces cerevisiae. Yeast:10.1002/yea.3016

*contributed equally

Introduction

Studies relating yeast genetic and phenotypic variability will help to increase our understanding of the high diversity of strains, particularly in what regards the highly variable wine strains. They should also elucidate genetic adaptations involved in phenotypes that are relevant for survival in stressful industrial environments. The increase in knowledge obtained in these studies may also contribute towards strain improvement strategies through breeding and genetic engineering (Dequin and Casaregola 2011; Roberts and Oliver 2011; Borneman et al. 2013). *S. cerevisiae* species consists of both "domesticated" and "wild" populations, whereby the genetic divergence is associated with both ecology and geography. Sequence comparison of 70 *S. cerevisiae* isolates confirmed the existence of five well defined lineages and some mosaics, suggesting the occurrence of two domestication events during the history of association with human activities, one for sake strains and one for wine yeasts (Liti et al. 2009; Schacherer et al. 2009b; Liti and Schacherer 2011). *S. cerevisiae* isolates associated with vineyards and wine production form a genetically differentiated group, distinct from 'wild' strains isolated from soil and oak tree habitats, and also from strains derived from other fermentations, such as palm wine and sake or clinical strains. Recent research indicates that wine strains were domesticated from wild *S. cerevisiae* (Fay and Benavides 2005a; Legras et al. 2007b), followed by dispersal, and the diversifying selection imposed after yeast expansion into new environments due to unique pressures lead to strain diversity (Diezmann and Dietrich 2009; Dunn et al. 2012; Borneman et al. 2013). The interactions between *S. cerevisiae* and humans are considered as a driver of yeast evolution and the development of genetically, ecologically and geographically divergent groups (Legras et al. 2007b; Goddard et al. 2010; Sicard and Legras 2011). The limited knowledge about the mechanisms responsible for the fixation of specific genetic variants due to ecological pressures can be extended by combining genetic and phenotypic characteristics. Recent studies show that groups of strains can be distinguished on the basis of specific traits that were shaped by the species´ population history. Wine and sake strains are phenotypically more diverse than would be expected from their genetic relatedness, and the contrary is the case for strains collected from oak-trees (Kvitek et al. 2008). Wine yeasts and other strains accustomed to growing in the presence of musts with high sugar concentrations

are able to efficiently ferment synthetic grape musts, contrary to isolates from oak trees or plants that occur in environments with low sugar concentrations. Commercial wine yeasts were differentiated by their fermentative performances as well as their low acetate production (Camarasa et al. 2011). West African population shared low-performance alleles conferring unique phenotypes regarding mitotic proliferation under different stress resistance environments. Other phenotypes differentiated lineages from Malaysia, North America and Europe, whereby the frequency of population specific traits could be mapped onto a corresponding population genomics tree based on low coverage genome sequence data (Warringer et al. 2011).

The global genetic architecture underlying phenotypic variation arising from populations adapting to different niches is very complex. Most phenotypic traits of interest in *S. cerevisiae* strains are quantitative, controlled by multiple genetic loci referred to as quantitative trait loci (QTL). Genome regions associated with a given trait can be detected by QTL analysis, using pedigree information or known population structure to make specific crosses for particular phenotypes. The crosses are then genotyped using single nucleotide polymorphisms (SNPs) or other markers across the whole genome and statistical associations of the linkage disequilibrium between genotype and phenotype are identified (Dequin and Casaregola 2011; Liti and Louis 2012; Swinnen et al. 2012; Salinas et al. 2012b; Borneman et al. 2013). QTL mapping was successfully applied to dissect phenotypes that are relevant in winemaking such as fermentation traits (Ambroset et al. 2011a) or aromatic compounds production (Katou et al. 2009; Steyer et al. 2012). QTLs that were relevant for oenological traits and wine metabolites were mapped to genes related to mitochondrial metabolism, sugar transport and nitrogen metabolism. Strong epistatic interactions were shown to occur between genes involved in succinic acid production (Salinas et al. 2012b). The genotype-phenotype landscape has also been explored by several studies using statistical and probabilistic models (O'Connor and Mundy 2009; MacDonald and Beiko 2010; Mehmood et al. 2011), as well as gene knockout approaches (Hillenmeyer et al. 2008).

Current methods to infer genomic variation and determine relationships between *S. cerevisiae* strains include microsatellite analyses (Legras et al. 2005a; Franco-Duarte et al. 2009; Muller and McCusker 2009; Richards et al. 2009), detection of genetic alterations using comparative genome hybridization - aCGH (Winzeler et al. 2003; Kvitek

et al. 2008; Carreto et al. 2008; Dunn et al. 2012), and SNPs detection by tiling arrays (Schacherer et al. 2009b).

In a previous work Franco-Duarte et al. (2009) evaluated the phenotypic and genetic variability of 103 *S. cerevisiae* strains that were isolated from vineyards of the Vinho Verde wine region (Northwest Portugal). A set of 11 polymorphic microsatellite loci were used and through subgroup discovery-based data mining successfully identified strains with similar genetic characteristics (microsatellite alleles) that exhibited similar, mostly taxonomic phenotypes, allowing also to make predictions about the phenotypic traits of strains. In the present study, we aim to established computational associations in a larger collection of diverse *S. cerevisiae* strains (172) obtained from worldwide geographical origins and distinct technological uses (winemaking, brewing, bakery, distillery, laboratory, natural, etc.). In the study we use 30 physiological traits, most of them being important from an oenological point of view.

Material and Methods

Genetic characterization

The established collection (Chapter III) was characterized genetically, in adition to the previous phenotypic characterization. For this, 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 was isolation as previously described (Schuller et al. 2004) and used for microsatellite analysis.

Genetic characterization was performed using eleven highly polymorphic *S. cerevisiae* specific microsatellite loci: ScAAT1, ScAAT2, ScAAT3, ScAAT4, ScAAT5, ScAAT6, ScYPL009c, ScYOR267c, C4, C5 and C11 (Field and Wills 1998; Techera et al. 2001; Pérez et al. 2001; Legras et al. 2005b; Schuller et al. 2007; Schuller et al. 2012). Multiplex PCR mixtures and cycling conditions were optimized and performed in 96-well PCR plates as previously described (Franco-Duarte et al. 2009).

Statistical analysis

We have estimated the number of repeats for the alleles from each locus based on the genome sequence of strain S288c available in the *Saccharomyces* Genome Database (http://www.yeastgenome.org) and the results obtained for the size of microsatellite amplicons of this strain. Principal component analysis (PCA), available in the The Unscrambler® X software (Camo) was used for microsatellite variability analysis. A set of standard predictive datamining methods, as implemented in the Orange data mining suite (Demšar et al. 2013) were used to study the relations between the genetic constitutions of strains and their geographical origins or technological applications. Alleles that were present in less than five strains were removed, and *k*-nearest neighbour algorithm (kNN) (Tan et al. 2006) was used for inference. The modelling approach was tested in 5-fold cross validation, each time fitting the model on 80% of the data and testing it on the remaining 20%. Results were reported in terms of cross-validated area under the receiver operating characteristics curve (AUC), which estimates the probability that the predictive model would correctly differentiate between distinct technological applications of the strains (Hanley and McNeil 1982). The strength of associations between microsatellites and specific phenotypes was scored using information gain ratio as implemented in the Orange data mining suite, using default parameters, and significant findings were confirmed by permutation tests and estimation of false discovery rate. Detailed description of the computation data analysis can be found in (Franco-Duarte et al. 2014).

Results

The *S. cerevisiae* collection described in chapter III, was used in the present chapter for genetic characterization. All 172 strains were characterized regarding allelic combinations for previously described microsatellites ScAAT1, ScAAT2, ScAAT3, ScAAT4, ScAAT5, ScAAT6, ScYPL009c, ScYOR267c, C4, C5 and C11 (Field and Wills 1998; Techera et al. 2001; Pérez et al. 2001; Legras et al. 2005b; Schuller et al. 2007; Schuller and Casal 2007; Schuller et al. 2012). As shown in **Table IV-1**, a total of 280 alleles were obtained, and microsatellites ScAAT1 and ScAAT5 were the most and least polymorphic with 39 and 6 alleles, respectively.

Table IV-1: Summary of the distribution of alleles (indicated in numbers of repetitions) among 172 *Saccharomyces cerevisiae* strains, from 11 microsatellite loci.

Microsatellite designation	Total number of alleles (range of allele sizes in number of repeats)	Most frequent alleles	Number of strains in which the allele was obtained	Most variable alleles (number of repetitions) identified by PCA (Fig. IV-2)	Percentage of most variable alleles among the total number of alleles per locus	References [1]
ScAAT1	39 (6-54)	24 16	27 21	17; 21; 26; 28; 29; 34	15	*a; b*
ScAAT2	18 (5-22)	15 16 14 13	58 33 34 21	6; 8; 12; 13; 14	28	*b*
ScAAT3	19 (3-49)	16 14 22	45 32 28	11; 14; 16; 17; 21; 22	32	*b; c*
ScAAT4	17 (1-27)	20 11	100 22	7; 9; 10; 11; 20; 21	35	*b*
ScAAT5	6 (2-49)	9 10 8	80 63 37	8; 9; 10; 11	67	*b*
ScAAT6	10 (12-44)	16 17	124 40	16; 17; 25; 26; 28	50	*b*
C4	9 (16-61)	21 24 22	52 44 31	20; 21; 22; 23; 24	56	*d*
C5	19 (3-38)	4 3 12 13	31 25 23 22	3; 12; 13	16	*d*
C11	18 (1-47)	13 14 24	42 24 28	15; 23; 24	17	*d*
ScYPL009c	13 (57-86)	80 81 82 79 65	47 45 28 23 20	55; 58; 69; 70; 71; 72	46	*a; c*
ScYOR267c	12 (37-100)	52 56	52 24	52; 56; 62; 63; 67	42	*a; c*

[1] *a* - Techera et al. 2001; *b* - Pérez et al. 2001 *c* - Field and Wills 1998; *d* - Legras et al. 2005

Figure IV-1 shows the principal component analysis (PCA) plot of the microsatellite data illustrating the genetic diversity of the collection. Some patterns of genetic relatedness between strains sharing the same technological origin became evident as shown in the panel A. Sake strains (●) were located in the right part of the PCA plot, due to larger sizes of alleles of loci ScYOR267c and C4. For this group of strains, we have identified nine unique alleles, from which three were present in more than one strain and belong to three different loci (ScAAT6, C4 and ScYOR267c). Strains from fermented beverages other than wine were separated by PC-2, being located in the upper part of the PCA plot, indicating that they share a combination between smaller alleles of microsatellite C4 and bigger alleles of ScYOR267c. These twelve strains are marked in the PCA plot inside the area surrounded by a dotted line. Twelve unique alleles were found for these strains, two of them (C4-58 and ScYPL009c- 57) being present in six of the twelve strains. On the contrary, the group of wine strains (both natural isolates and commercial strains), showed heterogeneous distribution across the two components, being preferentially located in the left side of the PCA plot. The nine clinical strains were distributed across both components with no discriminant results in any locus. The 172 strains (scores) were also segregated in the first two components of the PCA constructed from the allelic combination for 11 loci. Loci ScYOR267c and C4 had the highest weight in strain variability, followed by ScYPL009c and ScAAT4, although within a smaller extent (panel B).

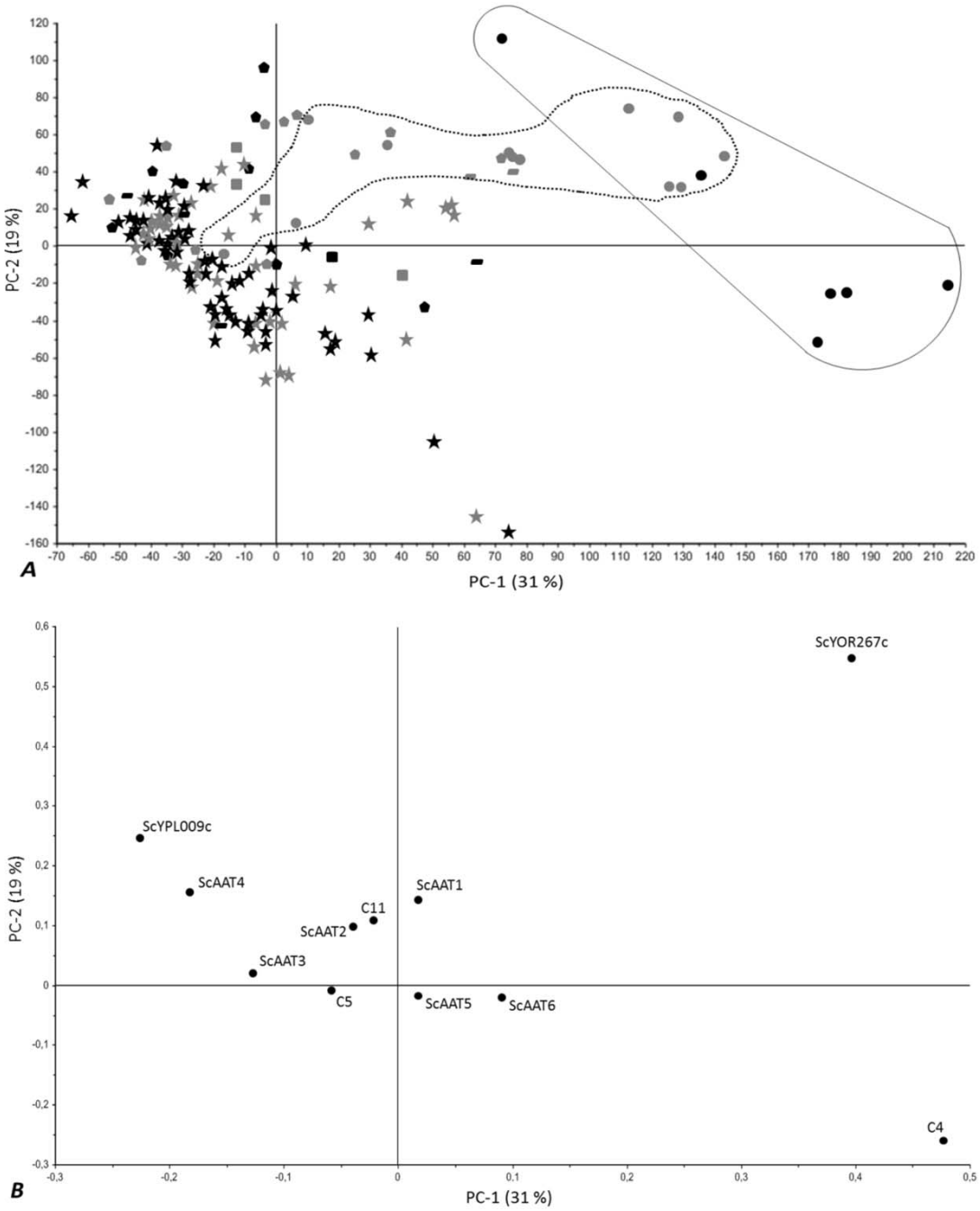


Figure IV-1: Principal component analysis of microsatellite data:

A: distribution of 172 strains according to their allelic combinations for 11 loci (scores); Symbols represent strains' technological applications or origin: ★ - wine and vine; ★ - commercial wine strain; ■ - clinical; ■ - natural isolates; ● - sake; ● - other fermented beverages; ⬟ - beer; ⬟ - bread; ▬ - laboratory; ▬ - unknown biological origin. Sake strains and strains from other fermented beverages are surrounded by full-lined and dotted lines, respectively.

B: contribution of microsatellite loci (loadings) to the separation of strains shown in panel A.

To reveal the weight of different alleles on the strains' genetic variability, the profile of the 11 microsatellites was represented for each strain as a vector where the values 0, 1 and 2 corresponded to the absence of an allele, the presence of a heterozygous allele or the presence of two copies of the allele, respectively. We assumed that all strains were diploid, because aneuploidy loci were rarely detected (< 3%). In addition, the DNA content of a representative set of homozygous strains corresponded to a diploid strain (flow cytometry analysis, data not shown). A total of 48160 data points were generated and the segregation of the 280 alleles in the two components of the PCA is shown in **Figure IV-2**. Alleles ScAAT4-20, ScAAT5-9 and ScAAT6-16 had highest weight in strain variability due to their positioning in the right and upper part of the PCA plot. Among the 11 microsatellite loci, 54 alleles were identified by PCA as contributing to the highest strain variability among 172 strains (**Table IV-1**). Loci ScAAT3, ScAAT4 and ScAAT5 were the ones with higher number of variable alleles (4), in opposition to loci ScAAT1, C5 and C11 with 1 allele each.

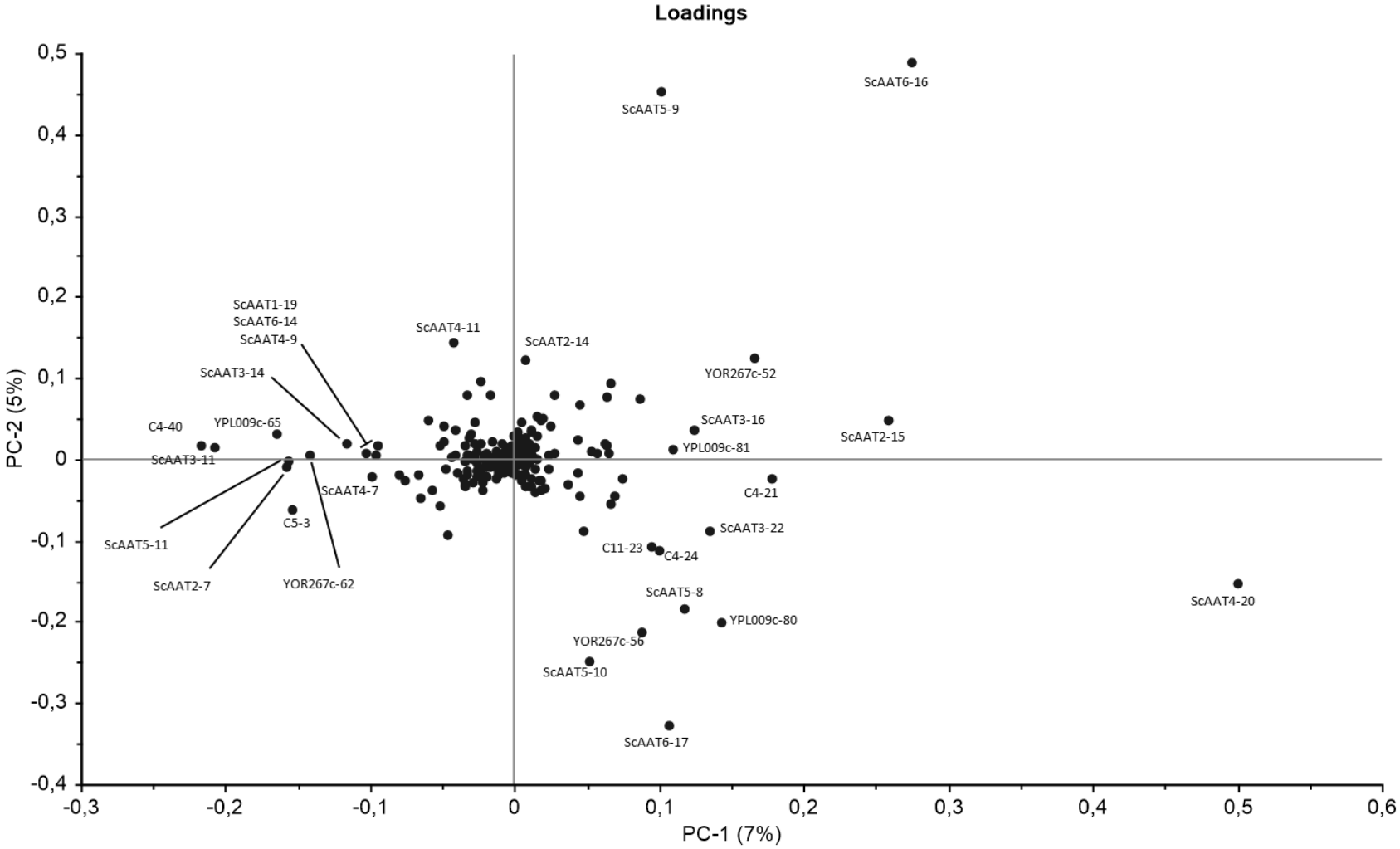


Figure IV-2: Principal component analysis of a Boolean matrix of 280 alleles from 11 microsatellites in 172 *Saccharomyces cerevisiae* strains.

In order to investigate further the relationships between strains, considering the microsatellite allelic variation found, we used the computational methods described in (Franco-Duarte et al., 2014). The results, summarized in the following paragraphs, showed that by using using computation models interesting assotiations could be made between the strains genetic and the phenotypic data (obtained in the previous chapter). In a first attempt we have examined the relations between strains technological group and the corresponding genotypes and scored them for their predictive value. Computational models were constructed to predict the strains' technological application or origin from microsatellite data. Details about the construction of the models and the data cleaning can be found in the methods section of (Franco-Duarte *et al.,* 2014). With these models, a confusion matrix was obtained (Table 2 – Franco-Duarte *et al*., 2014) with the *k*NN cross-validation classifications. Results show that for the strains derived from winemaking environments (commercial and natural wine strains), 47% and 72% of strains were correctly assigned, respectively. Interestingly, the majority of "false" assignments didn´t fall out of the wine strains group, occurring for commercial wine strains that were assigned to the natural wine strains (21 of 47 strains) or natural wine strains that were catalogued as commercial wine strains (16 of 74 strains). If all wine strains were grouped in one single category, the proportion of correct assignments would increase to 93% (112 of 121 strains). For the groups of strains isolated from sake, natural environments, other fermented beverages and bread, the proportion of correct assignments were 67%, 42%, 50% and 50% respectively. The high number of correct assignments even for small groups of strains, and a very high AUC score, both reinforce the validity of the modelling technique, confirming a strong relation between our genotype profiles and strain groups. On the other side and with only 22% of correct assignments, our approach was not successful on the identification of clinical strains, which was expected due to the absence of a common ancestor for this group and since pathogenic *S. cerevisiae* strains arise from different origins (Liti and Schacherer 2011).

Next, a major achievement obtained with the genetic and phenotypic characteriation results obtained in the present work, was the identification of subsets of strains sharing similar phenotypic results and allelic combinations. To test the associations between phenotypic results and microsatellite alleles we analysed pairwise relationships between corresponding variables (each microsatellite variable versus each phenotypic feature). Information gain ratio (IGR) was computed, between microsatellite predictor and binarized phenotypic response variable, and repeated again using permutated phenotypic data as described in the methods section of (Franco-Duarte *et al*., 2014). *p*-values were reported after correction using false discovery rate (FDR) procedure, and the pairs for which FDR was below 0.2 were identified (Figure 3 – Franco-Duarte *et al*., 2014). Significant associations were obtained between microsatellites ScAAT1, ScAAT2, ScAAT5, ScAAT6, ScYPL009c, C4 and C5, and for 13 phenotypic classes. For the classes in which significant associations with microsatellite alleles were found, between 1 and 8 relations were established with a particular microsatellite allele (numbers following black circles). For nine phenotypic tests and classes a single association was established: "40 ˚C = 1", "40 ˚C = 3", "SDS (0.01%, w/v) = 0", "$KHSO_3$ (150 mg/L) = 2", "Ethanol 10%, v/v (liquid medium)= 0", "Ethanol 10%, v/v (liquid medium)= 2", "Ethanol 10%, v/v (liquid medium)= 3", "Ethanol 12%, v/v + $Na_2S_2O_5$ 75 mg/L (solid medium) =1" and "wine supplemented with glucose 1% = 0". The phenotypes with the highest number of allelic associations were "$KHSO_3$ (300 mg/L) = 3" and "galactosidase activity = 1", with 8 associated alleles each. Twenty-two microsatellite alleles had an association with at least one phenotype. For two alleles, three significant associations were obtained (ScAAT2-13 and C4-21), being the highest number of associations with phenotypes (7) found for microsatellites ScAAT1 and ScAAT2, in opposition to ScAAT5, ScAAT6 and ScYPL009c with only 3 links established, each. These numbers are not related with the total number of alleles and the range of allele sizes shown in **Table IV-2**.

Discussion

In the present work, we aimed to assess if associations could be made between strains genotypes and phenotypes in a worldwide collection of 172 *S. cerevisiae* strains from different geographical origins and technological uses (winemaking, brewing, bakery, distillery, laboratory, natural, etc.). The established collection revealed high genetic diversity (**Figure IV-1, Figure IV-2 and Table IV-1**), with a total of 280 alleles obtained with 11 polymorphic microsatellites. Microsatellite ScAAT1 was the most polymorphic one with 39 alleles, followed by ScAAT3 and C5 with 19 alleles each, confirming the data of our previous study (Franco-Duarte et al. 2009). PCA components of **Figure IV-2** explains only a small part of the total variance (PC-1 – 7% and PC-2 – 5%) which seems to indicate that all the microsatellite alleles are important to differentiate between strains, but also revealed a group of 54 alleles that are the most relevant to explain variability among strains. Herein, we also observed some patterns of distribution according to the strains technological application or origin, when considering the PCA of genetic data, in particular for sake strains and strains from fermented beverages other than wine. Clinical strains, that are opportunistic environmental strains colonizing human tissues (Schacherer et al. 2007; Muller and McCusker 2009) didn´t show any discriminant distribution with PCA, which was expected, because they do not share a common ancestor (Liti and Schacherer 2011). Sake strains and strains obtained from fermented beverages other than wine showed some unique alleles in loci ScAAT6, C4, ScYOR267c and ScAAT1, ScAAT5, ScAAT6, C4, ScYPL009c, ScYOR267c respectively. These results highlight the existence of alleles that are representative of a specific technological group, which justifies the approach used in this research.

In our study, we demonstrate that strains' allelic combination and the respective technological application or origin (**Table IV-2**) are strongly related, as the later can be predicted from the proposed genotypic characterization. Regarding winemaking strains (both natural and commercial) the approach was able to predict the technological application or origin for 93% of the strains. The AUC score of the model was 0.802, between the values of an arbitrary and perfect classification (AUC=0.5 and 1.0, respectively) and can be considered as moderately high (Mozina et al. 2004). These results demonstrate the potential of the approach to predict the technological origin of a strain from the entire microsatellite

profile, even for groups of strains with small sample size (sake or bread, 6 and 4 strains, respectively). The genetic and phenotypic profile of strains obtained with 11 markers and 30 phenotypic tests was used to computationally score and rank genotype-phenotype associations. Associations were scored using information gain ratio (Quinlan 1986) and significant results were shown in form of *p*-value after false discovery rate procedure. Thirty two associations, representing thirteen phenotypic classes and 22 microsatellite alleles were significantly established. The phenotypic classes with more associations were related with high capacity to resist to the presence of $KHSO_3$ during fermentation, and to the galactosidase activity. These two phenotypes were associated with 8 alleles each. These results are valuable to select strains that are resistant to sulphur dioxide, an antioxidant and bacteriostatic agent used in vinification (Beech and Thomas 1985), and that were tested by the capacity of strains to grow in a medium supplemented with KHSO3. The association between 8 alleles and the strains moderate galactosidase activity, although not directly related with winemaking, could be also a beneficial criterion to choose *S. cerevisiae* strains capable of hydrolyse galactose, in alternative to the use of glucose as carbon source, pointing to an improved evolutionary capacity of these strains. The most polymorphic locus ScAAT1, revealed also the highest number of associations with phenotypes, but this was not observed for other polymorphic loci. Seven phenotype-genotype associations were found for each of the alleles ScAAT2–13 and C4–21, which can be considered as the most informative to predict strains' biotechnological potential regarding the associated phenotypes.

The prediction of the technological group from allelic combinations and the presence of statistically significant associations between phenotypes and alleles both demonstrate that computational approaches can be successfully used to relate genotype and phenotype of yeast strains. Microsatellite analysis revealed to be an efficient marker to evaluate genetic relatedness in yeasts and can be employed in the industry as a quick and cheap analysis. Although microsatellite analysis is the most accurate method for *S. cerevisiae* strain characterization, the 11 tested microsatellites are spread on only 9 chromosomes and might provide for a rather coarse representation of a genotype. Taking into account that the discovered associations apply to a smaller fraction of the genome, this study could be beneficially complemented with an extended search to monitor other genomic regions. These findings may become particularly important for the simplification of strain selection

programs, by partially replacing phenotypic screens through a preliminary selection based on the strain's microsatellite allelic combinations.

Chapter V

Integrating transcriptomics and metabolomics for the analysis of the aroma profiles of wine, sake, cachaça and laboratory Saccharomyces cerevisiae *strains*

The work presented in this chapter is being prepared to be submitted:

Inês Mendes, Isabelle Sanchez, Ricardo Franco-Duarte, Jean-Roch Mouret, Carole Camarasa, Dorit Schuller, Sylvie Dequin, Maria João Sousa (2015) *Integrating transcriptomic and metabolomic for the analysis of the aroma profiles of wine, sake, cachaça and laboratory* Saccharomyces cerevisiae *strains.* To be submitted.

Introduction

Wine flavour is the result of the interactions between grape must components and compounds originated from microbial metabolism. Grape must is constituted by three functional groups of compounds: nutrients, flavour precursors and flavour-active non-precursors. The action of yeasts on some of these compounds, will build the architecture of the wine flavour and their fermentation bouquet. Over the past 30 years, the huge increase in the understanding of *Saccharomyces cerevisiae* metabolism, namely of industrial yeast strains (Chambers and Pretorius 2010) has revealed its crucial role in the development of the wine secondary aroma, with higher alcohols, acetate esters and ethyl esters being the main aromatic compounds contributing to a floral and fruity aroma (Lambrechts and Pretorius 2000b). Generally, wine yeast strains can be responsible for "fruity", "floral", "neutral", or "cheesy"–"rancid" wine aromas, depending on their capacity to produce esters, higher alcohols, and volatile fatty acids (Cordente et al. 2012). The selection of the best wine yeast depends essentially on its oenological/phenotypic characteristics, such as fermentative rate, tolerance to ethanol and to SO_2, response to temperature, flocculent characteristics, the presence of killer factor, ethanol yield, malic acid metabolism and the production of several fermentation by-products, such as acetic acid, H_2S, higher alcohols, glycerol and acetaldehyde (Robinson 1994; Mannazzu et al. 2002; Schuller 2010; Bird 2013). Some studies showed that wine strains adapt to specific oenological environments during their selection for biotechnological purposes, which is reflected in their transcriptome, proteome and metabolome (Rossouw et al. 2008; Rossouw et al. 2009; Rossouw et al. 2010). Many researchers have studied the influence in the fermentation process of manipulating single genes through their deletion or over-expression, in order to clarify pathways involved in winemaking (Teixeira et al. 2009; Gómez-Pastor et al. 2010; López-Malo et al. 2013; Si et al. 2014). Nevertheless, the genomes of wine strains frequently show heterozygosity, SNPs and variation in gene copy number that makes difficult to perform whole-genome screenings (Pretorius 2000; Borneman et al. 2011; Borneman et al. 2012). On the other hand, transcriptome studies have been implemented using industrial yeast strains under winemaking conditions. These studies include gene expression analyses during alcoholic fermentation (Rossignol et al. 2003; Varela et al. 2005; Marks et al. 2008; Rossouw et al. 2009) and during exposure to a diversity of stresses such as high ethanol concentrations

(Alexandre et al. 2001), low temperature (Pizarro et al. 2008), and high-sugar concentrations (Erasmus et al. 2003). Gene expression is variable among wild-type yeast strains and it was shown that differences in gene expression during fermentation affected co-regulated genes and distinguished yeast strains (Carreto et al. 2011). Besides, winemaking strains deal better with stress-imposing environmental conditions and are able to manage nutrient deficiencies, such as nitrogen, in a more efficient and resourceful way suggesting a better adaptation to the specific stresses imposed. In order to understand the wine yeast aromatic profile, several metabolomics tools are available and are commonly used. The study of metabolome includes the analysis of a wide variety of chemical compounds, usually present at very low concentrations, which is a major barrier for appropriate bioanalytical approaches. The analysis of the metabolic profile has been performed using several analytical platforms, such as gas-chromatography (GC) or liquid-chromatography (LC) coupled to mass-spectroscopy (MS) (Birkemeyer et al. 2003; Kleijn et al. 2007; Fiehn 2008), capillary electrophoresis (CE) coupled to MS (Soga et al. 2003; Tanaka et al. 2007; Monton and Soga 2007; Ramautar et al. 2009), infrared and Raman spectroscopy (Ellis and Goodacre 2006), nuclear magnetic resonance (NMR) spectroscopy (Salek et al. 2007; Barton et al. 2008; Bjerrum et al. 2010) and direct injection MS (DIMS) (Allen et al. 2003; MacKenzie et al. 2008). GC-MS analysis has been one of the best accepted approaches to study wine metabolome, with several advantages: sensitivity, robustness, easiness of use, low cost and ample linear range (Villas-Bôas et al. 2005; Hollywood et al. 2006; Dettmer et al. 2007; Garcia et al. 2008). GC-MS combines advantages of both technologies: while MS provides individual mass spectra that can differentiate between chemically diverse metabolites, GC has high separation efficiency. The integration of the several "omic" approaches could be used to understand the variability existing within *S. cerevisiae* strains and to explore the molecular mechanisms underlying that variability.

In the present work we performed a comparative transcriptomic analysis of four *S. cerevisiae* strains from different origins and/or technological applications, wine, sake, cachaça and laboratory, at two time points during a must fermentation process and analysed the aroma profile of each strain in order to establish a correlation between gene expression and metabolite production.

Material and Methods

Yeast strains and culture media

Four *Saccharomyces cerevisiae* strains were used in this study, in particular the commercial strain Zymaflore® VL1 (Laffort oenologie®), the cachaça strain Z63 (kindly provided by Rogélio Brandão), the sake strain Z23 (K11, kindly provided by Gianni Liti) (Liti et al. 2009) and the laboratory strain S288c. Strains were chosen from a larger collection as being from heterogeneous origins in order to elucidate differences between the groups. Strains were grown at 28 °C, and routinely maintained at 4 °C on YPD plates containing 2% glucose (w/v), 2% peptone, 1% yeast extract and 2% agar, and in glycerol (30% v/v) stocks at -80 °C.

In this study, we used a natural must and a synthetic culture medium. The natural must was harvested in 2012 in the south of France (*Maccabeu*), flash-pasteurized and stored under sterile conditions. It contained 211 g/L of sugar and 213 mg/L of assimilable nitrogen. As a synthetic must, the MS300 (MS) medium (Bely et al. 1990) was used due to the fact that it mimics the grape musts to prepare the cells for fermentation. We inoculated 50 mL flasks containing 30 mL of YPD with cells from a Petri dish with YPD and incubated them overnight at 28 °C under stirring. Cells were then transferred to 1 L flasks containing 500 mL of MS medium in a final concentration of 2×10^6 cells/mL and incubated at 28 °C with continuous stirring. The fermentation cultures in MS medium were inoculated with 2×10^6 cells/mL in 1.1 L fermentors containing 900 mL of natural must.

Must fermentations

Fermentations were performed in 1.1 L fermenters (NH verre) equipped with a fermentor condenser, at 20 °C, stirred continuously (100 rpm) and linked to a mass flow meter that measured the CO_2 release rate online. CO_2 release was determined by automatic measurements of fermentor weight every 20 min. The rate of CO_2 production, dCO_2/dt, is the first derivative of the amount of CO_2 produced over time and was calculated automatically by polynomial smoothing of the CO_2 production curve (Sablayrolles et al. 1987). Fermentation experiments were performed in triplicate.

Metabolite analysis

Glucose, glycerol, ethanol, pyruvate, succinic, acetic and α-ketoglutaric acids levels were analysed by high-pressure liquid chromatography (HPLC), with an Rezex ROA - Organic Acid column (Phenomenex) at 45 °C. The column was eluted with 4 mM H_2SO_4 at a flow rate of 0.6 mL/min. Dual detection was performed with a refractometer and a UV detector (Agilent).

Volatile aroma compounds analysed by GC-MS after extraction as previously described (Rollero et al. 2014). Briefly, deuterated internal standards (100 μg/L) were added to samples (5 mL) before twice extraction using 1 mL of dichloromethane. The organic phases were dried over anhydrous sodium sulphate and concentrated under nitrogen flux. Extracts were analysed with a Hewlett Packard (Agilent Technologies, Santa Clara, California, USA) 6890 gas chromatograph coupled to a HP 5973 mass spectrometer.

RNA isolation and sample labelling

Cells were harvested by centrifugation (8000 rpm, 4 minutes at 4 °C), at two time points: 5 g/L and 50 g/L of CO_2 released. Total RNA was isolated using Trizol (TRI Reagent®, Sigma-Aldrich®), purified by isopropanol precipitation and then again with RNeasy kit (Qiagen). Cy3-labeled total RNA was synthesized with the One colour RNA Spike-In kit (Agilent Technologies) and purified with RNeasy kit (Qiagen). Quality and quantity of RNA were controlled at each step by spectrometry (NanoDrop 1000, Thermo Scientific) and using the Agilent Bioanalyzer 2100. Agilent gene expression microarrays 8x15k were used for the microarray hybridization, with one-colour method (Agilent Technologies, Santa Clara, CA, USA). Probe preparation and hybridization to Affymetrix Genechip microarrays were performed according to Affymetrix instructions. 600 ng of labelled RNA were hybridized for 17 h in 65 °C in a rotative hybridization oven (Corning) using the Expression Hybridization kit (Agilent Technologies, 5188–5242). Plates were washed with Expression Wash Buffer kit (Agilent Technologies, 5188–5325 5188–5326). Array pictures were analysed on a GenePix 4000B laser Scanner (Axon Instruments) and with the GenePix PRO7 software.

Statistical analysis

Statistical analyses were performed using R software, version 3.0.3 (The R Core Team 2013).To obtain a general overview of the production of volatiles compounds during the fermentation for each stage of fermentation (T1 and T2), principal component analysis (PCA) was performed using the FactoMineR package (Husson et al. 2012).

The limma package (Smyth and Speed 2003) was used to import and normalize the global microarray data (quantile method for normalization between arrays). For each studied time of released CO_2 (T1 and T2) and based on this normalized dataset of 6200 expression data for the 4 strains, we used a sparse partial least square – discriminant analysis (sPLS-DA), an exploratory approach in a supervised context in order to select the most important transcripts relative to the 4 strains (Lê Cao et al. 2011). We tuned the number of dimensions of the sPLS-DA to 2 and the number of variables to choose on these 2 dimensions to 400 (200 for each).

A functional analysis was performed on the selected transcripts by time point, in order to highlight significant functional groups according to the Gene Ontology (GO) process terms using the GeneCodis program with the FDR method at a p value cutoff of 0.05 (Nogales-Cadenas et al. 2009).

Always for each time point, a multivariate factorial analysis (MFA) was then performed to obtain an overview of the dataset, which consisted in 433 variables measured for 4 strains (S288c, VL1, cachaça, sake). The data set included a set of individuals described by two types of variables: the normalized expression of the 400 transcripts selected by the sPLA-DA according to the 4 strains, and the 33 volatile compounds produced during the fermentation by the 4 strains. The MFA takes into account the structure of the two groups of data and balances the influence of each group of variables. This enables the study of links between expression data and volatile compounds production (Husson et al. 2012).

Microarray data accession numbers. The complete data set is available through the Gene Expression Omnibus (GEO) database. The microarray description is under GEO accession number GPL16244.

Results and discussion

Fermentative profiles and metabolic characterization

Four *Saccharomyces cerevisiae* strains were used in this study; three isolated from different fermentative environments, namely cachaça Z63, sake Z23, the commercial wine yeast VL1, and the laboratory reference strain S288c. These strains were previously characterized genetically and phenotypically (Mendes et al. 2013; Franco-Duarte et al. 2014) and were selected from a larger yeast collection based on their dissimilarities (Mendes et al. 2013). Triplicate fermentations were carried out with each of the four strains using natural must Maccabeu. The fermentation performance of the strains is presented in **Figure V-1**, in which each curve represents the average debit of CO_2 from the three replicates for each strain. With the exception of the laboratory strain, for which a slower fermentation and a lower maximum fermentation rate were obtained, the remaining three strains present a similar fermentative profile with a V_{max} between 1.2 and 1.4 fg/L/h of CO_2 released.

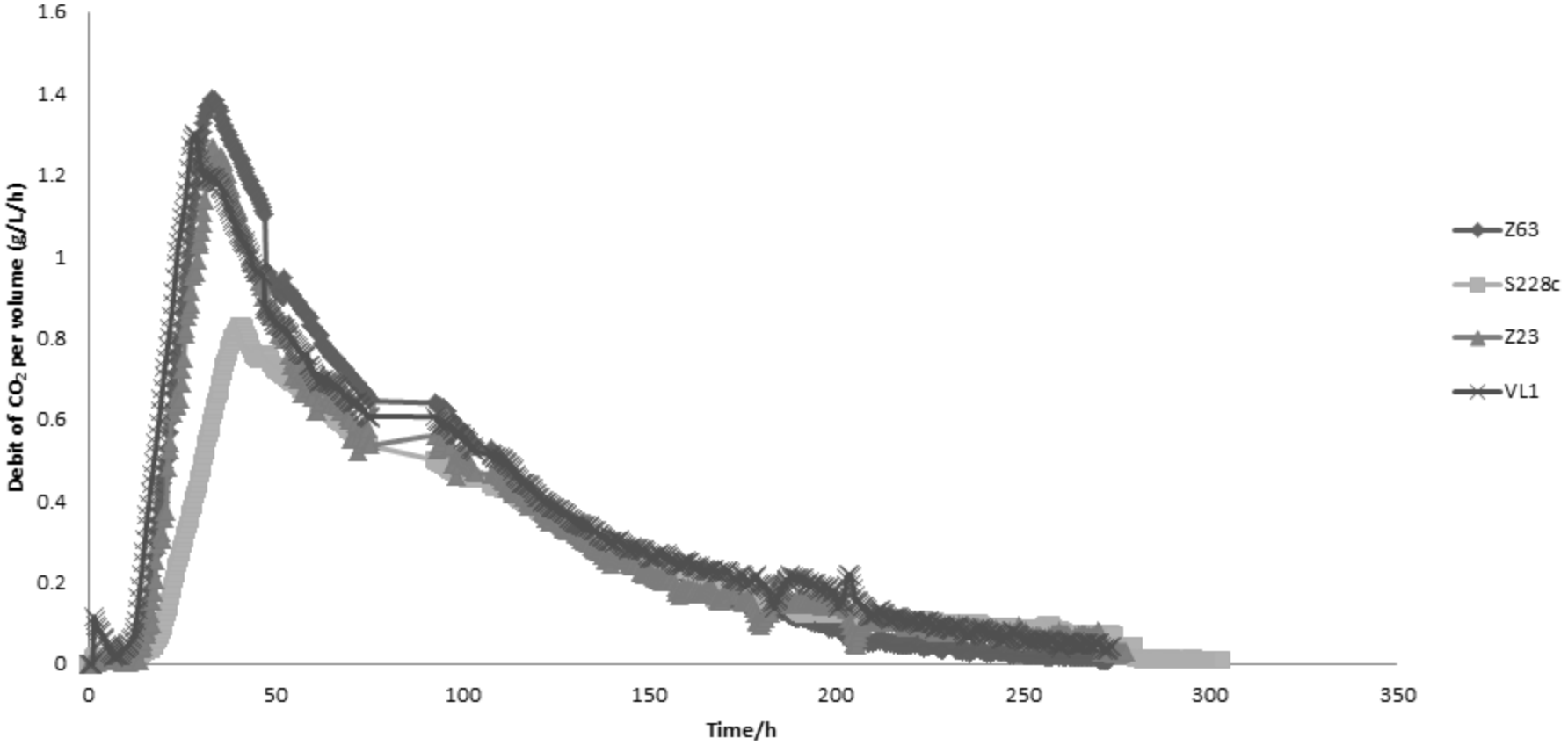


Figure V-1: Fermentation profiles of the four strains used in this study in respect to debit of CO_2 per volume (g/L/h) per time (h-hours). Values are the averages from 3 biological replicates. Fermentations were carried out at 20 ºC (100 rpm) using Maccabeu grape must.

