Yeast Proteome Variations Reveal Different Adaptive Responses to Grape Must Fermentation

Mélisande Blein-Nicolas,[1] Warren Albertin,[2,3] Benoît Valot,[1] Philippe Marullo,[3,4] Delphine Sicard,[5] Christophe Giraud,[6] Sylvie Huet,[7] Aurélie Bourgais,[2] Christine Dillmann,[5] Dominique de Vienne,[5] and Michel Zivy*,[8]

[1]INRA, PAPPSO, UMR 0320/8120 Génétique Végétale, Gif sur Yvette, France

[2]CNRS, UMR 0320/8120 Génétique Végétale, Gif-sur-Yvette, France

[3]Université Bordeaux Ségalen, UR Œnologie, Villenave d'Ornon, France

[4]SARCO/LAFFORT Group, Bordeaux, France

[5]Université Paris-Sud, UMR 0320/8120 Génétique Végétale, Gif sur Yvette, France

[6]CMAP, Ecole Polytechnique, Palaiseau, France

[7]INRA, Jouy-en-Josas, France

[8]CNRS, PAPPSO, UMR 0320/8120 Génétique Végétale, Gif sur Yvette, France

***Corresponding author:** E-mail: zivy@moulon.inra.fr.

Associate editor: Csaba Pal

Abstract

***Saccharomyces cerevisiae* and *S. uvarum* are two domesticated species of the *Saccharomyces* sensu stricto clade that diverged around 100 Ma after whole-genome duplication. Both have retained many duplicated genes associated with glucose fermentation and are characterized by the ability to achieve grape must fermentation. Nevertheless, these two species differ for many other traits, indicating that they underwent different evolutionary histories. To determine how the evolutionary histories of *S. cerevisiae* and *S. uvarum* are mirrored on the proteome, we analyzed the genetic variability of the proteomes of domesticated strains of these two species by quantitative mass spectrometry. Overall, 445 proteins were quantified. Massive variations of protein abundances were found, that clearly differentiated the two species. Abundance variations in specific metabolic pathways could be related to phenotypic traits known to discriminate the two species. In addition, proteins encoded by duplicated genes were shown to be differently recruited in each species. Comparing the strain differentiation based on the proteome variability to those based on the phenotypic and genetic variations further revealed that the strains of *S. uvarum* and some strains of *S. cerevisiae* displayed similar fermentative performances despite strong proteomic and genomic differences. Altogether, these results indicate that the ability of *S. cerevisae* and *S. uvarum* to complete grape must fermentation arose through different evolutionary roads, involving different metabolic pathways and duplicated genes.**

Key words: **adaptive evolution, genetic diversity, genome duplication, glucose fermentation, shotgun proteomics.**

Article

Introduction

To exploit plant, animal, or microbe populations, humans have applied strong directional selection that has led different taxa to independently gain similar morphological and physiological characteristics (Zeder et al. 2006; Purugganan and Fuller 2009). The case of crops is quite emblematic of this phenomenon called convergent evolution (Arendt and Reznick 2008), because many of them evolved several distinctive traits, such as increase in fruit and grain size, collectively referred to as the domestication syndrome (Harlan 1992). In yeasts, several species have been associated with humans as early as 7000 BC for the production of bread and fermented beverages (McGovern et al. 1996; McGovern et al. 2004; Legras et al. 2007). Two of them were domesticated to produce fruit juice beverage (Fay and Benavides 2005; Sicard and Legras 2011), *Saccharomyces cerevisiae* and *S. uvarum* (synonym *S. bayanus var. uvarum*, Pulvirenti et al. 2000; Nguyen and Gaillardin 2005; Libkind et al. 2011; Nguyen et al. 2011). Although most yeasts are inhibited by low pH and high sugar content of grape must and by the subsequent nutrient depletion and ethanol enrichment in wine (Rainieri and Pretorius 2000), domesticated strains of *S. cerevisiae* and *S. uvarum* are able to complete the grape must fermentation (Masneuf-Pomarède et al. 2010). Whether this common characteristic is related to similar metabolic functioning is currently unknown.

Outside their common ability to achieve grape must fermentation, *S. cerevisiae* and *S. uvarum* differ in their habitat and in a number of phenotypic traits indicating that they underwent different evolutionary histories. They are not adapted to the same growth temperature, *S. uvarum* being more psychrophilic than *S. cerevisiae* (Kishimoto and Goto 1995; Naumov 1996; Belloch et al. 2008). This aptitude to

© The Author 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

grow at low temperatures involves a conserved mechanism of response to cold aiming at the maintenance of membrane fluidity by increasing the concentration of unsaturated fatty acids in membrane lipids (Schade et al. 2004; D'Amico et al. 2006). Moreover, *S. uvarum* is less tolerant to high ethanol concentrations and confers a peculiar aromatic intensity to wine (Tosi et al. 2009; Masneuf-Pomarède et al. 2010; Gamero et al. 2011). In particular, *S. uvarum* produces higher amounts of 2-phenylethanol (2-PE) (Tosi et al. 2009; Masneuf-Pomarède et al. 2010). This aromatic alcohol is produced from phenylalanine (L-PHE) (Vuralhan et al. 2003) through the Ehrlich pathway (reviewed in Hazelwood et al. 2008). These phenotypic divergences are believed to reflect the adaptation of these two yeast species to different ecological niches (Sampaio and Gonçalves 2008; Salvadó et al. 2011). Accordingly, *S. cerevisiae* is associated with any type of wine, as well as with beer and spirits, whereas *S. uvarum* is preferentially found in white wines (Naumov, Masneuf, et al. 2000; Naumova et al. 2002; Rementeria et al. 2003; Demuyter et al. 2004) and is the major yeast involved in cider-making (Naumov et al. 2001).

Saccharomyces cerevisiae and *S. uvarum* both belong to the *Saccharomyces sensu stricto* complex (Naumov, James, et al. 2000; Kellis et al. 2003) and diverged after a whole-genome duplication (WGD) that occurred approximately 100 Ma (Wolfe and Shields 1997; Dietrich et al. 2004; Dujon et al. 2004; Kellis et al. 2004). The two species are distantly related, with approximately 80% of nucleotide identity within coding regions, whereas *S. cerevisiae* and its closest relative *S. paradoxus* are 90% identical (Kellis et al. 2003). Several studies showed that duplicated genes from WGD and other small-scale duplications were directly associated with the potential of the *Saccharomyces sensu stricto* complex to ferment glucose and/or grow anaerobically (Wolfe and Shields 1997; Piškur 2001; Thomson et al. 2005; Conant and Wolfe 2007; Merico et al. 2007), as well as to produce high rates of ethanol (Blank et al. 2005). In *S. cerevisiae,* the persistence of duplicated genes in the genome, in particular for those involved in the glucose fermentation, has been shown to be mainly explained by functional divergence (Ihmels et al. 2004; Kuepfer et al. 2005) and dosage effects (Papp et al. 2004; Conant and Wolfe 2007). The synteny between the genomes of *S. cerevisiae* and *S. uvarum* is well conserved, with many of the duplicated genes retained in both species (Bon et al. 2000; Llorente et al. 2000; Fischer et al. 2001) and 82% of predicted protein-coding genes of *S. cerevisiae* having an ortholog in *S. uvarum* (Scannell et al. 2011).

In this study, we were interested in determining how the evolutionary histories of strains of *S. cerevisiae* and *S. uvarum* are mirrored on the proteome and on the evolution of duplicated genes. Because of their central role in metabolism, transport, and signaling, proteins are a particularly relevant class of molecular components to apprehend the biological complexity behind phenotypes (Gstaiger and Aebersold 2009). Recently, mass-spectrometry based proteomics has gained increasing popularity because of its applications in evolutionary biology. Sequence-related proteomics data can indeed be used to better annotate genomes (Gupta et al. 2008) and have also proven to be valuable to perform phylogenetic analyses, particularly in the case of nonsequenced organisms (Wynne et al. 2010) and when marker genes are not available for classification (Bohlin et al. 2010). Moreover, quantitative proteomics has been recently used to study interspecific changes (Schrimpf et al. 2009; Weiss et al. 2010). Here, we used quantitative proteomics to analyze the genetic variability of nine *S. cerevisiae* strains and six *S. uvarum* strains grown in wine-making conditions convenient for both species. The strains originated from diverse media and diverse geographical origins. We showed massive variations of protein abundances between *S. cerevisiae* and *S. uvarum*, which were related to particular metabolic pathways and to a differential recruitment of proteins encoded by duplicated genes.

Results

Nine strains from *S. cerevisiae* and six strains from *S. uvarum* were analyzed in this study. The *S. cerevisiae* strains were isolated from diverse media (brewery, distillery, enology, and oak exudate; table 1) to maximize the genetic diversity within this species (Liti et al. 2009). The *S. uvarum* strains, originating from grape must fermentation, cider fermentation, or grape berries, were chosen to represent a wide part of the genetic diversity of the *S. uvarum* species (Masneuf-Pomarède I, personal communication).

All the 15 strains were anaerobically grown in white grape must. Fermentation kinetics of each strain is shown in supplementary figure S1, Supplementary Material online. Yeast samples were harvested at 30% of CO_2 release for proteomics analyses (supplementary fig. S2, Supplementary Material online). During fermentation, four fermentation kinetics parameters were recorded as well as two life-history traits known to be correlated to the fermentation ability and to the concentration of alcoholic fermentation products such as trehalose (Albertin et al. 2011; Wang et al. 2011). Three independent fermentations per strain were performed.

Variations of Protein Abundances Are Large and Clearly Differentiate the Species

The proteome of the yeast strains was analyzed by shotgun label-free quantitative proteomics. Peptides were quantified by integrating precursor ion peak areas. The quantification measurements obtained for each peptide as well as detailed information on all the peptides and all the proteins identified in all liquid chromatography-tandem mass spectrometry (LC-MS/MS) runs were deposited online using PROTICdb database (Ferry-Dumazet et al. 2005; Langella et al. 2007) at the following URL: http://moulon.inra.fr/protic/public. Abundances of 445 proteins were estimated from both shared and proteotypic peptides as described in Blein-Nicolas et al. (2012) (supplementary table S2, Supplementary Material online). The proteins presented a large range of abundance variation, with between-strains max/min ratios varying between 1.2 and 234.3 and roughly distributed around three modes (1.9, 8.9, and 28) (fig. 1*A*). A total of 325 (73.0%) proteins presented significant abundance changes (P value < 0.01 taking into account multiple testing,

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

Table 1. Origin of *Saccharomyces cerevisiae* and *S. uvarum* Strains.