In order to obtain a characterization of their metabolic profile, high-performance liquid chromatography (HPLC) and gas chromatography – mass spectrometry (GC-MS) analysis were performed with samples from two time points of fermentation: exponential phase (T1, 5 g/L of CO_2 released) and stationary phase (T2, 50 g/L CO_2 released). Thirty-eight compounds were quantified including 11 ethyl esters, 7 acetate esters, 3 organic acids, 5 higher alcohols, 9 volatile fatty acids, propanol, succinic acid, acetic acid, pyruvate and alpha-ketoglutarate (**Supplementary data S5**).

PCA analysis based on the compounds quantified both by HPLC and GC-MS (**Figure V-2**) showed intra-strain differences, with a discrimination of the laboratory strain from the other three strains at T1 (**Figure V-2A**) and T2 (**Figure V-2C**). Circles of correlation (**Figures V-2B, V-2D**) show the contribution of each quantified metabolic compounds to the separation of the strains in the scores plot. Only the first two components were considered, since they explain a high percentage of the variability found between isolates and between compounds: 83.6% and 84.3% for T1 and T2, respectively. At T1 (**Figures V-2A and V-2B**), a clear differentiation between laboratory strain and the other three strains was obtained according to the first axis. Productions of acetate esters (green) and of some higher alcohols (blue) were positive contributors to this axis while formation of medium chain fatty acids (hexanoic, octanoic and decanoic acids) was negatively involved. Strain Z63, having its origin in the fermentative beverage cachaça, distinguished along the second axis, by a higher production of ethyl decanoate, ethyl octanoate and ethyl butanoate compared with other tested strains.

At time-point T2, corresponding to the stationary phase of fermentation, a similar scenario was observed, with a clear separation of laboratory strain S288c from the others according to the first axis and a separation of strain Z23 (sake) from strains Z63 and VL1 along the second one. However, the major contributors to the two axes differed between the two time points. During the stationary phase, fermentation by strains Z63, Z23 and VL1 produced higher amounts of almost all metabolites assessed then the laboratory strain: acetate esters, ethyl esters, the majority of the acids apart from decanoic and propanoic acids and most of higher alcohols except propanol (first axis). From the three ethyl esters produced highly by cachaça strain at T1, only ethyl butanoate was again responsible for the separation of this strain from strains VL1 and Z63 (second axis).

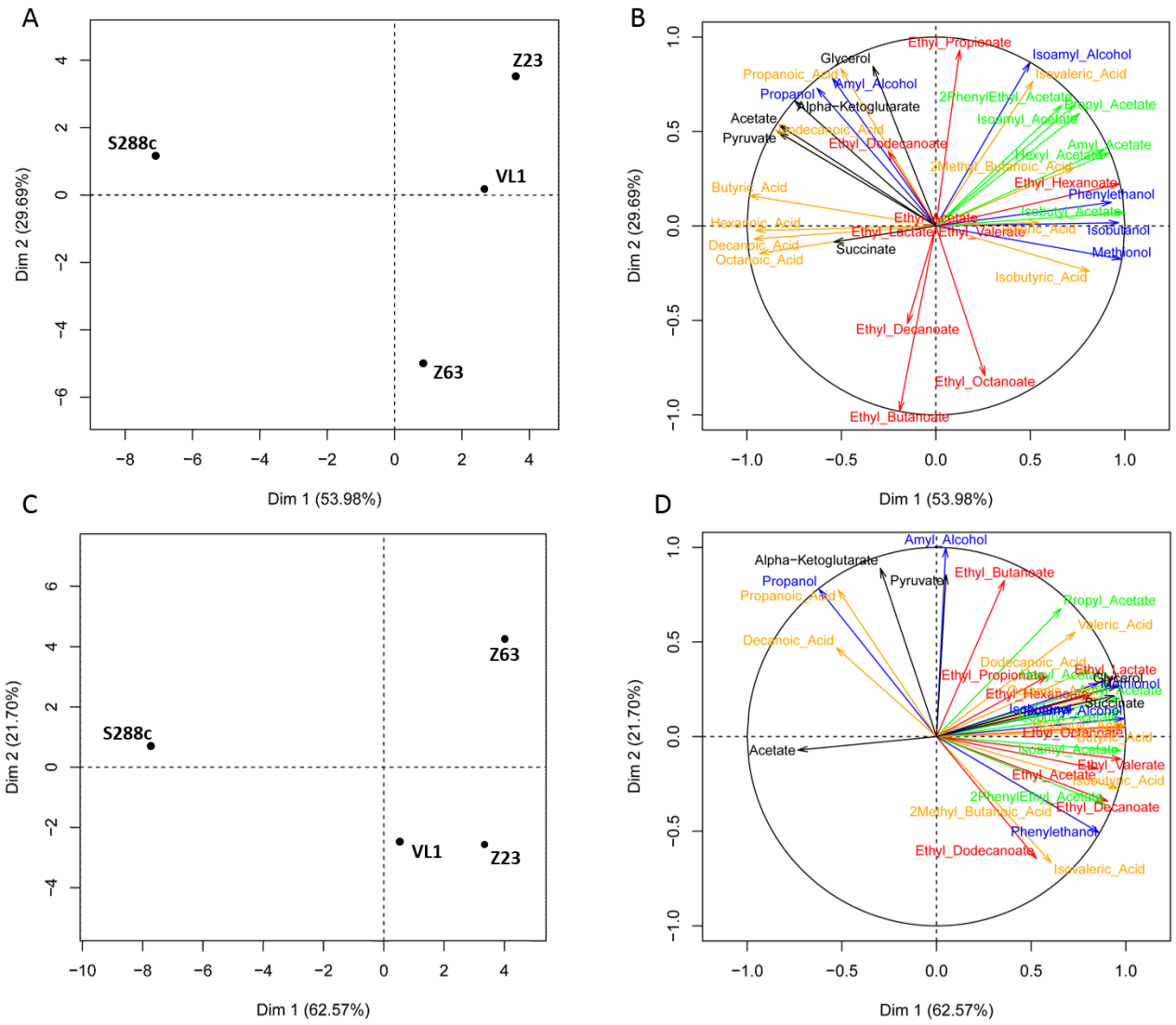


Figure V-2: Principal component analysis of GC-MS and HPLC data for the four strains tested:
A – four *S. cerevisiae* strains (scores) analysed by GC-MS and HPLC at T1 (5 g/L)
B – concentration of compounds detected by HPLC and GC-MS at T1 (5 g/L)
C - four *S. cerevisiae* strains (scores) analysed by GC-MS and HPLC at T2 (50 g/L)
D – concentration of compounds detected by HPLC and GC-MS at T2 (50 g/L)

Our results show that at the two time points considered in this work, the compounds contributing the most to the strains separation in comparison with S288c were the acetate and ethyl esters and the higher alcohols. It is well known that higher alcohols have positive effect on wine aroma as well (Swiegers et al. 2005b; Cordente et al. 2012). In the same way esters, produced by yeasts during alcoholic fermentation, have a significant influence on the fruity aromas of the final product, both in the case of ethyl fatty acid esters and acetate esters (Mason and Dufour 2000; Ribéreau-Gayon et al. 2000a). So, the results indicate that must fermentations carried with yeasts isolated from any of the three wild fermentative environments will be characterized by a higher development of the "yeast bouquet" and originate wines with much more complex aroma and flavour, than the laboratory strain used as reference. In addition, sake strain will give the wine aroma profile most close to the wine strain. In the case of volatile fatty acids, their concentration varied from 82 to 220 mg/L at T1 and 81 to 289 mg/L at T2, influencing also the PCA position of the analysed strains. The concentration of volatile acids is of particular relevance once they are 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 et al. 2011), mainly due to the inhibition of their esters hydrolysis.

Comparative transcriptomics

Comparative transcriptomics of the three *S. cerevisiae* strains isolated from the different fermentative environments in comparison with the reference yeast S288c was conducted using Affymetrix Genechip microarrays, containing 70 mer probes targeting the sequences of the laboratory strain S288c. mRNA samples were collected at the two time points T1 and T2, as explained in the previous section.

Tables V-1.1, V-1.2, V-2.1 and **V-2.2** summarize the main findings obtained with transcriptomic characterization of the three fermentation isolates, in comparison with laboratory strain S288c. Results were analysed using Funspec with Bonferroni correction ($p<0.05$), and down or upregulated genes are indicated for the three strains in comparison with S288c, both at T1 (**Tables V-1.1 and V-1.2**) and T2 (**Tables V-2.1 and V2.2**). Genes were categorized in accordance with MIPS Functional Catalogue (Ruepp et al. 2004), and the ones common to the three strains are underlined.

Table V-1.1: Categorization of genes with significantly decreased expression (Bonferroni p<0.05) in Z63, Z23 and VL1 strains in comparison to S288c, at T1 (5 g/L of CO_2 released). Genes common to the three strains are underlined.

MIPS functional category	Strain		
	Z63	**Z23**	**VL1**
pheromone response, mating-type determination, sex-specific proteins	*AFR1 ASG7 BAR1 DIG1 EXG1 FAR1 FUS1 FUS3 GIC2 GPA1 HO MFA1 MFA2 PHO81 PRY1 RDH54 SPA2 SST2 STE18 STE2 STE23 STE4 STE5 STE6 UBC4*	*AGA1 ASG7 BAR1 DIG1 FAR1 FUS1 FUS3 GIC2 GPA1 HO HSP82 MFA1 MFA2 PHO81 PRY1 RDH54 SST2 STE18 STE2 STE23 STE4 STE5 STE6*	*AFR1 ASG7 ASH1 BAR1 BEM1 CLN2 DIG1 FAR1 FUS1 FUS3 GFA1 GIC2 GPA1 HO MCK1 MFA1 MFA2 PHO81 PRY1 RDH54 SAG1 SAN1 SIR2 SST2 STE18 STE2 STE23 STE4 STE5 STE6*
degradation of asparagine	*ASP3-1 ASP3-2 ASP3-3 ASP3-4*	*ASP1 ASP3-1 ASP3-2 ASP3-3 ASP3-4*	*ASP1 ASP3-1 ASP3-2 ASP3-3 ASP3-4*
metabolism of aspartate	*ASP3-1 ASP3-2 ASP3-3 ASP3-4*	*ASP1 ASP3-1 ASP3-2 ASP3-3 ASP3-4*	*ASP1 ASP3-1 ASP3-2 ASP3-3 ASP3-4*

Table V-1.2: Categorization of genes with significantly increased expression (Bonferroni p<0.05) in Z63, Z23 and VL1 strains in comparison to S288c, at T1 (5 g/L of CO_2 released). Genes common to the three strains are underlined.

MIPS functional category	Strain		
	Z63	**Z23**	**VL1**
electron transport and membrane-associated energy conservation	*ATP20 COR1 COX1 COX13 COX3 COX5A COX6 COX7 CYB2 CYC1 CYC7 MCR1 NDE1 NDI1 QCR10 QCR2 QCR6 QCR7 QCR8 QCR9 RIP1*	*COB COX1 COX3 COX5A CYB2 CYC1 CYC7 NDE1 NDI1 QCR8 QCR9 RIP1*	*ATP20 COX1 COX13 COX5A COX6 COX7 CYB2 CYC1 CYC7 NDI1 QCR10 QCR7 QCR8 QCR9 RIP1*
aerobic respiration	*COR1 COX1 COX13 COX23 COX3 COX5A COX6 COX7 CYT1 ISF1 MAM33 MBR1 NDE1 NDI1 PET10 PET9 QCR10 QCR2 QCR6 QCR7 QCR8 QCR9 RIP1*	*COB COX1 COX16 COX23 COX3 COX5A CYT1 MAM33 NDE1 NDI1 QCR8 QCR9 RIP1 YDR115W*	*COX1 COX13 COX16 COX23 COX5A COX6 COX7 CYT1 MRPL1 NDI1 QCR10 QCR7 QCR8 QCR9 RIP1 YDR115W*
tetracyclic and pentacyclic triterpenes (cholesterin, steroids and hopanoids) metabolism	*ARE2 ERG1 ERG10 ERG13 ERG2 ERG27 ERG28 ERG5 ERG7 ERG9 HMG1 IDI1 MCR1 MVD1 NSG2 OSH6*	*ARE2 ERG1 ERG13 ERG2 ERG28 ERG5 ERG9 HMG1 IDI1 MVD1 NSG2 OSH6*	*ARE2 ERG1 ERG2 ERG27 ERG28 ERG5 ERG9 HMG1 IDI1 MVD1 NSG2*
mitochondrion	*CYB2 HSP10 MBR1 MDM35 MDV1 MNP1 MRM2 MRP2 MRP21 MRPL13 MRPL20 MRPL23 MRPL32 MRPL35 MRPL37 MRPL38 MRPL39 MRPL40 MRPL44 MRPL6 MRPL9 MRPS28 NDE1 NDI1 RSM25 TIM10 YMR31*	*CYB2 HSP10 MNP1 MRP13 MRP2 MRP21 MRPL10 MRPL13 MRPL19 MRPL20 MRPL23 MRPL27 MRPL32 MRPL35 MRPL37 MRPL38 MRPL39 MRPL4 MRPL40 MRPL44 MRPL6 MRPL8 MRPL9 MRPS16 MRPS28 NAM9 NDE1 NDI1 RML2 RSM18 RSM19 RSM25 RSM26 TIM10 YDR115W YMR31*	*CYB2 GET1 HSP10 MDM35 MDV1 MNP1 MRM2 MRP13 MRP2 MRP21 MRP49 MRPL1 MRPL10 MRPL13 MRPL20 MRPL23 MRPL27 MRPL32 MRPL35 MRPL36 MRPL37 MRPL38 MRPL39 MRPL40 MRPL44 MRPL49 MRPL6 MRPL9 MRPS16 MRPS28 NDI1 PET18 RSM18 RSM19 RSM25 RSM26 SAM37 TIM10 TIM12 TIM9 YDR115W YMR31*

Table V-1.2 (cont.)

MIPS functional category	Strain		
	Z63	**Z23**	**VL1**
ribosomal proteins	-	*MNP1 MRP13 MRP2 MRP21 MRPL10 MRPL13 MRPL19 MRPL20 MRPL23 MRPL27 MRPL32 MRPL35 MRPL37 MRPL38 MRPL39 MRPL4 MRPL40 MRPL44 MRPL6 MRPL8 MRPL9 MRPS16 MRPS28 NAM9 RML2 RPL19A RPL22A RPL34A RPL36A RPS10B RPS14B RPS17A RPS21B RPS24A RPS27A RSM18 RSM19 RSM25 RSM26 YDR115W YMR31*	*MNP1 MRP13 MRP2 MRP21 MRP49 MRPL1 MRPL10 MRPL13 MRPL20 MRPL23 MRPL27 MRPL32 MRPL35 MRPL36 MRPL37 MRPL38 MRPL39 MRPL40 MRPL44 MRPL49 MRPL6 MRPL9 MRPS16 MRPS28 RPL11A RPL11B RPL18B RPL19A RPL19B RPL20B RPL22A RPL23A RPL26A RPL27A RPL30 RPL33B RPL34A RPL35B RPL36A RPL38 RPL40A RPL43B RPL9B RPP1A RPS10A RPS10B RPS14B RPS16A RPS17A RPS18B RPS21A RPS21B RPS24A RPS24B RPS25A RPS26A RPS27A RPS30A RPS30B RPS6B RPS8B RPS9A RSM18 RSM19 RSM25 RSM26 SWS2 YDR115W YMR31*
fermentation	-	*AAD16 AAD4 AAD6 ADH7 ALD2 MSC7 YPL088W*	

Table V-2.1: Categorization of genes with significantly decreased expression (Bonferroni p<0.05) in Z63, Z23 and VL1 strains in comparison to S288c, at T2 (50 g/L of CO_2 released).

MIPS functional category	Strain		
	Z63	**Z23**	**VL1**
degradation of asparagine	-	*ASP1 ASP3-1 ASP3-2 ASP3-3 ASP3-4*	*ASP1 ASP3-1 ASP3-2 ASP3-3 ASP3-4*
metabolism of aspartate	-	*ASP1 ASP3-1 ASP3-2 ASP3-3 ASP3-4*	*ASP1 ASP3-1 ASP3-2 ASP3-3 ASP3-4*
ribosomal proteins	-	*MDN1 PIH1 RPL11A RPL11B RPL12A RPL13A RPL15A RPL16A RPL16B RPL22A RPL22B RPL23A RPL30 RPL32 RPL33B RPL34A RPL43A RPL8A RPS0B RPS11A RPS13 RPS18B RPS1B RPS24A RPS24B RPS27A RPS29B RPS4A RPS5 RPS6A*	-

Table V-2.2: Categorization of genes with significantly increased expression (Bonferroni p<0.05) in Z63, Z23 and VL1 strains in comparison to S288c, at T2 (50 g/L of CO_2 released).

MIPS functional category	Strain		
	Z63	**Z23**	**VL1**
electron transport and membrane-associated energy conservation	-	-	*ATP20 COR1 COX1 COX5A COX6 COX7 CYB2 CYC1 CYC7 NDI1 PMA2 QCR2 QCR7 RIP1*
tetracyclic and pentacyclic triterpenes (cholesterin, steroids and hopanoids) metabolism	*ARE2 ERG1 ERG10 ERG13 ERG2 ERG20 ERG24 ERG27 ERG28 ERG5 ERG6 ERG9 HMG1 IDI1 MVD1 NCP1*	-	*ARE2 ERG1 ERG10 ERG12 ERG13 ERG2 ERG20 ERG24 ERG25 ERG26 ERG27 ERG28 ERG5 ERG6 ERG7 ERG9 HMG1 IDI1 MVD1 NCP1*
mitochondrion	*CLU1 HOT13 HSP10 MDH1 MDM35 MRP2 MRP49 MRPL11 MRPL13 MRPL20 MRPL23 MRPL27 MRPL32 MRPL35 MRPL38 MRPL4 MRPL6 MRPL8 MRPS28 NDI1 PET18 PNT1*	-	-
fermentation	-	*AAD15 AAD3 AAD4 AAD6 ADH7 ALD2 ALD6*	*AAD15 AAD3 AAD4 AAD6 ADH6 ADH7 ALD6 MSC7*

Regarding time point 1, analysis of **Table V-1.1** shows that one group of genes related with the functions "pheromone response, mating-type determination, sex-specific proteins", was downregulated in all three strains. Since the 3 isolates used in the present work are diploid (Schuller et al. 2007; Liti et al. 2009; Franco-Duarte et al. 2014), and the laboratory strain S288c used for comparison is haploid (Goffeau et al. 1996), differences in ploidy could thus underlie the differences in expression of the genes related with the mating and the pheromone response. Genes involved in the degradation of asparagine/metabolism of aspartate (*ASP3-1*, *ASP3-2*, *ASP3-3* and *ASP3-4*) were also downregulated in the three isolates, and *ASP1* coding for cytosolic L-asparaginase was downregulated in Z23 and VL1 strains. This could be related with the fact that some *S. cerevisiae* strains, including some wine and sake strains, had lost the *ASP3* locus (League et al. 2012).

Genes with significantly increased expression at T1, include a group of genes related with tetracyclic and pentacyclic triterpenes metabolism (cholesterin, steroids and hopanoids) that was upregulated in the 3 strains comparatively to the laboratory strain (**Table V-1.2**). Most of these genes are involved in sterol synthesis namely ergosterol, which by contributing to the fluidity of the yeast membrane, allows a more efficient activity of membrane transporters and increased tolerance to ethanol (Alexandre et al. 1994), correlating with the superior fermentation performances of strains. The higher sterol biosynthesis could also divert acetyl CoA from fatty acid biosynthesis, so the lower levels of these genes in S288c strain could explain the higher production of medium chain fatty acids (MCFA) by this strain (**Figure V-2B**). Several genes involved in aerobic respiration, electron transport and mitochondrion were also upregulated in the three mentioned strains in comparison with S288c (**Table V-1.2**), suggesting a less strict glucose repression in the strains isolated from the fermentative environments. The higher respiratory capacity might also be associated with the higher production of fusel acids (**Figure V-2**), due to lower need to reoxidize NADH through the Ehrlich pathway (Cordente et al. 2012). Also, at T1, the increased expression in Z23 of genes related with aldehyde oxidation, namely *AAD4, AAD6, AAD16* and *ADH7,* could be related with the higher production of fusel alchools in this strain specialy of isoamylalcohol, phenylethanol, isobutanol and methionol (marked in blue in **Figure V-2B**).

Regarding time point T2 (**Table V-2.1**), there were no common downregulated genes in the three characterized strains. Genes related with ribosomal proteins were downregulated only in sake strain (**Table V-2.1**). The differences in the expression of these

genes, observed also at T1 for Z23 and VL1 strains, could be related with the different fermentative profile and the different metabolic stage of each strain, at this time point.

Regarding upregulated genes (**Table V-2.2**), a group of genes involved in the synthesis of sterols was still upregulated for the cachaça (Z63) and wine (VL1) strains. For the sake strain (Z23) these genes were similarly expressed when compared to the laboratory strain suggesting that sake strain could be in an earlier metabolic stage, in comparison with the other strains, requiring less sterol synthesis, which is in also agreement with the observed repression of ribosomal genes. Also at T2 it is visible that some genes upregulated in strains Z23 and VL1 (*ADH7*, *ADH6* and *AAD6)* are involved in the Ehrlich pathway and so related with the formation of specific compounds, such as higher alcohols. In accordance with these results, metabolic analysis showed an increase of the same higher alcohols for T2 in comparison with T1, namely: methionol, isobutanol, isoamyl alcohol and phenylethanol. The only alcohols that seem not to be included in this association are amylalcohol and propanol which were equal or less produced, respectively, in these strains in relation to S288c. The differential production of acetate esters by the two groups of strains (marked in orange in **Figures V-2B** and **V-2D**) could be related with the differences in expression of *ALD6*, which was overexpressed in strains Z23 and VL1. This gene is involved in the formation of acetic acid that can then be converted into acetyl-CoA and subsequently incorporated in acetate esters. As for the downregulated genes, at T2 there were no common upregulated genes for the three strains. This is opposite to the observed at T1 and may reflect that the differentiation of the strains, isolated from different fermentation processes, is especially important enduring the multistress stationary phase of fermentation where each strain developed different adaptive mechanisms in response to the specific fermentation conditions (Liti et al. 2009).

Combined transcriptomics and metabolomics analysis

Aiming to unravel new associations between genes and aromatic compounds production we next performed a combined analysis of transcriptomic and metabolic data sets. A supervised exploratory approach sPLS-DA was carried out from gene expression data in order to select the 400 most differential expressed genes (200 for each axis) at each time point (from the 6200 *S. cerevisiae* probes present in the microarray). At the two time points, multivariate factorial analysis (MFA) was then performed from expression levels of

the 400 chosen genes and the 33 metabolic variables (**Figures 3** and **4**). Genes clustered into four main groups together with metabolites, allowing a clear separation of the strains on the basis of their genes expression and metabolic profiles. GeneCodis (Carmona-Saez et al. 2007; Nogales-Cadenas et al. 2009; Tabas-Madrid et al. 2012) was used to determine biological annotations with statistical relevance associated with the genes present in each group (**Supplementary data S6 and S7**).

During the growth phase (T1, **Figure V-3**), the reference strain S288c differed from the other yeasts (sake, cachaça and wine strains) by an higher expression level of genes of group 3 associated with an important production of propanol, glycerol and medium chain fatty acid, and conversely, a lower expression of genes of group 1, connected with a limited formation of isobutanol, methionol, isobutylacetate and phenylethanol. Genes of group 1 were identified as coding for ribosomal proteins (*RPL14B, RPS24A, RPS25B, RPL30, RPS26B, MRPL23, RPS17B, RPL40B* and *RPL26A*), involved in the structural integrity of ribosome. The association of genes coding for ribosomal proteins, with the differential production of higher alcohols and the ester isobutyl acetate (**Supplementary data S6**), could suggest an impact of higher growth rates and on the production of these compounds. It is well known that the formation of higher alcohols depends of the reduction from the respective aldehyde with the oxidation of NADH into NAD^+ (Ehrlich 1907). Consequently, the need for rapid production of oxidised NAD^+ could have an important regulatory role in the formation of these compounds, explaining their higher formation by cachaça, wine and sake strains compared with the laboratory yeast. Regarding group 3, it contains genes associated with MAPK signalling pathway, cysteine and methionine metabolism and ABC transporters. The presence in this group of *ATM1*, coding for a mitochondrial exporter of Fe-S clusters and of genes from metabolism of cysteine, usually the limiting component in glutathione synthesis, suggests a more important response of S288c to oxidative stress compared with the other yeasts, generating a limitation of reductive power in this strain. This decrease may be the driving factor of the formation of several volatile fatty acids such as octanoic acid, decanoic acid, hexanoic acid, butyric acid and dodecanoic acid, which was increased in the laboratory strain. It is also tempting to speculate that *PDR5* may be involved in the export of the fatty acids. These differences between the laboratorial strain and strains from other environments were expected, and already shown previously (Camarasa et al. 2011), and constitute the basis for the strain selection to be used in this work, allowing the

better exploration of differences between phenotype and genotype, which was not possible with an homogenous group of strains. Interestingly, MFA revealed that cachaça yeast differentiated from the other strains by an increased production of ethyl esters, namely ethylbutanoate, ethyldecanoate and ethyloctanoate while VL1 and Z23 exhibited higher capacities of production of hexylacetate, propylacetate, 2-phenylethylacetate, amylalcohol, isovaleric acid, isoamylacetate, amylacetate, ethlpropionate, propanoic acid and isoamylalcohol (**Supplementary data S6**). In addition, genes that were more expressed specifically in Z63 are related with metabolism of butanoate, tyrosine, beta-alanine and fatty acids, and also associated with glycolysis and gluconeogenesis. Thus, the overexpression of genes involved in the butanoate and more general fatty acid metabolism, may directly explain the increased production of ethylbutanoate and of the other ethyl esters. Finally, no relevant biological annotation was found among the genes overexpressed in wine and sake yeast (group 4), pointing to a role of each of the genes individually.

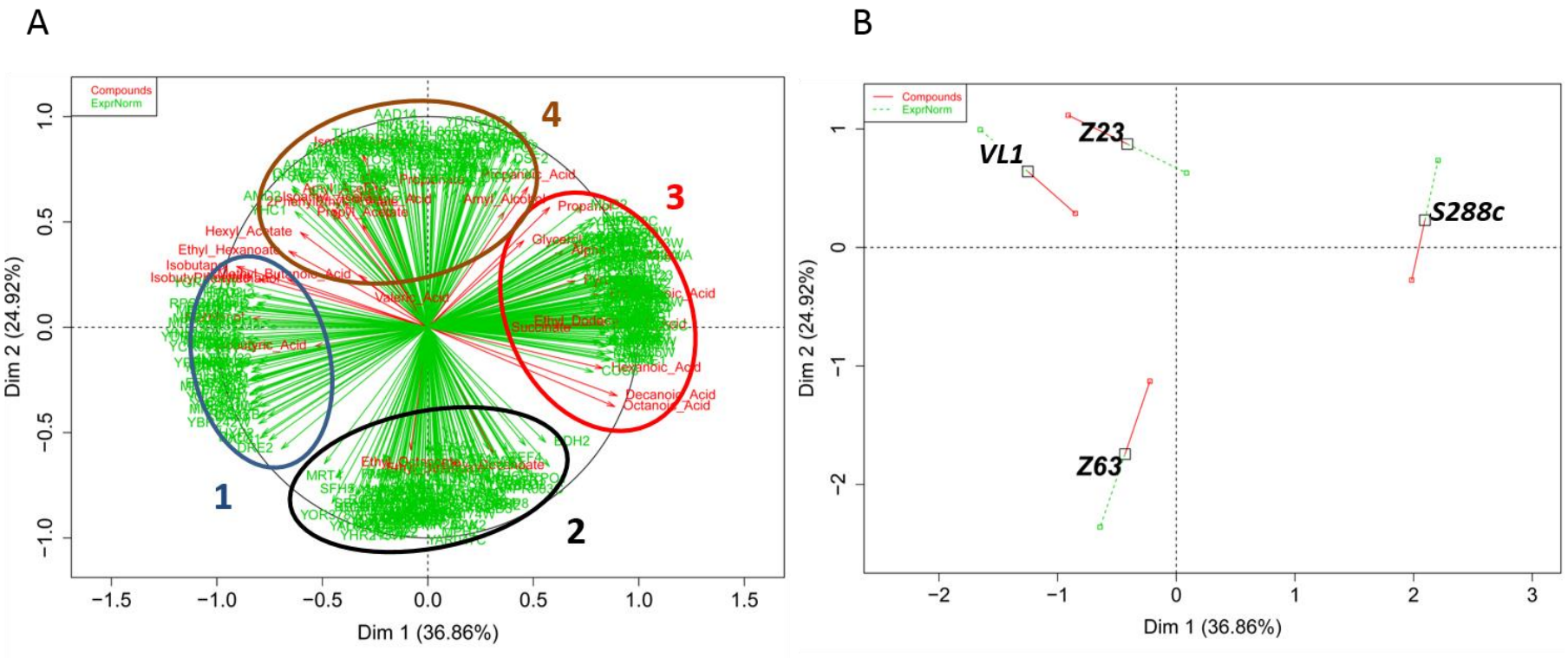


Figure V-3: Multivariate factorial analysis of GC-MS, HPLC and transcriptomic data for the four strains tested, at T1 (5 g/L). Circles 1-4 indicates groups of genes and compounds sharing similar results regarding their positioning in the image:

A – distribution of the quantified compounds (red) and genes (green)

B – distribution of the four tested strains.

At T2 (**Figure V-4**), a clear separation was also observed between strain S288c and the other strains, being this related with overexpression of genes from groups 1 and 2 *versus* downregulation of those of group 3 and 4 in the lab strain. In addition, S288c is characterised by an important formation of unpleasant or neutral compounds, in particular acids that contribute with unpleasant odors to wine. This is in accordance to the origin of S288c strain, not being associated with fermentative environments, in opposition to the other three tested strains. Genes from group 1, such as *TDH3, FBP26, SLT2, MIG2* and *GDH1,* which clustered with acids formation, were associated with central carbon metabolism and its regulation, cation transport and cell wall. Thus, the maintain of ionic homeostasis in the interaction with the environment may appear as a determining factor in the production of the unpleasant acids. Consequently, the manipulation of specific cation homeostasis and cell wall integrity pathway could be a way of avoiding/reducing their production. Genes from group 2 included once again the term "ribosomes" but associated with the formation of propanol, amylalcohol, alpha-ketoglutarate and pyruvate in addition to the production of higher alcohol, as evidenced at T1. The other biological annotations were associated with genes included purine or pyrimidine metabolism, and no clear scenario could be established between genes functions and the compounds produced. Genes from groups 3 and 4 were clearly related with the central carbon metabolism and formation of aroma compounds and are associated with marked increased concentrations of higher alcohols and ethyl and acetate esters for the fermentative yeasts, including several acetate and ethyl esters that contribute to the "floral" and "fruity" characteristics of wine (**Supplementary data S7**). Specifically, wine and cachaça strains were characterised by an overexpression of genes from group 3 combined with a downregulation of those of group 2. Group 3 included a set of 17 genes related with biosynthesis of secondary metabolites, which clearly related with the production of the metabolic compounds, being more specifically associated with the terms "steroid biosynthesis" "propanoate metabolism" (*ALD6*, *ACS2* and *ERG10*), "valine, leucine, isoleucine and lysine degradation" (*ALD6*, *ERG10*, *ERG13*), fatty acid metabolism (*FAA1*, *ALD6* and *ERG10*). This could be associated to an increase production of valeric acid but also succinate, methionol and isobutanol. Group 4 genes, which differentiated sake strain from the others, were mainly associated with the production of a high variety of acetate and ethyl ethers. Functional categories more significantly associated with this group of genes were c-compound metabolism and oxidation-reduction process.

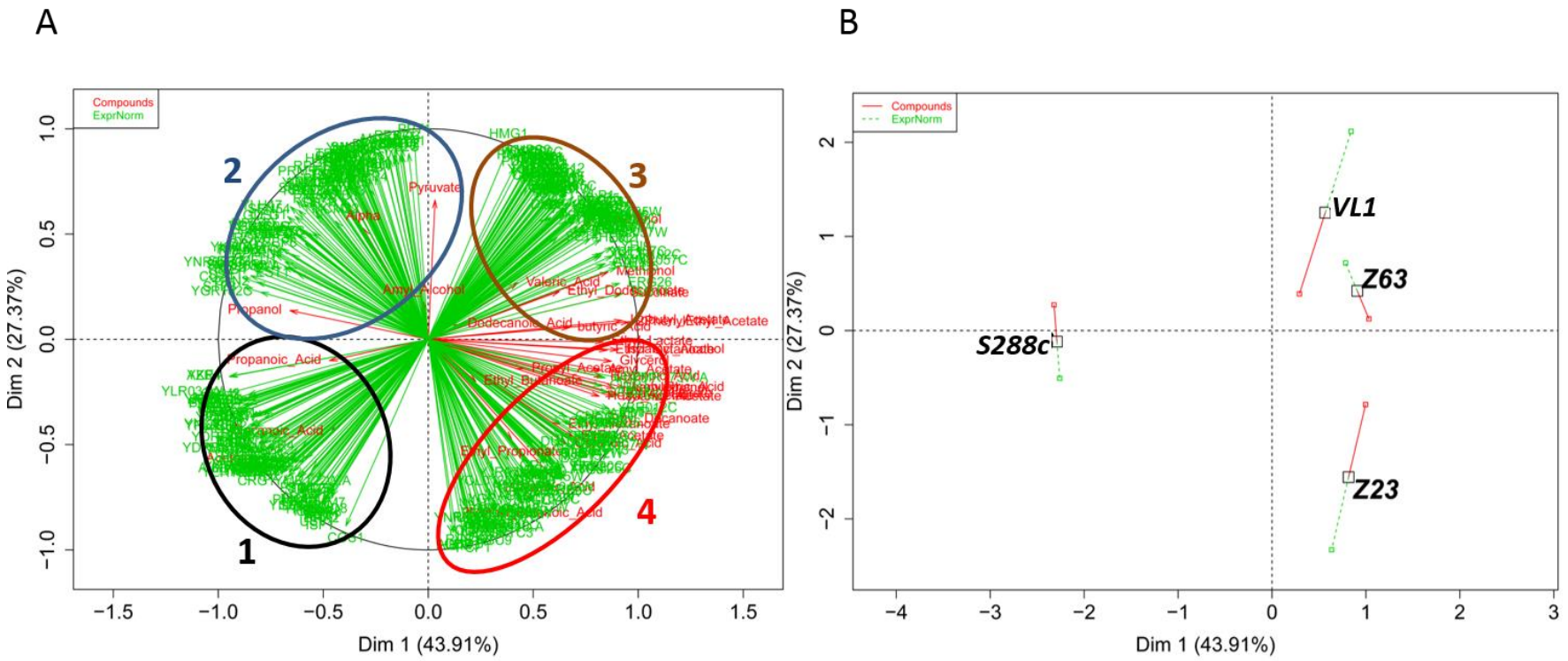


Figure V-4: Multivariate factorial analysis of GC-MS, HPLC and transcriptomic data for the four strains tested, at T2 (50 g/L). Circles 1-4 indicates groups of genes and compounds sharing similar results regarding their positioning in the image:

A – distribution of the quantified compounds (red) and genes (green)

B – distribution of the four tested strains.

Conclusions

The main objective of this work was the characterization and comparison of four *S. cerevisiae* strains, with different origins and technological applications, previously characterized genetically and phenotypically. Results have shown differences between the analysed strains, in terms of fermentative profile, gene expression and metabolite production. These differences were observed both at T1 (5 g/L of CO_2 released), in which a higher production of acetate esters and some higher alcohols was detected for strains Z63 (cachaça), Z23 (sake) and VL1 (wine), and at T2 (50 g/L of CO_2 released) with a higher production by the same strains not only of acetate esters and higher alcohols, as in T1, but also of ethyl esters and acids. These differences in metabolite production could be related with gene expression levels, with a significantly increased expression at T1 of genes associated with tetracycic and pentacyclic triterpenes metabolism, and also with aerobic respiration, electron transport and mitochondria in the three strains, in comparison with S288c. The higher mitochondrial activity may be associate with increased ketoacids production and so to the flavour compounds. The results presented also suggest growth rates and redox and cation homeostasis as major players in the differential production of metabolites.

The present work shed light in the understanding of differences between cachaça, sake and wine strains metabolism, which was not yet addressed in detail. Although the three strains are known for presenting a similar metabolism related with fermentative conditions, significant differences were found, mainly between cachaça and sake strains, in comparison with the wine strain. At T1 of fermentation, strain Z63 (cachaça) showed major differences from sake and wine strains, mainly regarding the production of ethyl esters ethyl decanoate and ethyl octanoate. These differences were associated with the expression of genes related with the metabolism of butanoate, tyrosine, beta-alanine and fatty acids. At T2, a different scenario was found in which the sake strain (Z23) had the most distinct behaviour. At this point this strain showed a higher production of several acetate and ethyl esters and an increase in the expression of genes of c-compound metabolism and oxidation-reduction process. On the contrary, wine and cachaça strains showed an upregulation of genes related with steroid biosynthesis, propanoate metabolism, valine, leucine, isoleucine and lysine degradation, and fatty acid metabolism.

In summary, the integration of several technologies (HPLC, GC-MS, microarrays) applied to fermentation results of four strains with diverse origins and technological applications, analysed using several data analysis methods (PCA, MFA) revealed successful to understand and clarify the genes and the pathways that lead to the formation of metabolic compounds that contribute to the wine aroma and flavour. The knowledge here obtained has the potential to be deeply explored and extended to other strains and other metabolic pathways, within an approach using aroma production as the primary selection criteria. The majority of the genes identified in this work as having their expression changed in correlation with the aroma compounds produced, play a central role in the metabolism of *S. cerevisiae*, namely *ADH6, ADH7*, *AAD6*, *ALD2*, *ALD6*, *FAA1*, *ACS2, ERG10* and *ERG13*. These genes are potential targets for gene deletion/overexpression programs using these and/or other strains, in order to better understand their role and their correlation with the aroma production network of *S. cerevisiae.* The information now obtained may be useful both for strain improvement and to drive the selection of yeast strains with improved aromatic properties.

Chapter VI

General Conclusions and Future Perspectives

Winemaking industry faces the need to strengthen the reputation of quality wine and to contribute to an improved wine image, especially concerning the Portuguese and European market. Winemaking industry could benefit massively from "system-level" research for the development of novel starter *Saccharomyces cerevisiae* strains, owing to the direct impact of yeast on several key areas such as flavour compounds synthesis. The market value of products derived from yeast fermentations, with particular emphasis in the use of *Saccharomyces cerevisiae* for wine production, is expected to increase much above the general market growth in the future. Winemaking represents today a multi-billion Euro industry, and all the research done, mainly regarding new methods for the study of systems biology, could have a huge impact in the development and modernization of wine production

Wine flavour is the result of yeast metabolism and must compounds interactions and only an integrated approach through systems biology characterization will allow a more comprehensive understanding of the enhanced flavour compounds synthesis. During must fermentation thousands of volatile aroma compounds are formed, with higher alcohols, acetate esters and ethyl esters being the main aromatic compounds contributing to a floral and fruity aroma. The action of yeast, in particular of *S. cerevisiae* strains, on the must components will build the architecture of the wine flavour and their fermentation bouquet.

In this thesis, *Saccharomyces cerevisiae* strains from different technological applications and geographical origins were used and a phenotypic, genetic, metabolomic and transcriptomic characterization was performed. The following paragraphs summarize the main findings of our research and also include personal perspectives for future approaches and for the application of the knowledge obtained.

A total of 172 *S. cerevisiae* strains were obtained worldwide from different technological applications or environments, constituting the core strain collection of this thesis. A complete phenotypic characterization was performed, using 30 traits that are important from an oenological point of view. The battery of phenotypic tests has revealed successful to illustrate the high strain diversity and three tests were pointed as contributing the most for strain variability: growth in the presence of potassium bisulphite, growth at 40 ºC and resistance to ethanol. Mathematical models were used and showed associations between the phenotypic profile of a strain and its technological group, and a simple and quicker method to identify a promising strain to be used in biotechnology (commercial strain) was developed allowing to greatly simplify strain selection procedures.

Our next goal was the genetic characterization of the strain collection. For this purpose we characterize the strains using amplification of 11 microsatellites, specific of *S. cerevisiae.* A high genetic variability was obtained, with the revelation of 280 alleles, being the microsatellite ScAAT1 the one contributing the most to intra-strain variability. Microsatellite amplification revealed to be an efficient method to characterize strain genetic diversity and the obtained microsatellite allelic profiles showed to be unique of each strain. Mathematical associations were found between microsatellite allelic profiles and the phenotype and the origin of the strains. The phenotypes associated with higher number of alleles were the capacity to resist to sulphur dioxide (tested by the capacity to grow in the presence of potassium bisulphite) and the presence of galactosidase activity. Results demonstrate once again the utility of these models as tools for preliminary yeast strain selection.