Species	Parental Strains	Monosporic Derivate	Collection/Supplier[a]	Isolation Origin	Area of Origin
S. uvarum	PM12	U1	ISVV[a]	Grape must fermentation	Jurançon, France
S. uvarum	PJP3	U2	ISVV[a]	Grape must fermentation	Sancerre, France
S. uvarum	Br6.2	U3	ADRIA NORMANDIE[a]	Cider fermentation	Normandie, France
S. uvarum	Br5.2	U5	ADRIA NORMANDIE[a]	Cider fermentation	Normandie, France
S. uvarum	Br20.1	U6	ADRIA NORMANDIE[a]	Cider fermentation	Normandie, France
S. uvarum	LC3	U7	ISVV[a]	Grape berries	Sancerre, France
S. cerevisiae	CLIB-382	B1	CIRM-Levures[a]	Brewery	Japan
S. cerevisiae	NRRL-Y-7327	B2	NRRL[a]	Brewery	Tibet
S. cerevisiae	CLIB-294	D1	CIRM-Levures[a]	Distillery	Cognac, France
S. cerevisiae	Alcotec 24	D2	Hambleton Bard[a]	Distillery	UK
S. cerevisiae	CLIB-328	E1	CIRM-Levures[a]	Enology	UK
S. cerevisiae	VL1	E3[b]	LAFFORT Oenologie[a]	Enology	Bordeaux, France
S. cerevisiae	F10	E4[b]	LAFFORT Oenologie[a]	Enology	Bordeaux, France
S. cerevisiae	VL3	E5	LAFFORT Oenologie[a]	Enology	Bordeaux, France
S. cerevisiae	YPS128	W1	SGRP[a]	Forest, oak exudate	Pennsylvania, USA

[a]ISVV, http://www.oenologie.u-bordeaux2.fr/; ADRIA NORMANDIE, http://www.adria-normandie.com; SGRP, http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html; LVBE, http://www.lvbe.uha.fr; CIRM-Levures, http://www.inra.fr/internet/Produits/cirmlevures; NRRL, http://nrrl.ncaur.usda.gov; Hambleton Bard, http://www.hambletonbard.com; LAFFORT Œnologie, http://www.laffort.com.

[b]E3 and E4 strains were previously referenced as GN and G-4A, respectively (Marullo et al. 2006, 2009).

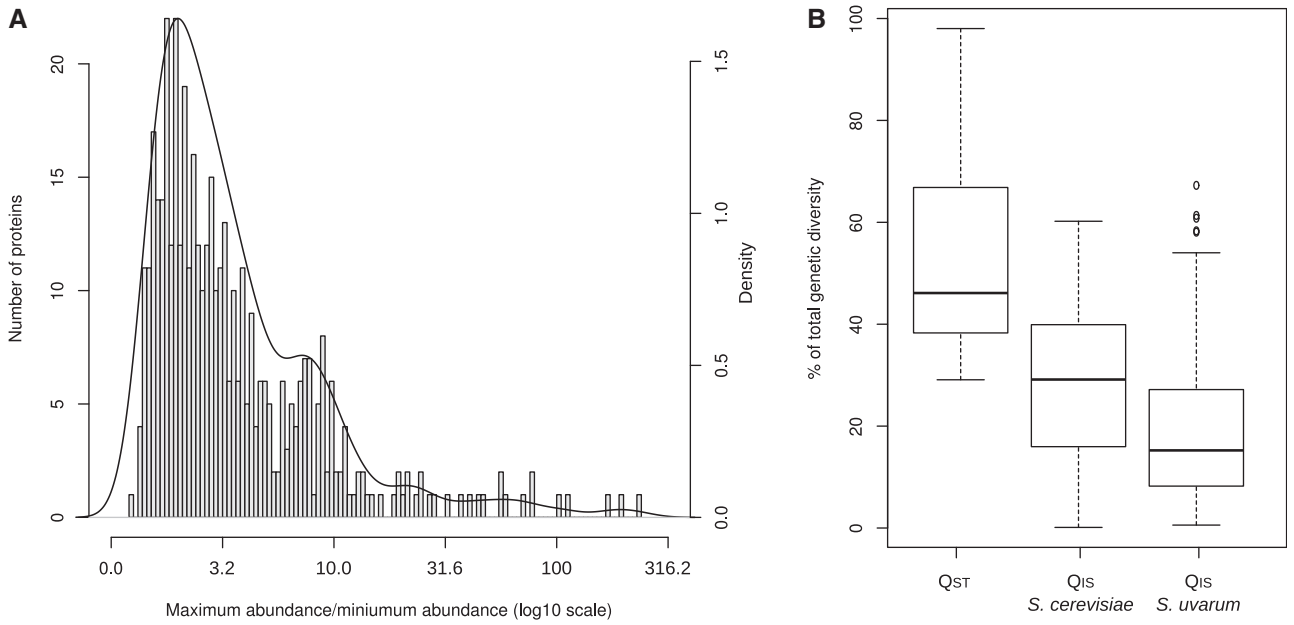

Fig. 1. Variations of protein abundances. (*A*) Distribution of maximum protein variations for 445 proteins quantified in at least four strains among nine *Saccharomyces cerevisiae* strains and six *S. uvarum* strains. Black line represents the density. (*B*) Box plot presenting the proportion of total genetic diversity explained by interspecies genetic diversity (Q_{ST}) and intraspecies genetic diversity (Q_{IS}) calculated for 401 proteins quantified in all the 15 strains (44 proteins were not considered because of missing data).

supplementary table S2, Supplementary Material online). Among them, 294 (66.1%) exhibited significant abundance changes between at least two strains, regardless of the species; 216 (50.0%) and 123 (28.5%) discriminated at least two strains within *S. cerevisiae* and within *S. uvarum*, respectively; 206 (51.4%) showed a significant difference between the two species. On average over all proteins, the degree of genetic differentiation in protein abundances was higher between species (Q_{ST} = 52.5%) than within species (Q_{IS} = 28.7% and 18.8% for *S. cerevisiae* and *S. uvarum*, respectively; fig. 1*B*). Moreover, the minimum Q_{ST} was 29.1%, which was greater than the average Q_{IS} values obtained for the two species. These results indicate that the variations of protein abundances were strongly structured according to the species (supplementary fig. S3, Supplementary Material online). This was confirmed by the heatmap based on the protein abundances showing that *S. cerevisiae* and *S. uvarum* clustered separately and were clearly isolated by two clusters of proteins (clusters A and D, fig. 2). The results also indicate that the proteome of the *S. cerevisiae* strains is more variable

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

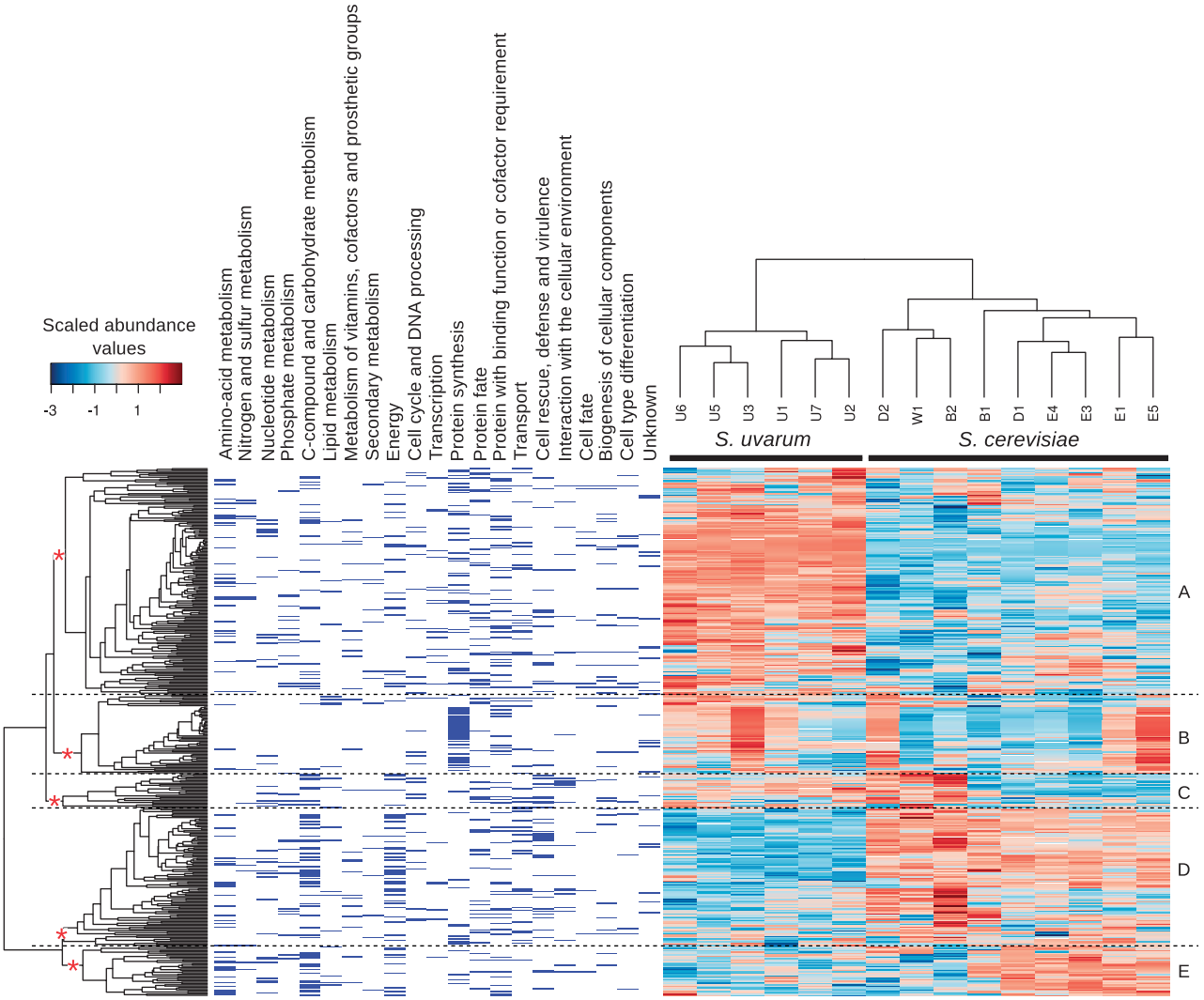

Fig. 2. Heatmap representation of the protein abundances. Each line corresponds to a protein and each column to a strain. Abundance values are indicated by the color key bar at the top left: low abundances are in blue and high abundances in red. Of 445 proteins, 401 proteins quantified in all the six *S. uvarum* strains and nine *S. cerevisiae* strains are represented (44 proteins were not included because of missing data). Membership of a protein to a functional category is shown on the left. Letters of the right indicate clusters of proteins exhibiting similar abundance patterns. These clusters were defined from the branches designated by a red star on the hierarchical clustering on the left. Note that cluster B contains mainly ribosomal proteins.

compared with the one of the *S. uvarum* strains (fig. 1*B*; supplementary fig. S3, Supplementary Material online). The *S. cerevisiae* strains clustered as two separate groups: one composed of the strains B2, D2, and W1 and the other one composed of the strains E1, E3, E4, E5, B1, and D1 (fig. 2). These two groups of strains showed opposite abundance patterns for two clusters of proteins (clusters C and E, fig. 2). To a lesser extent, within *S. uvarum*, cider strains were separated from wine strains.

Variations of Protein Abundances Are Related to Protein Functions

The 445 quantified proteins were classified in functional categories using the MIPS Functional Catalog (Ruepp et al. 2004; supplementary table S2, Supplementary Material online). Globally, all the functional categories contributed to the proteome differentiation between *S. cerevisiae* and *S. uvarum* (clusters A and D, fig. 2). Much of them also contributed to separate the *S. cerevisiae* strains B2, D2, and W1 from the other *S. cerevisiae* strains (clusters C and E, fig. 2). Remarkably, cluster B (fig. 2) was largely determined by ribosomal proteins that were particularly abundant in two genetically distant strains, U3 and E5.

Average protein abundances per functional category varied substantially between strains (fig. 3*A*). In particular, the three strains B2, D2, and W1 stood out from the other *S. cerevisiae* strains by exhibiting higher levels of proteins involved in cell rescue, defense and virulence, interaction with the cellular environment, and lipid metabolism (fig. 3*Ba*). Conversely, these three strains exhibited lower levels of proteins involved in amino acid, nitrogen, and sulfur metabolism (fig. 3*Ba*). In addition, the strains of *S. uvarum* presented higher levels of proteins involved in protein synthesis, amino acid, nitrogen, sulfur, and vitamin metabolism than *S. cerevisiae* (fig. 3*Bb*).

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

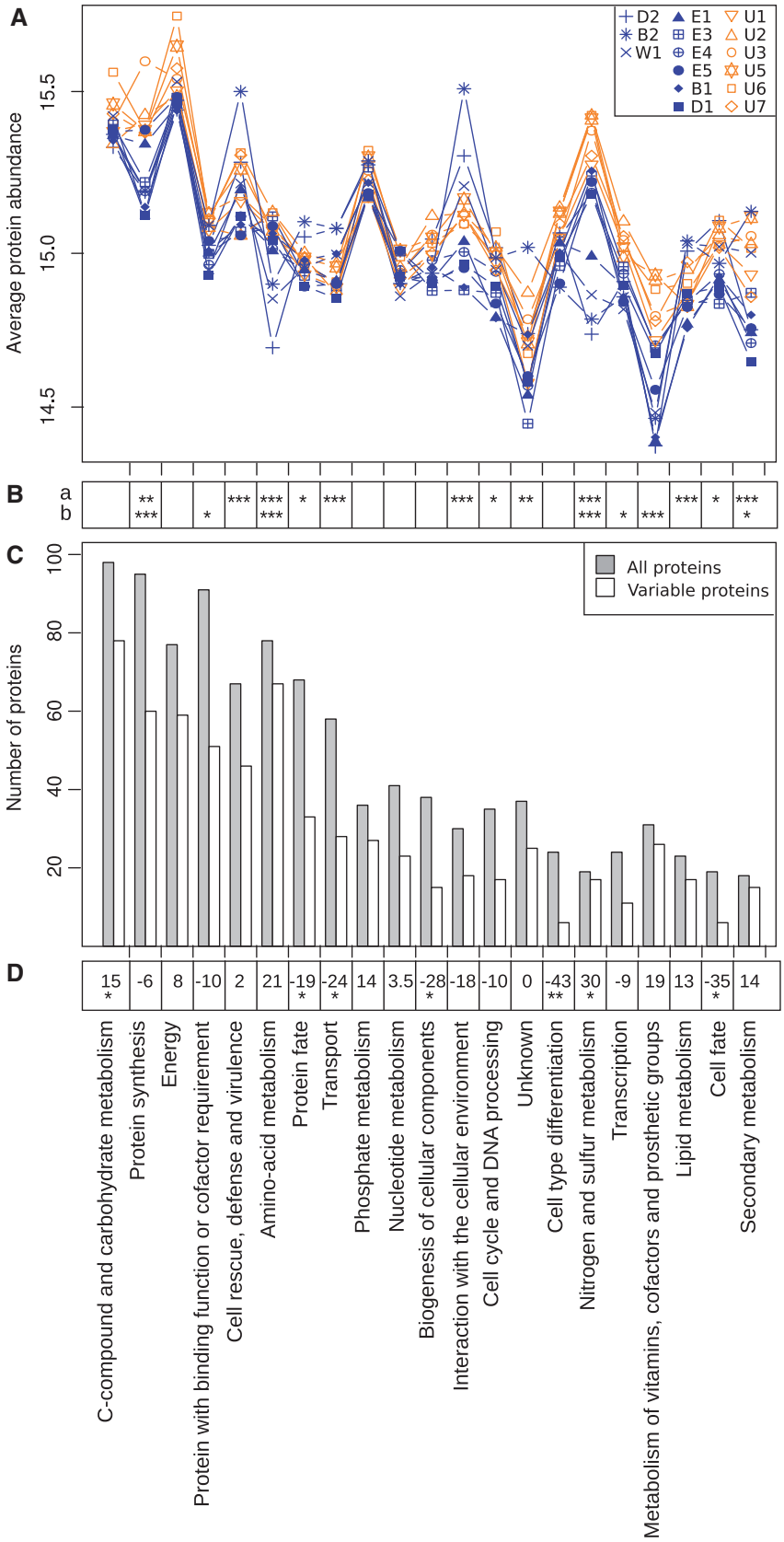

Fig. 3. Distribution of proteins by functional categories. (*A*) Average protein abundances in each functional category and each strain. (*B*) Significance of statistical tests comparing, for each functional category, average protein abundances between B2, D2, W1 and B1, D1, E1, E2, E3, and E5 (*a*) and between *S. cerevisiae* and *S. uvarum* (*b*). (*C*) Number of proteins in each functional category. One protein can belong to several functional categories. (*D*) Rate of enrichment (positive values) or depletion (negative values) for each functional categories of the set variable proteins compared with the whole set of quantified proteins. See Materials and Methods for the details of calculation. Asterisks indicate the significance of enrichment or depletion. *$5 \times 10^{-2} > P$ value $\geq 5 \times 10^{-3}$; **$5 \times 10^{-3} > P$ value $\geq 5 \times 10^{-4}$; ***$5 \times 10^{-4} > P$ value.

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

To better evaluate the contribution of the different functional categories to the genetic variation of protein abundance, we analyzed whether the proteins exhibiting significant abundance changes (called variable proteins hereafter) were significantly enriched (or depleted) for certain functional categories (fig. 3*C* and *D*). Variable proteins were enriched for C-compound, carbohydrate, nitrogen, and sulfur metabolism and depleted for protein fate, cell fate, transport, biogenesis of cellular components, and cell type differentiation (fig. 3*D*). It is worth noting that the number of variable proteins per functional category (fig. 3*C* and *D*) was not correlated to the genetic variation of the average protein abundance per category (fig. 3*A* and *B*). For example, carbohydrate metabolism was significantly enriched in variable proteins, but did not allow to discriminate the strains. Conversely, the category named "interaction with the cellular environment" was not significantly enriched in variable proteins, but clearly separated the D2–B2–W1 group from the other strains. These results illustrates the fact that, within a given functional category, the strains can be discriminated according to the average amount of proteins only if many proteins covary or if the variation of a major protein dominates the variations of the others.

To go further, we looked more closely at the proteins discriminating the B2–D2–W1 group from the other *S. cerevisiae* strains. The average abundances of 132 proteins were found to vary accordingly (supplementary table S2, Supplementary Material online). Consistent with the previous results, all but two proteins (LYS1 and ALD3) involved in amino acid biosynthesis were less abundant in B2, D2, and W1 than in other *S. cerevisiae* strains (fig. 4). In addition, B2, D2, and W1 exhibited more abundant proteins involved in stress response (HSP12, HSP26, HSP104, CTT1, TRX2, and SOD1), as well as in the metabolism of various sugars (HXK1, GLK1, SEC53, SUC2, and GRE3) and storage compounds such as glycerol (GPD1, DAK1, and GCY1), trehalose (TPS1), and glycogen (PGM2 and GPH1). Finally, several proteins of the purine biosynthetic pathway were less abundant in B2, D2, and W1.

We also examined the 206 proteins significantly varying between *S. cerevisiae* and *S. uvarum* (fig. 5). Three of them were involved in the biosynthesis of ergosterol (ERG6 and ERG20) and fatty acids (FAS2). These lipid compounds are important constituents of fungal membranes and are involved in a variety of cellular processes (reviewed in Parks et al. 1995; Tehlivets et al. 2007). They were more particularly shown to be implicated in the maintenance of membrane fluidity during cold stress (Nakagawa et al. 2002; Rodriguez-vargas et al. 2002; Schade et al. 2004). Interestingly, ERG6, ERG20, and FAS2 were more abundant in *S. uvarum*, in agreement with its psychrophilic aptitude (Kishimoto and Goto 1995; Naumov 1996; Belloch et al. 2008). Several proteins of the L-PHE biosynthesis and of the Ehrlich pathway, such as ARO4, ARO8, ARO10, and PDC5, were also more abundant in *S. uvarum*. This result is in agreement with the difference of 2-PE production between *S. uvarum* and *S. cerevisiae* (Antonelli et al. 1999; Tosi et al. 2009; Masneuf-Pomarède et al. 2010). The higher abundance of some proteins involved in the Ehrlich pathway may also modify the production of other higher alcohols such as *n*-propanol and amyl-alcohol, which involve the degradation of the branched-chain amino acids leucine, isoleucine, and valine. However, studies comparing strains of *S. cerevisiae* and *S. uvarum* in a wine context did not show clear differences in the production of higher alcohols other than 2-PE (Bertolini et al. 1996; Antonelli et al. 1999; Torriani et al. 1999). Several other metabolic pathways appeared to be particularly affected by interspecific variations. For example, proteins involved in the biosynthesis of sulfur amino acids were globally more abundant in *S. uvarum*, suggesting that these amino acids are produced in higher amounts in *S. uvarum* than in *S. cerevisiae*. Similarly, proteins of thiamine biosynthesis were more abundant in *S. uvarum* (supplementary fig. S4, Supplementary Material online). In nucleotide metabolism, enzymes at the beginning of the purine biosynthesis pathway were more abundant in *S. uvarum*, whereas those at the end were more abundant in *S. cerevisiae*. In carbohydrate metabolism, the first enzymes of the tricarboxylic acid cycle (CIT1, ACO1, IDH1, and IDH2) were more abundant in *S. cerevisiae*. In the glucose fermentation pathway, no tendency toward one or the other species was observed. Remarkably, isozymes (products of duplicated genes catalyzing the same reaction) in the glucose fermentation pathway did not covary. This was particularly striking within two isozyme families (PFK1/PFK2 and TDH1/TDH2/TDH3), where the most abundant isozyme was not the same in *S. cerevisiae* and *S. uvarum*. This result suggests that the two species use different sets of isozymes to ferment glucose.