After phenotypic and genetic characterization 4 isolates were chosen from the entire collection, as the most interesting to perform metabolic characterization, due to their phenotypic and genetic heterogeneity and their different technological origins: strain Z63 isolated from cachaça fermentations, strain Z23 used to produce the fermentative beverage sake, strain VL1 a commercial wine strain, and the laboratory strain S288c. The objective of this part of work was to better understand the molecular and metabolic bases of aroma production during a fermentation process, using comparative transcriptomic and metabolic analysis, to rationally identify new gene targets for improving aroma production. For this, we used several methodologies to characterize strain aroma production and associated genes, such as HPLC, GC-MS and microarrays, the results being analysed using mathematical approaches such as PCA and MFA. This approached has revealed successful to understand the pathways that lead to the formation of metabolic compounds, contributing to the wine aroma and flavour. The results confirmed differences between the three tested strains in comparison with the laboratory strain S288c, regarding fermentative profile, gene expression and metabolite production at two fermentation time points (5 and 50 g/L of CO_2 released). These results were in line with a previous work where wine strains were described to be phenotypically closer to sake strains then to laboratory strains (Camarasa et al. 2011). Strains from cachaça, sake and wine (VL1) showed a higher production of acetate esters, ethyl esters, acids and higher alcohols, in comparison with the laboratory strain. Phenotypic variation between *S. cerevisiae* strains caused by environmental exposure/adaptation was

shown previously (Camarasa et al. 2011; Brion et al. 2013; Barbosa et al. 2014). In this work we extended this knowledge by associating gene groups to the differential production of aromatic compounds by strains from four different environments: laboratory, and wine, cachaça and sake fermentations. Multiple factorial analyses combining data from gene expression and metabolic characterization, for both time points of fermentation, confirmed the separation of cachaça, sake and wine strains from the laboratory strain, as expected(Camarasa et al. 2011). Most relevant results showed that, at T1 of fermentation, cachaça strain differentiated from the other strains by an increased production of ethyl esters, namely ethylbutanoate, ethyldecanoate and ethyloctanoate while VL1 and sake strains exhibited higher capacities of production of hexylacetate, propylacetate, 2-phenylethylacetate, amylalcohol, isovaleric acid, isoamylacetate, amylacetate, ethlpropionate, propanoic acid and isoamylalcohol. Regarding gene expression, these differences were justified mainly by an overexpression of genes related with butanoate, tyrosine, beta-alanine and fatty acids in the cachaça strain. Conversely, at T2, differences were focused in the higher production by sake strain, in comparison with cachaça and wine strains, of several acetate and ethyl esters, together with an increase in the expression of genes of c-compound metabolism and oxidation-reduction process.

As a final viewpoint, data obtained in this thesis, will certainly be useful in biotechnological field such as for simplification of the laborious strain selection programmes, strain improvement, and to select yeast strains with improved aromatic properties.

In the future this work should be continued and new windows of opportunity are now open by this knowledge, in particular:

- Develop new methods to establish relations between gene expression patterns and metabolites formation pathways, in order to study new networks of genes with interest in winemaking and in particular in the aroma metabolome;
- To assess the impact of different must matrixes on the aroma and gene expression profiles of these strains;
- To construct genetically modified winemaking strains by allele replacement using genes implicated in flavour formation and evaluate at industrial scale their properties;

- Expand the current characterization to the proteome level in order to find associations with the transcriptome and metabolome of these strains;
- Give particular emphasis to the computational models focusing in predictive methodologies that by being adequate for high-throughput data analysis will be each time more mandatory in order to do research with this type of "omics" data.

Chapter VII

References

References

Agnolucci M, Scarano S, Santoro S, Sassano C, Toffanin a., Nuti M (2007) Genetic and phenotypic diversity of autochthonous Saccharomyces spp. strains associated to natural fermentation of "Malvasia delle Lipari." Lett Appl Microbiol 45:657–662.

Albertyn J, Hohmann S, Thevelein JM, Prior B a (1994) GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in *Saccharomyces cerevisiae*, and its expression is regulated by the high-osmolarity glycerol response pathway. Mol Cell Biol 14:4135–44.

Alexandre H, Ansanay-Galeote V, Dequin S, Blondin B (2001) Global gene expression during short-term ethanol stress in *Saccharomyces cerevisiae*. FEBS Lett 498:98–103.

Alexandre H, Rousseaux I, Charpentier C (1994) Ethanol adaptation mechanisms in *Saccharomyces cerevisiae*. Biotechnol Appl Biochem.20:173-83.

Allen J, Davey H, Broadhurst D, Heald J, Rowland J, Oliver S, Kell D (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. Nat Biotechnol. 21:692-6.

Ambroset C, Petit M, Brion C, Sanchez I, Delobel P, Guérin C, Chiapello H, Nicolas P, Bigey F, Dequin S, Blondin B (2011a) Deciphering the molecular basis of wine yeast fermentation traits using a combined genetic and genomic approach. G3 (Bethesda) 1:263–81.

Ambroset C, Petit M, Brion C, Sanchez I, Delobel P, Guerin C, Chiapello H, Nicolas P, Bigey F, Dequin S, Blondin B, Fay JC (2011b) Deciphering the Molecular Basis of Wine Yeast Fermentation Traits Using a Combined Genetic and Genomic Approach. G3 Genes|Genomes|Genetics 1:263–281.

Ayoub MJ, Legras JL, Saliba R, Gaillardin C (2006) Application of Multi Locus Sequence Typing to the analysis of the biodiversity of indigenous *Saccharomyces cerevisiae* wine yeasts from Lebanon. J Appl Microbiol 100:699–711.

Bakalinsky A, Snow F (1990) The chromosomal constitution of wine strains of *Saccharomyces cerevisiae.* Yeast. 6(5):367-82.

Bakhtiar R, Ramos L, Tse F (2002) High-throughput mass spectrometric analysis of xenobiotics in biological fluids. J Liq Chromatogr Relat Technol 25:507–540.

Baleiras Couto MM, Vogels JTWE, Hofstra H, Huis in't Veld JHJ, van der Vossen JMBM (1995) Random amplified polymorphic DNA and restriction enzyme analysis of PCR amplified rDNA in taxonomy: two identification techniques for food-borne yeasts. J Appl Bacteriol 79:525–535.

Barbosa C, Lage P, Vilela A, Mendes-Faia A, Mendes-Ferreira A (2014) Phenotypic and metabolic traits of commercial *Saccharomyces cerevisiae* yeasts. AMB Express 4:39.

Barnett J (1992) The taxonomy of the genus Saccharomyces meyen ex Reess: a short review for non taxonomists. Yeast 8:1–23.

Barre P, Vezinhet F, Dequin S, Blondin B (1993) Genetic improvement of wine yeasts. In: G. R. Fleet (ed) Wine microbiology and biotechnology. Harwood Academic Publisher, Newark, N.J., Chur, Switzerland.,

Barton RH, Nicholson JK, Elliott P, Holmes E (2008) High-throughput 1H NMR-based metabolic analysis of human serum and urine for large-scale epidemiological studies: validation study. Int J Epidemiol 37 Suppl 1:i31–40.

Bartowsky EJ, Henschke P a (2004) The "buttery" attribute of wine-diacetyl-desirability, spoilage and beyond. Int J Food Microbiol 96:235–52.

Beech W, Thomas S (1985) Action antimicrobienne de l'anhydride sulfureux. Bull l´OIV 58 652-653:564–581.

Beltran G, Torija MJ, Novo M, Ferrer N, Poblet M, Guillamón JM, Rozès N, Mas A (2002) Analysis of yeast populations during alcoholic fermentation: A six year follow-up study. Syst Appl Microbiol 25:287–93.

Bely M, Sablayrolles JM, Barre P (1990) Description of Alcoholic Fermentation Kinetics: Its Variability and Significance. Am J Enol Vitic 41:319–324.

Ben-Ari G, Zenvirth D, Sherman A, David L, Klutstein M, Lavi U, Hillel J, Simchen G (2006) Four linked genes participate in controlling sporulation efficiency in budding yeast. PLoS Genet 2:e195.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J R Stat Soc B 57:289–300.

Berg H, Filipello F, Hinreiner E, Webb A (1955) Evaluation of thresholds and minimum difference concentrations for various constituents of wines. 1. Water solutions of pure substances. Food Technol 9:23–26.

Bidenne C, Blondin B, Dequin S, Vezinhet F, Microbiologie L De, De P, Viala P, Technologie C De (1992) Analysis of the chromosomal DNA polymorphism of wine strains of *Saccharomyces cerevisiae*. Curr Genet 22:1–7.

Bird D (2013) Understanding wine technology - the science of wine explained.

Birkemeyer C, Kolasa A, Kopka J (2003) Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. J Chromatogr A 993:89–102.

Bisson L (1999) Stuck and sluggish fermentations. Am J Enol Vitic 50:107–119.

Bjerrum JT, Nielsen OH, Hao F, Tang H, Nicholson JK, Wang Y, Olsen J (2010) Metabonomics in ulcerative colitis: Diagnostics, biomarker identification, and insight into the pathophysiology. J Proteome Res 9:954–962.

Bleykasten-grosshans C, Friedrich A, Schacherer J (2013) Genome-wide analysis of intraspecific transposon diversity in yeast. BMC Genomics 14:399.

Blondin B, Dequin S (1998) Yeast, wine and genetic engineering.

Blondin B, Vezinhet F (1988) Identification de souches de levures oenologiques par leurs caryotypes obtenus en électrophorèse en champs pulsée. Rev française d'oenologie 28:7–11.

Borneman AR, Chambers PJ, Pretorius IS (2007) Yeast systems biology: modelling the winemaker's art. Trends Biotechnol 25:349–55.

Borneman AR, Desany B a, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ (2011) Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. PLoS Genet 7:e1001287.

Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ (2012) The genome sequence of the wine yeast VIN7 reveals an allotriploid hybrid genome with *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii* origins. FEMS Yeast Res 12:88–96.

Borneman AR, Forgan AH, Pretorius IS, Chambers PJ (2008) Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. FEMS Yeast Res 8:1185–95.

Borneman AR, Pretorius IS, Chambers PJ (2013) Comparative genomics: A revolutionary tool for wine yeast strain development. Curr Opin Biotechnol 24:192–199.

Bothwell JHF, Griffin JL (2011) An introduction to biological nuclear magnetic resonance spectroscopy. Biol Rev Camb Philos Soc 86:493–510.

Boulton R, Singleton V, Bisson L, Kunkee R (1998) Principles and Practices of Winemaking.

Boutou S, Chatonnet P (2007) Rapid headspace solid-phase microextraction/gas chromatographic/mass spectrometric assay for the quantitative determination of some of the main odorants causing off-flavours in wine. J Chromatogr A 1141:1–9.

Brandolini V, Tedeschi P, Capece a., Maietti a., Mazzotta D, Salzano G, Paparella a., Romano P (2002) Saccharomyces cerevisiae wine strains differing in copper resistance exhibit different capability to reduce copper content in wine. World J Microbiol Biotechnol 18:499–503.

Brion C, Ambroset C, Delobel P, Sanchez I, Blondin B (2014) Deciphering regulatory variation of THI genes in alcoholic fermentation indicate an impact of Thi3p on PDC1 expression. BMC Genomics 15:1085.

Brion C, Ambroset C, Sanchez I, Legras J-L, Blondin B (2013) Differential adaptation to multi-stressed conditions of wine fermentation revealed by variations in yeast regulatory networks. BMC Genomics 14:681.

Briones A, Ubeda J, Cabezudo M, Martin-Alvarez P (1995) Selection of spontaneous strains of *Saccharomyces cerevisiae* as starters in their viticultural area. Dev food Sci 37:1597–1622.

Bruns T, White T, Taylo J (1991) Fungal molecular systematics. Annu Rev Ecol Syst 22:525–564.

Buecher B, Cacheux W, Rouleau E, Dieumegard B, Mitry E, Lièvre A (2013) Role of microsatellite instability in the management of colorectal cancers. Dig Liver Dis 45:441–9.

Bundy JG, Papp B, Harmston R, Browne R a., Clayson EM, Burton N, Reece RJ, Oliver SG, Brindle KM (2007) Evaluation of predicted network modules in yeast metabolism using NMR-based metabolite profiling. Genome Res 17:510–519.

Cadez N, Raspor P, De Cock AW a M, Boekhout T, Smith MT (2002) Molecular identification and genetic diversity within species of the genera Hanseniaspora and Kloeckera. FEMS Yeast Res 1:279–289.

Čadež N, Zupan J, Raspor P (2010) The effect of fungicides on yeast communities associated with grape berries. FEMS Yeast Res 10:619–630.

Camarasa C, Grivet J, Dequin S (2003) Investigation by 13C-NMR and tricarboxylic acid (TCA) deletion mutant analysis of pathways for succinate formation in *Saccharomyces cerevisiae* during anaerobic fermentation. Microbiology 149:2669–2678.

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:e25147.

Cameron J, Loh E, Davis R (1979) Evidence for transposition of dispersed repetitive DNA families in yeast. Cell 16:739–751.

Cappello MS, Bleve G, Grieco F, Dellaglio F, Zacheo G (2004) Characterization of *Saccharomyces cerevisiae* strains isolated from must of grape grown in experimental vineyard. J Appl Microbiol 97:1274–1280.

Carle G, Olson M (1985) An electrophoretic karyotype for yeast. Proc Natl Acad Sci U S A 82:3756–3760.

Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A (2007) GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol 8:R3.

Carreto L, Eiriz MF, Domingues I, Schuller D, Moura GR, Santos MA (2011) Expression variability of co-regulated genes differentiates *Saccharomyces cerevisiae* strains. BMC Genomics 12:201.

Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D, Santos MA (2008) Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics 9:524.

Carro D, García-Martinez J, Pérez-Ortín JE, Piña B (2003) Structural characterization of chromosome I size variants from a natural yeast strain. Yeast 20:171–183

Carru C, Zinellu A, Galistu F, Sotgia S, Usai M, Pes G, Deiana L (2003) Ultra rapid capillary electrophoresis method for total plasma thiols measurement. Clin Chem 49:A36.

Cavalieri D, McGovern PE, Hartl DL, Mortimer R, Polsinelli M (2003) Evidence for *Saccharomyces cerevisiae* fermentation in ancient wine. J Mol Evol 57 Suppl 1:S226–32.

Chambers PJ, Pretorius IS (2010) Fermenting knowledge: the history of winemaking, science and yeast research.

Chatonnet P, Dubourdieu D, Boidron J-N, Lavigne V (1993) Synthesis of Volatile Phenols by *Saccharomyces cerevisiae* in wines. J Sci Food Agric 62:191–202.

Chen Q, Ding J, Cai J, Zhao J (2012) Rapid measurement of total acid content (TAC) in vinegar using near infrared spectroscopy based on efficient variables selection algorithm and nonlinear regression tools. Food Chem 135:590–595.

Choi MR, Gwak M, Yoo NJ, Lee SH (2015) Regional Bias of Intratumoral Genetic Heterogeneity of Apoptosis-Related Genes BAX, APAF1, and FLASH in Colon Cancers with High Microsatellite Instability.

Ciani M, Mannazzu I, Marinangeli P, Clementi F, Martini A (2004) Contribution of winery-resident *Saccharomyces cerevisiae* strains to spontaneous grape must fermentation. Antonie Van Leeuwenhoek 85:159–164.

Cocolin L, Pepe V, Comitini F, Comi G, Ciani M (2004) Enological and genetic traits of *Saccharomyces cerevisiae* isolated from former and modern wineries. FEMS Yeast Res 5:237–245.

Cohen A, Terabe S, Smith J, Karger B (1987) High-performance capillary electrophoretic separation of bases, nucleosides, and oligonucleotides: retention manipulation via micellar solutions and metal additives. Anal Chem 59:1021–1027.

Cordente AG, Curtin CD, Varela C, Pretorius IS (2012) Flavour-active wine yeasts. Appl Microbiol Biotechnol 96:601–18.

Corison C, Ough C, Berg H, Nelson K (1979) Must Acetic Acid and Ethyl Acetate as Mold and Rot Indicators in Grapes. Am J Enol Vitic 30:130–134.

Correia A, Sampaio P, Almeida J, Pais C (2004) Study of Molecular Epidemiology of Candidiasis in Portugal by PCR Fingerprinting of Candida Clinical Isolates Study of Molecular Epidemiology of Candidiasis in Portugal by PCR Fingerprinting of Candida Clinical Isolates. J Clin Microbiol 42:5889–5903.

Coulter AD, Godden PW, Pretorius IS (2004) Succinic acid – How it is formed, what is its effect on titratable acidity, and what factors influence its concentration in wine?

Culleré L, Escudero A, Cacho J, Ferreira V (2004) Gas chromatography-olfactometry and chemical quantitative study of the aroma of six premium quality spanish aged red wines. J Agric Food Chem 52:1653–60.

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

Czerny M, Christlbauer M, Christlbauer M, Fischer A, Granvogl M, Hammer M, Hartl C, Hernandez NM, Schieberle P (2008) Re-investigation on odour thresholds of key food aroma compounds and development of an aroma language based on odour qualities of defined aqueous odorant solutions. Eur Food Res Technol 228:265–273.

De Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol 9:67–103.

De Jong H, Geiselmann J, Hernandez C, Page M (2003) Genetic Network Analyzer: qualitative simulation of genetic regulatory networks. Bioinformatics 19:336–344.

Degré R, Thomas D, Ash J, Mailhiot K, Morin A, Dubord C (1989) Wine yeast strain identification. Am J Enol Vitic 40:309–315.

Demšar J, Curk T, Erjavec A, Gorup Č, Hočevar T, Milutinovič M, Možina M, Polajnar M, Toplak M, Starič A, Štajdohar M, Umek L, Žagar L, Žbontar J, Žitnik M, Zupan B (2013) Orange: data mining toolbox in python. J Mach Learn Res 14:2349–2353.

Demsar J, Zupan B, Leban G, Curk T (2004) Orange : From Experimental Machine Learning. Knowledge Discovery in Databases: PKDD 2004. Springer Berlin Heidelberg, pp 537–539

Demuyter C, Lollier M, Legras JL, Le Jeune C (2004) Predominance of *Saccharomyces uvarum* during spontaneous alcoholic fermentation, for three consecutive years, in an Alsatian winery. J Appl Microbiol 97:1140–1148.

Dequin S (2001) The potential of genetic engineering for improving brewing, wine-making and baking yeasts. Appl Microbiol Biotechnol 56:577–588.

Dequin S, Casaregola S (2011) The genomes of fermentative Saccharomyces. C R Biol 334:687–93.

Desselle F, Verset G, Polus M, Loius E, Van Daele D (2012) Lynch syndrome and microsatellite instability: a review. Rev Med Liege 67:638–643.

Dettmer K, Aronov PA, Hammock BD (2007) Mass spectrometry-based metabolomics. Mass Spectrom Rev 26:51–78. doi: 10.1002/mas.20108

Deutschbauer A, Davis R (2005) Quantitative trait loci mapped to single nucleotide resolution in yeast. Nat Genet 37:1333–1340.

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J, Faure S (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154.

Dickinson JR, Salgado LEJ, Hewlins MJE (2003) The catabolism of amino acids to long chain and complex alcohols in *Saccharomyces cerevisiae*. J Biol Chem 278:8028–34.

Diezmann S, Dietrich FS (2009) Saccharomyces cerevisiae: Population divergence and resistance to oxidative stress in clinical, domesticated and wild isolates. PLoS One.

Dow B, Ashley M (1998) High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. J Hered 89:62–70.

Dubois P (1994) Les arômes des vins et leurs défauts. Rev Française d' Oenologie 145:27–40.

Dubois P (1983) Volatile phenols in wine. Flavour of Distilled Beverages, Origins and Developments. pp 110–119

Dubourdieu D, Sokol A, Zucca J, Thalouarn P, Datte A, Aigle M (1984) Identification des souches de levures isolées de vins par l'analyse de leur ADN mitochondrial. Connais vigne vin 21:267–278.

Dubourdieu D, Torninaga T, Masneuf I, Des Gachons C, Murat M (2006) The role of yeasts in grape flavor development during fermentation: the example of Sauvignon blanc. Am J Enol Vitic 57:81–88.

Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 99:16144–16149.

Dunn B, Richter C, Kvitek DJ, Pugh T, Sherlock G (2012) Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome Res 22:908–924.

Dunn WB, Ellis DI (2005) Metabolomics: Current analytical platforms and methodologies. TrAC - Trends Anal Chem 24:285–294.

Eden A, Simchen G, Benvenisty N (1996) Two Yeast Homologs of ECA39, a Target for c-Myc Regulation, Code for Cytosolic and Mitochondrial Branched-chain Amino Acid Aminotransferases. J Biol Chem 271:20242–20245.

Efron B, Tibshirani R (1995) An Introduction to the Bootstrap. In: Grimshaw SD (ed) Technometrics. pp 340–341

Eglinton JM, Mcwilliam SJ, Fogarty MW, Francis IL, Kwiatkowski MJ, Høj PB, Henschke PA (2000) The effect of *Saccharomyces bayanus* -mediated fermentation on the chemical composition and aroma profile of Chardonnay wine. Aust New Zeal Wine Ind J 6:190–196.

Ehrlich F (1907) Uber das naturliche Isomere des Leucins. Berichte der Dtsch Chem Gesellschaft 40:2538–2562.

Ellis DI, Goodacre R (2006) Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. Analyst 131:875–85.

Enomoto K, Arikawa Y, Muratsubaki H (2002) Physiological role of soluble fumarate reductase in redox balancing during anaerobiosis in *Saccharomyces cerevisiae*.

Erasmus D, Vandermerwe G, Vanvuuren H (2003) Genome-wide expression analyses: Metabolic adaptation of to high sugar stress. FEMS Yeast Res 3:375–399.

Eriksson P, André L, Ansell R, Blomberg A, Adler L (1995) Cloning and characterization of GPD2, a second gene encoding sn-glycerol 3-phosphate dehydrogenase (NAD+) in *Saccharomyces cerevisiae*, and its comparison with GPD1. Mol Microbiol 17:95–107.

Eschenbruch R, Bonish P (1976) Production of sulphite and sulphide by low-and high-sulphite forming wine yeasts. Arch Microbiol 107:299–302.

Escudero a, Gogorza B, Melús M a, Ortín N, Cacho J, Ferreira V (2004) Characterization of the aroma of a wine from maccabeo. Key role played by compounds with low odor activity values. J Agric Food Chem 52:3516–24.

Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11:1591–1604.

Etiévant P (1991) Wine. In: Maarse H (ed) Volatile Compounds of Food and Beverages. Marcel Dekker Inc, New York, USA, pp 483–546

Fay JC, Benavides J a (2005a) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. PLoS Genet 1:66–71.

Fay JC, Benavides J a. (2005b) Evidence for domesticated and wild populations of saccharomyces cerevisiae. PLoS Genet 1:0066–0071.

Fedrizzi B, Magno F, Badocco D, Nicolini G, Versini G (2007) Aging effects and grape variety dependence on the content of sulfur volatiles in wine. J Agric Food Chem 55:10880–7.

Fernández-Espinar MT, López V, Ramón D, Bartra E, Querol a. (2001) Study of the authenticity of commercial wine yeast strains by molecular techniques. Int J Food Microbiol 70:1–10.

Fernandez-Ricaud L, Warringer J, Ericson E, Glaab K, Davidsson P, Nilsson F, Kemp GJL, Nerman O, Blomberg A (2007) PROPHECY - A yeast phenome database, update 2006. Nucleic Acids Res 35:463–467.

Fey SJ, Larsen PM (2001) 2D or not 2D. Curr Opin Chem Biol 5:26–33.

Fiehn O (2002) Metabolomics-the link between genotypes and phenotypes. Plant Mol Biol 48:155–71.

Fiehn O (2008) Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. Trends Anal Chem 27:261–269.

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. Proc Natl Acad Sci U S A 95:1647–1652.

Fleet GH (1998) Yeasts - What Reactions and Interactions Really Occur in Natural Habitats. Food Technol Biotechnol 36:285–289.

Ford G, Ellis EM (2002) Characterization of Ypr1p from *Saccharomyces cerevisiae* as a 2-methylbutyraldehyde reductase. Yeast 19:1087–96.

Foulet F, Nicolas N, Eloy O, Botterel F, Gantier J, Bretagne S, Costa J (2005) Microsatellite Marker Analysis as a Typing System for Candida glabrata Microsatellite Marker Analysis as a Typing System for *Candida glabrata*. J Clin Microbiol 43:4574–4579.

Fowles G (1992) Acids in grapes and wine: a review. J Wine Res 3:25–41.

Francis IL, Newton JL (2005) Determining wine aroma from compositional data. Aust J Grape Wine Res 11:114–126.

Franco-Duarte R, Mendes I, Gomes A, Santos M, de Sousa B, Schuller D (2011) Genotyping of *Saccharomyces cerevisiae* strains by interdelta sequence typing using automated microfluidics. Electrophoresis 32:1447– 1455.

Franco-Duarte R, Mendes I, Umek L, Drumonde-Neves J, Zupan B, Schuller D (2014) Computational models reveal genotype-phenotype associations in *Saccharomyces cerevisiae*.

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.

Freimer N, Sabatti C (2003) The human phenome project. Nat Genet 34:15–21.

Frezier V, Dubourdieu D (1992) Ecology of yeast strains *Saccharomyces cerevisiae* during spontaneous fermentation in Bordeaux winery. Am J Enol Vitic 43:375–380.

Fukuda K, Yamamoto N, Kiyokawa Y, Yanagiuchi T, Wakai1 Y, Kitamoto K, Inoue Y, Kimura A (1998) Brewing properties of sake yeast whose EST2 gene encoding isoamyl acetate-hydrolyzing esterase was disrupted.

Ganga M, Piñaga F, Vallés S, Ramón D, Querol A (1999) Aroma improving in microvinification processes by the use of a recombinant wine yeast strain expressing the *Aspergillus nidulans* xlnA gene. Int J Food Microbiol 47:171–178.

Garcia DE, Baidoo EE, Benke PI, Pingitore F, Tang YJ, Villa S, Keasling JD (2008) Separation and mass spectrometry in microbial metabolomics. Curr Opin Microbiol 11:233–9.

Garcia-Hermoso D, MacCallum DM, Lott TJ, Sampaio P, Serna MJB, Grenouillet F, Klaassen CHW, Bretagne S (2010) Multicenter collaborative study for standardization of *Candida albicans* genotyping using a polymorphic microsatellite marker. J Clin Microbiol 48:2578–2581.

Gawel R, Francis L, Waters EJ (2007a) Statistical correlations between the in-mouth textural characteristics and the chemical composition of Shiraz wines. J Agric Food Chem 55:2683–7.

Gawel R, Sluyter S, Waters E (2007b) The effects of ethanol and glycerol on the body and other sensory characteristics of Riesling wines. Aust J Grape Wine Res 13:38–45.

Gerke JP, Chen CTL, Cohen B a (2006) Natural isolates of *Saccharomyces cerevisiae* display complex genetic variation in sporulation efficiency. Genetics 174:985–97.

Giudici P, Pulvirenti A (2002) Molecular methods for identification of wine yeasts. In: Ciani M (ed) Biodiversity and Biotechnology of Wine Yeasts. p 3552

Goddard MR, Anfang N, Tang R, Gardner RC, 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. Environ Microbiol 12:63–73.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 Genes conveniently among the different interna- Old Questions and New Answers The genome . At the beginning of the se- of its more complex relatives in the eukary- cerevisiae has been completely sequenced *Schizosaccharomyces pombe* indicate.

Goldner MC, Zamora MC, Di Leo Lira P, Gianninoto H, Bandoni A (2009) Effect of Ethanol Level in the Perception of Aroma Attributes and the Detection of Volatile Compounds in Red Wine. J Sens Stud 24:243–257.

Gómez-Pastor R, Pérez-Torrado R, Cabiscol E, Ros J, Matallana E (2010) Reduction of oxidative cellular damage by overexpression of the thioredoxin TRX2 gene improves yield and quality of wine yeast dry active biomass. Microb Cell Fact 9:9.

Gonzalez S, Gallo L, Climent MD, Barrio E, Querol A (2007) Enological characterization of natural hybrids from *Saccharomyces cerevisiae* and S. kudriavzevii. Int J Food Microbiol 116:11–18.

González-Álvarez J, Blanco-Gomis D, Arias-Abrodo P, Díaz-Llorente D, Ríos-Lombardía N, Busto E, Gotor-Fernández V, Gutiérrez-Álvarez MD (2011) Characterization of hexacationic imidazolium ionic liquids as effective and highly stable gas chromatography stationary phases. J Sep Sci 273–279

Granchi L, Bosco M, Messini A, Vincenzini M (1999) Rapid detection and quantification of yeast species during spontaneous wine fermentation by PCR-RFLP analysis of the rDNA ITS region. J Appl Microbiol 87:949–956.

Granchi L, Ganucci D, Viti C, Giovannetti L, Vincenzini M (2003) Saccharomyces cerevisiae biodiversity in spontaneous commercial fermentations of grape musts with “adequate” and “inadequate” assimilable-nitrogen content. Lett Appl Microbiol 36:54–58.

Greig D, Leu J-Y (2009) Natural history of budding yeast. Curr Biol 19:R886–R890.

Guillamón JM, Barrio E, Querol A (1996) Characterization of Wine Yeast Strains of the Saccharomyces Genus on the Basis of Molecular Markers: Relationships Between Genetic Distance and Geographic or Ecological Origin. Syst Appl Microbiol 19:122–132.

Guth H (1997) Quantitation and sensory studies of character impact odorants of different white wine varieties. J Agric Food Chem 45:3027–3032.

Guth H, Sies A (2002) Flavour of wines: Towards an understanding by reconstitution experiments and an analysis of ethanol’s effect on odour activity of key compounds. Proceedings of the Eleventh Australian Wine Industry Technical Conference, Adelaide, Australia. Australian Wine Industry Technical Conference Inc.: Adelaide,SA. pp 128–139

Halket JM, Waterman D, Przyborowska AM, Patel RKP, Fraser PD, Bramley PM (2005) Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot 56:219–243.

Hallinan C, Saul D, Jiranek V (1999) Differential utilisation of sulfur compounds for H2S liberation by nitrogen-starved wine yeasts. Aust J Grape Wine Res 5:82–90.

Hanley J, McNeil B (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36.

Hauser M, Horn P, Tournu H, Hauser NC, Hoheisel JD, Brown AJP, Dickinson JR (2007) A transcriptome analysis of isoamyl alcohol-induced filamentation in yeast reveals a novel role for Gre2p as isovaleraldehyde reductase. FEMS Yeast Res 7:84–92.

Hazelwood L a, Daran J-M, van Maris AJ a, Pronk JT, Dickinson JR (2008) The Ehrlich pathway for fusel alcohol production: a century of research on Saccharomyces cerevisiae metabolism. Appl Environ Microbiol 74:2259–66.

Heinimann K (2013) Toward a Molecular Classification of Colorectal Cancer: The Role of MGMT. Front Oncol 3:1–3.

Henschke PA, Jiranek V (1991) Hydrogen sulfide formation during fermentation: Effect of nitrogen composition in model grape must. In: Rantz J (ed) Proceedings of the International Symposium on Nitrogen in Grapes and Wine. Davis, CA: American Society for Enology and Viticulture, Seattle, USA,

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Lee W, Proctor M, Onge RPS, Tyers M, Altman RB, Davis RW, Nislow C, Giaever G (2008) The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. Science (80-) 320:362–365.

Holgate A (1997) The Influence of Yeast Strain on the Acid Balance of Red Wine. Proceedings Advances in Juice Clarification and Yeast Inoculation, Melbourne, Australia (Australian Society of Viticulture and Oenology: Adelaide, SA). pp 39–41

Hollywood K, Brison DR, Goodacre R (2006) Metabolomics: current technologies and future trends. Proteomics 6:4716–23.

Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. Nat Rev Genet 11:855–866.

Hu XH, Wang MH, Tan T, Li JR, Yang H, Leach L, Zhang RM, Luo ZW (2007) Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*. Genetics 175:1479–87.

Husson F, Josse J, Le S, Mazet J (2012) FactoMineR: multivariate exploratory data analysis and data mining with R. R package version 1.18.

Ibáñez C, Pérez-Torrado R, Chiva R, Guillamón JM, Barrio E, Querol A (2014) Comparative genomic analysis of *Saccharomyces cerevisiae* yeasts isolated from fermentations of traditional beverages unveils different adaptive strategies. Int J Food Microbiol 171:129–35.

Iraqui I, Vissers S, Cartiaux M, Urrestarazu A (1998) Characterisation of *Saccharomyces cerevisiae* ARO8 and ARO9 genes encoding aromatic aminotransferases I and II reveals a new aminotransferase subfamily. Mol Gen Genet 238–248.

Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi a.-L, Barabási a L (2000) The large-scale organization of metabolic networks. Nature 407:651–654.

Jiranek V, Langridge P (1995) Regulation of Hydrogen Sulfide Liberation in Wine-Producing *Saccharomyces cerevisiae* Strains by Assimilable Nitrogen. Appl Environ Microbiol 61:461–467.

Jiranek V, Langridge P, Henschke P (1996) Determination of sulphite reductase activity and its response to assimilable nitrogen status in a commercial *Saccharomyces cerevisiae* wine yeast. J Appl Bacteriol 81:329–336.

Jiranek V, Langridge P, Henschke P (1995) Validation of bismuth-containing indicator media for predicting H2S -producing potential of *Saccharomyces cerevisiae* wine yeasts under enological conditions. Am J Enol Vitic 46:269–273.

Jobling M, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124.

Katou T, Namise M, Kitagaki H, Akao T, Shimoi H (2009) QTL mapping of sake brewing characteristics of yeast. J Biosci Bioeng 107:383–393.

Kent CF, Daskalchuk T, Cook L, Sokolowski MB, Greenspan RJ (2009) The Drosophila foraging gene mediates adult plasticity and gene-environment interactions in behaviour, metabolites, and gene expression in response to food deprivation. PLoS Genet 5:1–13.

Kim HS, Huh J, Fay JC (2009) Dissecting the pleiotropic consequences of a quantitative trait nucleotide. FEMS Yeast Res 9:713–22.

Kind T, Fiehn O (2007) Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8:105.

Kispal G, Steiner H, Court D a., Rolinski B, Lill R (1996) Mitochondrial and Cytosolic Branched-chain Amino Acid Transaminases from Yeast, Homologs of the myc Oncogene-regulated Eca39 Protein. J Biol Chem 271:24458–24464.

Kitano H (2002) Systems biology: a brief overview. Science 295:1662–4.

Kleijn RJ, Geertman J-M a, Nfor BK, Ras C, Schipper D, Pronk JT, Heijnen JJ, van Maris AJ a, van Winden W a (2007) Metabolic flux analysis of a glycerol-overproducing *Saccharomyces cerevisiae* strain based on GC-MS, LC-MS and NMR-derived C-labelling data. FEMS Yeast Res 7:216–31.

Kobayashi J, Shirao M, Nakazawa H (1998) Simultaneous determination of anions and cations in mineral water by capillary electrophoresis with a chelating agent. J Liq Chromatogr Relat Technol 21:1445–1456.

Koek M, Muilwijk B, van Stee L, Hankemeier T (2008) Higher mass loadability in comprehensive two-dimensional gas chromatography-mass spectrometry for improved analytical performance in metabolomics analysis. J Chromatogr 1186:420–429.

Koek MM, Muilwijk B, van der Werf MJ, Hankemeier T (2006) Microbial metabolomics with gas chromatography/mass spectrometry. Anal Chem 78:1272–81.

Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn MM a, Pronk JT, Slijper M, Heck AJR (2006) Proteome analysis of yeast response to various nutrient limitations. Mol Syst Biol 2:2006.0026.

Kondo T, Tezuka H, Ishii J, Matsuda F, Ogino C, Kondo A (2012) Genetic engineering to enhance the Ehrlich pathway and alter carbon flux for increased isobutanol production from glucose by *Saccharomyces cerevisiae*. J Biotechnol 159:32–7.

Kresnowati MT a P, van Winden W a, Almering MJH, ten Pierick a, Ras C, Knijnenburg T a, Daran-Lapujade P, Pronk JT, Heijnen JJ, Daran JM (2006) When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. Mol Syst Biol 2:49.

Kvitek DJ, Will JL, Gasch AP (2008) Variations in stress sensitivity and genomic expression in diverse *Saccharomyces cerevisiae* isolates. PLoS Genet 4:31–35.

Lambrechts M, Pretorius S (2000a) Yeast and its importance to wine aroma South African. J Enol Vitic 27:97–129.

Lambrechts MG, Pretorius IS (2000b) Yeast and its Importance to Wine Aroma - A Review. S Afr J Enol Vitic 21:97–129.

Lander E, Botstein S (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185.

Larroy C, Parés X, Biosca JA (2002) Characterization of a Saccharomyces cerevisiae NADP(H)-dependent alcohol dehydrogenase (ADHVII), a member of the cinnamyl alcohol dehydrogenase family. Eur J Biochem 269:5738–5745.

Lavallée F, Salvas Y, Lamy S, Thomas D, Degre R, Dulau L (1994) PCR and DNA fingerprinting used as quality control in the production of wine yeast strains. Am J Enol Vitic 45:86–91.

Le Berre E, Atanasova B, Langlois D, Etiévant P, Thomas-Danguin T (2007) Impact of ethanol on the perception of wine odorant mixtures. Food Qual Prefer 18:901–908.

Lê Cao K-A, Boitard S, Besse P (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics 12:253.

League GP, Slot JC, Rokas A (2012) The ASP3 locus in *Saccharomyces cerevisiae* originated by horizontal gene transfer from Wickerhamomyces. FEMS Yeast Res 12:859–63.

Lee SY, Knudsen FB (1985) Differentiation of brewery yeast strains by restriction endonuclease analysis of their mitochondrial DNA. J Inst Brew 91:169–173.

Legras JL, Karst F (2003) Optimisation of interdelta analysis for *Saccharomyces cerevisiae* strain characterisation. FEMS Microbiol Lett 221:249–255.

Legras JL, Merdinoglu D, Cornuet JM, Karst F (2007) Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Mol Ecol 16:2091–2102.

Legras J-L, Ruh O, Merdinoglu D, Karst F (2005a) Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. Int J Food Microbiol 102:73–83.

Legras J-L, Ruh O, Merdinoglu D, Karst F (2005b) Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. Int J Food Microbiol 102:73–83.

Lehmann R, Voelter W, Liebich H (1997) Capillary electrophoresis in clinical chemistry. J Chromatogr B Biomed Sci Appl 697:3–35.

Lilly M, Bauer FF, Lambrechts MG, Swiegers JH, Cozzolino D, Pretorius IS, Lilly M (2006a) The effect of increased yeast alcohol acetyltransferase and esterase activity on the flavour profiles of wine and distillates. Yeast 23:641–659.

Lilly M, Bauer FF, Styger G, Lambrechts MG, Pretorius IS (2006b) The effect of increased branched-chain amino acid transaminase activity in yeast on the production of higher alcohols and on the flavour profiles of wine and distillates. FEMS Yeast Res 6:726–43.

Lilly M, Lambrechts MG, Pretorius IS (2000) Effect of increased yeast alcohol acetyltransferase activity on flavor profiles of wine and distillates. Appl Environ Microbiol 66:744–53.

Linderholm A, Olineka T, Hong Y, Bisson L (2006) Allele Diversity among Genes of the Sulfate Reduction Pathway in Wine Strains of *Saccharomyces cerevisiae*. Am J Enol Vitic 57:431–440.

Linderholm AL, Findleton CL, Kumar G, Hong Y, Bisson LF (2008) Identification of genes affecting hydrogen sulfide formation in *Saccharomyces cerevisiae*. Appl Environ Microbiol 74:1418–27.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O´Kelly MJT, van Oudenaarden A, Barton DBH, Bailes E, Ba ANN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R, Louis EJ (2009) Population genomics of domestic and wild yeasts. Nature.

Liti G, Louis EJ (2012) Advances in Quantitative Trait Analysis in Yeast. PLoS Genet.

Liti G, Schacherer J (2011) The rise of yeast population genomics. C R Biol 334:612–619.

Liu S (2002) A review: malolactic fermentation in wine -- beyond deacidification. J Appl Microbiol 92:589–601.

Lommen A, an der Weg G, van Engelen M, Bor G, Hoogenboom L, Nielen M (2007) An untargeted metabolomics approach to contaminant analysis: pinpointing potential unknown compounds. Anal Chim Acta 584:43–49.

Loo J, Udseth H, Smith R (1989) Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. Anal Biochem 179:404–412.

Lopes C, Van Broock M, Querol A, Caballero A (2002) *Saccharomyces cerevisiae* wine yeast populations in a cold region in Argentinean Patagonia. A study at different fermentation scales. J Appl Microbiol 93:608–615.

López V, Querol A, Ramón D, Fernández-Espinar MT (2001) A simplified procedure to analyse mitochondrial DNA from industrial yeasts. Int J Food Microbiol 68:75–81.

López-Malo M, Chiva R, Rozes N, Guillamon JM (2013) Phenotypic analysis of mutant and overexpressing strains of lipid metabolism genes in *Saccharomyces cerevisiae*: implication in growth at low temperatures. Int J Food Microbiol 162:26–36.

Lu H, Liang Y, Dunn WB, Shen H, Kell DB (2008) Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. TrAC - Trends Anal Chem 27:215–227.

Lubbers S, Verret C, Voilley A (2001) The Effect of Glycerol on the Perceived Aroma of a Model Wine and a White Wine. Leb u-Technol 34:262–265.

Lynch M, Walsh B (1998) Quantitative trait loci. Genetics and analysis of quantitative traits. Sunderland: Sinauer, pp 319–532

MacDonald NJ, Beiko RG (2010) Efficient learning of microbial genotype-phenotype association rules. Bioinformatics 26:1834–40.

Mackay TFC (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35:303–339.

MacKenzie D a, Defernez M, Dunn WB, Brown M, Fuller LJ, de Herrera SRMS, Günther A, James S a, Eagles J, Philo M, Goodacre R, Roberts IN (2008) Relatedness of medically important strains of Saccharomyces cerevisiae as revealed by phylogenetics and metabolomics. Yeast 25:501–12.