Altogether, these results show that three groups of strains could be distinguished from marked quantitative differences of proteins related to functional differences: *S. cerevisiae* strains contrasted with *S. uvarum* strains, and within *S. cerevisiae*, B2, D2, and W1 contrasted with B1, D1, E1, E2, E3, and E5.

Duplicated Genes Are Involved in the Proteome Differentiation between Species

To know whether the isozyme specialization observed for the glucose fermentation pathway is a general result, we analyzed the abundance patterns of 35 isozymes encoded by duplicated genes and belonging to 17 families (table 2). For 15 of these families (88.2%), the isozymes were located in different protein clusters (fig. 2), indicating that they had different abundance patterns over the strains (table 2). For seven families (41.2%), the isozymes exhibited opposite abundance patterns in the two species, with one isozyme more abundant in *S. uvarum* (cluster A) and the other more abundant in *S. cerevisiae* (cluster D). Altogether, these results illustrate the fate of duplicated gene products in the process of species divergence. They may indicate that duplicates encoding isozymes gain specific expression patterns.

Overall Variations of Protein Abundances Are Consistent with Genetic Differences and to a Lesser Extent with Phenotypic Differences

We wondered whether the differences observed between the strains at the proteome level were related to the fermentative

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

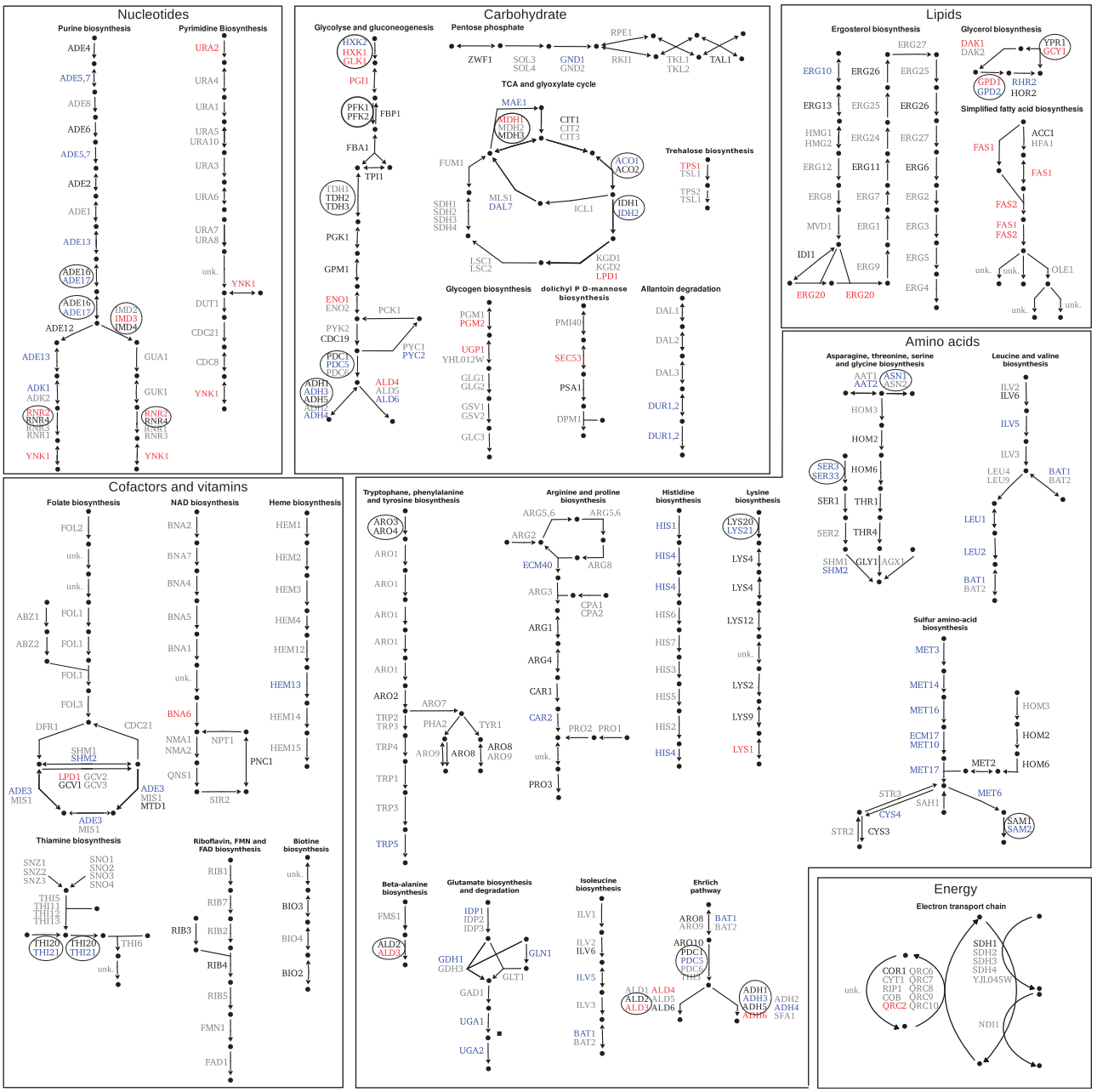

Fig. 4. Metabolic map of the proteins exhibiting significant abundance changes between two groups of *Saccharomyces cerevisiae* strains. Proteins that are significantly more and less abundant in the strains B2, D2, and W1 compared with E1, E2, E3, E5, B1, and D1 are represented in red and blue, respectively. The proteins exhibiting no significant abundance changes are shown in black; the proteins not quantified are shown in gray. Black circles indicate proteins encoded by duplicated genes.

phenotype and/or whether they reflected genetic differences acquired during the evolution. To analyze these relationships, we compared the strain differentiation based on proteome variability to those based on phenotypic variability and on genetic sequence polymorphism. Six phenotypic traits were measured during the fermentations: four fermentation traits (lag-phase duration, time to complete 30%, 50%, and 100% of the fermentation) and two life-history traits (population size and cell size at 30% of CO_2 release) known to be correlated to the fermentation ability and to the concentration of alcoholic fermentation products (Albertin et al. 2011, 2013; Wang et al. 2011). The genetic variability among the strains was assessed from 498 single nucleotide polymorphisms (SNPs) identified in six genes (*ACC1*, *ALA1*, *ADP1*, *GLN4*, *VSP13*, and *RPN2*) known to reveal sequence diversity among wine *S. cerevisiae* strains (Muñoz et al. 2009) and 2,681 peptides exhibiting qualitative variations (detected or not detected) assumed to result from single amino acid polymorphisms (SAPs). Because the individual analysis of SNPs and SAPs gave similar results (supplementary fig. S5, Supplementary Material online), we analyzed them jointly.

Based on phenotypic variations, two groups of strains could be distinguished according to their fermentative ability and cell size (fig. 6*A* and *B*). The first one contained slow

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

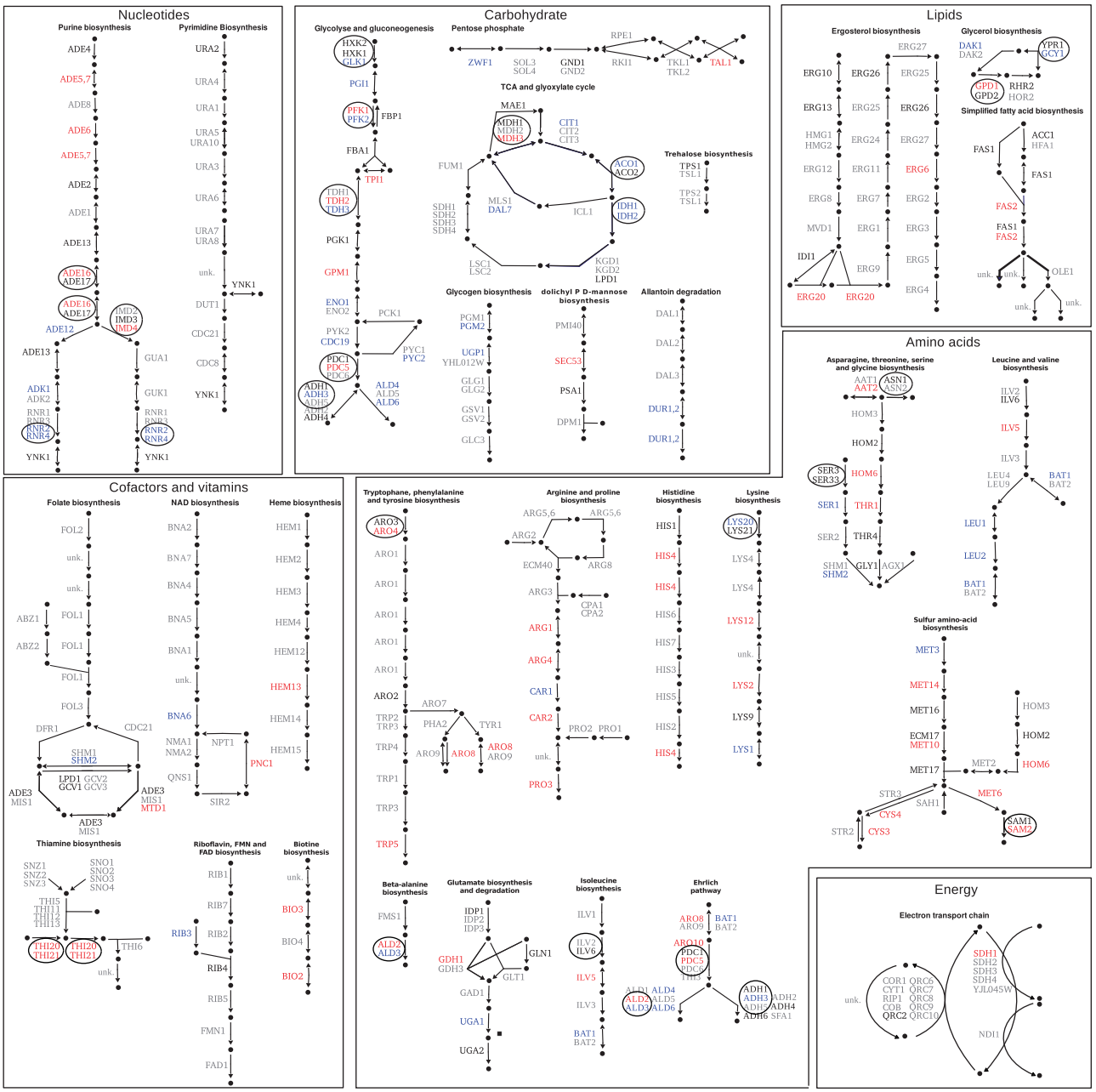

Fig. 5. Metabolic map of the proteins exhibiting significant abundance changes between *Saccharomyces cerevisiae* and *S. uvarum* strains. Proteins that are significantly more and less abundant in six *S. uvarum* strains compared with nine *S. cerevisiae* strains are represented in red and blue, respectively. The proteins exhibiting no significant abundance changes are shown in black; the proteins not quantified are shown in gray. Black circles indicate proteins encoded by duplicated genes.

fermenting strains with large cells (B1, B2, D2, W1, and E1). Among them, B2 and D2 exhibited particularly sluggish or stuck fermentations. The second one contained strains with small cells and large populations that fermented rapidly (D1, E3, E4, E5, and all the *S. uvarum* strains). Within this group, wine and cider strains of *S. uvarum* showed no difference in fermentative ability. Moreover, the *S. uvarum* strains exhibited larger population sizes than *S. cerevisiae* strains. The phenotypic variation between bad fermenting strains was higher than between well fermenting strains.

Although the phenotypic analysis grouped *S. uvarum* strains with *S. cerevisiae* strains exhibiting good fermentative performances, the sequence polymorphisms and the variations of protein abundances clearly separated the two species (fig. 6*C*–*F*). According to both analyses, the *S. cerevisiae* strains originating from enology (E strains) were closer to each other compared with the nonenological strains of the same species, including in particular B2, D2, and W1. Within *S. uvarum*, protein abundances separated wine and cider strains while sequences did not allow the observation of any structure.

Altogether, these results show that *S. cerevisiae* and *S. uvarum* reached similar fermentative performances despite strong genomic and proteomic differences, indicating that different metabolic roads have been selected for fermenting

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

Table 2. Distribution of Isozymes in the Seven Clusters of Proteins Defined on Figure 2.

Isozymes	Protein Clusters				
	A	B	C	D	E
GCY1/YPR1	X			X	
TSA1/TSA2	X			X	
ALD2/ALD3	X			X	
PDC1/PDC5	X			X	
TDH2/TDH3	X			X	
ACO1/ACO2	X			X	
ARO3/ARO4	X			X	
SAM1/SAM2	X				X
ADE16/ADE17	X				X
MDH1/MDH3	X		X		
IMD3/IMD4	X		X		
ADH1/ADH3				X	X
LYS20/LYS21				X	X
GLK1/HXK1/HXK2			X	X	X
GPD1/GPD2			X		X
SER3/SER33					X
THI20/THI21	X				

Note.—Xs are placed in the columns corresponding to the clusters where the isozymes were classified.

during the evolution of these two yeast species. They also highlight that B2, D2, and W1, the three *S. cerevisiae* strains expressing high levels of stress response proteins, presented poor fermentative performances and were genetically distant from the well-fermenting strains. This suggests that these three strains were not well adapted to the fermenting conditions, resulting in a stress state probably correlated to their poor fermenting performances.

Discussion

We report an original study of comparative proteomics that investigates how the evolutionary histories of *S. cerevisiae* and *S. uvarum* are mirrored on the proteome. We analyzed the genetic variability of the proteome of nine *S. cerevisiae* strains and six *S. uvarum* strains from diverse origins by label-free mass-spectrometry–based quantification. A total of 445 proteins were quantified, with a protein coverage comparable with the one obtained by Andrews et al. (2011) after optimization. Protein abundance changes were confidently detected by using a Bayesian hierarchical model that takes into account both shared and proteotypic peptides (Blein-Nicolas et al. 2012).

In our fermentation conditions (188 gl^{-1} of sugar, 18 °C), the strains of *S. uvarum* and a group of well fermenting strains of *S. cerevisiae* displayed similar fermentative performances that were not mirrored by the global variations of their proteome. Instead, we detected massive interspecific variations of protein abundances. These variations can be explained by genetic differences between the two species: mutations may have accumulated because of the divergence between *S. cerevisiae* and *S. uvarum*. Two results indicate that some of these protein variations are not selectively neutral: 1) the percentage of quantitative variations varied according to functional categories, whereas these variations would be randomly distributed in the absence of selection pressure; 2) some of the protein abundance changes were consistent with phenotypic traits known to discriminate the two species. Indeed, in *S. uvarum*, we found high amounts of proteins involved in the biosynthesis of ergosterol, fatty acids, and L-PHE biosynthesis and in the Ehrlich pathway, in agreement with its psychrophilic aptitude (Kishimoto and Goto 1995; Naumov 1996; Belloch et al. 2008) and its ability to produce high concentrations of 2-PE (Tosi et al. 2009; Masneuf-Pomarède et al. 2010). Interspecific variations of protein abundances can also be explained by genetic × environment interactions. Indeed, fermentations were carried out at 18 °C, that is, the standard temperature for white wine making. This temperature is a compromise between the optimal temperatures for *S. uvarum*, that is more adapted to lower temperatures and *S cerevisiae* that is more adapted to higher temperatures. The results would certainly have been different if the fermentations were carried out at a different temperature. Enzymes involved in sulfur amino acid biosynthesis and in thiamine biosynthesis were also found to be globally more abundant in *S. uvarum*. Sulfur amino acids and thiamine were shown to be related to the production of volatile sulfur compounds (Moreira et al. 2002; Linderholm et al. 2008; Bartra et al. 2010; Linderholm et al. 2010), another type of aromatic molecules with important impact on the sensory quality of wine (Swiegers and Pretorius 2007). However, the production of volatile sulfur compounds by *S. cerevisiae* and *S. uvarum* has never been formally compared and would require further investigation.

Interspecific variations of protein abundances were also related to a differential recruitment of isozymes, particularly in the glucose fermentation pathway, suggesting that *S. cerevisiae* and *S. uvarum* use different sets of proteins to ferment glucose. Further analysis of the abundance patterns of the isozymes present in our data set, most of which were involved in the carbohydrate and amino acid metabolisms, revealed that duplicated genes were involved in the interspecific proteome differentiation. Duplicated genes, and particularly those of central metabolism, were suggested to be involved in the adaptation to environmental changes (Gasch et al. 2000; Causton et al. 2001; Kwast et al. 2002) and to different ecological niches (Byrne and Wolfe 2005). In support of this hypothesis, duplicated genes were shown to diverge in function (Kuepfer et al. 2005) and sequence (Kellis et al. 2004). The differential recruitment of isozymes observed in our study between *S. cerevisiae* and *S. uvarum* is in agreement with the role of duplicated genes in adaptation and support the hypothesis that these two species are adapted to different ecological niches (Sampaio and Gonçalves 2008; Salvadó et al. 2011).

Within *S. cerevisiae*, well-fermenting strains differed from bad-fermenting strain on life-history traits. In particular, the former had smaller cells and larger maximum population sizes than the latter. Significant trade-offs were found between maximum population size and content of trehalose, a storage sugar though to be involved in yeast survivorship (Albertin et al. 2011). Among the bad-fermenting strains, B2, D2, and