Magwene PM, Kayıkçı Ö, Granek J a, Reininga JM, Scholl Z, Murray D (2011) Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 108:1987–92.

Manasatienkij C, Rangabpai C (2012) Clinical application of forensic DNA analysis: a literature review. J Med Assoc Thail 95:1357– 1363.

Mannazzu I, Clementi F, Ciani M (2002) Strategies and criteria for the isolation and selection of autochthonous starter. In: Ciani M (ed) Biodiversity and biotechnology of wine yeasts. pp 19–33

Manzanares P, Orejas M, Gil JV, Graaff LH De, Visser J, Ramón D (2003) Construction of a Genetically Modified Wine Yeast Strain Expressing the Aspergillus aculeatus rhaA Gene , Encoding an α -l-Rhamnosidase of Enological Interest Construction of a Genetically Modified Wine Yeast Strain Expressing the Aspergillus aculeatus rh. Appl Environ Microbiol.

Marks VD, Ho Sui SJ, Erasmus D, van der Merwe GK, Brumm J, Wasserman WW, Bryan J, van Vuuren HJJ (2008) Dynamics of the yeast transcriptome during wine fermentation reveals a novel fermentation stress response. FEMS Yeast Res 8:35–52.

Martineau B, Acree TE, Henick-Kling T (1995) Effect of wine type on the detection threshold for diacetyl. Food Res Int 28:139–143.

Martínez C, Gac S, Lavín A, Ganga M (2004) Genomic characterization of Saccharomyces cerevisiae strains isolated from wine-producing areas in South America. J Appl Microbiol 96:1161–1168.

Marullo P, Mansour C, Dufour M, Albertin W, Sicard D, Bely M, Dubourdieu D (2009) Genetic improvement of thermo-tolerance in wine *Saccharomyces cerevisiae* strains by a backcross approach. FEMS Yeast Res 9:1148–1160.

Marullo P, Yvert G, Bely M, Aigle M, Dubourdieu D (2007) Efficient use of DNA molecular markers to construct industrial yeast strains. FEMS Yeast Res 7:1295–1306.

Mas S, Villas-Boas S, Hansen M, Akesson M, Nielsen J (2007) A comparison of direct infusion MS and GC-MS for metabolic footprinting of yeast mutants. Biotechnol Bioeng 96:1014–1022.

Masneuf-Pomarède I, Le Jeune C, Durrens P, Lollier M, Aigle M, Dubourdieu D (2007) Molecular typing of wine yeast strains *Saccharomyces bayanus* var. uvarum using microsatellite markers. Syst Appl Microbiol 30:75–82.

Mason A, Dufour J (2000) Alcohol acetyltransferases and the significance of ester synthesis in yeast. Yeast 16:1287–1298.

McCullough MJ, Clemons K V., Farina C, McCusker JH, Stevens D a. (1998) Epidemiological investigation of vaginal *Saccharomyces cerevisiae* isolates by a genotypic method. J Clin Microbiol 36:557–562.

McCusker JH, Clemons K V, David A, Davis RW (1994) Genetic Characterization of Pathogenic *Saccharomyces cerevisiae* Isolates. Genetics 136:1261–1269.

Mehmood T, Martens H, Saebø S, Warringer J, Snipen L (2011) Mining for genotype-phenotype relations in Saccharomyces using partial least squares. BMC Bioinformatics 12:318.

Meilgaard M (1975) Flavor chemistry of beer: Part II: Flavor and threshold of 239 aroma volatiles. MBAA Tech Q 12:151–168.

Mendes I, Franco-Duarte R, Umek L, Fonseca E, Drumonde-Neves J, Dequin S, Zupan B, Schuller D (2013) Computational models for prediction of yeast strain potential for winemaking from phenotypic profiles. PLoS One 8:e66523.

Mestres M, Busto O, Guasch J (2000) Analysis of organic sulfur compounds in wine aroma. J Chromatogr A 881:569–581.

Michnick S, Roustan JL, Remize F, Barre P, Dequin S (1997) Modulation of glycerol and ethanol yields during alcoholic fermentation in *Saccharomyces cerevisiae* strains overexpressed or disrupted for GPD1 encoding glycerol 3-phosphate dehydrogenase. Yeast 13:783–93.

Mondello L, Tranchida P, Dugo P, Dugo G (2008) Comprehensive two-dimensional gas chromatography-mass spectrometry: A review. Mass Spectrom Rev 27:101–124.

Monton MRN, Soga T (2007) Metabolome analysis by capillary electrophoresis-mass spectrometry. J Chromatogr A 1168:237–46; discussion 236.

Moreira N, Mendes F, Pereira O, Guedes de Pinho P, Hogg T, Vasconcelos I (2002) Volatile sulphur compounds in wines related to yeast metabolism and nitrogen composition of grape musts. Anal Chim Acta 458:157–167.

Mozina M, Demšar J, Kattan M, Zupan B (2004) Nomograms for visualization of naive Bayesian classifier. Lecture Notes in Computer Science. pp 337–348

Muller L a H, McCusker JH (2009) Microsatellite analysis of genetic diversity among clinical and nonclinical *Saccharomyces cerevisiae* isolates suggests heterozygote advantage in clinical environments. Mol Ecol 18:2779–2786. x

Nadal D, Carro D, Fernández-Larrea J, Piña B (1999) Analysis and dynamics of the chromosomal complements of wild sparkling-wine yeast strains. Appl Environ Microbiol 65:1688–1695.

Nägele E, Vollmer M, Hörth P (2004) Improved 2D nano-LC/MS for proteomics applications: a comparative analysis using yeast proteome. J Biomol Tech 15:134–143.

Naumov G, Naumova E, Gaillardin C (1993) Genetic and Karyotypic Identification of Wine *Saccharomyces bayanus* Yeasts Isolated in France and Italy. Syst Appl Microbiol 16:274–279.

Navarro-Avino JP, Prasad R, Miralles VJ, Benito RM (1999) A Proposal for Nomenclature of Aldehyde Dehydrogenases in *Saccharomyces cerevisiae* and characterization of the Stress-Inducible ALD2 and ALD3 Genes. Yeast 15:829–842.

Ness F, Lavalee F, Dubordieu D, Aigle M, Dulau L (1993) Identification of yeast strains using the polymerase chain reaction. J Sci Food Agric 62:89–94.

Nieuwoudt H, Prior B, Pretorius S, Bauer F (2002) Glycerol in South African Table Wines: An Assessment of its Relationship to Wine Quality. S Afr J Enol Vitic 23:22–30.

Noble A, Bursick G (1984) The Contribution of Glycerol to Perceived Viscosity and Sweetness in White Wine. Am J Enol Vitic 35:110–112.

Nogales-Cadenas R, Carmona-Saez P, Vazquez M, Vicente C, Yang X, Tirado F, Carazo JM, Pascual-Montano A (2009) GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. Nucleic Acids Res 37:W317–22.

Nogami S, Ohya Y, Yvert G (2007) Genetic complexity and quantitative trait loci mapping of yeast morphological traits. PLoS Genet 3:e31.

Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S, Cambon B, Legras J-L, Wincker P, Casaregola S, Dequin S (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc Natl Acad Sci U S A 106:16333–8.

Nykänen L (1986) Formation and Occurrence of Flavor Compounds in Wine and Distilled Alcoholic Beverages. Am J Enol Vitic 37:84–96.

Nykänen L, Nykänen I, Suomalainen H, Hautvillicrs C (1977) Distribution of esters produced during sugar fermentation between the yeast cell and the medium. J Inst Brew 83:32–34.

O'Connor TD, Mundy NI (2009) Genotype-phenotype associations: Substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. Bioinformatics 25:94–100.

Ochiai N, Fujimura M, Oshima M, Motoyama T, Ichiishi A, Yamada-Okabe H, Yamaguchi I (2002) Effects of iprodione and fludioxonil on glycerol synthesis and hyphal development in *Candida albicans*. Biosci Biotechnol Biochem 66:2209–2215.

Paffetti D, Barberio C, Casalone E, Cavalieri D, Fani R, Fia G, Mori E, Polsinelli M (1995) DNA fingerprinting by random amplified polymorphic DNA and restriction fragment length polymorphism is useful for yeast typing. Res Microbiol 146:587–594.

Pan Z, Raftery D (2007) Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Anal Bioanal Chem 387:525–527.

Park SK, Boulton R, Noble AC (2000) Formation of hydrogen sulfide and glutathione during fermentation of white grape musts. Am J Enol Vitic 51:91–97.

Parts L, Cubillos F a., Warringer J, Jain K, Salinas F, Bumpstead SJ, Molin M, Zia A, Simpson JT, Quail M a., Moses A, Louis EJ, Durbin R, Liti G (2011) Revealing the genetic structure of a trait by sequencing a population under selection. Genome Res 21:1131–1138.

Pemberton TJ, DeGiorgio M, Rosenberg N a (2013) Population structure in a comprehensive genomic data set on human microsatellite variation. G3 Genes|Genomes|Genetics 3:891–907.

Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP (2003) Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC / LC - MS / MS) for Large-Scale Protein Analysis : The Yeast Proteome research articles. J Proteome Res 2:43–50.

Pérez MA, Gallego FJ, Martínez I, Hidalgo P (2001) Detection, distribution and selection of microsatellites (SSRs) in the genome of the yeast *Saccharomyces cerevisiae* as molecular markers. Lett Appl Microbiol 33:461–6.

Perez-Gonzalez J, Gonzalez R, Querol A, Sendra J, Ramon D (1993) Construction of a Recombinant Wine Yeast Strain Expressing Microvinification Processes. Appl Environ Microbiol 59:2801–2806.

Pérez-Ortín JE, Querol A, Puig S, Barrio E (2002) Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. Genome Res 12:1533–1539.

Perlstein EO, Ruderfer DM, Ramachandran G, Haggarty SJ, Kruglyak L, Schreiber SL (2006) Revealing complex traits with small molecules and naturally recombinant yeast strains. Chem Biol 13:319–27.

Perlstein EO, Ruderfer DM, Roberts DC, Schreiber SL, Kruglyak L (2007) Genetic basis of individual differences in the response to small-molecule drugs in yeast. Nat Genet 39:496–502.

Perpète P, Duthoit O, De Maeyer S, Imray L, Lawton AI, Stavropoulos KE, Gitonga VW, Hewlins MJE, Dickinson JR (2006) Methionine catabolism in *Saccharomyces cerevisiae*. FEMS Yeast Res 6:48–56.

Perrett D, Alfrzema L, Hows M, Gibbons J (1997) Capillary electrophoresis for small molecules and metabolites. Biochem Soc Trans 25:273.

Perrett D, Birch A, Ross G (1994) Capillary electrophoresis for peptides, including neuropeptides. Biochem Soc Trans 22:127–131.

Perrett D, Ross G (1992) Capillary electrophoresis - a powerful tool for biomedical analysis and research. Trends Anal Chem 11:156–163.

Pigeau GM, Inglis DL (2007) Response of wine yeast (*Saccharomyces cerevisiae*) aldehyde dehydrogenases to acetaldehyde stress during Icewine fermentation. J Appl Microbiol 103:1576–86.

Pigliucci M (2001) What is phenotypic plasticity? Phenotypic plasticity: beyond nature and nurture. The John Hopkins university press, pp 1–28

Pizarro F, Vargas FA, Agosin E (2007) A systems biology perspective of wine fermentations. Yeast 24:977–991.

Pizarro FJ, Jewett MC, Nielsen J, Agosin E (2008) Growth temperature exerts differential physiological and transcriptional responses in laboratory and wine strains of *Saccharomyces cerevisiae*. Appl Environ Microbiol 74:6358–68.

Pramateftaki P V., Lanaridis P, Typas M a. (2000) Molecular identification of wine yeasts at species or strain level: A case study with strains from two vine-growing areas of Greece. J Appl Microbiol 89:236–248.

Pretorius IS (2000) Tailoring wine yeast for the new millennium : novel approaches to the ancient art of winemaking. Yeast 16:675–729.

Primig M, Williams RM, Winzeler EA, Tevzadze GG, Conway AR, Hwang SY, Davis RW, Esposito RE (2000) The core meiotic transcriptome in budding yeasts. Nat Genet 26:415–423.

Pulvirenti A, Caggia C, Restuccia C, Gullo M, Giudici P (2001) DNA fingerprinting methods used for identification of yeasts isolated from Sicilian sourdoughs. Ann Microbiol 51:107–120.

Pulvirenti A, Solieri L, Gullo M, De Vero L, Giudici P (2004) Occurrence and dominance of yeast species in sourdough. Lett Appl Microbiol 38:113–117.

Querol A, Barrio E, Huerta T, Ramon D (1992) Molecular monitoring of wine fermentations conducted by active dry yeast strains. Appl Environ Microbiol 58:2948–2953.

Quesada M, Cenis J (1995) Use of random amplified polymorphic DNA (RAPD-PCR) in the characterization of wine yeasts. Am J Enol Vitic 46:204–208.

Quinlan J (1986) Induction of Decision Trees. Mach Learn 1 1:81–106.

Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes a, Walsh MC, Berden J a, Brindle KM, Kell DB, Rowland JJ, Westerhoff H V, van Dam K, Oliver SG (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol 19:45–50.

Rabilloud T (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. Proteomics 2:3–10.

Rachidi N, Barre P, Blondin B (1999) Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. Mol Gen Genet 261:841–50.

Radler F (1993) Yeast: Metabolism of organic acids. In: Fleet GH (ed) Wine Microbiology and Biotechnology.

Ramautar R, Somsen GW, de Jong GJ (2009) CE-MS in metabolomics. Electrophoresis 30:276–91.

Ramírez M, Peréz F, Regodón J a, Pere F, Rami M (1998) A Simple and Reliable Method for Hybridization of Homothallic Wine Strains of *Saccharomyces cerevisiae*. App 64:5039–5041.

Rankine B (1967) Influence of yeast strain and pH on pyruvic acid content of wines. J Sci Food Agric 18:41–44.

Rankine B (1968a) The importance of yeasts in determining the composition and quality of wines. Vitis 7:22–49.

Rankine B (1968b) Formation of α-ketoglutaric acid by wine yeasts and its oenological significance. J Sci Food Agric 19:624–627.

Rankine B (1963) Nature, origin and prevention of hydrogen sulphide aroma in wines. J Sci Food Agric 14:79–91.

Rankine B, Bridson D (1971) Glycerol in Australian Wines and Factors Influencing Its Formation. Am J Enol Vitic 23:6–12.

Rapp A, Versini G (1996) Influence of nitrogen compounds in grapes on aroma compounds of wines. Wein-Wissenschaft 51:193–203.

Ratnakumar S, Hesketh A, Gkargkas K, Wilson M, Rash BM, Hayes A, Tunnacliffe A, Oliver SG (2011) Phenomic and transcriptomic analyses reveal that autophagy plays a major role in desiccation tolerance in Saccharomyces cerevisiae. Mol Biosyst 7:139–149. doi: 10.1039/c0mb00114g

Rauhut D (1993) Yeasts – Production of sulfur compounds. In: Fleet G (ed) Wine microbiology and biotechnology. Harwood Academic Publishers, Chur, Switzerland, pp 183–223

Rauhut D, Kurbel H, Dittrich H, Grossmann M (1996) Properties and differences of commercial yeast strains with respect to their formation of sulfur compounds. Vitic Enol Sci 51:187–192.

Remize F, Andrieu E (2000) Engineering of the Pyruvate Dehydrogenase Bypass in *Saccharomyces cerevisiae* : Role of the Cytosolic Mg 2+ and Mitochondrial K+ Acetaldehyde Dehydrogenases Ald6p and Ald4p in Acetate Formation during Alcoholic Fermentation. Appl Environ Microbiol 66:3151–3159.

Remold SK, Lenski RE (2004) Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. Nat Genet 36:423–426. doi: 10.1038/ng1324

Ribéreau-Gayon P, Dubourdieu D, Doneche B, Lonvaud A (2000a) Handbook of Enology Volume 1 The Microbiology of Wine and Vinifications., 2nd edn. John Wiley & Sons Ltd

Ribéreau-Gayon P, Glories Y, Maujean A, Dubourdieu D (2000b) The chemistry of wine stabilisation and treatments. Handbook of Enology, Volume 2. John Wiley & Sons Ltd, Chichester, UK,

Richard GF, Hennequin C, Thierry A, Dujon B (1999) Trinucleotide repeats and other microsatellites in yeasts. Res Microbiol 150:589–602.

Richards KD, Goddard MR, Gardner RC (2009) A database of microsatellite genotypes for *Saccharomyces cerevisiae*. Antonie Van Leeuwenhoek 96:355–359.

Roberts IN, Oliver SG (2011) The yin and yang of yeast: biodiversity research and systems biology as complementary forces driving innovation in biotechnology. Biotechnol Lett 33:477–487.

Robinson J (1994) The Oxford companion to wine.

Rollero S, Bloem A, Camarasa C, Sanchez I, Ortiz-Julien A, Sablayrolles J-M, Dequin S, Mouret J-R (2014) Combined effects of nutrients and temperature on the production of fermentative aromas by *Saccharomyces cerevisiae* during wine fermentation. Appl Microbiol Biotechnol 99:2291–2304.

Romano A, Casaregola S, Torre P, Gaillardin C (1996) Use of RAPD and mitochondrial DNA RFLP for typing of *Candida zeylanoides* and *Debaryomyces hansenii* yeast strains isolated from cheese. Syst Appl Microbiol 19:255–264.

Rossignol T, Dulau L, Julien A, Blondin B (2003) Genome-wide monitoring of wine yeast gene expression during alcoholic fermentation. Yeast 20:1369–85.

Rossouw D, Bauer F (2009a) Comparing the transcriptomes of wine yeast strains: toward understanding the interaction between environment and transcriptome during fermentation.

Rossouw D, Bauer FF (2009b) Wine Science in the Omics Era : The Impact of Systems Biology on the Future of Wine Research. South African J Enol Vitic 30:101–109.

Rossouw D, Naes T, Bauer FF (2008) Linking gene regulation and the exo-metabolome: a comparative transcriptomics approach to identify genes that impact on the production of volatile aroma compounds in yeast. BMC Genomics 9:530.

Rossouw D, Olivares-Hernandes R, Nielsen J, Bauer FF (2009) Comparative transcriptomic approach to investigate differences in wine yeast physiology and metabolism during fermentation. Appl Environ Microbiol 75:6600–12.

Rossouw D, Van Den Dool AH, Jacobson D, Bauer FF (2010) Comparative transcriptomic and proteomic profiling of industrial wine yeast strains. Appl Environ Microbiol 76:3911–3923.

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65.

Roustan J, Sablayrolles J (2002) Impact of the addition of electron acceptors on the by-products of alcoholic fermentation. Enzyme Microb Technol 31:142–152.

Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Güldener U, Mannhaupt G, Münsterkötter M, Mewes HW (2004) The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Res 32:5539–45.

Sabate J, Cano J, Querol a., Guillamón JM (1998) Diversity of Saccharomyces strains in wine fermentations: Analysis for two consecutive years. Lett Appl Microbiol 26:452–455.

Sabino R, Sampaio P, Rosado L, Stevens DA, Clemons K V., Pais C (2010) New polymorphic microsatellite markers able to distinguish among Candida parapsilosis sensu stricto isolates. J Clin Microbiol 48:1677–1682.

Sablayrolles J, Barre P, Grenier P (1987) Design of laboratory automatic system for studying alcoholic fermentations in anisothermal enological conditions. Biotechnol Technol 1:181–184.

Sabon I, De Revel G, Kotseridis Y, Bertrand A (2002) Determination of volatile compounds in Grenache wines in relation with different terroirs in the Rhone Valley. J Agric Food Chem 50:6341–5.

Saerens SMG, Delvaux F, Verstrepen KJ, Van Dijck P, Thevelein JM, Delvaux FR (2008) Parameters affecting ethyl ester production by *Saccharomyces cerevisiae* during fermentation. Appl Environ Microbiol 74:454–61.

Saerens SMG, Delvaux FR, Verstrepen KJ, Thevelein JM (2010) Production and biological function of volatile esters in *Saccharomyces cerevisiae*. Microb Biotechnol 3:165–77.

Saerens SMG, Verstrepen KJ, Van Laere SDM, Voet ARD, Van Dijck P, Delvaux FR, Thevelein JM (2006) The *Saccharomyces cerevisiae* EHT1 and EEB1 genes encode novel enzymes with medium-chain fatty acid ethyl ester synthesis and hydrolysis capacity. J Biol Chem 281:4446–56.

Saint-Prix F, Bönquist L, Dequin S (2004) Functional analysis of the ALD gene family of *Saccharomyces cerevisiae* during anaerobic growth on glucose: The NADP+-dependent Ald6p and Ald5p isoforms play a major role in acetate formation. Microbiology 150:2209–2220.

Salek RM, Maguire ML, Bentley E, Rubtsov D V, Hough T, Cheeseman M, Nunez D, Sweatman BC, Haselden JN, Cox RD, Connor SC, Griffin JL (2007) A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. Physiol Genomics 29:99–108.

Salinas F, Cubillos F a., Soto D, Garcia V, Bergström A, Warringer J, Ganga MA, Louis EJ, Liti G, Martinez C (2012a) The Genetic Basis of Natural Variation in Oenological Traits in *Saccharomyces cerevisiae*. PLoS One.

Salinas F, Cubillos F, Soto D, Garcia V, Bergström A, Warringer J, Ganga MA, Louis EJ, Liti G, Martinez C (2012b) The genetic basis of natural variation in oenological traits in *Saccharomyces cerevisiae*. PLoS One 7:e49640.

Salinas F, Mandaković D, Urzua U, Massera A, Miras S, Combina M, Angelica Ganga M, Martínez C (2010) Genomic and phenotypic comparison between similar wine yeast strains of *Saccharomyces cerevisiae* from different geographic origins. J Appl Microbiol 108:1850–1858.

Salmon J (1987) L-Malic-acid permeation in resting cells of anaerobically grown *Saccharomyces cerevisiae*. Biochim Biophys Acta - Biomembr 901:30–34.

Salo P (1970) Determining the Odor Thresholds for Some Compounds in Alcoholic Beverages. J Food Sci 35:95–99.

Sampaio P, Gusmão L, Alves C, Pina-Vaz C, Amorim A, Pais C (2003) Highly Polymorphic Microsatellite for Identification of *Candida albicans* Strains. J Clin Microbiol 41:552–557.

Sampaio P, Sampaio P, Gusma L, Gusma L, Correia A, Correia A (2005) New Microsatellite Multiplex PCR for *Candida albicans* Strain Typing Reveals Microevolutionary Changes. J Clin Microbiol 43:3869–3876.

Schacherer J, Ruderfer DM, Gresham D, Dolinski K, Botstein D, Kruglyak L (2007) Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. PLoS One 2:1–7.

Schacherer J, Shapiro J, Ruderfer D, Kruglyak L (2009a) Comprehensive polymorphism survey elucidates population structure of *S. cerevisiae*. Nature 458:342–345.

Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009b) Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. Nature 458:342–345.

Schork NJ (1997) Genetics of complex disease: Approaches, problems, and solutions. Am J Respir Crit Care Med 156:103–109.

Schreiner M, Razzazi E, Luf W (2003) Determination of water-soluble vitamins in soft drinks and vitamin supplements using capillary electrophoresis. Nahrung - Food 47:243–247.

Schreirer P (1979) Flavour composition of wines. Crit Rev Food Sci Nutr 12:59–111.

Schuller D (2010) Better Yeast for Better Wine — Genetic Improvement of *Saccharomyces Cerevisiae* Wine Strains. In: Rai M, Koviks G (eds) Progress in Mycology.

Schuller D, Alves H, Dequin S, Casal M (2005) Ecological survey of Saccharomyces cerevisiae strains from vineyards in the Vinho Verde Region of Portugal. FEMS Microbiol Ecol 51:167–177.

Schuller D, Cardoso F, Sousa S, Gomes P, Gomes AC, Santos M a S, Casal M (2012) Genetic diversity and population structure of *Saccharomyces cerevisiae* strains isolated from different grape varieties and winemaking regions. PLoS One.

Schuller D, Casal M (2005) The use of genetically modified *Saccharomyces cerevisiae* strains in the wine industry. Appl Microbiol Biotechnol 68:292–304.

Schuller D, Casal M (2007) The genetic structure of fermentative vineyard-associated *Saccharomyces cerevisiae* populations revealed by microsatellite analysis. Antonie van Leeuwenhoek, Int J Gen Mol Microbiol 91:137–150.

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 625–636.

Schuller D, Valero E, Dequin S, Casal M (2004) Survey of molecular methods for the typing of wine yeast strains. FEMS Microbiol Lett 231:19–26.

Schure EG, Flikweert MT, Dijken JP Van, Pronk JT, Verrips CT (1998) Pyruvate Decarboxylase Catalyzes Decarboxylation of Branched-Chain 2-Oxo Acids but Is Not Essential for Fusel Alcohol Production by *Saccharomyces cerevisiae.*

Shirao M, Furuta R, Suzuki S, Nakasawa H, Fujita S, Maruyama T (1994) Determination of organic-acids in urine by capillary zone electrophoresis. J Chromatogr A 680:247–251.

Si T, Luo Y, Xiao H, Zhao H (2014) Utilizing an endogenous pathway for 1-butanol production in Saccharomyces cerevisiae. Metab Eng 1–9.

Sicard D, Legras J-L (2011) Bread, beer and wine: yeast domestication in the Saccharomyces sensu stricto complex. C R Biol 334:229–36.

Siebert TE, Smyth HE, Capone DL, Neuwöhner C, Pardon KH, Skouroumounis GK, Herderich MJ, Sefton M a, Pollnitz AP (2005) Stable isotope dilution analysis of wine fermentation products by HS-SPME-GC-MS. Anal Bioanal Chem 381:937–47.

Sinha H, Nicholson BP, Steinmetz LM, McCusker JH (2006) Complex genetic interactions in a quantitative trait locus. PLoS Genet 2:e13.

Smolka M, Zhou H, Aebersold R (2001) Quantitative Protein Profiling Using Two-dimensional Gel Electrophoresis, Isotope-coded Affinity Tag Labeling, and Mass Spectrometry. Mol Cell Proteomics 1:19–29.

Smyth GK, Speed T (2003) Normalization of cDNA microarray data. Methods 31:265–273.

Soga T, Heiger D (2000) Amino acid analysis by capillary electrophoresis electrospray ionization mass spectrometry. Anal Chem 72:1236–1241.

Soga T, Heiger D (1998) Simultaneous determination of monosaccharides in glycoproteins by capillary electrophoresis. Anal Biochem 261:73–78.

Soga T, Imaizumi M (2001) Capillary electrophoresis method for the analysis of inorganic anions, organic acids, amino acids, nucleotides, carbohydrates and other anionic compounds. Electrophoresis 22:3418–3425.

Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, Nishioka T (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res 2:488–94.

Soga T, Ueno Y, Naraoka H, Ohashi Y, Tomita M, Nishioka T (2002) Simultaneous determination of anionic intermediates for Bacillus subtilis metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. Anal Chem 74:2233–2239.

Soles R, Ough C, Kunkee R (1982) Ester concentration differences in wine fermented by various species and strains of yeasts. Am J Enol Vitic 33:94–98.

Spiropoulos A, Tanaka J, Flerianos I, Bisson L (2000) Characterization of hydrogen sulfide formation in commercial and natural wine isolates of Saccharomyces. Am J Enol Vitic 51:233–248.

Sponholz W (1993) Wine spoilage by microorganisms. In: Fleet G (ed) Wine Microbiology and Technology. Harwood Academic Publishing, Amsterdam, pp 395–420

Steinmetz LM, Sinha H, Richards DR, Spiegelman J, Oefner P, McCusker J, Davis R (2002) Dissecting the architecture of a quantitative trait locus in yeast. Nature 416:326–330.

Stelling J (2004) Mathematical models in microbial systems biology. Curr Opin Microbiol 7:513–8.

Steyer D, Ambroset C, Brion C, Claudel P, Delobel P, Sanchez I, Erny C, Blondin B, Karst F, Legras J-L (2012) QTL mapping of the production of wine aroma compounds by yeast. BMC Genomics 13:573.

Stratford M, Rose A (1986) Transport of sulfur dioxide by *Saccharomyces cerevisiae*. J Gen Microbiol 1–6.

Stratford M, Rose A (1985) Hydrogen Sulphide Production from Sulphite by *Succhuvumyces cevevisiae*. J Gen Microbiol 1417–1424.

Strauss C, Wilson B, Gooley P, Williams P (1986) Role of monoterpenes in grape and wine flavor. ACS Symposium series-American Chemical Society (USA).

Styger G, Jacobson D, Bauer FF (2011a) Identifying genes that impact on aroma profiles produced by *Saccharomyces cerevisiae* and the production of higher alcohols. Appl Microbiol Biotechnol 91:713–30.

Styger G, Prior P, Bauer F (2011b) Wine flavor and aroma. J Ind Microbiol Biotechnol 38:1145–59.

Sutherland CM, Henschke PA, Langridge P, de Barros Lopes M (2003) Subunit and cofactor binding of *Saccharomyces cerevisiae* sulfite reductase – towards developing wine yeast with lowered. Aust J Grape Wine Res 9:186–193.

Swiegers J, Chambers P, Pretorius I (2005a) Olfaction and taste: Human perception, physiology and genetics. Aust J Grape Wine Res 11:109–113.

Swiegers J, Pretorius I (2005) Yeast modulation of wine flavor. Adv Appl Microbiol 131–175.

Swiegers JH, Bartowsky EJ, Henschke P a., Pretorius IS (2005b) Yeast and bacterial modulation of wine aroma and flavour. Aust J Grape Wine Res 11:139–173.

Swiegers JH, Capone DL, Pardon KH, Elsey GM, Sefton MA, Francis IL, Pretorius IS (2007) Engineering volatile thiol release in *Saccharomyces cerevisiae* for improved wine aroma. Yeast 24:561–574. doi: 10.1002/yea

Swinnen S, Thevelein JM, Nevoigt E (2012) Genetic mapping of quantitative phenotypic traits in Saccharomyces cerevisiae. FEMS Yeast Res 12:215–27.

Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A (2012) GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. Nucleic Acids Res 40:W478–83.

Tan P, Steinbach M, Kumar V (2006) Introduction to Data Mining.

Tanaka Y, Higashi T, Rakwal R, Wakida S, Iwahashi H (2007) Quantitative analysis of sulfur-related metabolites during cadmium stress response in yeast by capillary electrophoresis-mass spectrometry. J Pharm Biomed Anal 44:608–13.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471.

Techera AG, Jubany S, Carrau FM, Gaggero C (2001) Differentiation of industrial wine yeast strains using microsatellite markers. Lett Appl Microbiol 33:71–75.

Teixeira MC, Raposo LR, Mira NP, Lourenço AB, Sá-Correia I (2009) Genome-wide identification of Saccharomyces cerevisiae genes required for maximal tolerance to ethanol. Appl Environ Microbiol 75:5761–72.

Templin M, Stoll D, Bachmann J, Joos T (2004) Protein Microarrays and Multiplexed Sandwich Immunoassays: What Beats the Beads?

Terabe S, Markuszewski M, Inoue N, Otsuka K, Nishioka T (2001) Capillary electrophoretic techniques toward the metabolome analysis. Pure Appl Chem 73:1563–1572.

The R Core Team (2013) R : A Language and Environment for Statistical Computing. 1:

Thoukis G, Ueda M, Wright D (1965) The formation of succinic acid during alcoholic fermentation. Am J Enol Vitic 16:1–8.

Tokuyama T, Kuraishi H, Aida K, Uemura T (1973) Hydrogen sulfide evolution due to a pantothenic acid deficiency in the yeast requiring this vitamin, with special reference to the effect of adenosine triphosphate on yeast cysteine desulfhydrase. J Gen Appl Microbiol 19:439–466.

Tominaga T, Blanchard L, Darriet P, Dubourdieu D (2000) A Powerful Aromatic Volatile Thiol, 2-Furanmethanethiol, Exhibiting Roast Coffee Aroma in Wines Made from Several Vitis vinifera Grape Varieties.

Tominaga T, Darriet P, Dubourdieu D (1996) Identification of 3-mercaptohexyl acetate in Sauvignon wine, a powerful aromatic compound exhibiting box-tree odor. Vitis 35:207–210.

Tominaga T, Furrer A, Henry R, Dubourdieu D (1998a) Identification of new volatile thiols in the aroma of Vitis vinifera L. var. Sauvignon blanc wines. Flavour Fragr J 13:159–162.

Tominaga T, Gachon C, Dubourdieu D (1998b) A New Type of Flavor Precursors in Vitis vinifera L. cv. Sauvignon Blanc: S-Cysteine Conjugates. J Agric Food Chem 46:5215–5219.

Tornai-Lehoczki J, Dlauchy D (2000) Delimination of brewing yeast strains using different molecular techniques. Int J Food Microbiol 62:37–45.

Ugliano M, Henschke P (2009) Yeasts and wine flavour. In: M.V. Moreno-Arribas and M.C. Polo E (ed) Wine Chemistry and Biochemistry. Springer Science and Business: New York.,

Vadali R V., Bennett GN, San KY (2004) Applicability of CoA/acetyl-CoA manipulation system to enhance isoamyl acetate production in *Escherichia coli*. Metab Eng 6:294–299.

Valero E, Cambon B, Schuller D, Casal M, Dequin S (2007) Biodiversity of Saccharomyces yeast strains from grape berries of wine-producing areas using starter commercial yeasts. FEMS Yeast Res 7:317–329.

Varela C, Cárdenas J, Melo F, Agosin E (2005) Quantitative analysis of wine yeast gene expression profiles under winemaking conditions. Yeast 22:369–83.

Vaz C, Sampaio P, Clemons K V., Huang YC, Stevens DA, Pais C (2011) Microsatellite multilocus genotyping clarifies the relationship of candida parapsilosis strains involved in a neonatal intensive care unit outbreak. Diagn Microbiol Infect Dis 71:159–162.

Vermeulen C, Gijs L, Collin S (2005) Sensorial contribution and formation pathways of thiols in foods: A review. Food Rev Int 21:69–137.

Verstrepen K, Chambers P, Pretorius I (2006) The Development of Superior Yeast Strains for the Food and Beverage Industries: Challenges, Opportunities and Potential Benefits. In: Querol A, Fleet GH (eds) Yeasts in Food and Beverages.

Verstrepen KJ, Laere SDM Van, Vanderhaegen BMP, Derdelinckx G, Dufour J, Pretorius IS, Winderickx J, Thevelein JM, Delvaux FR (2003) Expression Levels of the Yeast Alcohol Acetyltransferase Genes ATF1 , Lg-ATF1 , and ATF2 Control the Formation of a Broad Range of Volatile Esters. Appl Environ Microbiol 69:5228–5237.

Vezinhet F, Blondin B, Hallet J-N (1990) Chromosomal DNA patterns and mitochondrial DNA polymorphism as tools for identification of enological strains of *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 32:568–571.

Villas-Bôas SG, Moxley JF, Akesson M, Stephanopoulos G, Nielsen J (2005) High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. Biochem J 388:669–77.

Vuralhan Z, Luttik MAH, Tai SL, Boer VM, Morais MA, Schipper D, Almering MJH, Ko P, Dickinson JR, Daran J, Pronk JT (2005) Physiological Characterization of the ARO10-Dependent, Broad-Substrate-Specificity 2-Oxo Acid Decarboxylase Activity of Saccharomyces cerevisiae. Appl Environ Microbiol 71:3276–3284.

Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A (2003) High-resolution yeast phenomics resolves different physiological features in the saline response. Proc Natl Acad Sci U S A 100:15724–15729.

Warringer J, Zörgö E, Cubillos F a., Zia A, Gjuvsland A, Simpson JT, Forsmark A, Durbin R, Omholt SW, Louis EJ, Liti G, Moses A, Blomberg A (2011) Trait Variation in Yeast Is Defined by Population History. PLoS Genet 7:e1002111.

Werf MJ Van Der, Overkamp KM, Muilwijk B, Coulier L, Hankemeier T (2007) Microbial metabolomics: Toward a platform with full metabolome coverage. Anal Biochem 370:17–25.

Whiting G (1976) Organic acid metabolism of yeasts during fermentation of alcoholic beverages — a review. J Inst Brew 82:84–92.

Williams JG, Kubelik a R, Livak KJ, Rafalski J a, Tingey S V (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535.

Winterhalter P, Harmsen S, Trani F (1991) A C13-norisoprenoid gentiobioside from quince fruit. Phytochemistry 30:3021–3025.

Winzeler EA, Castillo-Davis CI, Oshiro G, Liang D, Richards DR, Zhou Y, Hartl DL (2003) Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics 163:79–89.

Wittmann C (2007) Fluxome analysis using GC-MS. Microb Cell Fact 6:6.

Xufre A, Albergaria H, Gírio F, Spencer-Martins I (2011) Use of interdelta polymorphisms of Saccharomyces cerevisiae strains to monitor population evolution during wine fermentation. J Ind Microbiol Biotechnol 38:127–132.

Yamamoto N, Amemiya H, Yokomori Y, Shimizu K, Totsuka A (1991) Electrophoretic karyotypes of wine yeasts. Am J Enol Vitic 42:358–363.

Yoshimoto H, Fukushige T, Yonezawa T, Sakai Y, Okawa K, Iwamatsu A, Sone H, Tamai Y (2001) Pyruvate Decarboxylase Encoded by the PDCl Gene Contributes , at Least Partially , to the Decarboxylation of a-ketoisocaproate for Isoamyl Alcohol Formation in *Saccharomyces cerevisiae*. J Biosci Bioeng 92:83–85.

Zambonelli C (1998) Microbiologia e biotecnologia dei vini. Edagricole, Bologna

Zambonelli C, Mutinelli P (1975) Biosynthesis of sulphur amino acids in *Saccharomyces cerevisiae*. I. Genetic analysis of leaky mutants of sulphite reductase.

Zerva L, Hollis RJ, Pfaller MA (1996) In Vitro Susceptibility Testing and DNA Typing of Saccharomyces cerevisiae Clinical Isolates. J Clin Microbiol 34:3031–3034.

Chapter VIII

Supporting material:

supplementary data

Supplementary data S1

Geographical origin and technological application/origin of the 172 *Saccharomyces cerevisiae* strains.

Strain Code	Geographical Origin	Technological application or origin	Provided by	(Liti *et al.* 2009)
Z1	France	Laboratory	Liti, G.	97 Y55
Z2	USA	Laboratory	Liti, G.	17 SK1
Z3	Italy	Clinical	Liti, G.	303 YJM978
Z4	Italy	Clinical	Liti, G.	304 YJM981
Z5	Italy	Clinical	Liti, G.	308 YJM975
Z6	UK	Clinical	Liti, G.	284 322134S
Z7	UK	Clinical	Liti, G.	287 378604X
Z8	UK	Clinical	Liti, G.	288 273614N
Z9	Finland	Natural isolate	Liti, G.	84 DBVPG1788
Z10	Netherlands	Natural isolate	Liti, G.	91 DBVPG1373
Z11	France	Commercial wine strain	Liti, G.	174 YIIc17_E5
Z12	Netherlands	Other fermented beverages	Liti, G.	155 DBVPG6040
Z13	Ireland	Beer	Liti, G.	248 NCYC361
Z14	USA	Natural isolate	Liti, G.	182 YPS606
Z15	USA	Natural isolate	Liti, G.	104 YPS128
Z16	Australia	Bread	Liti, G.	258 YS2
Z17	Netherlands	Bread	Liti, G.	259 YS4
Z18	Singapore	Bread	Liti, G.	262 YS9
Z19	USA	Wine and vine	Liti, G.	181 BC187
Z20	Malaysia	Natural isolate	Liti, G.	278 UWOPS03-461.4
Z21	Malaysia	Natural isolate	Liti, G.	279 UWOPS05-217.3
Z22	Malaysia	Natural isolate	Liti, G.	280 UWOPS05-227.2
Z23	Japan	Saké	Liti, G.	251 K11
Z24	Indonesia	Saké	Liti, G.	252 Y9
Z25	USA	Wine and vine	Liti, G.	345 RM11
Z26	Ethiopia	Bread	Liti, G.	92 DBVPG1853
Z27	Ivory Coast	Other fermented beverages	Liti, G.	253 Y12
Z28	West Africa	Other fermented beverages	Liti, G.	247 NCYC110
Z29	West Africa	Other fermented beverages	Liti, G.	60 DBVPG6044
Z30	Unknown geographical origin	Unknown biological origin	Liti, G.	3 DBVPG6765
Z31	Portugal	Unknown biological origin	Liti, G.	OV 382
Z32	Chile	Wine and vine	Liti, G.	220 L-1374
Z33	Chile	Wine and vine	Liti, G.	221 L-1528
Z34	Hawaii	Natural isolate	Liti, G.	271 UWOPS87-2421
Z35	Australia	Natural isolate	Liti, G.	150 DBVPG1106
Z36	Bahamas	Natural isolate	Liti, G.	270 UWOPS83-787.3
Z37	Portugal	Clinical	Carreto, L.	
Z38	Portugal	Clinical	Carreto, L.	
Z39	Portugal	Clinical	Carreto, L.	
Z40	Portugal – Bairrada	Wine and vine	Carreto, L.	
Z41	Portugal – Bairrada	Wine and vine	Carreto, L.	

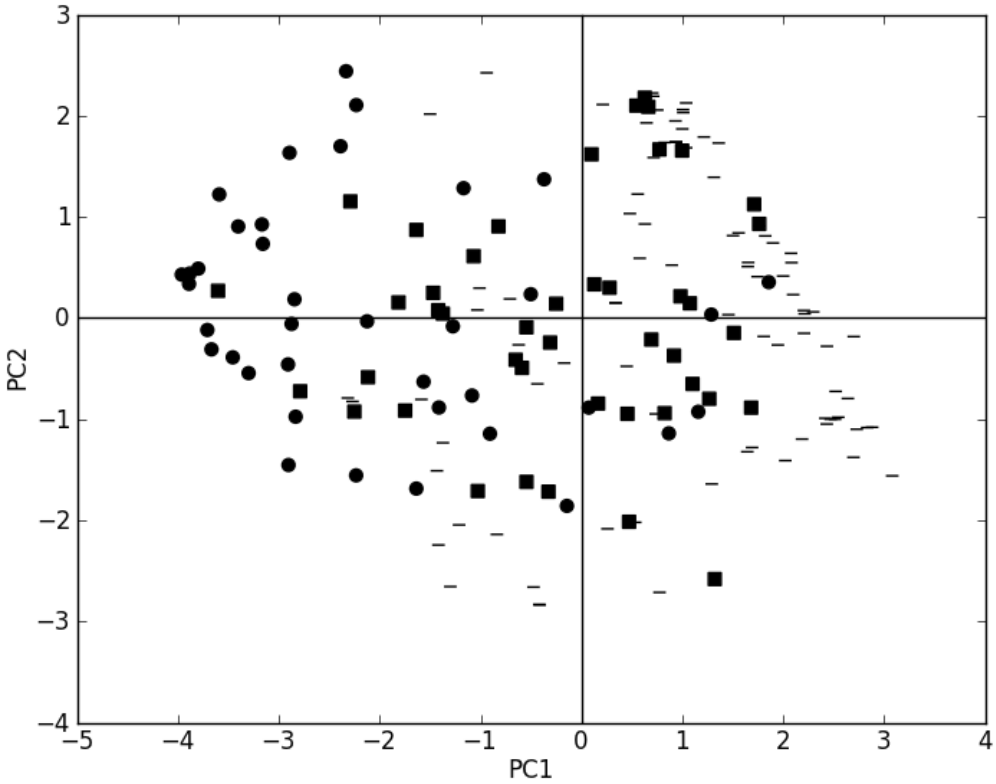
Strain Code	Geographical Origin	Technological application or origin	Provided by (Liti *et al.* 2009)
Z42	Portugal – Bairrada	Wine and vine	Carreto, L.
Z43	Portugal – Bairrada	Wine and vine	Carreto, L.
Z44	France	Commercial wine strain – MAC2338	
Z45	France - Rhône Valley	Commercial wine strain – JCY254 Lalvin	
Z46	France	Commercial wine strain – Fermol Rouge AEB	
Z47	USA	Commercial wine strain – Lalvin 522	
Z48	Japan	Saké	Goto-Yakamoto, N.
Z49	Japan	Saké	Goto-Yakamoto, N.
Z50	Japan	Saké	Goto-Yakamoto, N.
Z51	Japan	Saké	Goto-Yakamoto, N.
Z52	Unknown geographical origin	Natural isolate	Kurtzman, C.P.
Z53	Africa	Other fermented beverages	Kurtzman, C.P.
Z54	Indonesia	Natural isolate	Kurtzman, C.P.
Z55	West Africa	Other fermented beverages	Kurtzman, C.P.
Z56	French Guiana	Unknown biological origin	Kurtzman, C.P.
Z57	Turkey	Wine and vine	Kurtzman, C.P.
Z58	Indonesia	Other fermented beverages	Kurtzman, C.P.
Z59	Philippines	Other fermented beverages	Kurtzman, C.P.
Z60	Ivory Coast	Other fermented beverages	Kurtzman, C.P.
Z61	Brazil	Other fermented beverages	Brandão, R.
Z62	Brazil	Other fermented beverages	Brandão, R.
Z63	Brazil	Other fermented beverages	Brandão, R.
Z64	Turkey	Wine and vine	Huseyin, E.
Z65	Turkey	Wine and vine	Huseyin, E.
Z66	Turkey	Wine and vine	Huseyin, E.
Z67	Turkey	Wine and vine	Huseyin, E.
Z68	Turkey	Wine and vine	Huseyin, E.
Z69	Turkey	Wine and vine	Huseyin, E.
Z70	Turkey	Wine and vine	Huseyin, E.
Z71	Turkey	Wine and vine	Huseyin, E.
Z72	France	Wine and vine	
Z73	France	Wine and vine	
Z74	France	Wine and vine	
Z75	France	Wine and vine	
Z76	France	Wine and vine	
Z77	France	Wine and vine	
Z78	France	Wine and vine	
Z79	France	Wine and vine	
Z80	France	Wine and vine	
Z81	France	Wine and vine	
Z82	France	Wine and vine	
Z83	France	Wine and vine	
Z84	France	Wine and vine	

Strain Code	Geographical Origin	Technological application or origin	Provided by	(Liti *et al.* 2009)
Z85	France – Bordeaux	Commercial wine strain – VL3		
Z86	Unknown geographical origin	Laboratory – S288c		
Z87	Unknown geographical origin	Unknown biological origin		
Z88	Portugal – Vinho Verde	Wine and vine		
Z89	Portugal – Vinho Verde	Wine and vine		
Z90	Portugal – Vinho Verde	Wine and vine		
Z91	Portugal – Vinho Verde	Wine and vine		
Z92	Portugal – Vinho Verde	Wine and vine		
Z93	Portugal – Vinho Verde	Wine and vine		
Z94	Portugal – Vinho Verde	Wine and vine		
Z95	Portugal – Vinho Verde	Wine and vine		
Z96	Portugal – Vinho Verde	Wine and vine		
Z97	Portugal – Vinho Verde	Wine and vine		
Z98	Portugal – Vinho Verde	Wine and vine		
Z99	Portugal – Vinho Verde	Wine and vine		
Z100	Portugal – Vinho Verde	Wine and vine		
Z101	Portugal – Vinho Verde	Wine and vine		
Z102	Portugal – Vinho Verde	Wine and vine		
Z103	Portugal – Vinho Verde	Wine and vine		
Z104	Portugal – Vinho Verde	Wine and vine		
Z105	Portugal – Vinho Verde	Wine and vine		
Z106	Portugal – Bairrada	Wine and vine		
Z107	Portugal – Bairrada	Wine and vine		
Z108	Portugal – Bairrada	Wine and vine		
Z109	Portugal – Bairrada	Wine and vine		
Z110	Portugal – Bairrada	Wine and vine		
Z111	Portugal – Vinho Verde	Wine and vine		
Z112	Portugal – Vinho Verde	Wine and vine		
Z113	Portugal – Vinho Verde	Wine and vine		
Z114	Portugal – Vinho Verde	Wine and vine		
Z115	Portugal – Vinho Verde	Wine and vine		
Z116	Portugal – Vinho Verde	Wine and vine		
Z117	Portugal – Vinho Verde	Wine and vine		
Z118	Portugal – Vinho Verde	Wine and vine		
Z119	Portugal – Vinho Verde	Wine and vine		
Z120	Portugal – Vinho Verde	Wine and vine		
Z121	Portugal – Vinho Verde	Wine and vine		
Z122	Portugal – Vinho Verde	Wine and vine		
Z123	Portugal – Vinho Verde	Wine and vine		
Z124	Portugal – Vinho Verde	Wine and vine		
Z125	Portugal – Vinho Verde	Wine and vine		
Z126	Portugal – Vinho Verde	Wine and vine		
Z128	Portugal – Vinho Verde	Wine and vine		
Z129	Portugal – Vinho Verde	Wine and vine		

Strain Code	Geographical Origin	Technological application or origin	Provided by	(Liti *et al.* 2009)
Z130	Minho	Commercial wine strain – Lalvin QA23		
Z131	Sangiovese (=grape variety)	Commercial wine strain – Lalvin BM 45		
Z132	France – Bordelais	Commercial wine strain – Maurivin AWRI R2		
Z133	France – Vallée du Rhône	Commercial wine strain – Lalvin ICV D80		
Z134	France – Languedoc	Commercial wine strain – K1		
Z135	South Africa – Stellenbosch	Commercial wine strain – Anchor Vin13		
Z136	France – Vallée du Rhône	Commercial wine strain – ICV D47		
Z137	France – Languedoc	Commercial wine strain – ICV D254		
Z138	Spain – Valencia	Commercial wine strain – Enolevure K34		
Z139	France – Champagne	Commercial wine strain – Uvaline BL		
Z140	France – Val de Loire	Commercial wine strain – Uvaline Arôme		
Z141	France – Champagne	Commercial wine strain – Maurivin PDM		
Z142	France – Bordeaux-Gironde	Commercial wine strain – Zymaflore		
Z143	France – Limoux Languedoc	Commercial wine strain – Vitilevure Chardonnay		
Z144	France – Bordeaux-Gironde	Commercial wine strain – Zymaflore		
Z145	France – Bordelais	Commercial wine strain – Zymaflore F10		
Z146	France – Bordeaux-Gironde	Commercial wine strain – Zymaflore F15		
Z147	Portugal – Dão	Commercial wine strain – Zymaflore QD145		
Z148	Portugal – Bairrada	Commercial wine strain – Zymaflore BA11		
Z149	Unknown geographical origin	Commercial wine strain – Siha 3		
Z150	Unknown geographical origin	Commercial wine strain – Siha 6		
Z151	Germany – Pfalz	Commercial wine strain – Siha 7		
Z152	Germany – Baden	Commercial wine strain – Siha 8		
Z153	Unknown geographical origin	Commercial wine strain – Fermol Premier		
Z154	Unknown geographical origin	Commercial wine strain – Fermol Reims Champagne		
Z155	Unknown geographical origin	Commercial wine strain – Uvaferm 228		
Z156	France – Alsace	Commercial wine strain – Uvaferm CS 2		
Z157	France – Champagne	Commercial wine strain – Lalvin EC1118		
Z158	France – Burgund	Commercial wine strain – Lalvin Bourgoblanc Cy3079		
Z159	Unknown geographical origin	Commercial wine strain – ALB		
Z160	France – Vallée du Rhône	Commercial wine strain – Uvaferm L 2056		
Z161	France – Alsace	Commercial wine strain – Fermichamp		
Z162	France – Champagne	Commercial wine strain – Fermicru LS2		
Z163	South Africa – Stellenbosch	Commercial wine strain – Anchor Vin 13		
Z164	France – Narbonne	Commercial wine strain – Uvaferm 71 B		
Z165	France – Bordeaux	Commercial wine strain – Uvaferm BDX		
Z166	France – Bourgogne	Commercial wine strain – Levuline BRG		
Z167	France – Rhone Valley	Commercial wine strain – Lalvin ICV D254		
Z168	France – Rhone Valley	Commercial wine strain – Lalvin ICV D47		
Z169	Unknown geographical origin	Commercial wine strain – Danstil 493 EDV		
Z184	France	Commercial wine strain – VL1		
Z185	Portugal – Bairrada	Wine and vine		

Strain Code	Geographical Origin	Technological application or origin	Provided by	(Liti *et al.* 2009)
Z186	Portugal – Bairrada	Wine and vine		
Z187	Portugal – Douro	Wine and vine		

Supplementary data S2



PCA representation of the three strain clusters, obtained with *k*-means clustering algorithm. The symbols represent the belonging of the 172 strains shown in the phenotypic data PCA (Figure III-2B) to each cluster: ● – cluster 1 (38 strains); ▬ – cluster 2 (90 strains); ■ – cluster 3 (44 strains).

Supplementary data S3

Statistical *p*-values (adjusted) of associations between phenotypic classes and microsatellite alleles. Shaded cells indicate significant associations (false discovery rate below 0.2).

	18 ˚C = 0	18 ˚C = 1	40 ˚C = 0	40 ˚C = 1	40 ˚C = 2	40 ˚C = 3	H_2S production = 1	H_2S production = 2	H_2S production = 3
ScAAT1-16	0.181	0.181	0.098	0.065	0.186	0.172	0.097	0.129	0.113
ScAAT1-22	0.117	0.117	0.043	0.007	0.123	0.039	0.133	0.090	0.157
ScAAT1-24	0.020	0.019	0.030	0.136	0.105	0.055	0.095	0.147	0.168
ScAAT1-27	0.136	0.108	0.034	0.078	0.181	0.050	0.102	0.067	0.114
ScAAT1-31	0.025	0.023	0.045	0.137	0.177	0.113	0.197	0.098	0.136
ScAAT1-32	0.046	0.080	0.092	0.094	0.124	0.029	0.108	0.093	0.119
ScAAT2-13	0.161	0.182	0.015	0.176	0.089	0.003	0.099	0.197	0.170
ScAAT2-14	0.010	0.011	0.069	**0.003**	0.149	0.006	0.072	0.039	0.081
ScAAT2-15	0.107	0.093	0.054	0.167	0.080	0.027	0.060	0.038	0.083
ScAAT2-16	0.160	0.127	0.019	0.120	0.180	0.072	0.197	0.143	0.166
ScAAT3-14	0.072	0.066	0.066	0.176	0.133	0.153	0.147	0.137	0.172
ScAAT3-16	0.145	0.128	0.176	0.185	0.186	0.170	0.174	0.096	0.145
ScAAT3-22	0.068	0.120	0.034	0.128	0.186	0.048	0.197	0.131	0.138
ScAAT4-11	0.102	0.101	0.185	0.095	0.145	0.144	0.173	0.120	0.073
ScAAT4-20	0.103	0.119	0.185	0.031	0.153	0.134	0.091	0.167	0.097
ScAAT5-8	0.076	0.081	0.175	0.147	0.186	0.186	0.126	0.030	0.016
ScAAT5-9	0.025	0.025	0.049	0.069	0.140	0.016	0.009	0.088	0.181
ScAAT5-10	0.096	0.101	0.052	0.083	0.142	0.062	0.040	0.023	0.042
ScAAT5-21	0.025	0.044	0.024	0.012	0.180	0.010	0.078	0.014	0.035
ScAAT5-22	0.104	0.148	0.131	0.048	0.099	0.033	0.155	0.152	0.175
ScAAT6-16	0.033	0.058	0.018	0.013	0.182	0.010	0.059	0.029	0.083
ScAAT6-17	0.135	0.168	0.121	0.042	0.099	0.031	0.156	0.181	0.198
C4-21	0.173	0.178	0.184	0.179	0.054	0.044	0.163	0.198	0.182
C4-22	0.141	0.129	0.185	0.082	0.184	0.183	0.147	0.198	0.171
C4-24	0.134	0.162	0.074	0.142	0.041	**0.003**	0.030	0.057	0.162
C5-4	0.030	0.027	0.003	0.170	0.186	0.040	0.122	0.079	0.100
C5-5	0.115	0.127	0.039	0.070	0.176	0.054	0.063	0.068	0.089
C5-10	0.150	0.142	0.101	0.038	0.088	0.177	0.110	0.165	0.198
C5-12	0.172	0.172	0.185	0.091	0.133	0.047	0.173	0.125	0.136
C5-13	0.120	0.173	0.026	0.159	0.148	0.098	0.130	0.183	0.150
C5-18	0.141	0.142	0.127	0.186	0.036	0.027	0.197	0.119	0.080
C11-13	0.125	0.139	0.100	0.123	0.127	0.065	0.093	0.087	0.102
C11-24	0.115	0.111	0.093	0.108	0.184	0.083	0.060	0.014	0.041
C11-25	0.153	0.166	0.174	0.175	0.010	0.011	0.138	0.095	0.124
ScYOR267c-52	0.182	0.182	0.105	0.128	0.159	0.091	0.048	0.059	0.136
ScYOR267c-63	0.096	0.095	0.185	0.094	0.106	0.118	0.174	0.033	0.043
ScYPL009c-79	0.090	0.086	0.103	0.148	0.186	0.133	0.040	0.092	0.090
ScYPL009c-80	0.056	0.049	0.119	0.180	0.042	0.032	0.174	0.154	0.135
ScYPL009c-81	0.070	0.062	0.160	0.021	0.106	0.050	0.117	0.066	0.115
ScYPL009c-82	0.115	0.111	0.141	0.129	0.032	0.042	0.149	0.118	0.073

	$CuSO_4$ = 0	$CuSO_4$ = 1	Cycloheximide (0.05 μg/mL) = 3	Cycloheximide (0.1 μg/mL) = 2	Cycloheximide (0.1 μg/mL) = 3	Ethanol 10% (v/v) (LM) = 0	Ethanol 10% (v/v) (LM) = 1
ScAAT1-16	0.058	0.096	0.173	0.194	0.194	**0.001**	0.179
ScAAT1-22	0.011	0.013	0.136	0.004	0.007	0.077	0.147
ScAAT1-24	0.188	0.183	0.143	0.066	0.007	0.110	0.160
ScAAT1-27	0.152	0.152	0.194	0.149	0.102	0.162	0.053
ScAAT1-31	0.176	0.106	0.152	0.082	0.072	0.085	0.190
ScAAT1-32	0.076	0.075	0.057	0.162	0.135	0.189	0.070
ScAAT2-13	0.188	0.071	0.149	0.077	0.076	0.044	0.036
ScAAT2-14	0.009	0.015	0.103	0.033	0.012	0.129	0.107
ScAAT2-15	0.004	0.008	0.056	0.120	0.141	0.047	0.115
ScAAT2-16	0.110	0.108	0.020	0.022	0.013	0.039	0.064
ScAAT3-14	0.169	0.188	0.050	0.063	0.074	0.070	0.190
ScAAT3-16	0.188	0.184	0.088	0.082	0.069	0.122	0.129
ScAAT3-22	0.055	0.044	0.161	0.081	0.058	0.027	0.073
ScAAT4-11	0.090	0.103	0.112	0.012	0.037	0.189	0.150
ScAAT4-20	0.155	0.173	0.155	0.053	0.078	0.148	0.106
ScAAT5-8	0.146	0.143	0.090	0.009	0.008	0.095	0.077
ScAAT5-9	0.094	0.117	0.194	0.103	0.082	0.156	0.062
ScAAT5-10	0.175	0.179	0.178	0.064	0.043	0.127	0.038
ScAAT5-21	0.188	0.177	0.078	0.020	0.030	0.084	0.041
ScAAT5-22	0.044	0.067	0.114	0.067	0.050	0.190	0.082
ScAAT6-16	0.185	0.188	0.097	0.029	0.048	0.080	0.073
ScAAT6-17	0.064	0.093	0.128	0.065	0.047	0.190	0.108
C4-21	0.116	0.113	0.146	0.065	0.096	0.099	0.165
C4-22	0.017	0.021	0.059	0.069	0.044	0.111	0.145
C4-24	0.166	0.114	0.175	0.172	0.194	0.181	0.112
C5-4	0.122	0.125	0.059	0.165	0.118	0.098	0.190
C5-5	0.126	0.090	0.174	0.118	0.106	0.040	0.163
C5-10	0.087	0.088	0.144	0.107	0.164	0.190	0.139
C5-12	0.140	0.113	0.017	0.040	0.056	0.190	0.057
C5-13	0.164	0.115	0.173	0.038	0.026	0.169	0.190
C5-18	0.031	0.030	0.003	0.034	0.018	0.089	0.190
C11-13	0.043	0.057	0.155	0.005	0.004	0.168	0.029
C11-24	0.004	0.007	0.065	0.046	0.029	0.189	0.160
C11-25	0.153	0.177	0.139	0.126	0.128	0.023	0.190
ScYOR267c-52	0.064	0.100	0.164	0.073	0.052	0.170	0.067
ScYOR267c-63	0.055	0.088	0.114	0.026	0.047	0.090	0.191
ScYPL009c-79	0.182	0.153	0.194	0.105	0.176	0.040	0.014
ScYPL009c-80	0.093	0.070	0.174	0.181	0.194	0.028	0.158
ScYPL009c-81	0.088	0.071	0.194	0.108	0.156	0.138	0.190
ScYPL009c-82	0.040	0.073	0.161	0.008	0.015	0.177	0.047

	Ethanol 10% (v/v) (LM) = 2	Ethanol 10% (v/v) (LM) = 3	Ethanol 14% (v/v) (LM) = 0	Ethanol 14% (v/v) (LM) = 1	Ethanol 14% (v/v) (LM) = 2	Ethanol 6% (v/v) (LM) = 2
ScAAT1-16	0.067	0.040	0.023	0.139	0.033	0.037
ScAAT1-22	0.125	0.152	0.123	0.108	0.039	0.156
ScAAT1-24	0.166	0.061	0.160	0.081	0.161	0.157
ScAAT1-27	0.072	0.062	0.050	0.153	0.122	0.027
ScAAT1-31	0.178	0.172	0.154	0.135	0.054	0.028
ScAAT1-32	0.107	0.006	0.078	0.109	0.015	0.050
ScAAT2-13	0.124	0.007	0.025	0.191	0.032	0.016
ScAAT2-14	0.140	0.027	0.007	0.146	0.009	0.160
ScAAT2-15	0.011	0.111	0.117	0.084	0.159	0.132
ScAAT2-16	0.035	0.119	0.080	0.125	0.102	0.010
ScAAT3-14	0.056	0.034	0.019	0.191	0.013	0.182
ScAAT3-16	0.010	0.061	0.147	0.100	0.166	0.132
ScAAT3-22	0.021	0.085	0.157	0.191	0.161	0.008
ScAAT4-11	0.129	0.191	0.191	0.067	0.038	0.159
ScAAT4-20	0.146	0.027	0.029	0.102	0.087	0.021
ScAAT5-8	0.036	0.026	0.011	0.120	0.024	0.115
ScAAT5-9	0.051	0.004	0.106	0.156	0.137	0.010
ScAAT5-10	0.066	0.016	0.158	0.136	0.103	0.045
ScAAT5-21	0.051	0.005	0.068	0.129	0.122	0.032
ScAAT5-22	0.107	0.016	0.061	0.158	0.128	0.086
ScAAT6-16	0.029	0.005	0.092	0.162	0.103	0.022
ScAAT6-17	0.076	0.013	0.063	0.144	0.117	0.062
C4-21	0.112	0.003	0.084	0.148	0.048	0.111
C4-22	0.073	0.028	0.044	0.166	0.033	0.084
C4-24	0.044	0.012	0.110	0.035	0.125	0.154
C5-4	0.133	0.135	0.078	0.037	0.191	0.019
C5-5	0.039	0.035	0.022	0.169	0.020	0.121
C5-10	0.051	0.084	0.050	0.152	0.064	0.130
C5-12	0.142	0.057	0.183	0.113	0.119	0.159
C5-13	0.167	0.162	0.057	0.022	0.131	0.020
C5-18	0.109	0.191	0.017	0.041	0.074	0.163
C11-13	0.123	0.026	0.088	0.191	0.096	0.126
C11-24	0.157	0.118	0.074	0.171	0.030	0.063
C11-25	0.003	0.023	0.094	0.056	0.083	0.070
ScYOR267c-52	0.083	0.131	0.145	0.104	0.056	0.144
ScYOR267c-63	0.041	0.014	0.094	0.178	0.044	0.151
ScYPL009c-79	**0.0005**	0.019	0.079	0.097	0.024	0.180
ScYPL009c-80	0.072	0.098	0.070	0.049	0.141	0.162
ScYPL009c-81	0.060	0.005	0.069	0.149	0.078	0.189
ScYPL009c-82	0.011	**0.0008**	0.006	0.157	0.019	0.022

	Ethanol 6% (v/v) (LM) = 3	Ethanol 12% (v/v) + $Na_2S_2O_5$ (50 mg/L)	Ethanol 14% (v/v) + $Na_2S_2O_5$ (50 mg/L)	Ethanol 16% (v/v) + $Na_2S_2O_5$ (50 mg/L)	Galactosidase activity = 1
ScAAT1-16	0.033	0.198	0.198	0.175	0.141
ScAAT1-22	0.154	0.150	0.148	0.138	0.066
ScAAT1-24	0.149	0.022	0.085	0.143	0.091
ScAAT1-27	0.021	0.071	0.075	0.198	**0.003**
ScAAT1-31	0.025	0.072	0.056	0.153	0.124
ScAAT1-32	0.046	0.134	0.102	0.198	**0.001**
ScAAT2-13	0.013	0.076	0.098	0.198	**0.002**
ScAAT2-14	0.176	0.018	0.026	0.160	0.167
ScAAT2-15	0.136	0.091	0.093	0.071	0.133
ScAAT2-16	0.040	0.017	0.067	0.199	**0.001**
ScAAT3-14	0.176	0.086	0.103	0.199	0.045
ScAAT3-16	0.125	0.075	0.182	0.199	0.062
ScAAT3-22	0.036	0.144	0.091	0.199	0.004
ScAAT4-11	0.189	0.110	0.198	0.113	0.084
ScAAT4-20	0.011	0.058	0.104	0.053	0.025
ScAAT5-8	0.116	0.168	0.198	0.146	0.015
ScAAT5-9	0.012	0.096	0.071	0.152	0.035
ScAAT5-10	0.048	0.150	0.062	0.149	0.020
ScAAT5-21	0.032	0.067	0.074	0.152	0.025
ScAAT5-22	0.074	0.154	0.181	0.146	0.006
ScAAT6-16	0.024	0.099	0.108	0.130	0.026
ScAAT6-17	0.054	0.143	0.171	0.145	0.007
C4-21	0.087	0.090	0.085	0.111	**0.0002**
C4-22	0.081	0.016	0.145	0.199	0.092
C4-24	0.120	0.054	0.067	0.110	0.028
C5-4	0.012	0.016	0.022	0.068	0.166
C5-5	0.098	0.107	0.122	0.134	**0.0009**
C5-10	0.130	0.164	0.198	0.144	0.051
C5-12	0.159	0.035	0.033	0.088	0.0148
C5-13	0.009	0.077	0.121	0.050	**0.002**
C5-18	0.162	0.156	0.154	0.140	0.104
C11-13	0.144	0.165	0.051	0.013	0.004
C11-24	0.069	0.165	0.105	0.142	0.072
C11-25	0.054	0.052	0.047	0.141	0.085
ScYOR267c-52	0.155	0.138	0.037	0.060	0.149
ScYOR267c-63	0.177	0.083	0.130	0.114	0.135
ScYPL009c-79	0.151	0.023	0.021	0.112	0.146
ScYPL009c-80	0.163	0.100	0.183	0.199	**0.002**
ScYPL009c-81	0.189	0.055	0.066	0.155	0.046
ScYPL009c-82	0.021	0.117	0.157	0.082	0.200

	Galactosidase activity = 2	Galactosidase activity = 3	Iprodion (0.05 mg/mL) = 2	Iprodion (0.05 mg/mL) = 3	Iprodion (0.1 mg/mL) = 2	Iprodion (0.1 mg/mL) = 3	KCl (0.75 M) = 2	KCl (0.75 M) = 3	$KHSO_3$ (150 mg/L) = 0
ScAAT1-16	0.081	0.018	0.019	0.018	0.066	0.121	0.150	0.177	0.107
ScAAT1-22	0.124	0.200	0.014	0.042	0.192	0.192	0.187	0.187	0.108
ScAAT1-24	0.029	0.079	0.167	0.104	0.192	0.192	0.161	0.179	0.011
ScAAT1-27	0.031	0.050	0.180	0.180	0.192	0.192	0.179	0.151	0.027
ScAAT1-31	0.178	0.098	0.170	0.191	0.192	0.192	0.083	0.071	0.105
ScAAT1-32	0.053	0.017	0.116	0.116	0.192	0.193	0.116	0.150	0.018
ScAAT2-13	0.042	0.151	0.059	0.059	0.177	0.164	0.187	0.187	0.009
ScAAT2-14	0.167	0.153	0.037	0.053	0.039	0.034	0.075	0.086	0.115
ScAAT2-15	0.150	0.091	0.184	0.191	0.053	0.032	0.170	0.162	0.078
ScAAT2-16	0.019	0.059	0.127	0.131	0.192	0.158	0.143	0.130	0.008
ScAAT3-14	0.124	0.114	0.171	0.171	0.192	0.193	0.169	0.145	0.095
ScAAT3-16	0.200	0.117	0.074	0.049	0.102	0.092	0.082	0.138	0.121
ScAAT3-22	0.031	0.183	0.134	0.133	0.139	0.111	0.069	0.063	0.075
ScAAT4-11	0.129	0.041	0.093	0.092	0.071	0.064	0.078	0.084	0.106
ScAAT4-20	0.073	0.121	0.119	0.100	0.073	0.043	0.106	0.085	0.109
ScAAT5-8	0.170	0.101	0.048	0.061	0.147	0.109	0.065	0.120	0.157
ScAAT5-9	0.054	0.141	0.117	0.146	0.048	0.037	0.175	0.144	0.055
ScAAT5-10	0.185	0.027	0.117	0.124	0.132	0.154	0.012	0.020	0.090
ScAAT5-21	0.120	0.060	0.152	0.142	0.164	0.183	0.005	0.008	0.010
ScAAT5-22	0.159	0.067	0.138	0.144	0.112	0.078	0.164	0.181	0.095
ScAAT6-16	0.109	0.081	0.126	0.125	0.182	0.193	0.020	0.034	0.007
ScAAT6-17	0.184	0.046	0.130	0.130	0.112	0.077	0.149	0.163	0.079
C4-21	0.008	0.158	0.051	0.032	0.134	0.114	0.107	0.124	0.180
C4-22	0.184	0.087	0.145	0.139	0.089	0.065	0.143	0.145	0.123
C4-24	0.178	0.076	0.039	0.036	0.175	0.122	0.143	0.134	0.097
C5-4	0.128	0.057	0.130	0.137	0.104	0.080	0.052	0.060	0.069
C5-5	0.057	0.095	0.110	0.105	0.161	0.134	0.006	0.007	0.068
C5-10	0.045	0.132	0.075	0.077	0.105	0.094	0.123	0.090	0.080
C5-12	0.009	0.019	0.151	0.151	0.127	0.129	0.175	0.187	0.139
C5-13	0.160	0.035	0.013	0.054	0.160	0.148	0.134	0.135	0.086
C5-18	0.139	0.052	0.127	0.191	0.137	0.139	0.014	0.014	0.108
C11-13	0.182	0.030	0.045	0.048	0.096	0.079	0.130	0.127	0.045
C11-24	0.028	0.091	0.082	0.082	0.138	0.126	0.014	0.015	0.163
C11-25	0.167	0.121	0.011	0.012	0.087	0.085	0.187	0.187	0.160
ScYOR267c-52	0.185	0.178	0.070	0.131	0.192	0.181	0.047	0.039	0.080
ScYOR267c-63	0.200	0.149	0.113	0.114	0.171	0.169	0.116	0.085	0.131
ScYPL009c-79	0.097	0.150	0.095	0.103	0.121	0.097	0.051	0.058	0.167
ScYPL009c-80	0.045	0.061	0.036	0.031	0.164	0.168	0.103	0.103	0.087
ScYPL009c-81	0.061	0.055	0.068	0.057	0.164	0.180	0.025	0.037	0.153
ScYPL009c-82	0.042	0.054	0.023	0.019	0.093	0.101	0.047	0.035	0.119

	$KHSO_3$ (150 mg/L) = 1	$KHSO_3$ (150 mg/L) = 2	$KHSO_3$ (150 mg/L) = 3	$KHSO_3$ (300 mg/L) = 0	$KHSO_3$ (300 mg/L) = 1	$KHSO_3$ (300 mg/L) = 2	$KHSO_3$ (300 mg/L) = 3	NaCl (1.5 M) = 0	NaCl (1.5 M) = 1
ScAAT1-16	0.172	0.194	0.083	0.012	0.195	0.082	0.159	0.071	0.111
ScAAT1-22	0.132	0.094	0.158	0.085	0.136	0.063	**0.002**	0.065	0.093
ScAAT1-24	0.064	0.194	0.014	0.006	0.151	0.151	**0.001**	0.059	0.110
ScAAT1-27	0.103	0.079	0.012	**0.001**	0.150	0.165	0.006	0.089	0.051
ScAAT1-31	0.066	0.053	0.043	0.100	0.083	0.124	0.015	0.156	0.187
ScAAT1-32	0.041	0.140	0.092	0.004	0.036	0.165	**0.002**	0.123	0.086
ScAAT2-13	0.099	0.195	0.012	**0.002**	0.124	0.138	**0.0006**	0.165	0.148
ScAAT2-14	0.084	0.118	0.085	0.075	0.115	0.007	0.003	0.118	0.165
ScAAT2-15	0.076	0.063	0.007	0.032	0.128	0.010	**0.003**	0.029	0.151
ScAAT2-16	0.043	0.154	0.016	0.067	0.026	0.141	0.033	0.075	0.052
ScAAT3-14	0.058	0.195	0.089	0.068	0.088	0.129	0.158	0.099	0.045
ScAAT3-16	0.116	0.139	0.135	0.089	0.070	0.089	0.105	0.179	0.185
ScAAT3-22	0.097	0.167	0.106	0.073	0.021	0.133	0.004	0.048	0.094
ScAAT4-11	0.172	0.115	0.195	0.140	0.083	0.028	0.043	0.122	0.079
ScAAT4-20	0.035	0.131	0.018	0.029	0.058	0.016	0.051	0.003	0.020
ScAAT5-8	0.039	0.126	0.032	0.034	0.195	0.099	0.010	0.111	0.181
ScAAT5-9	0.028	0.161	0.019	0.021	0.064	0.054	0.057	0.126	0.081
ScAAT5-10	0.059	0.156	0.018	0.038	0.068	0.080	0.065	0.080	0.076
ScAAT5-21	0.017	0.065	0.004	0.003	0.043	0.040	0.011	0.057	0.068
ScAAT5-22	0.055	0.195	0.047	**0.0003**	0.123	0.031	**0.0003**	0.021	0.055
ScAAT6-16	0.009	0.071	0.006	0.004	0.062	0.041	0.011	0.046	0.044
ScAAT6-17	0.046	0.195	0.063	**0.002**	0.113	0.031	**0.0004**	0.024	0.055
C4-21	0.013	0.093	0.034	**0.002**	0.126	0.155	**0.0008**	0.037	0.068
C4-22	0.046	**0.001**	0.030	0.019	0.177	0.164	0.026	0.018	0.082
C4-24	0.046	0.043	0.005	0.009	0.074	0.195	0.004	0.173	0.185
C5-4	0.078	0.136	0.049	0.055	0.109	0.163	0.036	0.020	0.030
C5-5	0.022	0.043	0.008	0.005	0.147	0.069	0.0083	0.123	0.128
C5-10	0.194	0.102	0.034	0.163	0.195	0.077	0.196	0.017	0.032
C5-12	0.079	0.072	0.055	0.017	0.195	0.179	0.028	0.155	0.183
C5-13	0.130	0.053	0.012	0.006	0.111	0.116	0.013	0.183	0.176
C5-18	0.115	0.064	0.023	0.096	0.086	0.155	0.151	0.066	0.053
C11-13	0.052	0.095	0.062	0.195	0.114	0.069	0.115	0.187	0.168
C11-24	0.060	0.179	0.149	0.040	0.151	0.109	0.118	0.124	0.172
C11-25	0.137	0.084	0.122	0.144	0.160	0.137	0.113	0.048	0.116
ScYOR267c-52	0.014	0.177	0.052	0.048	0.132	0.116	0.017	0.073	0.038
ScYOR267c-63	0.112	0.141	0.163	0.181	0.140	0.099	0.081	0.072	0.049
ScYPL009c-79	0.131	0.108	0.029	0.033	0.156	0.075	0.023	0.026	0.027
ScYPL009c-80	0.056	0.167	0.085	0.038	0.021	0.164	0.024	0.179	0.131
ScYPL009c-81	0.045	0.075	0.020	0.023	0.171	0.183	0.009	0.097	0.081
ScYPL009c-82	0.122	0.039	0.058	0.195	0.151	0.118	0.099	0.017	0.022

	NaCl (1.5 M) = 2	pH 2 = 0	pH 2 = 1	Procymidon (0.1 mg/mL) = 2	Procymidon (0.1 mg/mL) = 3	SDS (0.01% w/v) = 0	SDS (0.01% w/v) = 1	Ethanol 12% (v/v) + $Na_2S_2O_5$ (75 mg/L) = 0	Ethanol 12% (v/v) + $Na_2S_2O_5$ (75 mg/L) = 1
ScAAT1-16	0.187	0.102	0.123	0.042	0.046	0.107	0.116	0.162	0.172
ScAAT1-22	0.138	0.158	0.158	0.137	0.135	0.108	0.109	0.100	0.132
ScAAT1-24	0.113	0.042	0.071	0.170	0.193	0.162	0.171	0.074	0.123
ScAAT1-27	0.119	0.060	0.068	0.119	0.105	0.107	0.114	0.092	0.044
ScAAT1-31	0.069	0.091	0.098	0.193	0.193	0.122	0.189	0.138	0.152
ScAAT1-32	0.143	0.024	0.028	0.143	0.121	0.036	0.035	0.199	0.199
ScAAT2-13	0.112	0.089	0.101	0.112	0.101	0.165	0.165	0.177	0.199
ScAAT2-14	0.126	0.051	0.069	0.193	0.193	0.065	0.068	0.109	0.135
ScAAT2-15	0.005	0.020	0.024	0.173	0.174	0.184	0.189	0.053	0.045
ScAAT2-16	0.159	0.047	0.065	0.193	0.193	0.166	0.174	0.134	0.093
ScAAT3-14	0.074	0.055	0.092	0.169	0.169	0.049	0.049	0.199	0.089
ScAAT3-16	0.052	0.168	0.166	0.155	0.143	0.159	0.123	0.109	0.090
ScAAT3-22	0.104	0.186	0.171	0.080	0.076	0.171	0.172	0.199	0.200
ScAAT4-11	0.113	0.100	0.099	0.176	0.174	0.178	0.178	0.071	0.098
ScAAT4-20	0.031	0.056	0.079	0.042	0.025	0.009	0.008	0.140	0.120
ScAAT5-8	0.074	0.081	0.106	0.168	0.168	0.168	0.120	0.091	0.039
ScAAT5-9	0.060	0.023	0.037	0.094	0.139	0.007	0.006	0.085	0.131
ScAAT5-10	0.148	0.056	0.058	0.056	0.045	0.022	0.029	0.131	0.105
ScAAT5-21	0.187	0.080	0.098	0.141	0.117	0.031	0.043	0.061	0.077
ScAAT5-22	0.101	0.037	0.050	0.146	0.137	0.188	0.189	0.088	0.110
ScAAT6-16	0.17	0.083	0.114	0.152	0.117	0.011	0.018	0.124	0.141
ScAAT6-17	0.100	0.041	0.061	0.145	0.110	0.183	0.189	0.076	0.107
C4-21	0.094	0.122	0.133	0.175	0.146	0.142	0.157	0.070	0.076
C4-22	0.030	0.049	0.033	0.041	0.070	0.086	0.140	0.155	0.157
C4-24	0.169	0.182	0.185	0.110	0.107	0.106	0.117	0.175	0.200
C5-4	0.187	0.186	0.184	0.064	0.059	0.188	0.189	0.005	**0.002**
C5-5	0.102	0.028	0.132	0.084	0.082	**0.002**	0.016	0.127	0.125
C5-10	0.188	0.017	0.012	0.144	0.144	0.142	0.163	0.106	0.111
C5-12	0.176	0.058	0.075	0.175	0.174	0.158	0.157	0.061	0.078
C5-13	0.112	0.186	0.160	0.140	0.194	0.188	0.189	0.169	0.105
C5-18	0.024	0.063	0.063	0.193	0.056	0.046	0.046	0.109	0.116
C11-13	0.135	0.045	0.049	0.041	0.034	0.047	0.051	0.101	0.042
C11-24	0.104	0.033	0.038	0.170	0.161	0.014	0.015	0.199	0.200
C11-25	0.041	0.180	0.186	0.015	0.013	0.188	0.161	0.087	0.137
ScYOR267c-52	0.156	0.058	0.038	0.031	0.008	0.064	0.073	0.180	0.140
ScYOR267c-63	0.178	0.032	0.027	0.178	0.177	0.086	0.086	0.096	0.112
ScYPL009c-79	0.173	0.156	0.154	0.112	0.084	0.089	0.024	0.061	0.081
ScYPL009c-80	0.073	0.079	0.100	0.193	0.174	0.188	0.189	0.129	0.174
ScYPL009c-81	0.128	0.125	0.147	0.128	0.123	0.165	0.170	0.127	0.134
ScYPL009c-82	0.142	0.087	0.077	0.095	0.094	0.025	0.022	0.022	0.076

	Wine supplemented with glucose (0.5% w/v) = 0	Wine supplemented with glucose (0.5% w/v) = 1	Wine supplemented with glucose (0.5% w/v) = 2	Wine supplemented with glucose (1% w/v) = 2	Wine supplemented with glucose (1% w/v) = 1	Wine supplemented with glucose (1% w/v) = 2
ScAAT1-16	0.040	0.161	0.025	0.006	0.014	0.121
ScAAT1-22	0.158	0.153	0.049	0.196	0.111	0.092
ScAAT1-24	0.170	0.183	0.157	0.169	0.092	0.098
ScAAT1-27	0.087	0.120	0.179	0.064	0.101	0.197
ScAAT1-31	0.179	0.148	0.171	0.136	0.109	0.016
ScAAT1-32	0.037	0.094	0.196	0.022	0.006	0.079
ScAAT2-13	0.054	0.105	0.091	0.110	0.118	0.016
ScAAT2-14	0.005	0.078	0.005	**0.001**	0.010	0.018
ScAAT2-15	0.100	0.028	0.089	0.171	0.197	0.164
ScAAT2-16	0.035	0.026	0.032	0.052	0.137	0.086
ScAAT3-14	0.058	0.038	0.130	0.097	0.128	0.159
ScAAT3-16	0.166	0.149	0.196	0.154	0.159	0.079
ScAAT3-22	0.196	0.146	0.156	0.121	0.059	0.052
ScAAT4-11	0.150	0.196	0.104	0.050	0.063	0.148
ScAAT4-20	0.026	0.036	0.148	0.013	0.023	0.146
ScAAT5-8	0.053	0.044	0.051	0.154	0.153	0.101
ScAAT5-9	0.095	0.184	0.057	0.166	0.067	0.004
ScAAT5-10	0.070	0.126	0.042	0.168	0.096	0.099
ScAAT5-21	**0.002**	0.008	0.097	0.017	0.050	0.060
ScAAT5-22	0.170	0.172	0.087	0.127	0.184	0.077
ScAAT6-16	**0.002**	0.015	0.062	0.028	0.038	0.147
ScAAT6-17	0.196	0.163	0.104	0.157	0.197	0.077
C4-21	0.114	0.063	0.142	0.156	0.176	0.132
C4-22	0.126	0.140	0.010	0.132	0.182	0.133
C4-24	0.196	0.154	0.129	0.059	0.104	0.084
C5-4	0.173	0.163	0.196	0.132	0.167	0.052
C5-5	0.072	0.075	0.178	0.143	0.047	0.061
C5-10	0.142	0.091	0.133	0.024	0.030	0.150
C5-12	0.196	0.147	0.162	0.197	0.161	0.086
C5-13	0.066	0.074	0.135	0.060	0.034	0.089
C5-18	0.088	0.037	0.155	0.067	0.049	0.197
C11-13	0.144	0.167	0.034	0.118	0.184	0.071
C11-24	0.051	0.121	0.053	0.090	0.100	0.197
C11-25	0.180	0.196	0.196	0.134	0.197	0.125
ScYOR267c-52	0.137	0.062	0.096	0.035	0.061	0.102
ScYOR267c-63	0.031	0.180	0.036	0.018	0.062	0.168
ScYPL009c-79	0.092	0.097	0.133	0.138	0.134	0.169
ScYPL009c-80	0.132	0.169	0.152	0.119	0.060	0.113
ScYPL009c-81	0.088	0.119	0.166	0.139	0.147	0.104
ScYPL009c-82	0.036	0.153	0.021	0.027	0.173	0.009