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

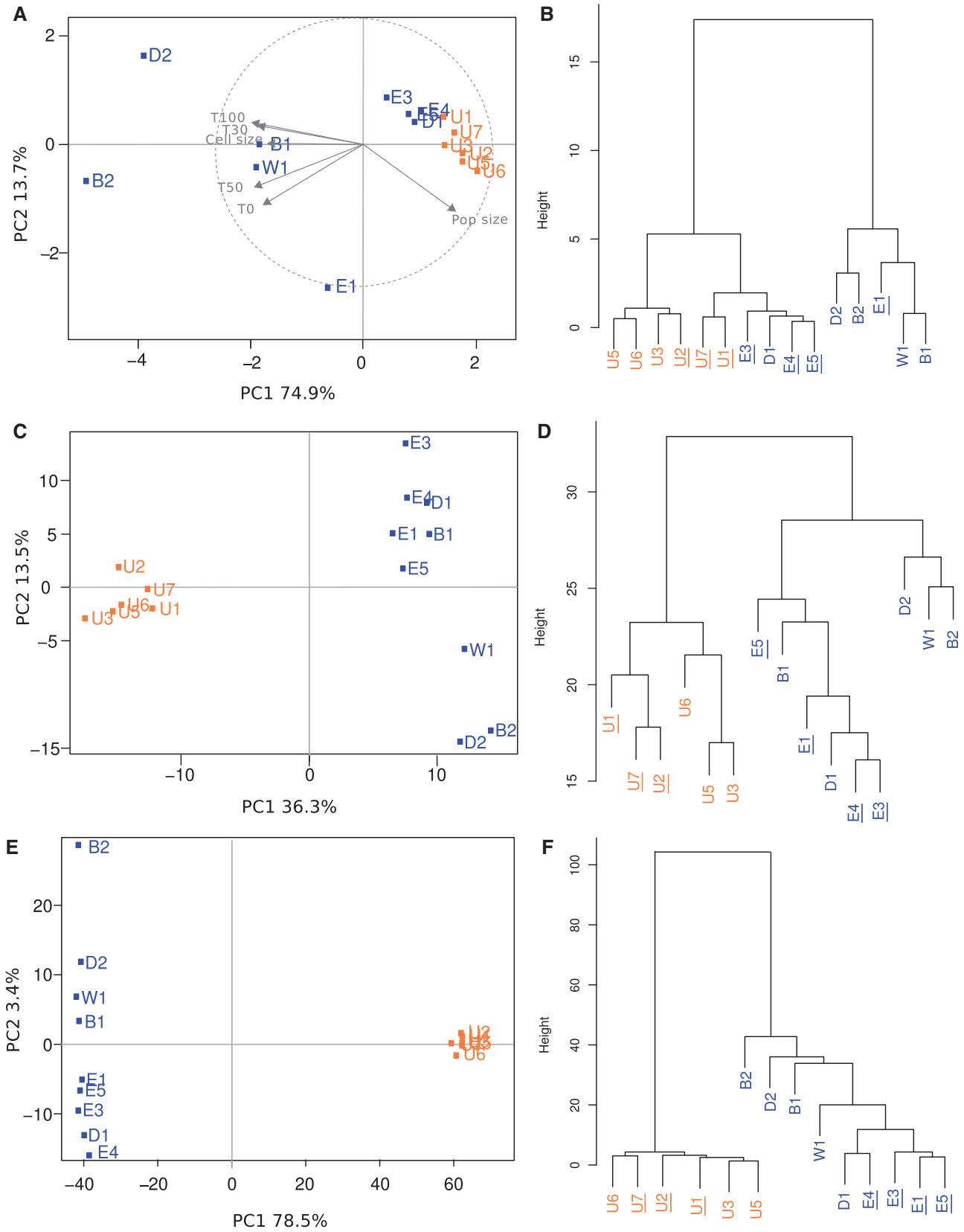

Fig. 6. Integrative analysis of the strain differentiation. Differentiation of six strains of *Saccharomyces uvarum* (green) and nine strains of *S. cerevisiae* (red) was based on the phenotypic variability assessed from lag phase time (T0), times to complete 30%, 50%, and 100% of fermentation (T30, T50, and T100, respectively), cell size, and population size (pop size) at 30% of CO_2 release (*A* and *B*), on the proteome variability assessed from abundances of 401 proteins (*C* and *D*) and on sequence variability inferred from 498 SNPs and 2,681 SAPs (*E* and *F*). *A*, *C*, and *E* are principal component analyses and *B*, *D*, and *F* are phylogenetic trees obtained by hierarchical clustering.

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

W1 formed a homogeneous group isolated from the rest of *S. cerevisiae* strains at the proteomic level. These strains were particularly characterized by highly abundant proteins involved in the biosynthesis of storage compounds, such as glycerol, trehalose, and glycogen. These results thus confirm the trade-off between the maximum population size and the storage of cell resources. The biosynthesis of storage compounds is known to be induced during the general stress response (Gasch et al. 2000; Causton et al. 2001). Moreover, B2, D2, and W1 exhibited high levels of proteins involved in the oxidative stress response and high levels of proteins carrying the stress responsive element in their promoter, such as GPH1, HXK1, TPS1, TSL1, PGM2, CTT1, HSP12, HSP26, and HSP104 (Moskvina et al. 1998). These results suggest that B2, D2, and W1 were in a stress state. Accordingly, the ability of yeasts to grow and complete fermentation was shown to be correlated to their resistance to stress (Ivorra et al. 1999; Zuzuarregui and del Olmo 2004a). In particular, Zuzuarregui and del Olmo (2004b) showed that stress response genes were more expressed in strains with severe fermentative problems, with sometimes high expression maintained during fermentation. These authors suggested that the strains fermenting poorly were unable to restore the physiological nonstressed condition.

In conclusion, the results presented in this study showed that *S. cerevisiae* and *S. uvarum* reached comparable abilities to complete grape must fermentation by different proteome evolutionary roads, involving different metabolic pathways and isozymes. Indeed, interspecies proteome variations affected several metabolic pathways, some of which were related to phenotypic characteristics known to differ between the two species. In addition, isozymes were differently recruited by each species, suggesting that duplicated genes underwent different evolution and display nowadays divergent expression patterns in the two species. We hypothesize that such divergent expression of duplicated genes is related to the adaptation of each species to their different ecological niches. Finally, in a more general perspective, this study showed that the variations of protein abundances can be related to phenotypes in a context of natural genetic variability. This reveals the interest of proteomic studies to identify the genes associated to phenotypes in the absence of perturbation factors such as mutants or environmental stress. This shows that the proteome may be used in the future as a predictor of the phenotype.

Materials and Methods

Yeast Strains

Nine diploid strains of *S. cerevisiae* and six diploid strains of *S. uvarum* isolated from different geographical locations and from either natural or food processing origins (table 1) were studied. For each strain, one meiospore was isolated with a micromanipulator (Singer MSM Manual, Singer Instrument, Somerset, UK). For two heterothallic (*ho*/*ho*) strains, Alcotec 24 and NRRL-Y-7327, the isolated haploid meiospore was diploidized via transient expression of the HO endonuclease: the strains were transformed with the pHS2 plasmid (kindly given by S. Himanshu) using lithium acetate (Gietz and Schiestl 1991). After diploidization, the plasmid was eliminated from the strains through recurrent cultures on YPD (Yeast extract Peptone Dextrose) medium. All the other strains were homothallic (*HO*/*HO*), so that the isolated meiospores gave rise to fully homozygous diploid strains through mating type switch and further fusion of opposite mating-type cells. The fully homozygous diploid meiosporic derivates were used for further alcoholic fermentation (table 1). All strains were grown at 30 °C in YPD medium containing 1% yeast extract (Difco Laboratories, Detroit, MI), 1% bactopeptone (Difco), 2% glucose, supplemented or not with 2% agar.

Strain Genotyping

Yeast cell culture and DNA extraction were performed as described by Albertin et al. (2009). Six genes (*ACC1*, *ALA1*, *ADP1*, *GLN4*, *VSP13*, and *RPN2*) known to reveal sequence diversity among wine *S. cerevisae* strains (Muñoz et al. 2009) were amplified by using the protocol previously described (Albertin et al. 2011). Both strands of PCR products were sequenced using Sanger method. The sequences were aligned with Clustal X program (Thompson et al. 1997; Larkin et al. 2007).

Alcoholic Fermentation in Grape Must

White grape must was obtained from Sauvignon grapes, harvested in vineyards in Bordeaux area (2009 vintage). Tartaric acid precipitation was stabilized and turbidity was adjusted to 100 NTU (nephelometric turbidity unit) before storage at −20 °C. Sugar concentration was 188 gl^{-1} and the indigenous yeast population, estimated by YPD-plate counting after must thawing, was less than 20 CFU (colony-forming unit) per ml. Precultures of each strains were run in half-diluted must filtered through a 0.45 -µm nitrate-cellulose membrane (24 °C, 150 RPM [rounds per minute]). After 24 h, yeast population sizes were measured using a particle counter (Z2 Coulter Counter, Beckman Coulter, Villepinte, France). One million cells per milliliter was then sampled from pre-cultures and added to a final volume of 125 ml of Sauvignon must. After inoculation, the musts were transferred into 125-ml glass reactors locked to maintain anaerobiosis. Fermentations were run simultaneously for the 15 strains (18 °C, 300 RPM). The amount of CO_2 released was regularly determined by measurement of glass reactor weight loss and used to represent the fermentation kinetics of each strain (supplementary fig. S1, Supplementary Material online). During fermentation, four kinetics parameters were recorded (supplementary fig. S2, Supplementary Material online): the lag phase duration (h), the times (h) to release 30% and 50% of CO_2 (minus lag phase duration), and the time (h) required to achieve the fermentation (meaning 100% of CO_2 released or all initial sugar consumed). Samples were harvested at 30% of CO_2 release to perform proteomics analyses and to measure maximum population size (cells per ml) and cell size (mean diameter, µm) using a particle counter (Z2 Coulter Counter, Beckman Coulter) (supplementary fig. S2, Supplementary Material online). The time points at which strains were

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

sampled are indicated in supplementary table S1, Supplementary Material online. Fermentations were repeated three times independently.

Protein Extraction and Digestion

Proteomic analyses were performed at 30% of CO_2 release. At this time, all strains had reached their maximum population size and performed alcoholic fermentation without growing. Five milliliters of fermentative media was sampled (the rest of the culture was maintained in fermentation) and centrifugated (5 min, 2,750 × g). The pellets were rinsed two times with 5 ml of water, frozen in liquid nitrogen and stored at −80 °C until protein extraction. Yeast cell pellets were suspended in 500 µl of an ice-cold solution of acetone containing 10% of trichloroactetic acid and 0.07% β-mercaptoethanol and ground for 5 min with 200 µl of glass beads for protein extraction and precipitation. After transferring the homogenate to new tubes to remove glass beads, 700 µl of the extraction solution were added. Protein extracts were incubated 1 h at −20 °C, centrifugated for 10 min at 14,000 rpm and washed in 1.2 ml of 0.07% β-mercaptoethanol in acetone. Centrifugation and washing were repeated three times. After the last washing, proteins were dried in a vacuum centrifuge, solubilized in 300 µl of a solution containing 6 M of urea, 2 M of thiurea, 2% of CHAPS and 30 mM Tris–HCl 30 mM (pH 8.8) and centrifugated for 10 min at 14,000 rpm. Protein concentration was determined using PlusOne 2-D Quant Kit (GE Healthcare, Little Chalfont, UK) and adjusted to 4 µg μl^{-1}. After a 10 times dilution in 50 mM of ammonium bicarbonate, proteins were reduced 1 h in 100 mM dithiothreitol, alkylated 1 h in 40 mM iodoacetamide and digested overnight at 37 °C with 1/50 (w/w) trypsin (Promega, Madison, WI). Digestion was stopped by adding 0.4% of TFA (trifluoroacetic acid).

LC-MS/MS Analysis

LC-MS/MS analysis were performed using an Ultimate 3000 LC system (Dionex, Sunnyvale, CA) connected to an LTQ Orbitrap mass spectrometer (Thermo Electron, Waltham, MA). A 700 ng of protein digest was loaded onto a PepMap C18 precolumn (0.3 × 5 mm, 100 Å, 5 µm; Dionex) at 15 µl min^{-1} and desalted with 0.08% TFA and 2% acetonitrile (ACN). After 3 min, the precolumn was connected to a PepMap C18 nanocolumn (0.075 × 15 cm, 100 Å, 3 µm). Buffers were 0.1% formic acid, 3% ACN (A) and 0.1% formic acid, and 80% ACN (B). Peptides were separated at 40 °C using a linear gradient from 4% to 36% buffer B for 60 min at 300 nl min^{-1}. One run took 90 min, including the regeneration step at 100% buffer B and the equilibration step at 100% buffer A.