Supplementary data S4

Concentration (mg/L) of aromatic compounds determined by GC-MS for the four *Saccharomyces cerevisiae* strains at two time-points

Strain	**Z63**		**S288c**		**Z23**		**VL1**	
Time points	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
Ethyl acetate	n.d.	13.54 ± 2.79	n.d.	7.044 ± 4.738	n.d.	17.42 ± 2.96	n.d.	10.26 ± 3.43
Ethyl propionate	0.082 ± 0.011	0.255 ± 0.096	n.d.	0.182 ± 0.075	0.111 ± 0.015	0.267 ± 0.045	0.109 ± 0.024	0.147 ± 0.053
Propyl acetate	0.013 ± 0.001	0.130 ± 0.034	0.012 ± 0.001	0.076 ± 0.039	0.019 ± 0.004	0.100 ± 0.017	0.014 ± 0.003	0.071 ± 0.011
2-methylpropyl acetate	0.016 ± 0.0002	0.194 ± 0.060	n.d.	0.032 ± 0.018	0.024 ± 0.004	0.150 ± 0.030	0.019 ± 0.003	0.147 ± 0.023
Ethyl butanoate	0.007 ± 0.001	0.202 ± 0.059	0.003 ± 0.004	0.134 ± 0.059	n.d.	0.144 ± 0.015	0.002 ± 0.004	0.087 ± 0.013
Propanol	4.21 ± 0.021	14.22 ± 2.17	7.89 ± 1.25	17.119 ± 1.845	6.125 ± 0.491	7.05 ± 0.45	6.225 ± 0.438	7.57 ± 0.24
Ethyl 2-methylbutanoate	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Ethyl 3-methylbutanoate	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
2-methylpropanol	5.00 ± 0.020	1.873 ± 0.438	2.79 ± 0.58	10.214 ± 0.478	5.817 ± 0.233	17.17 ± 1.87	6.388 ± 0.273	28.22 ± 1.38
2-methylbutylacetate	0.001 ± 0.0002	28.17 ± 5.03	n.d.	0.005 ± 0.001	0.002 ± 0.0002	0.010 ± 0.002	0.0012 ± 0.0004	0.0078 ± 0.0015
3-methylbutylacetate	0.064 ± 0.006	0.011 ± 0.003	0.04 ± 0.01	0.756 ± 0.272	0.195 ± 0.019	2.047 ± 0.337	0.108 ± 0.034	1.473 ± 0.270
Ethyl valerate	n.d.	1.873 ± 0.438	n.d.	0.024 ± 0.011	n.d.	0.074 ± 0.012	n.d.	0.053 ± 0.010
Isoamyl alcohol	24.19 ± 1.37	141.2 ± 20.3	31.13 ± 5.98	92.872 ± 2.905	46.28 ± 2.44	132.49 ± 10.43	38.766 ± 0.504	125.59 ± 14.48
Ethyl hexanoate	0.063 ± 0.019	0.756 ± 0.160	0.040 ± 0.001	0.439 ± 0.159	0.083 ± 0.005	0.775 ± 0.049	0.073 ± 0.011	0.474 ± 0.112
2-methylbutanol	0.035 ± 0.005	0.029 ± 0.003	0.049 ± 0.009	0.024 ± 0.001	0.043 ± 0.007	0.020 ± 0.002	0.044 ± 0.004	0.0195 ± 0.0021
Hexyl acetate	0.113 ± 0.006	0.288 ± 0.034	0.025 ± 0.035	0.111 ± 0.040	0.273 ± 0.037	0.258 ± 0.019	0.171 ± 0.046	0.167 ± 0.021
Ethyl lactate	n.d.	0.690 ± 0.177	n.d.	0.308 ± 0.032	n.d.	0.549 ± 0.051	n.d.	0.513 ± 0.103
Ethyl octanoate	0.169 ± 0.078	1.508 ± 0.323	0.069 ± 0.007	0.728 ± 0.139	0.096 ± 0.019	1.439 ± 0.046	0.069 ± 0.028	1.238 ± 0.261
Propanoic acid	0.084 ± 0.015	0.211 ± 0.043	0.198 ± 0.047	0.231 ± 0.034	0.160 ± 0.014	0.167 ± 0.015	0.153 ± 0.015	0.144 ± 0.014
Isobutyric acid	0.20 ± 0.007	0.295 ± 0.035	0.116 ± 0.003	0.124 ± 0.028	0.200 ± 0.037	0.344 ± 0.027	0.155 ± 0.016	0.288 ± 0.043
Butyric acid	0.184 ± 0.016	0.598 ± 0.045	0.252 ± 0.004	0.476 ± 0.102	0.176 ± 0.023	0.565 ± 0.031	0.180 ± 0.013	0.584 ± 0.058
Ethyl decanoate	0.04 ± 0.01	0.644 ± 0.127	0.033 ± 0.003	0.373 ± 0.023	0.033 ± 0.004	0.810 ± 0.099	0.019 ± 0.003	0.618 ± 0.104
3-methylbutanoic acid	0.09 ± 0.0001	0.219 ± 0.023	0.095 ± 0.002	0.188 ± 0.038	0.139 ± 0.028	0.350 ± 0.032	0.101 ± 0.009	0.255 ± 0.056
2-methylbutanoic acid	0.06 ± 0.002	0.134 ± 0.015	0.050 ± 0.002	0.115 ± 0.025	0.078 ± 0.018	0.179 ± 0.022	0.057 ± 0.005	0.122 ± 0.030
Methionol	0.07 ± 0.002	0.295 ± 0.060	0.032 ± 0.005	0.035 ± 0.003	0.070 ± 0.011	0.157 ± 0.020	0.068 ± 0.011	0.226 ± 0.053
Valeric acid	0.002 ± 0.003	0.007 ± 0.002	0.002 ± 0.002	0.003 ± 0.002	0.002 ± 0.0004	0.004 ± 0.002	0.003 ± 0.004	0.005 ± 0.001
2-phenylethyl acetate	0.02 ± 0.023	0.282 ± 0.032	0.009 ± 0.001	0.063 ± 0.010	0.088 ± 0.020	0.320 ± 0.017	0.043 ± 0.014	0.347 ± 0.037
Hexanoic acid	0.74 ± 0.110	4.133 ± 0.304	0.976 ± 0.003	2.500 ± 0.526	0.650 ± 0.148	3.99 ± 0.52	0.570 ± 0.032	3.55 ± 0.60
ethyl dodecanoate	0.01 ± 0.001	0.097 ± 0.040	0.012 ± 0.003	0.048 ± 0.012	0.014 ± 0.002	0.138 ± 0.029	0.003 ± 0.0002	0.209 ± 0.051
2-phenylethanol	1.59 ± 0.029	5.535 ± 0.866	0.954 ± 0.057	2.312 ± 0.307	1.988 ± 0.220	7.95 ± 0.74	1.524 ± 0.084	6.71 ± 0.66
Octanoic acid	0.54 ± 0.031	3.671 ± 0.422	0.941 ± 0.021	2.333 ± 0.452	0.254 ± 0.053	3.73 ± 0.55	0.100 ± 0.028	2.58 ± 0.46

Strain	Z63		S288c		Z23		VL1	
Time points	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
Dodecanoic acid	n.d.	0.031 ± 0.021	0.025 ± 0.012	0.028 ± 0.011	0.009 ± 0.002	0.029 ± 0.008	0.001 ± 0.002	0.030 ± 0.007
Glycerol	1718.37 ± 401.81	5319.69 66.95	2067.32 ± 54.02	3889.33 ± 88.99	2097.71 ± 28.53	4979.84 ± 206.59	1772.08 ± 13.29	4534.07 ± 80.17
Succinate	362.37 ± 88.14	665.36 ± 16.34	366.96 ± 6.49	383.79 ± 7.94	357.82 ± 3.88	558.63 ± 36.22	327.17 ± 3.62	578.86 ± 8.58
Acetate	83.60 ± 18.43	88.38 ± 2.08	217.46 ± 5.13	284.42 ± 0.99	134.80 ± 5.93	197.28 ± 15.84	91.51 ± 3.31	71.99 ± 10.31
Pyruvate	69.64 ± 16.74	65.72 ± 0.19	87.51 ± 1.61	n.d.	73.25 ± 2.50	53.42 ± 3.19	68.93 ± 1.01	n.d.
Alpha-ketoglutarate	117.66 ± 41.75	125.74 ± 1.18	172.23 ± 2.90	121.34 ± 3.43	143.99 ± 1.93	102.58 ± 5.79	132.37 ± 1.23	112.75 ± 3.78

Supplementary data S5

Group	Genes	Function	Metabolic compounds obtained in MFA analysis
1	*RPL14B* *RPS24A* *RPS25B* *RPL30* *RPS26B* *MRPL23* *RPS17B* *RPL40B* *RPL26A*	Ribosome	Isobutanol Methionol Isobutylacetate Phenylethanol
2	*BDH2* *AAD10* *GAD1*	Butanoate metabolism	Ethylbutanoate Ethyldecanoate Ethyloctanoate
	ADH2 *ALD3* *SFA1*	Tyrosine metabolism	
	ALD4 *ADH2* *ALD3* *SFA1*	Glycolysis / Gluconeogenesis	
	ALD4 *GAD1* *ALD3*	beta-Alanine metabolism	
	ALD4 *ADH2* *SFA1*	Fatty acid metabolism	
3	*STE5* *STE4* *SLT2* *STE2* *MFA1* *GSC2* *FUS1*	MAPK signaling pathway - yeast	Octanoic acid Decanoic acid Hexanoic acid Butyric acid Ethyldodecanoate Dodecanoic acid Pyruvate Acetate Alphacetoglutarate Glycerol Propanol
	TUM1 *SAM1* *MET2* *CYS3*	Cysteine and methionine metabolism	
	ATM1 *PDR5*	ABC transporters	
4	There are not annotations significantly enriched		Hexylacetate propylacetate 2-phenylethylacetate amylalcohol isovaleric acid isoamylacetate amylacetate ethlpropionate propanoic acid isoamylalcohol

Supplementary data S6

Group	Genes	Function/discription	Metabolic compounds obtained in MFA
1	*RAS1*	GTPase involved in G-protein signaling in adenylate cyclase activation	Acetate Decanoic acid Propanoic acid
	PMC1	Vacuolar Ca2+ ATPase involved in depleting cytosol of Ca2+ ions	
	ENA5	Protein with similarity to P-type ATPase sodium pumps	
	GPI18	Functional ortholog of human PIG-V	
	YPS3	Aspartic protease	
	SPS100	Protein required for spore wall maturation	
	YDR543C	Dubious open reading frame	
	CRH1	Chitin transglycosylase	
	YDR034W-B	Predicted tail-anchored plasma membrane protein	
	COS1	Protein of unknown function	
	HXT5	Hexose transporter with moderate affinity for glucose;	
	CAT8	Zinc cluster transcriptional activator	
	VPS36	Component of the ESCRT-II complex	
	TIS11	mRNA-binding protein expressed during iron starvation	
	SPI1	GPI-anchored cell wall protein involved in weak acid resistance	
	PHM7	Protein of unknown function	
	SCS3	Protein required for inositol prototrophy	
	ENA2	P-type ATPase sodium pump	
	YAR028W	Putative integral membrane protein	
	MEP1	Ammonium permease	
	COS6	Protein of unknown function	
	YKE4	Zinc transporter	
	ARN2	Transporter	
	YLR031W	Putative protein of unknown function	
	YNL114C	Dubious open reading frame	
	PDE1	Low-affinity cyclic AMP phosphodiesterase	
	PCS60	Oxalyl-CoA synthetase	
	SUI1	Translation initiation factor eIF1	
	ZAP1	Zinc-regulated transcription factor	
	AZR1	Plasma membrane transporter of the major facilitator superfamily; involved in resistance to azole drugs such as ketoconazole and fluconazole	
	HOP1	Meiosis-specific protein required for chromosome synapsis	
	ISF1	Serine-rich, hydrophilic protein	
	BDH1	NAD-dependent (R,R)-butanediol dehydrogenase	
	GRE1	Hydrophilin essential in desiccation-rehydration process	
	YPS3	Aspartic protease	
	SPS19	Peroxisomal 2,4-dienoyl-CoA reductase	
	YNL285W	Dubious open reading frame	
	YIL014C-A	Putative protein of unknown function	
	YDR262W	Putative protein of unknown function	

1 (cont.)	*YER188W*	Dubious open reading frame
	IMA3	Alpha-glucosidase
	YMR122W-A	Protein of unknown function;
	MEP3	Ammonium permease of high capacity and low affinity
	CPR6	Peptidyl-prolyl cis-trans isomerase (cyclophilin)
	FLO1	Lectin-like protein involved in flocculation
	DDR2	Multi-stress response protein
	YLR030W	Putative protein of unknown function
	CRG1	S-AdoMet-dependent methyltransferase involved in lipid homeostasis
	NQM1	Transaldolase of unknown function
	GID8	Subunit of GID Complex, binds strongly to central component Vid30p
	ERR2	Enolase, a phosphopyruvate hydratase; catalyzes the conversion of 2-phosphoglycerate to phosphoenolpyruvate
	DIG1	MAP kinase-responsive inhibitor of the Ste12p transcription factor; involved in the regulation of mating-specific genes and the invasive growth pathway
	YPL025C	Dubious open reading frame
	MIG2	Zinc finger transcriptional repressor; cooperates with Mig1p in glucose-induced repression of many genes; under low glucose conditions Mig2p relocalizes to mitochondrion, where it interacts with Ups1p and antagonizes mitochondrial fission factor, Dnm1p, indicative of a role in mitochondrial fusion or regulating morphology
	USV1	Putative transcription factor containing a C2H2 zinc finger; mutation affects transcriptional regulation of genes involved in growth on non-fermentable carbon sources
	AQY1	Spore-specific water channel; mediates the transport of water across cell membranes, developmentally controlled
	ASH1	Component of the Rpd3L histone deacetylase complex
	YEL074W	Dubious open reading frame
	GDH1	NADP(+)-dependent glutamate dehydrogenase; synthesizes glutamate from ammonia and alpha-ketoglutarate
	INP2	Peroxisome-specific receptor important for peroxisome inheritance
	CCH1	Voltage-gated high-affinity calcium channel
	TDH3	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), isozyme 3
	GIP1	Meiosis-specific regulatory subunit of the Glc7p protein phosphatase
	SIP2	One of three beta subunits of the Snf1 kinase complex; involved in the response to glucose starvation
	BDF1	Protein involved in transcription initiation
	FBP26	Fructose-2,6-bisphosphatase, required for glucose metabolism
	HHT1	Histone H3
	STE23	Metalloprotease
	PEX22	Putative peroxisomal membrane protein
	SLT2	Serine/threonine MAP kinase
	YMR295C	Protein of unknown function that associates with ribosomes;
	HTA1	Histone H2A
	GIS3	Protein of unknown function
	ASR1	Ubiquitin ligase that modifies and regulates RNA Pol II
	RLM1	MADS-box transcription factor
	IMP2'	Transcriptional activator involved in maintenance of ion homeostasis

2	*PRS3R PA12 RPC37 RPB5 RPC40 IMD3 RPC19*	Purine metabolism	Propanol Amylalcohol Alphacetoglutarate Pyruvate
	RPS24B RPL28 RPL12B RPL30 RPL16A RPS7B RPL12A RPL22B RPL21B RPL15A RPL43A RPL8A RPL32	Ribosome	
	RPA12 RPC37 RPB5 RPC40 RPC19	Pyrimidine metabolism RNA polymerase	
3	*ALD6 IMD2 ACS2 ERG6 ERG12 MET8 COQ2 ERG10 ERG9 ERG24 ERG11 MVD1 SAM2 ERG20 HEM2 HMG1 ERG13*	Biosynthesis of secondary metabolites	Succinate Ethyldodecanoate Valeric acid Methionol Isobutanol
	ERG6 ERG2 ERG5 ERG27 ERG26 ERG9 ERG24 ERG11 ARE2	Steroid biosynthesis	
	ALD6 ACS2 ERG10	Pyruvate metabolism	
	ERG10 ERG13	Synthesis and degradation of ketone bodies	
	ALD6 ERG10 ERG13	Valine, leucine and isoleucine degradation	
	ERG12 ERG10 MVD1 ERG20 HMG1 ERG13	Terpenoid backbone biosynthesis	
	ALD6 ACS2 ERG10	Propanoate metabolism	
	FAA1 ALD6 ERG10	Fatty acid metabolism	
	ALD6 ERG10	Lysine degradation	

4	GOR1	Glyoxylate reductase	2-methylbutanoic acid Isovaleric acid Octanoic acid Ethylpropionate Ethylactete Ethylhexanoate Ethyldecanoate Isoamyl acetate Hexylacetate Ethylvalerate Phenylethanol Isobutyric acid Ethylbutanoate Hexanoic acid Amylacetate Propylacetate Glycerol Ethyloctanoate Isoamyl alcohol Ethylactate Butyric acid 2-phenylethyl acetate Isobutyl acetate
	ALD2	Cytoplasmic aldehyde dehydrogenase	
	OPI10	Protein with a possible role in phospholipid biosynthesis	
	HBT1	Shmoo tip protein, substrate of Hub1p ubiquitin-like protein	
	PCP1	Mitochondrial serine protease	
	YCR100C	Putative protein of unknown function	
	RPN3	Essential non-ATPase regulatory subunit of the 26S proteasome lid	
	YBR012C	Dubious open reading frame	
	DOG2	2-deoxyglucose-6-phosphate phosphatase	
	RPN9	Non-ATPase regulatory subunit of the 26S proteasome	
	CHA1	Catabolic L-serine (L-threonine) deaminase	
	GPD1	NAD-dependent glycerol-3-phosphate dehydrogenase; key enzyme of glycerol synthesis, essential for growth under osmotic stress	
	PIN4	Protein involved in G2/M phase progression and response to DNA damage	
	GRE3	Aldose reductase	
	EMI2	Non-essential protein of unknown function	
	YNR065C	Protein of unknown function	
	GUD1	Guanine deaminase	
	YBR284W	Putative metallo-dependent hydrolase superfamily protein	
	COQ6	Putative flavin-dependent monooxygenase	
	RFS1	Protein of unknown function	
	DDI3	Protein of unknown function	
	RTC3	Protein of unknown function involved in RNA metabolism	
	PNC1	Nicotinamidase that converts nicotinamide to nicotinic acid	
	GPM2	Homolog of Gpm1p phosphoglycerate mutase; converts 3-phosphoglycerate to 2-phosphoglycerate in glycolysis	
	YMR173W-A	Dubious open reading frame	
	GLG1	Glycogenin glucosyltransferase	
	GCY1	Glycerol dehydrogenase; involved in an alternative pathway for glycerol catabolism used under microaerobic conditions	
	CDC53	Cullin; structural protein of SCF complexes (which also contain Skp1p, Cdc34p, Hrt1p and an F-box protein) involved in ubiquitination	
	YBR285W	Putative protein of unknown function	
	GSP2	GTP binding protein (mammalian Ranp homolog); involved in the maintenance of nuclear organization, RNA processing and transport	
	ADH2	Glucose-repressible alcohol dehydrogenase II	
	IMA1	Major isomaltase (alpha-1,6-glucosidase/alpha-methylglucosidase)	
	GPI12	ER membrane protein involved in the second step of GPI anchor assembly	
	YNL034W	Putative protein of unknown function	
	THI20	Trifunctional enzyme of thiamine biosynthesis, degradation and salvage	
	CLB2	B-type cyclin involved in cell cycle progression	
	ADD37	Protein of unknown function; involved in ER-associated protein degradation	
	YFL052W	Putative zinc cluster protein that contains a DNA binding domain	
	RVS167	Actin-associated protein with roles in endocytosis and exocytosis	

4 (cont.)	UBP11	Ubiquitin-specific protease
	SLM1	Phosphoinositide PI4,5P(2) binding protein, forms a complex with Slm2p
	SOR1	Sorbitol dehydrogenase
	UPC2	Sterol regulatory element binding protein
	CWP1	Cell wall mannoprotein that localizes to birth scars of daughter cells
	PUP3	Beta 3 subunit of the 20S proteasome
	HXT8	Protein of unknown function with similarity to hexose transporters
	YCL049C	Protein of unknown function; localizes to membrane fraction
	OPT1	Proton-coupled oligopeptide transporter of the plasma membrane
	XDJ1	Chaperone with a role in facilitating mitochondrial protein import
	YIR020C	Protein of unknown function
	YGR107W	Dubious open reading frame
	YNL193W	Putative protein of unknown function

Chapter XI

Supporting material:

published papers

OPEN ACCESS Freely available online

PLOS ONE

Computational Models for Prediction of Yeast Strain Potential for Winemaking from Phenotypic Profiles

Inês Mendes[1☯], Ricardo Franco-Duarte[1☯], Lan Umek[2,3], Elza Fonseca[1], João Drumonde-Neves[1,4], Sylvie Dequin[5], Blaz Zupan[3], Dorit Schuller[1]*

1 CBMA (Centre of Molecular and Environmental Biology)/Department of Biology/University of Minho, Braga, Portugal, **2** Faculty of Administration, University of Ljubljana, Ljubljana, Slovenia, **3** Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia, **4** Research Center for Agricultural Technology – Department of Agricultural Sciences, University of Azores, Ponta Delgada, São Miguel, Azores, Portugal, **5** INRA (Institut National de la Recherche), UMR1083, Sciences pour l'Enologie, Montpellier, France

Abstract

Saccharomyces cerevisiae strains from diverse natural habitats harbour a vast amount of phenotypic diversity, driven by interactions between yeast and the respective environment. In grape juice fermentations, strains are exposed to a wide array of biotic and abiotic stressors, which may lead to strain selection and generate naturally arising strain diversity. Certain phenotypes are of particular interest for the winemaking industry and could be identified by screening of large number of different strains. The objective of the present work was to use data mining approaches to identify those phenotypic tests that are most useful to predict a strain's potential for winemaking. We have constituted a *S. cerevisiae* collection comprising 172 strains of worldwide geographical origins or technological applications. Their phenotype was screened by considering 30 physiological traits that are important from an oenological point of view. Growth in the presence of potassium bisulphite, growth at 40°C, and resistance to ethanol were mostly contributing to strain variability, as shown by the principal component analysis. In the hierarchical clustering of phenotypic profiles the strains isolated from the same wines and vineyards were scattered throughout all clusters, whereas commercial winemaking strains tended to co-cluster. Mann-Whitney test revealed significant associations between phenotypic results and strain's technological application or origin. Naïve Bayesian classifier identified 3 of the 30 phenotypic tests of growth in iprodion (0.05 mg/mL), cycloheximide (0.1 μg/mL) and potassium bisulphite (150 mg/mL) that provided most information for the assignment of a strain to the group of commercial strains. The probability of a strain to be assigned to this group was 27% using the entire phenotypic profile and increased to 95%, when only results from the three tests were considered. Results show the usefulness of computational approaches to simplify strain selection procedures.

Citation: Mendes I, Franco-Duarte R, Umek L, Fonseca E, Drumonde-Neves J, et al. (2013) Computational Models for Prediction of Yeast Strain Potential for Winemaking from Phenotypic Profiles. PLoS ONE 8(7): e66523. doi:10.1371/journal.pone.0066523

Editor: Joseph Schacherer, University of Strasbourg, France

Received January 27, 2013; **Accepted** May 6, 2013; **Published** July 16, 2013

Funding: Inês Mendes and Ricardo Franco-Duarte are recipients of a fellowship from the Portuguese Science Foundation, FCT (SFRH/BD/74798/2010, SFRH/BD/48591/2008, respectively) and João Drumonde-Neves is recipient of a fellowship from the Azores government (M3.1.2/F/006/2008 (DRCT)). Financial support was obtained from FEDER funds through the program COMPETE and by national funds through FCT by the projects FCOMP-01-0124-008775 (PTDC/AGR-ALI/103392/2008) and PTDC/AGR-ALI/121062/2010. Lan Umek and Blaz Zupan acknowledge financial support from Slovene Research Agency (P2-0209). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dschuller@bio.uminho.pt

☯ These authors contributed equally to this work.

Introduction

Most European wine producers use commercial starter yeasts to guarantee the reproducibility and the predictability of wine quality. The advantages of fermentations containing *Saccharomyces cerevisiae* starter cultures relies on the fact that they are rapid and produce wine with desirable organoleptic characteristics through successive processes and harvests [1,2]. In these fermentations the winemaker has control over the microbiology of the process, because it is expected that the inoculated yeast strain predominates and suppresses the indigenous flora. Currently, there are about 200 commercial *S. cerevisiae* winemaking strains available, and it is a common practice among wineries to use commercial starter yeasts that were obtained in other winemaking regions.

S. cerevisiae strains from diverse natural habitats harbour a vast amount of phenotypic diversity [3], driven by interactions between yeast and the respective environment. In grape juice fermentations, strains are exposed to a wide array of biotic and abiotic stressors [4], which may lead to strain selection and generate naturally arising strain diversity. Outside the wineries, this diversifying selection occurs due to unique pressures imposed after expansion into new habitats [5–9]. This agrees with findings showing that wine and sake strains are phenotypically more diverse than would be expected from their genetic relatedness [10].

Recent phylogenetic analyses of *S. cerevisiae* strains showed that the species as a whole consists of both "domesticated" and "wild" populations. DNA sequence analysis revealed that domesticated strains derived from two independent clades, corresponding to strains from winemaking and sake. "Wild" populations are mostly associated with oak trees, nectars or insects [11–13]. Although some *S. cerevisiae* strains are specialized for the production of

alcoholic beverages, they were derived from natural populations that were not associated with industrial fermentations. This was proposed once that the oldest lineages and the majority of variation were found in strains from sources unrelated to wine production [14].

The phenotypic diversity of *S. cerevisiae* strains has been explored for decades in strain selection programmes to choose the ones that enhance the wine's sensorial characteristics and confer typical attributes to specific wines. These strains are used as commercial ones by winemakers to efficiently ferment grape musts and produce desirable metabolites, associated with reduced off-flavours [15,16]. Strain selection approaches are mentioned in many studies aiming to characterize *S. cerevisiae* isolates obtained from winemaking regions worldwide. The most relevant physiological tests refer to fermentation rate and optimum fermentation temperature, stress resistance (ethanol, osmotic and acidic), killer phenotype, sulphur dioxide (SO_2) tolerance and production, hydrogen sulphide (H_2S) production, glycerol and acetic acid production, synthesis of higher alcohols (e.g. isoamyl alcohol, n-propanol, isobutanol), β-galactosidase and proteolytic enzyme activity, copper resistance, foam production and flocculation [17].

In our previous work [18] we evaluated the phenotypic and genetic variability of 103 *S. cerevisiae* strains from the *Vinho Verde* wine region (Northwest Portugal). We then applied several data mining procedures to estimate a strain's phenotypic behaviour based on its genotypic data. We used mainly taxonomic tests and strains from winemaking environments of one geographical origin. This study was, to our best knowledge, the first attempt to computationally associate genotypic and phenotypic data of *S. cerevisiae* strains. We used subgroup discovery techniques to successfully identify strains with similar genetic characteristics (microsatellite alleles) that exhibited similar phenotypes.

Within the present study we expanded the strain collection to 172 isolates from worldwide geographical origins and technological groups (wine, bread, sake, etc.) and included 30 tests with biotechnological relevance for the selection of winemaking strains.

Our objective was to gain a deeper understanding of the phenotypic diversity of a global strain collection and to infer computational models that predict the biotechnological potential or geographic origin of a strain from its phenotypic profile.

Results

Phenotypic characterization of the strain collection

A *Saccharomyces cerevisiae* collection was constituted with 172 strains obtained from different geographical origins as shown in the map in Figure 1. As detailed in Table S1 (supplementary data), the technological applications or environments from where the strains were derived were: wine and vine (74 isolates), commercial wine strains (47 isolates), other fermented beverages (12 isolates), other natural environments – soil woodland, plants and insects (12 isolates), clinical (9 isolates), sake (6 isolates), bread (4 isolates), laboratory (3 isolates), beer (1 isolate), and four isolates with unknown origin.

A phenotypic screen was devised to evaluate strain-specific patterns for a set of physiological tests, including also tests that are important for winemaking strain selection. The first group of tests were performed in microplates using supplemented grape must, whereas a high reproducibility was obtained between experimental replicates. The second set of tests consisted in the evaluation of growth in solid culture media (BiGGY medium, Malt Extract Agar supplemented with ethanol and sodium metabisulfite). Galactosidase activity was evaluated by growth evaluation using Yeast Nitrogen Base supplemented with galactose, as indicated in the materials and methods section. After incubation, growth was evaluated by visual scoring (solid media) or by A_{640} determination (liquid media). Table 1 summarizes the number of strains belonging to each of the phenotypic classes. Similarities between strains were evident, but each strain showed a unique phenotypic profile.

A total of 5160 phenotypic data points were obtained, from 172 strains and 30 tests. The concentrations of the added compounds were chosen to obtain a wide range of tolerance patterns. As expected, all strains grew well at 30°C, contrary to the growth at 40°C, where a large phenotypic diversity was observed. Most strains were able to grow well at pH 8, contrarily to the pH value of 2. As expected, cellular growth decreased with increasing concentrations of ethanol (6–14% v/v, liquid media), whereas only five isolates were able to grow well at the highest ethanol concentration of 14% (v/v). When ethanol was combined with sodium metabisulfite in solid culture media, growth was reduced with increasing concentrations of ethanol (12 to 18%, v/v) or sodium metabisulfite (50–100 mg/L). Resistance to sulphur dioxide, which is an antioxidant and bacteriostatic agent used in vinification, was tested by growth in the presence of wine must supplemented with potassium bisulphite ($KHSO_3$). For the concentrations of 150 and 300 mg/L, 101 and 67 strains achieved the highest class of growth, respectively. Resistance to the fungicides iprodion, procymidon and to cycloheximide was rather high at the indicated concentrations. Hydrogen sulphide production was tested using BiGGY medium. The majority of the strains were intermediate H_2S producers with the exception of one strain (from the group of wine and vine strains) that did not produce H_2S.

A global view of strain's phenotypic diversity is shown in Figures 2 and S1. Principal component analysis (PCA) of phenotypic data (Figure 2) show the segregation of all 172 strains (scores) and the loadings for phenotypic variables in the first two PCA components. The phenotypes responsible for the highest strain variability (Figure 2a) were associated with growth patterns in the presence of potassium bisulphite ($KHSO_3$), at 40°C, in a finished wine supplemented with glucose (0.5%, w/v), and resistance to ethanol in liquid media (10 and 14%, v/v). PC-1 (31%) and PC-2 (15%) explained 46% of strain variability and segregated strains by phenotypic behaviour into some patterns, as shown in Figure 2b. The group of sake strains (dark dot) and the group of natural strains (dark square), tended to be separated by the second component, accumulating in the lower part of the PCA, indicating that they were influenced by the presence of ethanol in the medium (higher resistance), and by the growth in the presence of potassium bisulphite (300 mg/L, lower resistance). Strains isolated from vines or wine (dark star) showed a heterogeneous phenotypic behaviour since they were dispersed throughout the PCA plot for both components. A similar tendency was observed for commercial strains (light star); however, the majority of strains tended to concentrate in the upper part of the PCA, indicative of a trend to higher $KHSO_3$ resistance and lower ethanol resistance. The nine clinical strains were distributed in both PCA components, showing no discriminant results in any of the phenotypic tests.

UPGMA (Unweighted Pair Group Method with Arithmetic Mean) algorithm was used to hierarchical cluster the 172 strains. The dissimilarity between two strains was measured using Euclidean distance (Figure S1). The combined phenotypes of wine strains did not separate this group of strains that were rather scattered throughout all the clusters. Commercial strains (light star) tended to be more predominant in the clusters shown in the lower

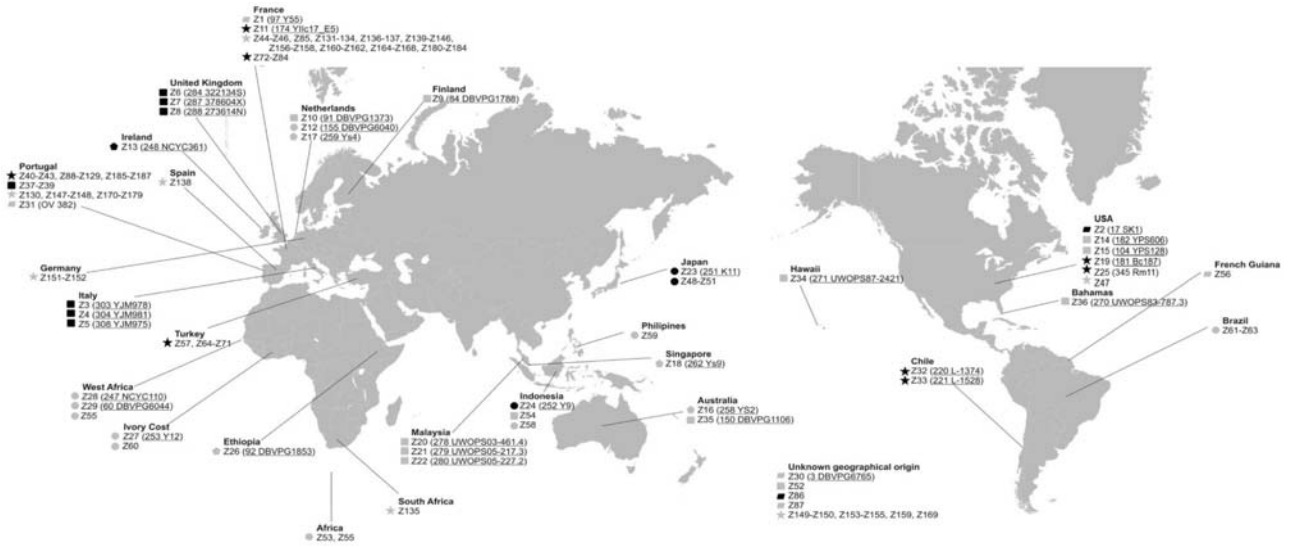


Figure 1. Geographical location of 172 yeast strains. Underlined identifiers indicate the original designation of sequenced strains [12]. Symbols represents the strains technological applications or origin: black star – wine and vine; grey star – commercial wine strain; black square – clinical; grey square – natural isolates; black circle – sake; grey circle – other fermented beverages; black pentagon – beer; grey pentagon- baker; black rectangle – laboratory; grey rectangle – unknown biological origin.
doi:10.1371/journal.pone.0066523.g001

part of the dendogram, where some of the clusters are constituted only by commercial strains.

We further analysed phenotypic diversity through *k*-means clustering algorithm. Using silhouette score [19] we identified 3 distinct clusters (Table 2), composed of 38, 90 and 44 strains respectively. The phenotypes that most distinguished the strains, as indicated by high values of information gain to classify strains into clusters, were growth at the highest and lowest temperature tested (18 and 40°C). Cluster 2 was constituted of strains that didn't grow at both 18 and 40°C, whereas cluster 1 and 3 included strains that grew at both temperatures, but with more pronounced growth at 40°C, in particular for strains of cluster 3. Other tests that were also relevant for the cluster separation included growth in the presence of NaCl (1.5 M), $KHSO_3$ (150 and 300 mg/L), ethanol 6% (v/v) and at pH 2. The strain cluster membership is displayed in the phenotypic data PCA visualization (supplementary Figure S2).

Statistical analysis

The number of strains belonging to each group of technological applications or environment varies between 1 and 74. To assess a possible influence of a sample bias, due to an unequal number of representatives from each group, we determined the 95% confidence intervals for average Manhattan distance [20] between two strains in a selected group (composed by at least 5 strains). The distance was estimated based on the strain's entire phenotypic profile. The lower and upper bound of each confidence interval were determined by percentiles of average distances for 10000 bootstraps samples. For example, with this analysis we show that while the group of commercial strains (47 isolates) includes 31 commercial strains isolated in France, this should not bias our statistical analysis on utility of strains. Namely, the 95% confidence interval for average distances between pairwise combinations of commercial strains from France (6.37, 8.01) overlaps with the confidence interval of commercial strains from other geographical origins (4.97, 8.13). The inclusion of a high number of strains from France does not change the limits of the confidence interval of the group of commercial strains. A similar result was observed for the group of wine and vine strains that includes numerous strains from Portugal: the 95% confidence interval for average distances between pairwise combinations of strains from Portugal (8–12, 9.83) overlaps with the same interval for wine and vine strains from other geographical locations (8.06, 9.59).

Mann-Whitney test is mostly used to identify statistically significant associations between two data sets in which data instances in each group are measured on ordinal level and when there is an unequal number of members in the classes to be compared. This test was used to search for relationships between phenotypic results for the 172 strains, and their shared geographical origin or technological application group. After the dichotomization of variables (geographical origin and technological application or origin), Mann-Whitney test was performed for each phenotypic variable and *p*-values were computed and further adjusted using Bonferroni correction. Statistical analysis using Mann-Whitney test revealed 300 associations between phenotypes and technological application or origin of strains, whereas statistical significance was found for 11 associations (Bonferroni adjusted *p*-value lower than 0.1). For each phenotypic test, we computed the probability of each phenotypic class (0–3) according to its contribution to the observed association. The most significant associations between a phenotypic class and a technological group are reported in Table 3. Two associations were found for the resistance to iprodion, whereas class 3 and 2 were associated with strains collected from wine/vineyards and commercial strains, respectively. Capacity to grow in the presence of potassium bisulphite (150 mg/mL, classes 2 and 3) was associated with commercial wine strains. Natural isolates (87%–89%) were associated with class 2 of growth in wine supplemented with glucose, both at 0.5 and 1% (w/v), contrarily to 57% of commercial strains that were unable to grow in wine supplemented with glucose (0.5%, w/v). The lower ability of commercial strains to grow at higher ethanol concentrations was also supported by the finding of one significant association for absent growth (class 0) in liquid medium containing ethanol (14%, v/v).

Table 1. Number of strains belonging to different phenotypic classes, regarding values of optical density (Class 0: $A_{640} = 0.1$; Class 1: $0.2 < A_{640} > 0.4$; Class 2: $0.5 < A_{640} > 1.0$; Class 3: $A_{640} > 1.0$), growth patterns in solid media, or colour change in BiGGY medium.

Phenotypic test	Type of medium	Phenotypic class of growth			
		0	1	2	3
30°C	liquid (must)	0	0	3	168
18°C	liquid (must)	51	120	1	0
40°C	liquid (must)	28	14	80	50
pH 2	liquid (must)	101	68	3	0
pH 8	liquid (must)	0	0	19	153
KCl (0.75 M)	liquid (must)	0	2	146	24
NaCl (1.5 M)	liquid (must)	84	79	9	0
$CuSO_4$ (5 mM)	liquid (must)	124	45	3	0
SDS (0.01% w/v)	liquid (must)	139	32	1	0
Ethanol 6% (v/v)	liquid (must)	0	2	36	134
Ethanol 10% (v/v)	liquid (must)	17	28	85	42
Ethanol 14% (v/v)	liquid (must)	82	35	50	5
Ethanol 12% (v/v)	solid (MEA)	150	20	1	1
Ethanol 12% (v/v) + $Na_2S_2O_5$ (75 mg/L)	solid (MEA)	159	14	0	0
Ethanol 12% (v/v) + $Na_2S_2O_5$ (100 mg/L)	solid (MEA)	169	3	0	0
Ethanol 14% (v/v) + $Na_2S_2O_5$ (50 mg/L)	solid (MEA)	148	24	0	0
Ethanol 16% (v/v) + $Na_2S_2O_5$ (50 mg/L)	solid (MEA)	163	9	0	0
Ethanol 18% (v/v) + $Na_2S_2O_5$ (50 mg/L)	solid (MEA)	165	7	0	0
$KHSO_3$ (150 mg/L)	liquid (must)	34	11	26	101
$KHSO_3$ (300 mg/L)	liquid (must)	57	19	29	67
Wine supplemented with glucose (0.5% w/v)	liquid	103	45	24	0
Wine supplemented with glucose (1% w/v)	liquid	115	41	16	0
Iprodion (0.05 mg/mL)	liquid (must)	1	0	28	143
Iprodion (0.1 mg/mL)	liquid (must)	1	1	13	157
Procymidon (0.05 mg/mL)	liquid (must)	0	0	7	165
Procymidon (0.1 mg/mL)	liquid (must)	1	0	9	162
Cycloheximide (0.05 μg/mL)	liquid (must)	3	0	7	162
Cycloheximide (0.1 μg/mL)	liquid (must)	2	1	19	150
H_2S production	solid (BiGGY)	1	11	105	55
Galactosidase activity	liquid (YNB)	0	21	98	53

MEA: Malt Extract Agar.

doi:10.1371/journal.pone.0066523.t001

About half of the strains included in the groups shared the inability to grow in must containing SDS (0.01%, w/v) and $CuSO_4$ (5 mM), but grew well in cycloheximide-supplemented must (76% of strains, class 2). An identical approach was undertaken to find associations between the phenotypic results and the geographical origin of strains, but no statistically relevant results were obtained (data not shown).

Prediction of technological group based on phenotypic results

Our next objective was to construct a model that would predict strain's technological group from its phenotypic profile. *K*-nearest neighbour algorithm (*k*NN) and naïve Bayesian classifiers [21], as implemented in the Orange data mining software were used for modelling.

The predictive performance of both classifiers was evaluated in terms of area under the Receiver-Operating-Characteristics (ROC) curve, using 5-fold cross validation [22]. Table 4 shows the confusion matrix of naïve Bayesian classifications in test data sets of cross-validation; *k*NN results are not shown, as these were similar for both modelling techniques. Cross validated AUC score was 0.70. Correct assignments were found for the larger groups of commercial wine strains and strains obtained from wine and vineyards, where 36 (77%) and 54 (73%) strains respectively, were accurately allocated. The same computational technique was also used to explore which phenotypes mostly contributed to the assignment of a strain to the commercial wine group. Figure 3 represents a nomogram that shows naïve Bayesian classifier results [23]. Three phenotypes were considered by the classifier as the ones contributing more positively to build the model, having the remaining ones a smaller impact. To predict the commercial potential of a strain, the contribution of each phenotype was scored in the scale from −100 to 100, and the individual scores were summed-up to read-out the probability of the predicted class. For the present data set, growth in must containing the fungicide

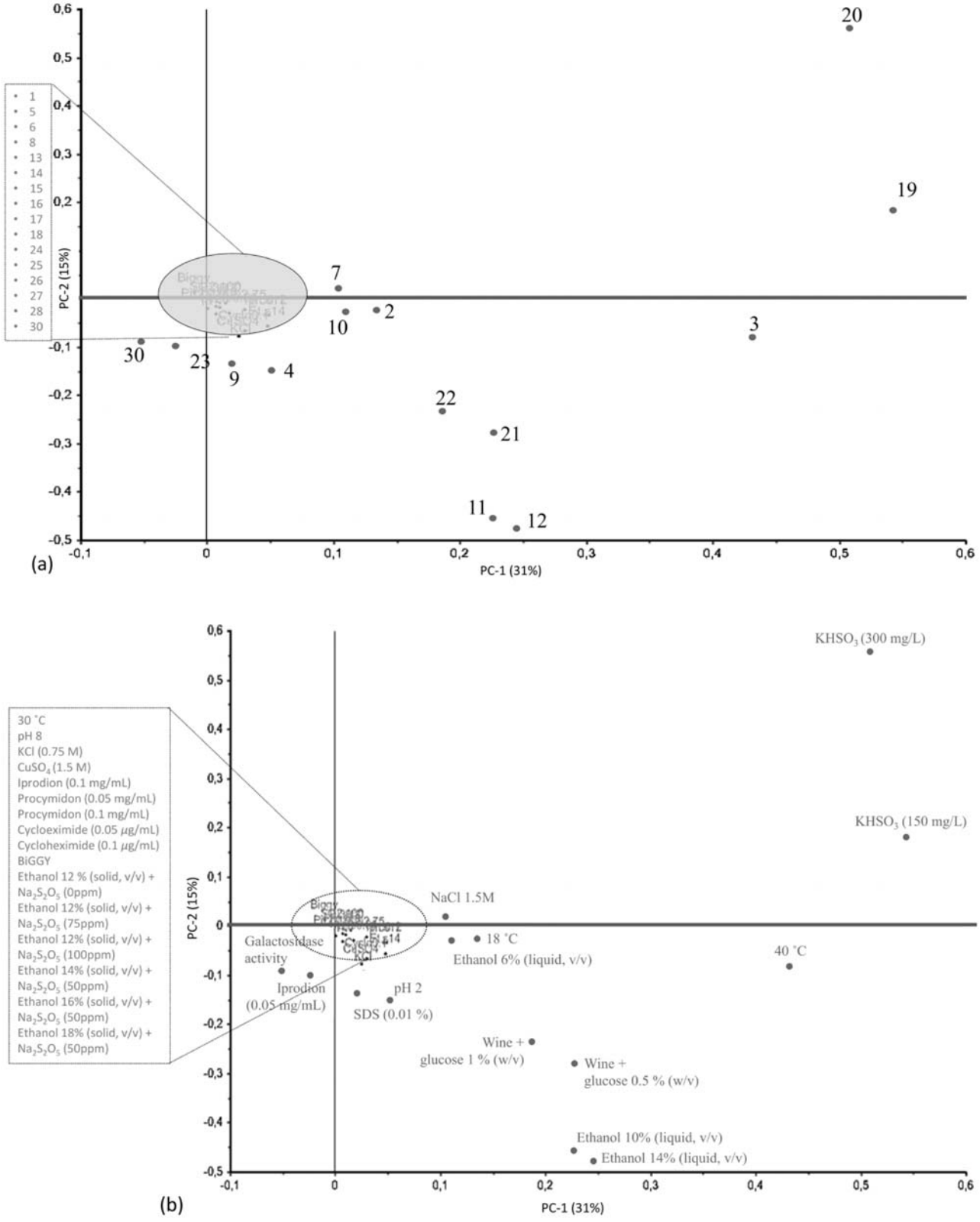


Figure 2. Principal component analysis of phenotypic data for 172 strains. (**a**) −30 phenotypic tests (loadings). Numbers indicate phenotypic tests, as mentioned in Table 1: (1) −30°C; (2) −18°C; (3) −40°C; (4) – pH 2; (5) – pH 8; (6) – KCl (0.75 M); (7) – NaCl (1.5 M); (8) – $CuSO_4$ (1.5 M); (9) – SDS (0.01%); (10) – ethanol 6% (v/v) liquid medium; (11) – ethanol 10% (v/v) liquid medium; (12) – ethanol 14% (v/v) liquid medium; (13) – ethanol 12% (v/v) solid medium; (14) – ethanol 12% (v/v) solid medium + $Na_2S_2O_5$ (75 mg/L); (15) – ethanol 12% (v/v) solid medium + $Na_2S_2O_5$ (100 mg/L); (16) – ethanol 14% (v/v) solid medium + $Na_2S_2O_5$ (50 mg/L); (17) – ethanol 16% (v/v) solid medium + $Na_2S_2O_5$ (50 mg/L); (18) – ethanol

18% (v/v) solid medium + $Na_2S_2O_5$ (50 mg/L); (19) – $KHSO_3$ (150 mg/L); (20) – $KHSO_3$ (300 mg/L); (21) – wine supplemented with glucose 0.5% (w/v); (22) – wine supplemented with glucose 1% (w/v); (23) – Iprodion (0.05 mg/mL); (24) – Iprodion (0.1 mg/mL); (25) – Procymidon (0.05 mg/mL); (26) – Procymidon (0.1 mg/mL); (27) – Cycloheximide (0.05 μg/mL); (28) – Cycloheximide (0.1 μg/mL); (29) – H_2S production; (30)– galactosidase activity. (**b**) – 172 strains (scores) distribution. Symbols represents the strains technological applications or origin: black star – wine and vine; grey star – commercial wine strain; black square – clinical; grey square – natural isolates; black circle – sake; grey circle – other fermented beverages; black pentagon – beer; grey pentagon- baker; black rectangle – laboratory; grey rectangle – unknown biological origin.
doi:10.1371/journal.pone.0066523.g002

iprodion (0.05 mg/mL), in cycloheximide (0.1 μg/mL) and in the presence of potassium bisulphite (150 mg/mL) were the three features with the most relevant contribution for the mathematical assignment of a strain to the commercial group (Figure 3a). The probability of a strain to be assigned to the group of commercial strains is 0.27 (27%) when considering the strains entire phenotypic profile and increases to 0.95 (95%) when only the three phenotypic results mentioned in Figure 3a are taken into consideration, as shown in the probability scale present in Figure 3b.

Discussion

Within our previous work [18] we developed computational techniques to relate the genotypes and phenotypes of 103 *Saccharomyces cerevisiae* strains from a winemaking region. The isolates were characterized regarding their allelic combinations for 11 microsatellites and phenotypic screens included mainly taxonomic criteria but also some tests with biotechnological relevance. Subgroups were found for strains sharing allelic combinations and specific phenotypes such as low ethanol resistance, growth at 30°C and growth in media containing galactose, raffinose or urea. Herein, we aim to extend the work to a phenotypically mostly heterogeneous strain collection of 172 *S. cerevisiae* isolates from worldwide origins, to computationally relate the phenotype with the strain's geographical origins and to make predictions about a strain's biotechnological potential based on phenotypic data. The group of phenotypic tests used herein was based on approaches that are generally applied for the selection of *S. cerevisiae* winemaking strains [17].

The collection of 172 strains from worldwide geographical origins revealed a high phenotypic diversity (Figures 2, S2 and Table 2), which is in agreement with previous studies [3,10,18,24–27]. A significantly higher phenotypic diversity was observed in the present study compared to our results from 2009 using 103 Portuguese wine yeast strains [18]. In particular, the inclusion of new tests compared to our previous study allowed a more detailed analysis of the phenotypic variability of strains associated with winemaking environments. Recent studies aimed to describe the elements that shaped the genomes of *S. cerevisiae* strains, suggesting that populations comprise distinct domesticated and natural groups, as well as mosaics within these groups, based on the strain origin and application [12,28,29]. Clinical isolates for example, do not derive from a common ancestor, but rather represent multiple events in which environmental strains opportunistically colonize humans [28,30].

Table 2. Phenotypic tests mostly contributing for the division of strains into three clusters, in terms of information gain, obtained with *k*-means clustering algorithm.

Phenotypic test	Information gain	Cluster		
		1	2	3
18°C	0,33	1	0	1
40°C	0,33	2	0	3
NaCl (1.5M)	0,26	0	0	1
$KHSO_3$ (300 mg/L)	0,23	3	0	3
Ethanol 6% (v/v) – liquid medium	0,23	3	2	3
pH 2	0,21	0	0	1
$KHSO_3$ (150 mg/L)	0,21	3	0	3
Total number of strains		38	90	44

Numbers in the last three columns represent the most characteristic value in terms of phenotypic class of strains included in the clusters, for the mentioned phenotypic tests.
doi:10.1371/journal.pone.0066523.t002

Genetic rearrangements and intra-strain variation is characteristic for this species [31,32], which might explain the rather high phenotypic variability that was described in recent studies. Camarasa [3] showed that some phenotypes (resistance to high sugar concentrations, ability to complete fermentation and low acetate production) were able to distinguish groups of strains according to their ecological niches, providing evidence for phenotypic evolution driven by environmental adaptation. This high phenotypic variation in stressful conditions was also revealed by Kvitek *et al.*, showing the existence of unique features shared by strains from similar habitats [10]. Our data are in agreement with the previously mentioned studies regarding the high phenotypic diversity. They also confirm the findings of Legras and co-workers [33], that found populational substructures of *S. cerevisiae* strains according to their technological application or origin, using multilocus microsatellite typing. In the work of Legras only 28% of the diversity was associated with geographical origins, which suggests local domestication events. We herein investigated the utility of data mining to improve our understanding of relations between phenotypes and the strains technological application or origin. The developed models can also be useful to optimize screening tests and to find commercial wine yeast candidates from strain collections.

Using Mann-Whitney test, 11 significant associations were found between a particular phenotypic result and a technological application or origin of the strains (Table 3). The most significant results were found for the resistance to iprodion, growth in potassium bisulphite and in wine supplemented with glucose. Iprodion is a dicarboximide contact fungicide used to control a wide variety of fungal pests on vegetables, ornamentals, pome and stone fruit, root crops, cotton and sunflowers. *S. cerevisiae* shows a higher resistance to this fungicide than other yeast species such as *Candida albicans*. In this species iprodion stimulates glycerol synthesis and inhibits the cell growth for several days, contrarily to *S. cerevisiae* where a low toxicity was observed [34,35]. Our results showed that iprodion resistance (0.05 mg/mL) was higher in strains from wine and vineyards compared to commercial wine strains. The higher iprodion resistance among strains obtained from wineries and vineyards might be explained by the evolution of this trait upon recurrent exposure, which does not apply for commercial wine strains that are added to clarified musts that should not contain this fungicide. The low ethanol resistance of commercial wine strains in liquid media containing 14% (v/v)

Table 3. Relevant associations (adjusted $p<0.1$) between phenotypic results and strain's technological application or origin, obtained using Mann-Whitney test and after Bonferroni correction.

Phenotypic test	Class of phenotypic result	Technological group/origin	Adjusted *p*-value	% of strains sharing positive association *
Iprodion (0.05 mg/mL)	2	Commercial	3.24×10^{-8}	82.0
Iprodion (0.05 mg/mL)	3	Wine and vine	0.015	56.4
$KHSO_3$ (150 mg/L)	2, 3	Commercial	0.001	59.3
Wine supplemented with glucose (0.5%, w/v)	0	Commercial	0.075	57.0
Wine supplemented with glucose (0.5%, w/v)	2	Natural isolate	0.002	87.2
Wine supplemented with glucose (1%, w/v)	2	Natural isolate	0.041	89.5
Ethanol 14% (v/v) – liquid medium	0	Commercial	0.004	64.5
Cycloheximide (0.1 µg/mL)	2	Commercial	0.007	75.6
Procymidon (0.1 mg/mL)	2	Other fermented beverages	0.005	92.4
SDS (0.01%, w/v)	0	Commercial	0.078	45.3
$CuSO_4$ (5 mM)	0	Commercial	0.075	50.6

*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.
doi:10.1371/journal.pone.0066523.t003

ethanol was somehow unexpected, because these strains are usually selected for high ethanol resistance. This could be explained by the fact that the mathematical relations were observed for ethanol concentrations above the values that usually occur in wines (10–13%, v/v). Results showed also that commercial strains tended to a better growth in media containing potassium bisulphite, a compound used as wine antiseptic and antioxidant, reflecting also an adaptive mechanism among this group of strains.

We found that the large phenotypic variability between strains could be associated with the technological application or origin of the strains (Table 3) rather than their geographical origin, once that no relevant relations were considered for the last analysis. The naïve Bayesian classifier was used to assign a strain to their technological application or origin group, based on their phenotypic profile (Table 4). This association was achieved for the majority of strains belonging to the commercial and wine and vine groups (77% and 73% respectively). The cross-validated performance of this method yielded an AUC score of 0.70, that is

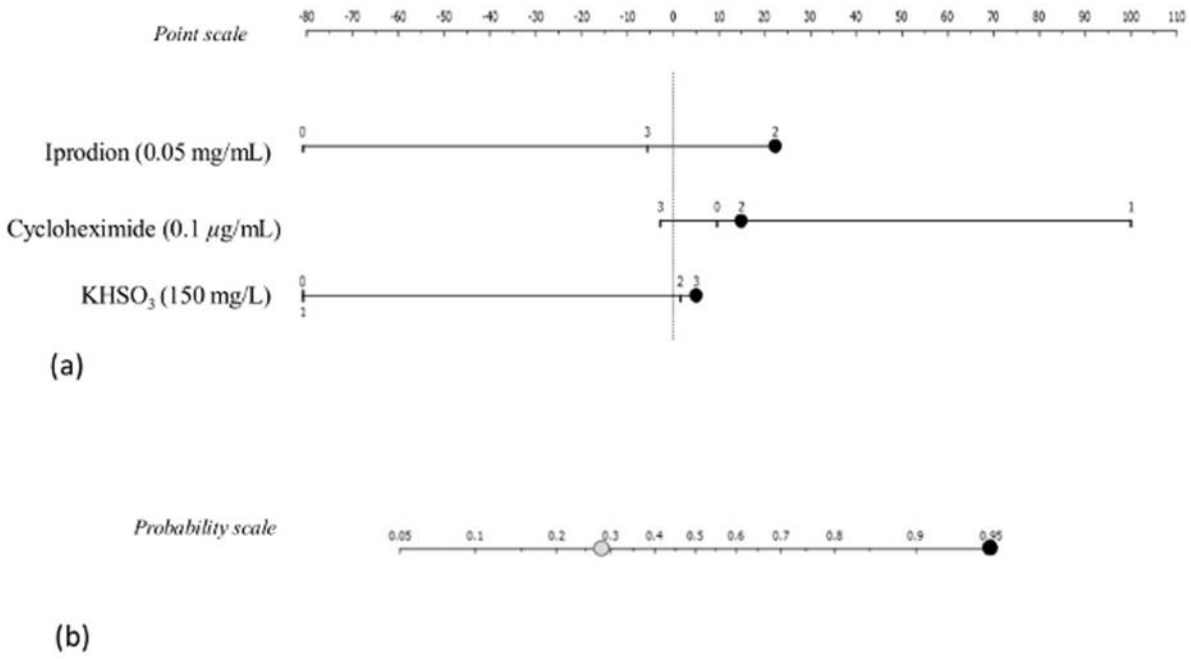


Figure 3. Nomogram showing naïve Bayesian classifier results for the prediction of commercial strains based on phenotypic classes of growth for each test. (**a**) Performance of three phenotypic tests that contributed in a positive way to predict commercial strains; (**b**) Probability of predicting commercial strains when considering the entire phenotypic profile (grey circle), or only the three phenotypic tests mentioned in panel (a) by the blue dots (black circle).
doi:10.1371/journal.pone.0066523.g003

Table 4. Confusion matrix indicating the technological application or origin prediction of 172 strains and their predictions as obtained with naïve Bayesian classifier (AUC = 0.70).

			Predicted technological application or origin									
		Total number of strains	Beer	Bread	Clinical	Commercial wine strain	Laboratory	Natural isolate	Other fermented beverages	Sake	Unknown biological origin	Wine and vine
Real technological application or origin	Beer	1	**0** (0%)	0	0	0	0	1	0	0	0	0
	Bread	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	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
	Wine and vine	74	0	1	3	8	1	2	3	1	1	**54** (73%)

doi:10.1371/journal.pone.0066523.t004

considered as moderate [22] and lies in between the values of an arbitrary and perfect classification (AUC = 0.5 and 1.0, respectively). Poor results were obtained for the remaining groups, which is due to the corresponding small number of isolates. These results demonstrate the potential of the predictive models to classify strains based on results of phenotypic screens.

Bayesian classifier used the strains phenotypic profiles for prediction of commercial strains, and identified 3 of the 30 phenotypic tests (growth in musts containing iprodion (0.05 mg/mL), cycloheximide (0.1 μg/mL) or potassium bisulphite (150 mg/mL)) as the ones providing more information for the assignment of strains to the commercial group. When using only 3 tests, rather than the entire phenotypic profile, the probability of a strain to be classified as commercial increases significantly (from 27% to 95%).

In conclusion, our results demonstrate the usefulness of computational approaches to describe phenotypic variability among groups of *S. cerevisiae* strains that also might occur as adaptive mechanisms in specific environments. The herein developed models can make predictions about the biotechnological potential of strains and simplify the selection of candidate strains to be used as commercial wine strains.

Materials and Methods

Strain collection

A *Saccharomyces cerevisiae* strain collection was constituted, comprising 172 strains with different geographical origins and technological applications or origins (Figure 1 and Table S1 – supplementary data). This collection includes strains used for winemaking (commercial and natural isolates that were obtained from winemaking environments), brewing, bakery, distillery (sake, cachaça) and ethanol production, laboratory strains and also strains from particular environments (e.g. pathogenic strains, isolates from fruits, soil and oak exudates). The complete genome sequence of thirty strains is currently available [12] (their original strain code is mentioned in the map of Figure 1). All strains were coded (Zn) and stored at −80°C in cryotubes containing 1 mL glycerol (30% v/v).

Phenotypic characterization

Phenotypic screening was performed considering a wide range of physiological traits that are also important from an oenological point of view.

In a first set of phenotypic tests, strains were inoculated into replicate wells of 96-well microplates. Isolates were grown overnight in YPD medium (yeast extract 1% w/v, peptone 1% w/v, glucose 2% w/v), and the optical density (A_{640}) was then determined and adjusted to 1.0. After washing with peptone (1% w/v), 15 μL of this suspension were inoculated in quadruplicate in microplate wells containing 135 μL of white grape must of the variety Loureiro, to a cellular density of 5×10^6 cells/mL (A_{640} = 0.1). Final optical density was determined after 22 h (30°C, 200 rpm) in a microplate spectrophotometer. All microplates were carefully sealed with parafilm, and no evaporation was observed for incubation temperatures of 30°C and 40°C. As shown in Table 1, this approach included the following tests: growth at various temperatures (18, 30 and 40°C), evaluation of ethanol resistance (6, 10 and 14%, v/v), tolerance to several stress conditions caused by extreme pH values (2 and 8), osmotic/saline stress (0.75 M KCl and 1.5 M NaCl). Growth was also assessed in the presence of potassium bisulfite ($KHSO_3$, 150 and 300 mg/L), copper sulphate ($CuSO_4$, 5 mM), sodium dodecyl sulphate (SDS, 0.01%, w/v), the fungicides iprodion (0.05 and 0.1 mg/mL) and

procymidon (0.05 and 0.1 mg/mL), as well as cycloheximide (0.05 and 0.1 mg/mL). These tests were carried out using Loureiro grape must supplemented with the mentioned compounds. The growth in finished wines was determined by adding glucose (0.5 and 1%, w/v) to a commercial white wine (12.5% v/v alcohol content). Galactosidase activity was evaluated by adding galactose (5% w/v) to Yeast Nitrogen Base (YNB, Difco^{TM}, Ref. 239210), using test tubes with 5 mL culture medium and 5×10^6 cells/mL, followed by 5 to 6 days of incubation at 26°C.

Other tests were performed using solid media. Overnight cultures were prepared as previously described, adjusted to an optical density (A_{640}) of 10.0 and washed. One µl of this suspension was placed on the surface of the culture media mentioned below. Hydrogen sulphide production was evaluated using BiGGY medium (SIGMA-ALDRICH, Ref. 73608) [36], followed by incubation at 27°C for 3 days. The colony colour, which represents the amount of H_2S produced was then analysed, attributing a score from 0 (no colour change) to 3 (dark brown colony). Ethanol resistance (12%, v/v) and the combined resistance to ethanol (12, 14, 16 and 18%, v/v) and sodium bisulphite ($Na_2S_2O_5$; 75 and 100 mg/L) was evaluated by adding the mentioned compounds to Malt Extract Agar (MEA, SIGMA-ALDRICH, Ref. 38954), and growth was visually scored after incubation (2 days at 27°C).

All phenotypic results were assigned to a class between 0 and 3 (0: no growth (A_{640} = 0.1) or no visible growth on solid media or no colour change of the BiGGY medium; 3: at least 1.5 fold increase of A_{640}, extensive growth on solid media or a dark brown colony formed in the BiGGY medium; scores 1 and 2 corresponded to the respective intermediate values) as shown in table S2.

Data analysis

The phenotypic variability was evaluated by principal component analysis (PCA), available in the Unscrambler X software (Camo). The BioNumerics software (Applied Maths) was used for clustering, dendogram drawing and calculation of cophenetic correlation coefficients. Mann-Whitney test was applied to the phenotypic data set, including Bonferroni correction, to find relevant associations between phenotypic data and the strain's technological or geographical origin. A set of standard predictive data-mining methods, such as naïve Bayesian classifier and *k* nearest-neighbours algorithm [21], as implemented in the Orange data mining suite [37,38], were used for the inference of prediction models. For prediction scoring, area under the receiver operating characteristics (ROC) curve (AUC) was used [22], which estimates the probability that the predictive model would correctly differentiate between distinct locations or distinct technological application or origins, given the associated pairs of strains.

Supporting Information

Figure S1 Phenotypic variation of 172 strains under 30 growth conditions. Strains are organized according to UPGMA-based hierarchical clustering (cophenetic correlation factor = 0.75), using Euclidean distance correlation to estimate phenotypic profile similarities. Symbols represents the strains technological applications or origin: black star – wine and vine; grey star – commercial wine strain; black square – clinical; grey square – natural isolates; black circle – sake; grey circle – other fermented beverages; black pentagon – beer; grey pentagon-baker; black rectangle – laboratory; grey rectangle – unknown biological origin.
(TIF)

Figure S2 PCA representation of the three strain clusters, obtained with *k*-means clustering algorithm. The symbols represent the belonging of the 172 strains shown in the phenotypic data PCA (Figure 2b) to each cluster: circles – cluster 1 (38 strains); lines – cluster 2 (90 strains); squares – cluster 3 (44 strains).
(TIF)

Table S1 Origin and technological application of the 172 *Saccharomyces cerevisiae* strains.
(DOCX)

Table S2
(XLSX)

Acknowledgments

The authors would like to thank all the researchers that kindly provided yeast strains: Gianni Liti, Institute of Genetics UK, Laura Carreto, CESAM and Biology Department Portugal, Goto-Yamamoto, NRIB Japan, Cletus Kurtzman, Microbial Properties Research USA, Rogelio Brandao, Laboratório de Fisologia e Bioqui?mica de Microorganismos Brazil, Huseyin Erten, Cukurova University Turkey.

Author Contributions

Conceived and designed the experiments: IM RD DS. Performed the experiments: IM RD EF JN. Analyzed the data: RD IM LU. Contributed reagents/materials/analysis tools: DS BZ. Wrote the paper: RD IM DS. Revised the final manuscript: SD.

References

1. Fleet GH (1998) Yeasts – What reactions and interactions really occur in natural habitats. Food Technol. Biotechnol. 36: 285–289.
2. Schuller D (2010) Better yeast for better wine – genetic improvement of *Saccharomyces cerevisiae* wine strains. In: Rai M, Koevics G, editors. Progress in mycology. Jodhpur: Scientific Publishers (India). 1–51.
3. 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: e25147.
4. Bisson LF (1999) Stuck and sluggish fermentations. Am J Enol Vitic 50: 107–119.
5. Frezier V, Dubourdieu D (1992) Ecology of yeast strains *Saccharomyces cerevisiae* during spontaneous fermentation in Bordeaux winery. Am J Enol Vitic 43: 375–380.
6. Lopes CA, Broock M Van, Querol A, Caballero AC (2002) *Saccharomyces cerevisiae* wine yeast populations in a cold region in Argentinean Patagonia. A study at different fermentation scales. J Appl Microbiol 93: 608–615.
7. Sabate J, Cano J, Querol A, Guillamo JM (1998) Diversity of *Saccharomyces* strains in wine fermentations: analysis for two consecutive years. Lett Appl Microbiol 26: 452–455.
8. Schuller D, Alves H, Dequin S, Casal M (2005) Ecological survey of *Saccharomyces cerevisiae* strains from vineyards in the Vinho Verde Region of Portugal. FEMS Microbiol Ecol 51: 167–177.
9. Valero E, Cambon B, Schuller D, Casal M, Dequin S (2007) Biodiversity of *Saccharomyces* yeast strains from grape berries of wine-producing areas using starter commercial yeasts. FEMS Yeast Res 7: 317–329.
10. Kvitek DJ, Will JL, Gasch AP (2008) Variations in stress sensitivity and genomic expression in diverse *Saccharomyces cerevisiae* isolates. PLoS Genet 4: 31–35.
11. Greig D, Leu JY (2009) Natural history of budding yeast. Curr Biol 19: 886–890.
12. Liti G, Carter DM, Moses AM, Warringer J, Parts L, et al. (2009) Population genomics of domestic and wild yeasts. Nature 458: 337–341. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2659681&tool=pmcentrez&rendertype=abstract. Accessed 2 March 2012.
13. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. Nature 458: 342–345.
14. Fay JC, Benavides J (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. PLoS Genet 1: 66–71.

15. Briones AI, Ubeda JF, Cabezudo MD, Martin-Alvarez P (1995) Selection of spontaneous strains of Saccharomyces cerevisiae as starters in their viticultural area. In: Charalambous G, editor. Food flavours: generation, analysis and process influence. Amsterdam: Elsevier Science. 1597–1622.
16. Ramirez M, Perez F, Regodon JA (1998) A simple and reliable method for hybridization of homothallic wine strains of *Saccharomyces cerevisiae*. Appl Environ Microbiol 64: 5039–5041.
17. Mannazzu I, Clementi F, Ciani M (2002) Strategies and criteria for the isolation and selection of autochthonous starter. In: Ciani M, editor. Biodiversity and biotechnology of wine yeasts. Trivandrum: Research Signpost. 19–35.
18. 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.
19. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20: 53–65. Available: http://linkinghub.elsevier.com/retrieve/pii/0377042787901257.
20. Grimshaw SD, Efron B, Tibshirani RJ (1995) An Introduction to the Bootstrap. Technometrics 37: 341.
21. Tan P, Steinbach M, Kumar V (2006) Introduction to data mining. Pearson Ed. Boston: Pearson Addison Wesley.
22. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143: 29–36.
23. Mozina M, Demsar J, Kattan M, Zupan B (2004) Nomograms for visualization of naive Bayesian classifier. Lecture Notes in Computer Science 3202: 337–348.
24. Agnolucci M, Scarano S, Santoro S, Sassano C, Toffanin A, et al. (2007) Genetic and phenotypic diversity of autochthonous Saccharomyces spp. strains associated to natural fermentation of "Malvasia delle Lipari". Lett Appl Microbiol 45: 657–662.
25. Brandolini V, Tedeschi P, Capece A, Maietti A, Mazzotta D, et al. (2002) *Saccharomyces cerevisiae* wine strains differing in copper resistance exhibit different capability to reduce copper content in wine. World J Microbiol Biotechnol 18: 499–503.
26. Salinas F, Mandakovic D, Urzua U, Massera A, Miras S, et al. (2010) Genomic and phenotypic comparison between similar wine yeast strains of *Saccharomyces cerevisiae* from different geographic origins. J Appl Microbiol 108: 1850–1858.
27. Cubillos F a, Zia A, Gjuvsland A, Jared T, Warringer J, et al. (2011) Trait variation in yeast is defined by population history. PLoS Genet 7: e1002111. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid = 3116910&tool = pmcentrez&rendertype = abstract. Accessed 8 November 2012.
28. Schacherer J, Ruderfer DM, Gresham D, Dolinski K, Botstein D, et al. (2007) Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. PLoS One 2: e322.
29. Goddard MR, Anfang N, Tang R, Gardner RC, 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. Environ Microbiol 12: 63–73. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd = Retrieve&db = PubMed&dopt = Citation&list_uids = 19691498.
30. Muller LH, McCusker JH (2009) Microsatellite analysis of genetic diversity among clinical and nonclinical Saccharomyces cerevisiae isolates suggests heterozygote advantage in clinical environments. Mol Ecol 18: 2779–2786.
31. Schuller D, Pereira L, Alves H, Cambon B, Dequin S, et al. (2007) Genetic characterization of commercial *Saccharomyces cerevisiae* isolates recovered from vineyard environments. Yeast 24: 625–636.
32. Dunn B, Levine RP, Sherlock G (2005) Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. BMC genomics 6: 53. doi:10.1186/1471-2164-6-53.
33. Legras J-L, Merdinoglu D, Cornuet J-M, Karst F (2007) Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Mol Ecol 16: 2091–2102.
34. Chiai NO, Ujimura MF, Shima MO, Otoyama TM, Chiishi AI, et al. (2002) Efects of iprodione and fludioxonil on glycerol synthesis and hyphal development in Candida albicans. Biosci Biotechnol Biochem 66: 2209–2215.
35. Cadez N, Zupan J, Raspor P (2010) The effect of fungicides on yeast communities associated with grape berries. FEMS Yeast Res 10: 619–630.
36. Jiranek V, Langridge P, Henschke PA (1995) Validation of bismuth-containing indicator media for predicting H2S-producing potential of Saccharomyces cerevisiae wine yeasts under enological conditions. Am J Enol Vitic 46: 269–273.
37. Curk T, Demsar J, Xu Q, Leban G, Petrovic U, et al. (2005) Microarray data mining with visual programming. Bioinformatics 21: 396–398.
38. Demsar J, Zupan B, Leban G (2004) Orange: from experimental machine learning to interactive data mining. White Paper (www.ailab.si/orange), Faculty of Computer and Information Science, University of Ljubljana.

Yeast
Yeast 2014; **31**: 265–277.
Published online 26 May 2014 in Wiley Online Library
(wileyonlinelibrary.com) **DOI:** 10.1002/yea.3016

Research Article

Computational models reveal genotype–phenotype associations in *Saccharomyces cerevisiae*

Ricardo Franco-Duarte[1†], Inês Mendes[1†], Lan Umek[2,3], João Drumonde-Neves[1,4], Blaz Zupan[3] and Dorit Schuller[1*]

[1]*Centre of Molecular and Environmental Biology (CBMA), Department of Biology, University of Minho, Braga, Portugal*
[2]*Faculty of Administration, University of Ljubljana, Slovenia*
[3]*Faculty of Computer and Information Science, University of Ljubljana, Slovenia*
[4]*Research Centre for Agricultural Technology, Department of Agricultural Sciences, University of Azores, Angra do Heroísmo, Portugal*

**Correspondence to:*
D. Schuller, Centre of Molecular and Environmental Biology (CBMA), Department of Biology, University of Minho, Braga, Portugal.
E-mail: dorit.schuller@gmail.com
†These authors contributed equally to this study.

Received: 26 November 2013
Accepted: 10 April 2014

Abstract

Genome sequencing is essential to understand individual variation and to study the mechanisms that explain relations between genotype and phenotype. The accumulated knowledge from large-scale genome sequencing projects of *Saccharomyces cerevisiae* isolates is being used to study the mechanisms that explain such relations. Our objective was to undertake genetic characterization of 172 *S. cerevisiae* strains from different geographical origins and technological groups, using 11 polymorphic microsatellites, and computationally relate these data with the results of 30 phenotypic tests. Genetic characterization revealed 280 alleles, with the microsatellite ScAAT1 contributing most to intrastrain variability, together with alleles 20, 9 and 16 from the microsatellites ScAAT4, ScAAT5 and ScAAT6. These microsatellite allelic profiles are characteristic for both the phenotype and origin of yeast strains. We confirm the strength of these associations by construction and cross-validation of computational models that can predict the technological application and origin of a strain from the microsatellite allelic profile. Associations between microsatellites and specific phenotypes were scored using information gain ratios, and significant findings were confirmed by permutation tests and estimation of false discovery rates. The phenotypes associated with higher number of alleles were the capacity to resist to sulphur dioxide (tested by the capacity to grow in the presence of potassium bisulphite) and the presence of galactosidase activity. Our study demonstrates the utility of computational modelling to estimate a strain technological group and phenotype from microsatellite allelic combinations as tools for preliminary yeast strain selection.

Keywords: ***Saccharomyces cerevisiae*; microsatellite; phenotypic characterization; data mining; nearest-neighbour classifier**

Introduction

Large-scale genome-sequencing projects of *Saccharomyces cerevisiae* strains are essential to understand individual variation and to study the mechanisms that explain relations between genotype and phenotype. Revealing such associations will help to increase our understanding of genetic and phenotypic strain diversity, which is particularly high in the case of winemaking strains. Relational studies of genetic and phenotypic variability should help to decipher genotype–phenotype relationships and elucidate genetic adaptations involved in phenotypes that are relevant to thrive in stressful industrial environments. They should also contribute towards strain improvement strategies through breeding and genetic engineering, taking into consideration the diversity of the wild strains (Borneman *et al.*, 2013; Dequin and Casaregola, 2011; Roberts and Oliver, 2011).

Recent phylogenetic analyses of *S. cerevisiae* strains showed that the species as a whole consists of both 'domesticated' and 'wild' populations, whereby the genetic divergence is associated with both ecology and geography. Sequence comparison of 70 *S. cerevisiae* isolates confirmed the existence of five well-defined lineages and some mosaics, suggesting the occurrence of two domestication events during the history of association with human activities, one for sake strains and one for wine yeasts (Liti and Schacherer, 2011; Liti *et al.*, 2009; Schacherer *et al.*, 2009). *S. cerevisiae* isolates associated with vineyards and wine production form a genetically differentiated group, distinct from 'wild' strains isolated from soil and oak-tree habitats, and also from strains derived from other fermentations, such as palm wine and sake or clinical strains. Recent research indicates that wine strains were domesticated from wild *S. cerevisiae* (Fay and Benavides, 2005; Legras *et al.*, 2007), followed by dispersal, and the diversifying selection imposed after yeast expansion into new environments due to unique pressures led to strain diversity (Borneman *et al.*, 2013; Diezmann and Dietrich, 2009; Dunn *et al.*, 2012). The interactions between *S. cerevisiae* and humans are considered drivers of yeast evolution and the development of genetically, ecologically and geographically divergent groups (Goddard *et al.*, 2010; Legras *et al.*, 2007; Sicard and Legras, 2011). The limited knowledge about the mechanisms responsible for the fixation of specific genetic variants due to ecological pressures can be extended by combining genetic and phenotypic characteristics. Recent studies show that groups of strains can be distinguished on the basis of specific traits that were shaped by the species' population history. Wine and sake strains are phenotypically more diverse than would be expected from their genetic relatedness, and the contrary is the case for strains collected from oak trees (Kvitek *et al.*, 2008). Wine yeasts and other strains accustomed to growing in the presence of musts with high sugar concentrations are able to efficiently ferment synthetic grape musts, contrary to isolates from oak trees or plants that occur in environments with low sugar concentrations. Commercial wine yeasts were differentiated by their fermentative performances as well as their low acetate production (Camarasa *et al.*, 2011). West African population shared low-performance alleles conferring unique phenotypes regarding mitotic proliferation under different stress-resistance environments. Other phenotypes differentiated lineages from Malaysia, North America and Europe, in which the frequency of population-specific traits could be mapped onto a corresponding population genomics tree based on low-coverage genome sequence data (Warringer *et al.*, 2011).

The global genetic architecture underlying phenotypic variation arising from populations adapting to different niches is very complex. Most phenotypic traits of interest in *S. cerevisiae* strains are quantitative, controlled by multiple genetic loci referred to as quantitative trait loci (QTLs). Genome regions associated with a given trait can be detected by QTL analysis, using pedigree information or known population structure to make specific crosses for particular phenotypes. The crosses are then genotyped using single nucleotide polymorphisms (SNPs) or other markers across the whole genome and statistical associations of the linkage disequilibrium between genotype and phenotype are identified (Borneman *et al.*, 2013; Dequin and Casaregola, 2011; Liti and Louis, 2012; Salinas *et al.*, 2012; Swinnen *et al.*, 2012). QTL mapping was successfully applied to dissect phenotypes that are relevant in winemaking, such as fermentation traits (Ambroset *et al.*, 2011) or aromatic compounds production (Katou *et al.*, 2009; Steyer *et al.*, 2012). QTLs that were relevant for oenological traits and wine metabolites were mapped to genes related to mitochondrial metabolism, sugar transport and nitrogen metabolism. Strong epistatic interactions were shown to occur between genes involved in succinic acid production (Salinas *et al.*, 2012). The genotype–phenotype landscape has also been explored by several studies using statistical and probabilistic models (MacDonald and Beiko, 2010; Mehmood *et al.*, 2011; O'Connor and Mundy, 2009), as well as gene knockout approaches (Hillenmeyer *et al.*, 2008).

Current methods to infer genomic variation and determine relationships between *S. cerevisiae* strains include microsatellite analyses (Franco-Duarte *et al.*, 2009; Legras *et al.*, 2005; Muller and McCusker, 2009; Richards *et al.*, 2009), detection of genetic alterations using comparative genome hybridization (aCGH) (Carreto *et al.*, 2008; Dunn *et al.*, 2012; Kvitek *et al.*, 2008; Winzeler *et al.*, 2003) and SNPs detection by tiling arrays (Schacherer *et al.*, 2009).

DOI: 10.1002/yea

Within our previous work (Franco-Duarte *et al*., 2009) we evaluated the phenotypic and genetic variability of 103 *S. cerevisiae* strains that were isolated from vineyards of the Vinho Verde wine region (north-west Portugal). We used a set of 11 polymorphic microsatellite loci and, through subgroup discovery-based data mining, successfully identified strains with similar genetic characteristics (microsatellite alleles) that exhibited similar, mostly taxonomic phenotypes, allowing us also to make predictions about the phenotypic traits of strains. Within this study, we aim to investigate whether such computational associations can be established in a larger collection of 172 diverse *S. cerevisiae* strains obtained from worldwide geographical origins and distinct technological uses (winemaking, brewing, bakery, distillery, laboratory, natural, etc.). In the study we use 30 physiological traits, most of them being important from an oenological point of view.

Materials and methods

Strain collection and phenotypic characterization

The *S. cerevisiae* strain collection used in this work consists of 172 strains of different geographical origins and technological applications or origins (see supporting information, Table S1, strains Z1–Z187). The collection includes strains used for winemaking (commercial and natural isolates that were obtained from winemaking environments), brewing, bakery, distillery (sake, cachaça) and ethanol production, laboratory strains and also strains from particular environments (e.g. pathogenic strains, isolates from fruits, soil and oak exudates). The collection further includes a set of sequenced strains (Liti *et al*., 2009). All strains were stored at –80ºC in cryotubes containing 1 ml glycerol (30% v/v).

Phenotypic screening was performed considering a wide range of physiological traits that are also important from an oenological point of view. In a first set of phenotypic tests, strains were inoculated into replicate wells of 96-well microplates. Isolates were grown overnight in YPD medium (yeast extract 1% w/v, peptone 1% w/v, glucose 2% w/v) and the optical density (A_{640}) was then determined and adjusted to 1.0. After washing with peptone water (1% w/v), 15 µl of this suspension were inoculated in quadruplicate in microplate wells containing 135 µl white grape must of the variety Loureiro, supplemented with the compounds mentioned below. The initial cellular density was 5×10^6 cells/ml ($A_{640} = 0.1$) and the final optical density was determined in a microplate spectrophotometer after 22 h of incubation (30ºC, 200 rpm). All microplates were carefully sealed with parafilm, and no evaporation was observed for incubation temperatures of 30ºC and 40ºC. As summarized in Table S2 (see supporting information), this approach included the following tests: growth at various temperatures (18ºC, 30ºC and 40ºC), evaluation of ethanol resistance (6%, 10% and 14% v/v) and tolerance to several stress conditions caused by extreme pH values (2 and 8), osmotic/saline stress (0.75 M KCl and 1.5 M NaCl). Growth was also assessed in the presence of potassium bisulphite ($KHSO_3$, 150 and 300 mg/l), copper sulphate ($CuSO_4$, 5 mM), sodium dodecyl sulphate (SDS, 0.01% w/v), the fungicides iprodion (0.05 and 0.1 mg/ml) and procymidon (0.05 and 0.1 mg/ml), as well as cycloheximide (0.05 and 0.1 mg/ml). The growth in finished wines was determined by adding glucose (0.5 and 1% w/v) to a commercial white wine (12.5% v/v alcohol). Galactosidase activity was evaluated by adding galactose (5% w/v) to Yeast Nitrogen Base (YNB, Difco™, cat. no. 239210), using test tubes with 5 ml culture medium and the same initial cell concentration (5×10^6 cells/ml), followed by 5–6 days of incubation at 26ºC and subsequent visual evaluation of growth. Other tests were performed using solid media. Overnight cultures were prepared as previously described, adjusted to an optical density (A_{640}) of 10.0 and washed; 1 µl of this suspension was placed on the surface of the culture media mentioned below. Hydrogen sulphide production was evaluated using BiGGY medium (Sigma-Aldrich, cat. no. 73608) (Jiranek *et al*., 1995), followed by incubation at 27ºC for 3 days. The colony colour, which represents the amount of H_2S produced, was then analysed, attributing a score from 0 (no colour change) to 3 (dark brown colony). Ethanol resistance (12% v/v) and the combined resistance to ethanol (12%, 14%, 16% and 18% v/v) and sodium bisulphite ($Na_2S_2O_5$; 75 and 100 mg/l) was evaluated by adding the mentioned compounds to Malt Extract Agar (MEA; Sigma-Aldrich, cat. no. 38954) and growth

DOI: 10.1002/yea

was visually scored after incubation (2 days at 27ºC). All phenotypic results were assigned to a class between 0 and 3 before the statistical analysis (0, no growth in liquid media ($A_{640} = 0.1$) or no visible growth on solid media; 3, $A_{640} \geq 1.0$, extensive growth on solid media or a dark brown colony formed in the BiGGY medium; scores 1 and 2 corresponded to A_{640} of 0.2–0.4 and 0.5–1.0, respectively, and to intermediate values of growth and colour changes in solid medium and BiGGY medium), as shown in Table S2 (see supporting information).