Ionization was performed with a 1.3-kV spray voltage applied to an uncoated capillary probe (10 -µm tip inner diameter; New Objective). Peptide ions were analyzed using Xcalibur v2.0.7 (Thermo Electron) with the following data-dependent acquisition steps: 1) FTMS scan on Orbitrap (mass-to-charge ratio [m/z] 300 to 1,300, 15,000 resolution, profile mode), 2) MS/MS on the LTQ ($qz = 0.25$, activation time = 30 ms, and collision energy = 35%; centroid mode). Step 2 was repeated for the two major ions detected in step 1. Dynamic exclusion was set to 90 s. Xcalibur raw datafiles were transformed to mzXML open source format using ReadW software (v4.3.1, http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW). During transformation, MS data were centroided.

MS Data Availability

The raw MS output files were deposited online using PROTICdb database (Ferry-Dumazet et al. 2005; Langella et al. 2007) at the following URL: http://moulon.inra.fr/protic/public.

Protein Identification

A custom FASTA format database containing 10,851 entries was constructed from the translations of all systematically named ORFs of *S. cerevisiae* and *S. uvarum* downloaded from the *Saccharomyces* Genome Database (SGD project, http://www.yeastgenome.org/, versions dated 5 October 2010 and 16 December 2003, respectively). This database was searched with X!Tandem (v2010.01.01.4; http://www.thegpm.org/TANDEM/). Unique labels were attributed to proteins encoded by orthologous genes. A contaminant database containing the sequences of standard contaminants and the sequences of 16 proteins of *Vitis vinifera* previously identified in extracts of yeast grown in grape must was also interrogated. The decoy database comprised the reverse protein sequences of the custom database. X!tandem settings were as follows. Enzymatic cleavage was declared as a trypsin digestion with one possible misscleavage. Carboxyamidomethylation of cysteine residues and oxidation of methionine residues were set to static and possible modifications, respectively. Precursor mass precision was set to 20 ppm. Fragment mass tolerance was 0.5 Th. A refinement search was added with the same settings, except that semi-trypsic peptides and protein N-ter acetylations were also searched. Only peptides with an *E* value smaller than 0.1 were reported.

Identified proteins were filtered and sorted by using the X!Tandem pipeline (v2.2.6, http://pappso.inra.fr/bioinfo/xtandempipeline/). Criteria used for protein identification were 1) at least two different peptides identified with an *E* value smaller than 0.05; 2) a protein *E* value (product of unique peptide *E* values) smaller than 10^{-4}. These criteria led to a false discovery rate estimated by using the decoy database less than 1% for peptide and protein identification. Proteins were sorted in groups and subgroups depending on the list of peptides by which they were identified: groups gathered proteins sharing at least one peptide and subgroups gathered proteins that could not be distinguished based on their list of peptides. More than 96% of the subgroups contained only one protein. The remaining 4% were generally conserved proteins, such as ribosomal proteins. Groups corresponded in general to proteins encoded by paralogous genes.

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

Peptide Quantification and Filtering Intensity Data

Peptides were quantified based on extracted ion chromatograms using MassChroQ software version 1.0.1, with the parameters given in supplementary file S1, Supplementary Material online. Peptide intensity data were filtered as follows. We first removed 1) dubious peptides for which standard deviation of retention time was more than 20 s; 2) peptides × strain combinations quantified in only one replicate; 3) peptides quantified in less than four strains; 4) subgroup-specific peptides when, within a group, each subgroup was represented by only one specific peptide and more than five shared peptides (quantification was therefore performed at the group level, from shared peptides uniquely). We then removed the peptides corresponding to subgroups quantified with more than 80% of species-specific peptides. The effects of species-specific peptides on the measured intensities were indeed confunded with a species effect in cases where the differentiation between the two species affected the abundance of a protein. Considering the statistical model used to estimate protein abundances and described below, highly confounding effects would have biased the estimation of protein abundances. Species-specific peptides were identified by searching the sequences of all quantified peptides in the protein sequence databases of *S. cerevisiae* and *S. uvarum* used for protein identification. Species-specific peptides were those absent in one of the two databases. Finally, we removed a group containing 11 heat shock proteins sharing peptides because of complex peptide–protein relationships. A total of 1,672 peptides were removed from the data set that finally contained 4,960 peptides belonging to 416 sub-groups and 29 peptides belonging to four groups. For simplification, these groups and subgroups were regarded as 445 distinct proteins. To take into account possible global quantitative variations between LC-MS runs, normalization was performed. For each LC-MS run, the ratio of all peptide values to their value in the chosen reference LC-MS run was computed. Normalization was performed by dividing peptide values by the median value of peptide ratios.

Detection of Protein Abundance Changes

Protein abundance changes were detected using the statistical framework described in Blein-Nicolas et al. (2012). Briefly, protein abundances were estimated from both shared and proteotypic peptides by using the following model:

$$\log(I_{itr}) = \log\left(\sum_k \delta_{ik}\ \exp(\theta_{kt})\right) + D_i + B_r + C_{tr} + \varepsilon_{itr} \tag{1}$$

where I_{itr} is the intensity measured for peptide i in replicate r and strain t; $\delta_{ik} = 1$ if the peptide i belongs to the protein k, else $\delta_{ik} = 0$; θ_{kt} is the natural logarithm of the abundance of protein k in strain t; $B_r \sim N(0,\ \sigma_B^2)$ is an error due to the biological variation of replicate r; $C_{tr} \sim N(0,\ \sigma_C^2)$ is an error due to the technical variation of sample tr; $D_i \sim N(0,\ \sigma_D^2)$ is an error due to the LC-MS response (or peptide effect) of peptide i; $\varepsilon_{itr} \sim N(0,\ \sigma_\varepsilon^2)$ is the residual error.

Because of shared peptides, this model is nonlinear and the estimation of its parameters was performed in a Bayesian hierarchical framework. Protein abundance changes were detected with a multiple test procedure. As several couples of strains and several proteins were tested, P values were adjusted for multiple testing by a Benjamini–Hochberg procedure (Benjamini and Hochberg 1995).

SAP Detection

SAP detection was based on the qualitative variations of the peptides. A peptide was considered as present in a strain if it was quantified in at least one of the three replicates, otherwise it was noted absent. Two main reasons can explain the absence of a peptide in a strain: a genetic variation resulting in an amino acid change or a quantitative variation below the detection threshold of the mass-spectrometer. To get rid as far as possible of the quantitative variations, we used a Student's t-test to compare the intensity distribution of the peptides exhibiting presence/absence variations to the one of the other peptides. Peptides with the lowest average intensities were removed from the set of present/absent peptides until the Student's t-test become nonsignificant. The remaining peptides were considered as SAP markers.

Data Analysis

For each protein, Q_{ST} was calculated following Bonnin et al. (1996):

$$Q_{ST} = \sigma_{esp} + \sigma_{cer} + \sigma_{uva} \tag{2}$$

where σ_{esp} is the variance between species, σ_{cer} is the variance within *S. cerevisiae*, and σ_{uva} is the variance within *S. uvarum*. Q_{IS} for *S. cerevisiae* and *S. uvarum* were derived from Q_{ST} formula as follows:

$$Q_{IScerevisiae} = \sigma_{cer}/\left(\sigma_{esp} + \sigma_{cer} + \sigma_{uva}\right) \tag{3}$$

$$Q_{ISuvarum} = \sigma_{uva}/\left(\sigma_{esp} + \sigma_{cer} + \sigma_{uva}\right). \tag{4}$$

The rate of enrichment (or depletion) for the functional category i of the proteins that changed in abundance was calculated as follows:

$$R_i = \frac{(V_i/V_{tot}) - (N_i/N_{tot})}{(N_i/N_{tot})} \tag{5}$$

where N_i is the number of proteins in the functional category i; N_{tot} is the total number of quantified proteins; V_i is the number of variable proteins in the functional category i; and V_{tot} is the total number of variable proteins.

To determine whether R_i was significantly different from zero, we performed a Fisher's exact test. As several functional categories were tested, we adjusted P values by a Benjamini–Hochberg procedure (Benjamini and Hochberg 1995).

Hierarchical clustering analyses were performed using Euclidean distances and unweighted pair group averages as the aggregation method. All data analyses and graphical representations were performed using R v2.10.1 (R Development Core Team 2009). Metabolic pathways were reproduced from SGD.

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

Supplementary Material

Supplementary file S1, tables S1 and S2, and figure S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Agence Nationale de la Recherche (ANR-08-ALIA-09 HeterosYeast). The authors thank Isabelle Masneuf-Pomarède and Marina Bely for strong scientific support regarding *S. uvarum* species and fermentation conduction under enological conditions.

References

Albertin W, Marullo P, Aigle M, Bourgais A, Bely M, Dillmann C, De Vienne D, Sicard D. 2009. Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. *J Evol Biol.* 22:2157–2170.

Albertin W, Marullo P, Aigle M, Dillmann C, De Vienne D, Bely M, Sicard D. 2011. Population size drives industrial *Saccharomyces cerevisiae* alcoholic fermentation and is under genetic control. *Appl Environ Microbiol.* 77:2772–2784.

Albertin W, Marullo P, Bely M, et al. (13 co-authors). 2013. Linking post-translational modifications and variation of phenotypic traits. *Mol Cell Proteomics.* 12:720–735.

Andrews GL, Dean RA, Hawkridge AM, Muddiman DC. 2011. Improving proteome coverage on a LTQ-Orbitrap using design of experiments. *J Am Soc Mass Spectrom.* 22:773–783.

Antonelli A, Castellari L, Zambonelli C, Carnacini A. 1999. Yeast influence on volatile composition of wines. *J Agric Food Chem.* 47: 1139–1144.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol.* 23:26–32.

Bartra E, Casado M, Carro D, Campamà C, Piña B. 2010. Differential expression of thiamine biosynthetic genes in yeast strains with high and low production of hydrogen sulfide during wine fermentation. *J Appl Microbiol.* 109:272–281.