Genetic characterization

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) for 36 h at 28ºC (160 rpm), DNA isolation was performed as previously described (Schuller *et al.*, 2004) and used for microsatellite analysis.

Genetic characterization was performed using 11 highly polymorphic *S. cerevisiae*-specific microsatellite loci: ScAAT1, ScAAT2, ScAAT3, ScAAT4, ScAAT5, ScAAT6, ScYPL009c, ScYOR267c, C4, C5 and C11 (Field and Wills, 1998; Legras *et al.*, 2005; Perez *et al.*, 2001; Schuller *et al.*, 2007, 2012; Techera *et al.*, 2001). Multiplex PCR mixtures and cycling conditions were optimized and performed in 96-well PCR plates, as previously described (Franco-Duarte *et al.*, 2009).

Data analysis

We have estimated the number of repeats for the alleles from each locus based on the genome sequence of strain S288c available in the *Saccharomyces* Genome Database (http://www.yeastgenome.org) and the results obtained for the size of microsatellite amplicons of this strain.

Principal component analysis (PCA), available in the The Unscrambler® X software (Camo), was used for microsatellite variability analysis. A set of standard predictive data-mining methods, as implemented in the Orange data mining suite (Demsar *et al.*, 2013), were used to study the relations between the genetic constitutions of strains and their geographical origins or technological applications. Alleles that were present in less than five strains were removed, and the k nearest-neighbour algorithm (kNN) (Tan *et al.*, 2006) was used for inference. The modelling approach was tested in five-fold cross-validation, each time fitting the model on 80% of the data and testing it on the remaining 20%. Results were reported in terms of cross-validated area under the receiver operating characteristics curve (AUC), which estimates the probability that the predictive model would correctly differentiate between distinct technological applications of the strains (Hanley and McNeil, 1982).

The strength of associations between microsatellites and specific phenotypes was scored using information gain ratio, as implemented in the Orange data-mining suite. Significant findings were confirmed by permutation tests and estimation of false-discovery rate. Data was first preprocessed to filter out features with only a single, constant value, in which the distribution was too skewed, or when more than 95% of strains shared the same value. This was done for both microsatellite and phenotypic data. The filtering procedure reduced our dataset to retain 40 of the initial 295 microsatellite features and 60 of the initial 83 phenotypic ones. We then considered the resulting dataset to test $40 \times 60 = 2400$ associations between microsatellites and phenotypes. Information gain (IG) (Quinlan, 1986), also popularly referred to as 'mutual information', is a measure of mutual dependence of two random variables. In the present study we used it to assess the influence of an independent variable, X, on a dependent class variable, Y. IG tells us how much information we gain about Y by knowing the value of X. If the class variable Y can take l distinct values, $y_1, y_2, \ldots, y_l$, we can define its entropy by:

$$H(Y) = \sum_{j=1}^{l} P\big(Y = y_j\big) * \log_2\big(P\big(Y = y_j\big)\big) \qquad (1)$$

Here, P is a probability measure. The entropy $H(Y)$ measures the unpredictability of a random variable Y that represents the amount of information required to answer the question, 'what is the value of Y?'. By knowing the value of independent variable X one can reduce this uncertainty if the dependent and independent variables are related. Suppose that $X = x_i$, where x_i is one of k distinct values $x_1, x_2, \ldots, x_k$ that variable X can take. By replacing the probability $P(Y = y_j)$ in equation 1, with conditional probability $P(Y = y_j | X = x_i)$, we define a conditional entropy $H(Y|X = x_i)$ of Y, assuming that

DOI: 10.1002/yea

the value of X is x_i. By knowing the distribution of X, i.e. by knowing the probabilities $P(X = x_i)$ for all $i = 1,\ldots, k$, we can define a conditional entropy of Y, given the variable X:

$$H(Y|X) = \sum_{i=1}^{k} H(Y|X = x_i) * P(X = x_i) \quad (2)$$

The reduction of uncertainty from $H(Y)$ to $H(Y|X)$ is called information gain and is defined as the difference $IG(X) = H(Y) - H(Y|X)$. If this difference is normalized by $H(X)$, the entropy of the variable X, the ratio is called information gain ratio (IGR). This score was first introduced in $\frac{IG(X)}{H(X)}$ Quinlan (1986) in order to avoid overestimation of multi-valued variables. IGR(X) ranges from 0, where knowing the value of X provides no information about Y, to 1 in cases where X and Y are perfectly correlated. To compute IGR we need to estimate the unconditional and conditional probabilities from the data; in the present work these probabilities were estimated with relative frequencies. For computation of IGR, Orange software (v. 2.7.1) was used. Each IGR estimate was compared to its null distribution, obtained from 100 000 computations of IGR for that particular feature combination on permuted data. We then tested the null hypothesis (IGR = 0) and obtained p values as proportions of permutation experiments where IGR $\geq$ the score obtained from original dataset. The permutation procedure was repeated for all microsatellite–phenotype pairs and the computed p values were corrected using the false-discovery rate procedure (FDR) (Benjamini and Hochberg, 1995). We here report on pairs of correlated microsatellites and phenotypic features with FDR < 0.2.

Results

Strain collection and genetic characterization

A *S. cerevisiae* collection was constituted, including 172 strains from different geographical origins and technological origins, as follows: wine and vine (74 isolates), commercial wine strains (47 isolates), other fermented beverages (12 isolates), other natural environments – soil woodland, plants and insects (12 isolates), clinical (nine isolates), sake (six isolates), bread (four isolates), laboratory (three isolates), beer (one isolate) and four isolates of unknown origin (see supporting information, Table S1).

All 172 strains were genetically characterized regarding allelic combinations for the previously described microsatellites ScAAT1, ScAAT2, ScAAT3, ScAAT4, ScAAT5, ScAAT6, ScYPL009c, ScYOR267c, C4, C5 and C11 (Field and Wills, 1998; Legras *et al.*, 2005; Perez *et al.*, 2001; Schuller and Casal, 2007; Schuller *et al.*, 2007, 2012; Techera *et al.*, 2001). As shown in Table 1, a total of 280 alleles was obtained; microsatellites ScAAT1 and ScAAT5 were the most and the least polymorphic, with 39 and 5 alleles, respectively. The genetic diversity of the collection is illustrated on the principal component analysis (PCA) plot in Figure 1. Some patterns of genetic relatedness between strains sharing the same technological origin became evident, as shown in Figure 1A. Sake strains (black dots) were located in the right part of the PCA plot, due to the larger sizes of alleles for loci ScYOR267c and C4. For this group of strains, we identified nine unique alleles, where three were present in more than one strain and belonged to three different loci (ScAAT6, C4 and ScYOR267c). Strains from fermented beverages other than wine were separated by PC-2, being located in the upper part of the PCA plot, indicating that they share a combination between smaller alleles of microsatellite C4 and bigger alleles of ScYOR267c. These 12 strains are marked in the PCA plot inside the area surrounded by a dotted line. Twelve unique alleles were found for these strains, two of them (C4-58 and ScYPL009c-57) being present in six of the 12 strains. On the contrary, the group of wine strains (both natural isolates and commercial strains) showed heterogeneous distribution across the two components, being preferentially located in the left side of the PCA plot. The nine clinical strains were distributed across both components, with no discriminant results in any locus. The 172 strains (scores) were also segregated in the first two components of the PCA constructed from the allelic combination for 11 loci. Loci ScYOR267c and C4 had the highest weight in strain variability, followed by ScYPL009c and ScAAT4, although within a smaller extent (Figure 1B).

To reveal the weight of different alleles on the genetic variability of the strains, the profile of the 11 microsatellites was represented for each strain as a vector where the values 0, 1 and 2 corresponded

DOI: 10.1002/yea

Table 1. Summary of the distribution of alleles (indicated in numbers of repetitions) among 172 *Saccharomyces cerevisiae* strains from 11 microsatellite loci

Microsatellite designation	Total number of alleles (range of allele sizes in number of repeats)	Most frequent alleles	Number of strains in which the allele was obtained	Most variable alleles (number of repetitions) identified by PCA (Figure 2)	Percentage of most variable alleles among the total number of alleles per locus	References*
ScAAT1	39 (6–54)	24	27	19	15	A, B
		16	21			
ScAAT2	18 (5–22)	15	58	7, 14, 15	28	
		16	33			
		14	34			
		13	21			
ScAAT3	19 (3–49)	16	45	11, 14, 16, 22	32	B, C
		14	32			
		22	28			
ScAAT4	17 (1–27)	20	100	7, 9, 11, 20	35	B
		11	22			
ScAAT5	6 (2–49)	9	80	8, 9, 10, 11	67	B
		10	63			
		8	37			
ScAAT6	10 (12–44)	16	124	14, 16, 17	50	B
		17	40			
C4	9 (16–61)	21	52	21, 24, 40	56	D
		24	44			
		22	31			
C5	19 (3–38)	4	31	3	16	D
		3	25			
		12	23			
		13	22			
C11	18 (1–47)	13	42	23	17	D
		14	24			
		24	28			
ScYPL009c	13 (57–86)	80	47	65, 80, 81	46	A, C
		81	45			
		82	28			
		79	23			
		65	20			
ScYOR267c	12 (37–100)	52	52	52, 56, 62	42	A, C
		56	24			

*A, Techera *et al.*, 2001; B, Perez *et al.*, 2001; C, Field and Wills, 1998; D, Legras *et al.*, 2005.

to the absence of an allele, the presence of a heterozygous allele and the presence of two copies of the allele, respectively. We assumed that all strains were diploid, because aneuploid loci were rarely detected ($< 3\%$). In addition, the DNA content of a representative set of homozygous strains corresponded to a diploid strain (flow-cytometry analysis, data not shown). A total of 48 160 data points were generated and the segregation of the 280 alleles in the two components of the PCA is shown in Figure 2. Alleles ScAAT4-20, ScAAT5-9 and ScAAT6-16 have the highest weight in strain variability, due to their positioning in the right and upper part of the PCA plot. Among the 11 microsatellite loci, 30 alleles were identified by PCA as contributing to the highest strain variability among 172 strains (Table 1). Loci ScAAT3, ScAAT4, and ScAAT5 were the ones with the higher number of variable alleles (four), in opposition to loci ScAAT1, C5 and C11 with 1 allele each.

Prediction of the technological group based on microsatellite alleles

We examined the relations between strains' technological groups and the corresponding genotypes

DOI: 10.1002/yea

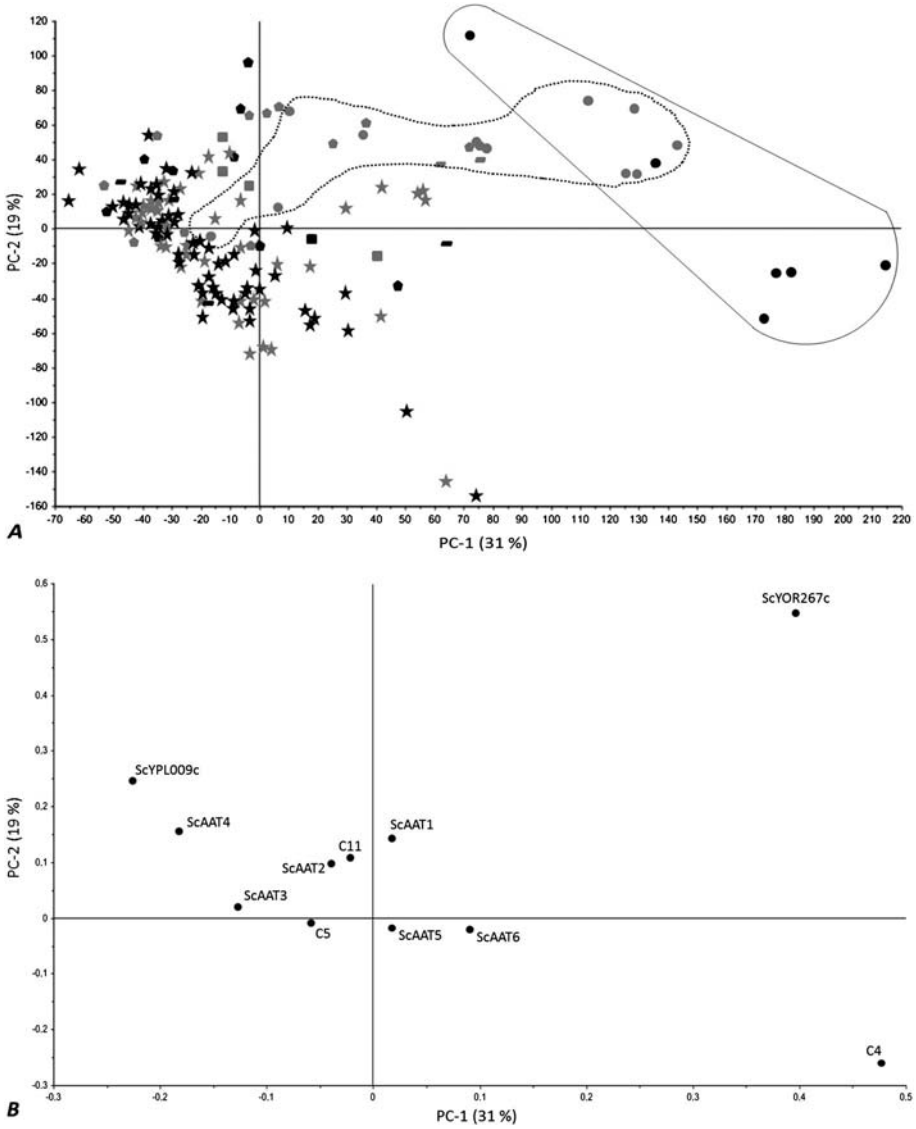


Figure 1. Principal component analysis of microsatellite data. (A) Distribution of 172 strains according to their allelic combinations for 11 loci (scores): symbols represent the strains technological applications or origin: ★, wine and vine; ★, commercial wine strain; ■, beer; ■, baker; ●, sake; ●, other fermented beverages; ⬟, clinical; ⬟, natural isolates; ▬, laboratory; ▬, unknown biological origin. Sake strains and strains from other fermented beverages are surrounded by unbroken and dotted lines, respectively. (B) Contribution of microsatellite loci (loadings) to the separation of strains shown in (A)

and scored them for their predictive value. Computational models were constructed to either predict the strains' technological application or origin from microsatellite data. Alleles that were present in less than five strains were removed, reducing the total number of alleles from 280 to 153. In 71% of the cases, the removed alleles were present in only one or two strains. The *k* nearest-neighbour (*k*NN) algorithm was used for inference, as implemented in the Orange data-mining software. A good prediction model was obtained in terms of both area under the receiver-operating-characteristics curve (AUC) (Hanley and McNeil, 1982) and classification accuracy (0.8018 and 0.547, respectively). Table 2 shows the confusion matrix of the *k*NN cross-validation classifications, where the report on averaged posterior AUCs estimated only on the test data that are not included in

DOI: 10.1002/yea

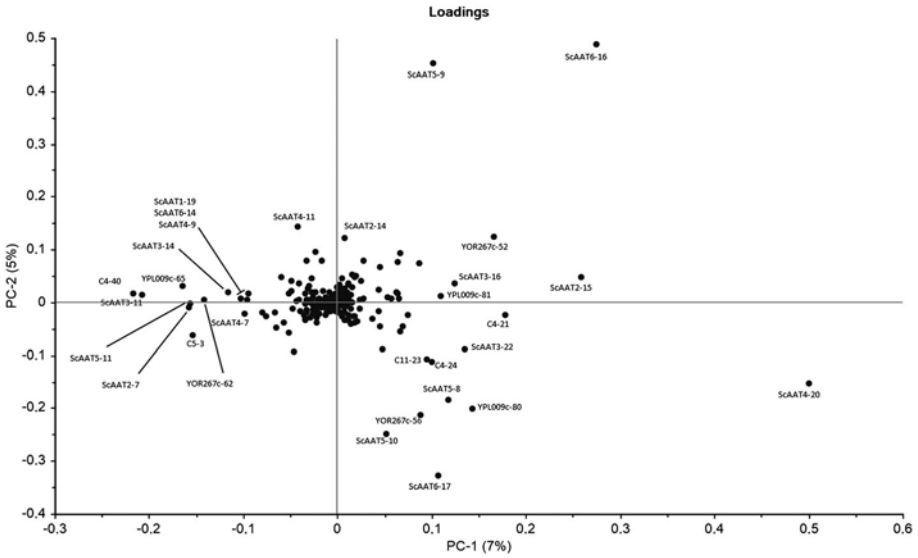


Figure 2. Principal component analysis of a Boolean matrix of 280 alleles from 11 microsatellites in 172 *Saccharomyces cerevisiae* strains

the training of the model. For the strains derived from winemaking environments (commercial and natural wine strains), 47% and 72% of strains, respectively, were correctly assigned. Interestingly, the majority of 'false' assignments did not fall out of the wine strains group, occurring for commercial wine strains that were assigned to the natural wine strains (21 of 47 strains) or natural wine strains that were catalogued as commercial wine strains (16 of 74 strains). If all wine strains were grouped in one single category, the proportion of correct assignments would increase to 93% (112 of 121 strains). For the groups of strains isolated from sake, natural environments, other fermented beverages and bread, the proportion of correct assignments were 67%, 42%, 50% and 50%, respectively, which is rather high considering the relatively small number of isolates included in these groups (6, 12, 12 and 4, respectively). The high number of correct assignments, even for small groups of strains, and a very high AUC score both reinforce the validity of the modelling technique, confirming a strong relation between our genotype profiles and strain groups. On the other side, and with only 22% of correct assignments, our approach was not successful in identification of clinical strains, which was expected due the absence of a common ancestor for this group, and because pathogenic *S. cerevisiae* strains arise from different origins (Liti and Schacherer, 2011).

Associations between microsatellites and phenotypes

The 172 *S. cerevisiae* strains were characterized phenotypically, considering 30 physiological traits that are important from an oenological point of view, in four replicates, measuring A_{640} after 22 h of growth. A high reproducibility was obtained between the four replicates, with the average standard deviation (SD) = 0.08. Results were catalogued with a number between 0 and 3 [0, no growth in liquid media ($A_{640} = 0.1$) or no visible growth on solid media or no colour change of the BiGGY medium; 3, at least 1.5-fold increase of A_{640}, extensive growth on solid media or a dark brown colony formed in the BiGGY medium; scores 1 and 2 corresponded to the respective intermediate values], resulting in a total of 5160 data points, as summarized in Table S2 (see supporting information). Our objective was to identify subsets of strains sharing similar phenotypic results and allelic combinations. To test the associations between phenotypic results and microsatellite alleles, we analysed pairwise relationships between corresponding variables (each microsatellite variable vs each phenotypic feature). First we binarized all phenotypic features in order to analyse the relationship more precisely (which phenotypic value is associated with a certain microsatellite), then the constant features (shared by $> 95\%$ of strains)

DOI: 10.1002/yea

Table 2. Confusion matrix indicating the technological origin prediction of 172 strains, obtained with *k*-nearest neighbor algorithm (kNN) applied to microsatellite data, in comparison with their real technological origins

Real technological application or origin	Total number of strains	Predicted technological application or origin									
		Wine and vine	Commercial (wine)	Natural	Other fermented beverages	Clinical	Sake	Bread	Unknown	Laboratory	Beer
Wine and vine	74	53 (72%)	16	2	0	1	0	0	2	0	0
Commercial (wine)	47	21	22 (47%)	0	0	2	0	2	0	0	0
Natural	12	2	2	5 (42%)	2	0	0	0	0	0	1
Other fermented beverages	12	0	3	1	6 (50%)	0	1	1	0	0	0
Clinical	9	2	1	1	1	2 (22%)	0	2	0	0	0
Sake	6	0	0	0	2	0	4 (67%)	0	0	0	0
Bread	4	1	0	0	0	1	0	2 (50%)	0	0	0
Unknown	4	3	0	1	0	0	0	0	0 (0%)	0	0
Laboratory	3	1	0	0	2	0	0	0	0	0 (0%)	0
Beer	1	0	0	1	0	0	0	0	0	0	0 (0%)

AUC = 0.802; Classification accuracy = 0.547.

were removed. Information gain ratio (IGR) was computed between microsatellite predictor and binarized phenotypic response variable, and repeated again using permutated phenotypic data, as described in Materials and methods; *p* values were reported after correction using the false-discovery rate (FDR) procedure, and the pairs for which FDR was < 0.2 are marked in Figure 3. In Table S3 (see supporting information) the exact FDR-adjusted *p* values are shown for associations between all phenotypic and genetic data. Significant associations were obtained between microsatellites ScAAT1, ScAAT2, ScAAT5, ScAAT6, YPL009c, C4 and C5, and for 13 phenotypic classes. For the phenotypic classes in which significant associations were found with microsatellite alleles, between one and eight associations were found with a particular microsatellite allele (number following black circles). For nine phenotypic tests and classes, a single association was established: '40°C = 1', '40°C = 3', 'SDS (0.01% w/v) = 0', '$KHSO_3$ (150 mg/l) = 2', 'ethanol 10% v/v (liquid medium) = 0', 'ethanol 10% v/v (liquid medium) = 2', 'ethanol 10% v/v (liquid medium) = 3', 'ethanol 12% v/v + $Na_2S_2O_5$ 75 mg/l (solid medium) = 1' and 'wine supplemented with glucose 1% = 0'. The phenotypes with the highest number of allelic associations were '$KHSO_3$ (300 mg/l) = 3' and 'galactosidase activity = 1', with eight associated alleles each. In terms of microsatellite alleles, 22 alleles had an association with at least one phenotype. For two alleles, three significant associations were obtained (ScAAT2-13 and C4-21), being the highest number of associations with phenotypes (seven) found for microsatellites ScAAT1 and ScAAT2, in opposition to ScAAT5, ScAAT6 and YPL009c, with only three associations each established. These numbers are not related to the total number of alleles and the range of allele sizes shown in Table 2.

Discussion

In our previous work (Franco-Duarte *et al*., 2009) we developed a method to computationally associate the genotype and phenotype of 103 *S. cerevisiae* strains, mainly from the Vinho Verde winemaking region, using microsatellite data obtained with 11 polymorphic markers and phenotypic data from a set of 24 taxonomic tests. Herein, we aimed to

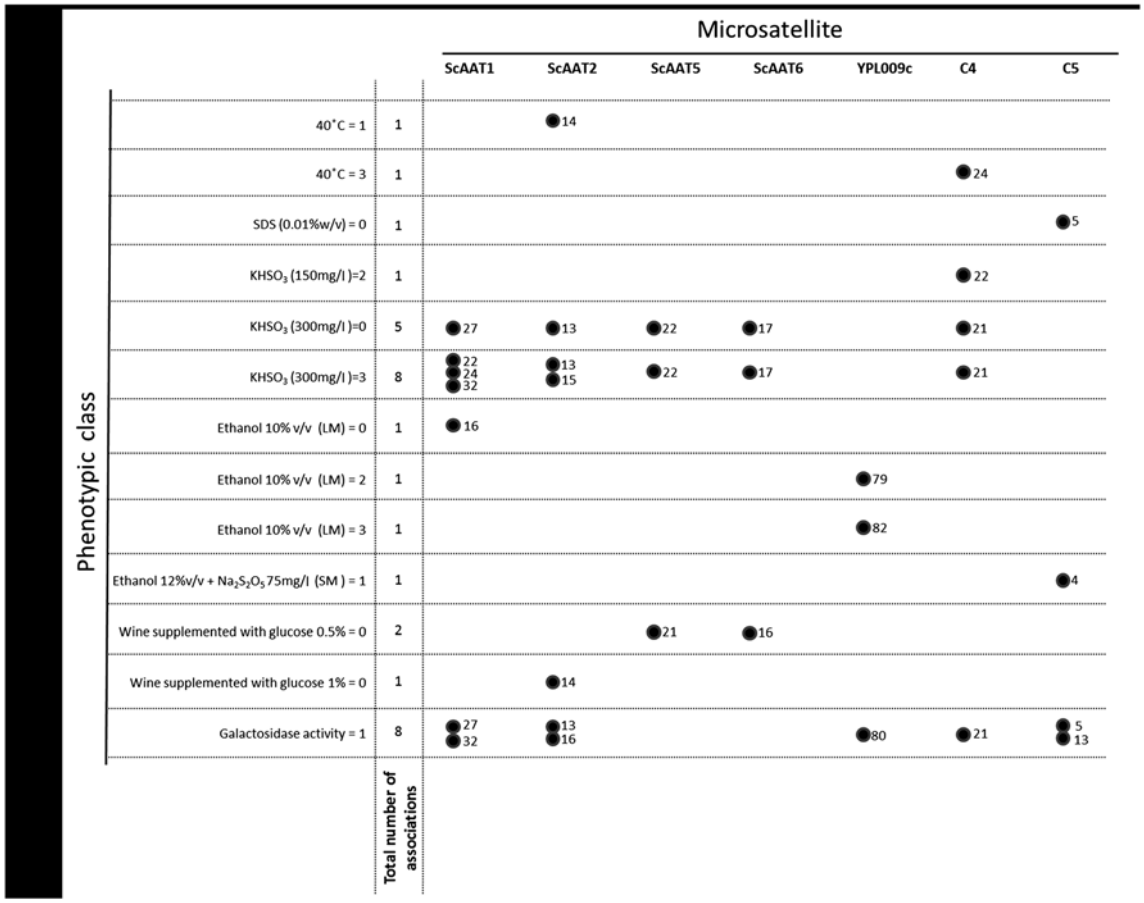


Figure 3. Significant associations (black circles) between microsatellites and phenotypes, obtained with Orange data-mining software. Each association was calculated between a microsatellite allele (numbers following black circles) of the microsatellite represented at the top, and a phenotypic class (0–3). Marked associations refer to significant *p* values obtained after false-discovery rate correction (*p* value after FDR < 0.2), using information gain ratio associations compared against data from permutation test (for details, see Materials and methods)

investigate whether such associations could be established in a worldwide collection of 172 *S. cerevisiae* strains from different geographical origins and technological uses (winemaking, brewing, bakery, distillery, laboratory, natural, etc.). We considered 30 physiological traits that are mainly used in *S. cerevisiae* winemaking strain selection programmes (Mannazzu *et al.*, 2002). Phenotypic analysis revealed a high diversity, similar to other studies that showed high diversity within domesticated and natural populations of *S. cerevisiae*, describing also mosaic strains, depending on their origin and application (Agnolucci *et al.*, 2007; Brandolini *et al.*, 2002; Camarasa *et al.*, 2011; Goddard *et al.*, 2010; Kvitek *et al.*, 2008; Liti *et al.*, 2009; Salinas *et al.*, 2010; Schacherer *et al.*, 2009; Warringer *et al.*, 2011). In addition, we showed significant associations between phenotypic results and strains' technological applications or origins using the Mann–Whitney test (Mendes *et al.*, 2013). Part of the high phenotypic variability and intrastrain variation can also be explained by the existence of genetic rearrangements that are characteristic for *S. cerevisiae*, being particularly high in the case of winemaking strains (Schuller *et al.*, 2007). Large-scale genome sequencing projects are now under way to provide data for an in-depth understanding of relationships between genotype and phenotype.

The collection of 172 *S. cerevisiae* strains obtained from different geographical origins and technological groups also revealed high genetic diversity (Figures 1, 2, Table 1), with a total of 280 alleles obtained with 11 polymorphic

DOI: 10.1002/yea

microsatellites. PCA components of Figure 2 explain only a small part of the total variance (PC-1, 7%; PC-2, 5%), which seems to indicate that all the microsatellite alleles are important to differentiate between strains, but also revealed a group of 54 alleles that are the most relevant to explain variability among strains. Microsatellite ScAAT1 was the most polymorphic one, with 39 alleles, followed by ScAAT3 and C5 with 19 alleles each, confirming the data of our previous study (Franco-Duarte *et al*., 2009). Herein we also observed some patterns of distribution according to the strains' technological applications or origins, when considering the PCA of genetic data, in particular for sake strains and strains from fermented beverages other than wine. Clinical strains, which are opportunistic environmental strains colonizing human tissues (Muller and McCusker, 2009; Schacherer *et al*., 2007), did not show any discriminant distribution with PCA, which was expected because they do not share a common ancestor (Liti and Schacherer, 2011). Sake strains and strains obtained from fermented beverages other than wine showed some unique alleles in loci ScAAT6, C4, ScYOR267c and ScAAT1, ScAAT5, ScAAT6, C4, ScYPL009c, ScYOR267c, respectively. These results highlight the existence of alleles that are representative of a specific technological group, which justifies the approach used in this research.

Regarding microsatellite distributions in human populations (5795 individuals and 645 microsatellite loci), multidimensional scaling detected 240 intrapopulation and 92 interpopulation pairs regarding genetic and geographical relatedness (Pemberton *et al*., 2013). In our study we demonstrate that a strain's allelic combination and the respective technological application or origin (Table 2) are strongly related, as the latter can be predicted from the proposed genotypic characterization. Regarding winemaking strains (both natural and commercial), the approach was able to predict the technological application or origin for 93% of the strains. The AUC score of the model was 0.802, between the values of an arbitrary and perfect classification (AUC = 0.5 and 1.0, respectively) and can be considered as moderately high (Mozina *et al*., 2004). These results demonstrate the potential of the approach to predict the technological origin of a strain from the entire microsatellite profile, even for groups of strains with small sample size (sake or bread, six and four strains, respectively).

The genetic and phenotypic profile of strains obtained with 11 markers and 30 phenotypic tests was used to computationally score and rank genotype–phenotype associations. Associations were scored using information gain ratio (Quinlan, 1986) and significant results were shown in form of p value after the false-discovery rate procedure. Thirty-two associations, representing 13 phenotypic classes and 22 microsatellite alleles, were significantly established. The phenotypic classes with more associations were related to high capacity to resist to the presence of $KHSO_3$ during fermentation, and to galactosidase activity; these two phenotypes were associated with eight alleles each. These results are valuable to select strains that are resistant to sulphur dioxide, an antioxidant and bacteriostatic agent used in vinification (Beech and Thomas, 1985), and that were tested by the capacity of strains to grow in a medium supplemented with $KHSO_3$. The association between eight alleles and the strains' moderate galactosidase activity, although not directly related to winemaking, could be also a beneficial criterion to choose *S. cerevisiae* strains capable to hydrolyse galactose, an alternative to the use of glucose as carbon source, pointing to an improved evolutionary capacity of these strains. The most polymorphic locus, ScAAT1, also revealed the highest number of associations with phenotypes, but this was not observed for other polymorphic loci. Seven phenotype–genotype associations were found for each of the alleles ScAAT2–13 and C4-21, which can be considered as the most informative to predict strains biotechnological potential regarding the associated phenotypes.

The prediction of the technological group from allelic combinations and the presence of statistically significant associations between phenotypes and allele both demonstrate that computational approaches can be successfully used to relate genotype and phenotype of yeast strains. Microsatellite analysis revealed to be an efficient marker to evaluate genetic relatedness in yeasts and can be employed in the industry as a quick and cheap analysis. Although microsatellite analysis is the most accurate method for *S. cerevisiae* strain characterization, the 11 microsatellites are spread on only nine chromosomes and might provide for a rather coarse representation of a genotype. Taking into account that the discovered associations apply to smaller fraction of the genome, this study could

DOI: 10.1002/yea

be beneficially complemented with additional markers of other genomic regions. These findings may become particularly important for the simplification of strain selection programmes, by partially replacing phenotypic screens through a preliminary selection based on the strain's microsatellite allelic combinations.

Acknowledgement

Ricardo Franco-Duarte and Inês Mendes are the recipients of fellowships from the Portuguese Science Foundation (FCT; Grant Nos SFRH/BD/74798/2010 and SFRH/BD/48591/2008, respectively) and João Drumonde-Neves is the recipient of a fellowship from the Azores Government (Grant No. M3.1.2/F/006/2008; DRCT). Financial support was obtained from FEDER funds through the programme COMPETE and by national funds through FCT by Project Nos FCOMP-01-0124-008775 (PTDC/AGR-ALI/103392/2008) and PTDC/AGR-ALI/121062/2010. Lan Umek and Blaz Zupan acknowledge financial support from the Slovene Research Agency (Grant No. P2-0209). The authors would like also to thank all the researchers who kindly provided yeast strains: Gianni Liti, Institute of Genetics, UK; Laura Carreto, CESAM and Biology Department, Portugal; Goto Yamamoto, NRIB, Japan; Cletus Kurtzman, Microbial Properties Research, USA; Rogelio Brandao, Laboratório de Fisologia e Bioquímica de Microorganismos, Brazil; and Huseyin Erten, Cukurova University, Turkey.

References

Agnolucci M, Scarano S, Santoro S, *et al.* 2007. Genetic and phenotypic diversity of autochthonous *Saccharomyces* spp. strains associated to natural fermentation of 'Malvasia delle Lipari'. *Lett Appl Microbiol* **45**: 657–662.

Ambroset C, Petit C, Brion C, *et al.* 2011. Deciphering the molecular basis of wine yeast fermentation traits using a combined genetic and genomic approach. *G3 (Bethesda)* **1**: 263–281.

Beech W, Thomas S. 1985. Action antimicrobienne de l'anhydride sulfureux. *Bull l'OIV* **58**: 564–581.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* **57**: 289–300.

Borneman AR, Pretorius IS, Chambers PJ. 2013. Comparative genomics: a revolutionary tool for wine yeast strain development. *Curr Opin Biotechnol* **24**: 1–8.

Brandolini V, Tedeschi P, Capece A, *et al.* 2002. *Saccharomyces cerevisiae* wine strains differing in copper resistance exhibit different capability to reduce copper content in wine. *World J Microbiol Biotechnol* **18**: 499–503.

Camarasa C, Sanchez I, Brial P, *et al.* 2011. Phenotypic landscape of *Saccharomyces cerevisiae* during wine fermentation: evidence for origin-dependent metabolic traits. *PLoS One* **6**: e25147.

Carreto L, Eiriz MF, Gomes AC, *et al.* 2008. Comparative genomics of wild type yeast strains unveils important genome diversity. *BMC Genom* **9**: 524.

Demsar J, Curk T, Erjave A, *et al.* 2013. Orange: data mining toolbox in Python. *J Mach Learn Res* **14**: 2349–2353.

Dequin S, Casaregola S. 2011. The genomes of fermentative *Saccharomyces*. *C R Biol* **334**: 687–693.

Diezmann S, Dietrich FS. 2009. *Saccharomyces cerevisiae*: population divergence and resistance to oxidative stress in clinical, domesticated and wild isolates. *PLoS One* **4**: e5317.

Dunn B, Richter C, Kvitek DJ, *et al.* 2012. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Res* **22**: 908–924.

Fay JC, Benavides JA. 2005. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet* **1**: 66–71.

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. *Proc Natl Acad Sci U S A* **95**: 1647–1652.

Franco-Duarte R, Umek L, Zupan B, *et al.* 2009. Computational approaches for the genetic and phenotypic characterization of a *Saccharomyces cerevisiae* wine yeast collection. *Yeast* **26**: 675–692.

Goddard MR, Anfang N, Tang R, *et al.* 2010. A distinct population of *Saccharomyces cerevisiae* in New Zealand: evidence for local dispersal by insects and human-aided global dispersal in oak barrels. *Env Microbiol* **12**: 63–73.

Hanley JA, McNeil BJ. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**: 29–36.

Hillenmeyer ME, Fung E, Wildenhaim J, *et al.* 2008. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**: 362–365.

Jiranek V, Langridge P, Henschke PA. 1995. Validation of bismuth-containing indicator media for predicting H_2S-producing potential of *Saccharomyces cerevisiae* wine yeasts under enological conditions. *Am J Enol Vitic* **46**: 269–273.

Katou T, Namise M, Kitagaki H, *et al.* 2009. QTL mapping of sake brewing characteristics of yeast. *J Biosci Bioeng* **107**: 383–393.

Kvitek DJ, Will JL, Gasch AP. 2008. Variations in stress sensitivity and genomic expression in diverse *Saccharomyces cerevisiae* isolates. *PLoS Genet* **4**: 31–35.

Legras JL, Merdinoglu D, Cornuet JM, *et al.* 2007. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol Ecol* **16**: 2091–2102.

Legras JL, Ruh O, Merdinoglu D, *et al.* 2005. Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. *Int J Food Microbiol* **102**: 73–83.

Liti G, Carter DM, Moses AM, *et al.* 2009. Population genomics of domestic and wild yeasts. *Nature* **458**: 337–341.

Liti G, Louis EJ. 2012. Advances in quantitative trait analysis in yeast. *PLoS Genet* **8**: e1002912.

Liti G, Schacherer J. 2011. The rise of yeast population genomics. *C R Biol* **334**: 612–619.

MacDonald NJ, Beiko RG. 2010. Efficient learning of microbial genotype–phenotype association rules. *Bioinformatics* **26**: 1834–1840.

Mannazzu I, Clementi F, Ciani M. 2002. Strategies and criteria for the isolation and selection of autochthonous starter. In *Biodiversity and Biotechnology of Wine Yeasts*, Ciani M (ed.). Research Signpost: Trivandrum; 19–35.

Mehmood T, Martens H, Saebø S, *et al.* 2011. Mining for genotype–phenotype relations in *Saccharomyces* using partial least squares. *BMC Bioinformat* **12**: 318.

Mendes I, Franco-Duarte R, Umek L, *et al.* 2013. Computational models for prediction of yeast strain potential for winemaking from phenotypic profiles, Schacherer J (ed). *PLoS One* **8**: e66523.

Mozina M, Demsar J, Kattan M, *et al.* 2004. Nomograms for visualization of naive Bayesian classifier. *Lect Notes Comput Sci* **3202**: 337–348.

Muller LLH, McCusker JHJ. 2009. Microsatellite analysis of genetic diversity among clinical and nonclinical *Saccharomyces cerevisiae* isolates suggests heterozygote advantage in clinical environments. *Mol Ecol* **18**: 2779–2786.

O'Connor TD, Mundy NI. 2009. Genotype–phenotype associations: substitution models to detect evolutionary associations between phenotypic variables and genotypic evolutionary rate. *Bioinformatics* **25**: 94–100.

Pemberton TJ, DeGiorgio M, Rosenberg NA. 2013. Population structure in a comprehensive genomic data set on human microsatellite variation. *G3 Genet* **3**: 891–907.

Perez MA, Gallego FJ, Martinez I, *et al.* 2001. Detection, distribution and selection of microsatellites (SSRs) in the genome of the yeast *Saccharomyces cerevisiae* as molecular markers. *Lett Appl Microbiol* **33**: 461–466.

Quinlan J. 1986. Induction of decision trees. *Mach Learn* **1**: 81–106.

Richards KD, Goddard MR, Gardner RC. 2009. A database of microsatellite genotypes for *Saccharomyces cerevisiae*. *Antonie Van Leeuwenhoek* **96**: 355–359.

Roberts IN, Oliver SG. 2011. The yin and yang of yeast: biodiversity research and systems biology as complementary forces driving innovation in biotechnology. *Biotechnol Lett* **33**: 477–487.

Salinas F, Cubillos F, Soto D, *et al.* 2012. The genetic basis of natural variation in oenological traits in *Saccharomyces cerevisiae. PLoS One* **7**: e49640.

Salinas F, Mandakovic D, Urzua U, *et al.* 2010. Genomic and phenotypic comparison between similar wine yeast strains of *Saccharomyces cerevisiae* from different geographic origins. *J Appl Microbiol* **108**: 1850–1858.

Schacherer J, Ruderfer DM, Gresham D, *et al.* 2007. Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS One* **2**: e322.

Schacherer J, Shapiro JA, Ruderfer DM, *et al.* 2009. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**: 342–345.

Schuller D, Cardoso F, Sousa S, *et al.* 2012. Genetic diversity and population structure of *Saccharomyces cerevisiae* strains isolated from different grape varieties and winemaking regions. *PLoS One* **7**: e32507.

Schuller D, Casal M. 2007. The genetic structure of fermentative vineyard-associated *Saccharomyces cerevisiae* populations revealed by microsatellite analysis. *Antonie Van Leeuwenhoek* **91**: 137–150.

Schuller D, Pereira L, Alves H, *et al.* 2007. Genetic characterization of commercial *Saccharomyces cerevisiae* isolates recovered from vineyard environments. *Yeast* **24**: 625–636.

Schuller D, Valero E, Dequin S, *et al.* 2004. Survey of molecular methods for the typing of wine yeast strains. *FEMS Microbiol Lett* **231**: 19–26.

Sicard D, Legras JL. 2011. Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C R Biol* **334**: 229–236.

Steyer D, Ambroset C, Brion C, *et al.* 2012. QTL mapping of the production of wine aroma compounds by yeast. *BMC Genom* **13**: 573.

Swinnen S, Thevelein JM, Nevoigt E. 2012. Genetic mapping of quantitative phenotypic traits in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **12**: 215–227.

Tan P, Steinbach M, Kumar V. 2006. *Introduction to Data Mining*. Pearson Addison Wesley: Boston, MA.

Techera AG, Jubany S, Carrau FM, *et al.* 2001. Differentiation of industrial wine yeast strains using microsatellite markers. *Lett Appl Microbiol* **33**: 71–75.

Warringer J, Zorgo E, Cubillos F, *et al.* 2011. Trait variation in yeast is defined by population history. *PLoS Genet* **7**: e1002111.

Winzeler EA, Castillo-Davis CI, Oshiro G, *et al.* 2003. Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics* **163**: 79–89.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's website.

Table S1. Geographical origin and technological application/origin of the 172 *Saccharomyces cerevisiae* strains

Table S2. Number of strains belonging to different phenotypic classes, regarding values of optical density, growth patterns in solid media or colour change in BiGGY medium

Table S3. Statistical *p* values of associations between all the phenotypic classes and microsatellite alleles