Belloch C, Orlic S, Barrio E, Querol A. 2008. Fermentative stress adaptation of hybrids within the *Saccharomyces* sensu stricto complex. *Int J Food Microbiol.* 122:188–195.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc.* 57:289–300.

Bertolini L, Zambonelli C, Giudici P, Castellari L. 1996. Higher alcohol production by cryotolerant *Saccharomyces* strains. *Am J Enol Vitic.* 47:343–345.

Blank L, Lehmbeck F, Sauer U. 2005. Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res.* 5:545–558.

Blein-Nicolas M, Xu H, De Vienne D, Giraud C, Huet S, Zivy M. 2012. Including shared peptides for estimating protein abundances: a significant improvement for quantitative proteomics. *Proteomics* 12: 2797–2801.

Bohlin J, Snipen L, Cloeckaert A, Lagesen K, Ussery D, Kristoffersen A, Godfroid J. 2010. Genomic comparisons of *Brucella* spp. and closely related bacteria using base compositional and proteome based methods. *BMC Evol Biol.* 10:249.

Bon E, Neuvéglise C, Casaregola S, Artiguenave F, Wincker P, Aigle M, Durrens P. 2000. Genomic exploration of the hemiascomycetous yeasts: 5. *Saccharomyces bayanus* var. *uvarum*. *FEBS Lett.* 487:37–41.

Bonnin I, Prosperi JM, Olivieri I. 1996. Genetic markers and quantitative genetic variation in *Medicago truncatula* (*Leguminosae*): a comparative analysis of population structure. *Genetics* 143:1795–1805.

Byrne KP, Wolfe KH. 2005. The yeast gene order browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* 15:1456–1461.

Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Landers ES, Young RA. 2001. Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell.* 12: 323–337.

Conant GC, Wolfe KH. 2007. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol Syst Biol.* 3:129.

D'Amico S, Collins T, Marx J-C, Feller G, Gerday C. 2006. Psychrophilic microorganisms: challenges for life. *EMBO Rep.* 7:385–389.

Demuyter C, Lollier M, Legras J-L, 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.

Dietrich FS, Voegeli S, Brachat S, et al. (14 co-authors). 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 304:304–307.

Dujon B, Sherman D, Fischer G, et al. (67 co-authors). 2004. Genome evolution in yeasts. *Nature* 430:35–44.

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

Ferry-Dumazet H, Houel G, Montalent P, et al. (12 co-authors). 2005. PROTICdb: a web-based application to store, track, query, and compare plant proteome data. *Proteomics* 5:2069–2081.

Fischer G, Neuvéglise C, Durrens P, Gaillardin C, Dujon B. 2001. Evolution of gene order in the genomes of two related yeast species. *Genome Res.* 11:2009–2019.

Gamero A, Hernández-Orte P, Querol A, Ferreira V. 2011. Effect of aromatic precursor addition to wine fermentations carried out with different *Saccharomyces* species and their hybrids. *Int J Food Microbiol.* 147:33–44.

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell.* 11: 4241–4257.

Gietz RD, Schiestl RH. 1991. Applications of high efficiency lithium acetate transformation of intact yeast cells using single-stranded nucleic acids as carrier. *Yeast* 7:253–263.

Gstaiger M, Aebersold R. 2009. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet.* 10:617–627.

Gupta N, Benhamida J, Bhargava V, et al. (15 co-authors). 2008. Comparative proteogenomics: combining mass spectrometry and comparative genomics to analyze multiple genomes. *Genome Res.* 18: 1133–1142.

Harlan JR. 1992. Crops & man. Madison (WI): American Society of Agronomy, Crop Science Society of America.

Hazelwood LA, Daran J-M, Van Maris AJA, Pronk JT, Dickinson JR. 2008. The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Appl Environ Microbiol.* 74:2259–2266.

Ihmels J, Levy R, Barkai N. 2004. Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat Biotech.* 22: 86–92.

Ivorra C, Pérez-Ortín JE, Del Olmo M. 1999. An inverse correlation between stress resistance and stuck fermentations in wine yeasts. A molecular study. *Biotechnol Bioeng.* 64:698–708.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428:617–624.

Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423:241–254.

Kishimoto M, Goto S. 1995. Growth temperatures and electrophoretic karyotyping as tools for practical discrimination of *Saccharomyces bayanus* and *Saccharomyces cerevisiae*. *J Gen Appl Microbiol.* 41: 239–247.

Kuepfer L, Sauer U, Blank LM. 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* 15:1421–1430.

Kwast KE, Lai L-C, Menda N, James DT, Aref S, Burke PV. 2002. Genomic analyses of anaerobically induced genes in *Saccharomyces*

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

cerevisiae: functional roles of Rox1 and other factors in mediating the anoxic response. *J Bacteriol.* 184:250–265.

Langella O, Zivy M, Joets J. 2007. The PROTICdb database for 2-DE proteomics. *Methods Mol Biol.* 355:279–303.

Larkin MA, Blackshields G, Brown NP, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.

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.

Libkind D, Hittinger CT, Valerio E, Goncalves C, Dover J, Johnston M, Goncalves P, Sampaio JP. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Nat Acad Sci U S A.* 108:14539–14544.

Linderholm A, Dietzel K, Hirst M, Bisson LF. 2010. Identification of MET10-932 and characterization as an allele reducing hydrogen sulfide formation in wine strains of *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 76:7699–7707.

Linderholm AL, Findleton CL, Kumar G, Hong Y, Bisson LF. 2008. Identification of genes affecting hydrogen sulfide formation in *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 74:1418–1427.

Liti G, Carter DM, Moses AM, et al. (26 co-authors). 2009. Population genomics of domestic and wild yeasts. *Nature* 458:337–341.

Llorente B, Malpertuy A, Neuvéglise C, et al. (24 co-authors). 2000. Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. *FEBS Lett.* 487:101–112.

Marullo P, Bely M, Masneuf-Pomarède I, Pons M, Aigle M, Dubourdieu D. 2006. Breeding strategies for combining fermentative qualities and reducing off-flavor production in a wine yeast model. *FEMS Yeast Res.* 6:268–279.

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.

Masneuf-Pomarède I, Bely M, Marullo P, Lonvaud-Funel A, Dubourdieu D. 2010. Reassessment of phenotypic traits for *Saccharomyces bayanus* var. *uvarum* wine yeast strains. *Int J Food Microbiol.* 139:79–86.

McGovern PE, Glusker DL, Exner LJ, Voigt MM. 1996. Neolithic resinated wine. *Nature* 381:480–481.

McGovern PE, Zhang J, Tang J, et al. (13 co-authors). 2004. Fermented beverages of pre- and proto-historic China. *Proc Nat Acad Sci U S A.* 101:17593–17598.

Merico A, Sulo P, Piskur J, Compagno C. 2007. Fermentative lifestyle in yeasts belonging to the *Saccharomyces* complex. *FEBS J.* 274: 976–989.

Moreira N, Mendesa F, Pereirab O, Guedes de Pinhoa P, Hogga 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.

Moskvina E, Schüller C, Maurer CT, Mager WH, Ruis H. 1998. A search in the genome of *Saccharomyces cerevisiae* for genes regulated via stress response elements. *Yeast* 14:1041–1050.

Muñoz R, Gómez A, Robles V, Rodríguez P, Cebollero E, Tabera L, Carrascosa AV, Gonzalez R. 2009. Multilocus sequence typing of oenological *Saccharomyces cerevisiae* strains. *Food Microbiol.* 26:841–846.

Nakagawa Y, Sakumoto N, Kaneko Y, Harashima S. 2002. Mga2p is a putative sensor for low temperature and oxygen to induce OLE1 transcription in *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun.* 291:707–713.

Naumova ES, Antunovics Z, Sipiczki M, Naumov GI. 2002. *Saccharomyces bayanus* var. *uvarum* in Tokaj wine-making of Slovakia and Hungary. *Appl Microbiol Biotech.* 59:727–730.

Naumov G. 1996. Genetic identification of biological species in the *Saccharomyces sensu stricto* complex. *J Indust Microbiol Biotech.* 17: 295–302.

Naumov G, Masneuf I, Naumova ES, Aigle M, Dubourdieu D. 2000. Association of *Saccharomyces bayanus* var. *uvarum* with some French wines: genetic analysis of yeast populations. *Res Microbiol.* 151:683–691.

Naumov G, Nguyen H-V, Naumova E, Michel A, Aigle M, Gaillardin C. 2001. Genetic identification of *Saccharomyces bayanus* var. *uvarum*, a cider-fermenting yeast. *Int J Food Microbiol.* 65:163–171.

Naumov GI, James SA, Naumova ES, Louis EJ, Roberts IN. 2000. Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *Int J Syst Evol Microbiol.* 5:1931–1942.

Nguyen H-V, Gaillardin C. 2005. Evolutionary relationships between the former species *Saccharomyces uvarum* and the hybrids *Saccharomyces bayanus* and *Saccharomyces pastorianus*; reinstatement of *Saccharomyces uvarum* (Beijerinck) as a distinct species. *FEMS Yeast Res.* 5:471–483.

Nguyen H-V, Legras J-L, Neuvéglise C, Gaillardin C. 2011. Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380T. *PLoS One* 6:e25821.

Papp B, Pál C, Hurst LD. 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429: 661–664.

Parks LW, Smith SJ, Crowley JH. 1995. Biochemical and physiological effects of sterol alterations in yeast—a review. *Lipids* 30:227–230.

Piškur J. 2001. Origin of the duplicated regions in the yeast genomes. *Trends Genet.* 17:302–303.

Pulvirenti A, Nguyen H-V, Caggia C, Giudici P, Rainieri S, Zambonelli C. 2000. *Saccharomyces uvarum*, a proper species within *Saccharomyces sensu stricto*. *FEMS Microbiol Lett.* 192:191–196.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. *Nature* 457:843–848.

R Development Core Team. 2009. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Rainieri S, Pretorius IS. 2000. Selection and improvement of wine yeasts. *Ann Microbiol.* 31:15–31.

Rementeria A, Rodriguez JA, Cadaval A, Amenabar R, Muguruza JR, Hernando FL, Sevilla MJ. 2003. Yeast associated with spontaneous fermentations of white wines from the "Txakoli de Bizkaia" region (Basque Country, North Spain). *Int J Food Microbiol.* 86: 201–207.

Rodriguez-vargas S, Estruch F, Randez-Gil F. 2002. Gene expression analysis of cold and freeze stress in baker's yeast. *Appl Environ Microbiol.* 68:3024–3030.

Ruepp A, Zollner A, Maier D, et al. (11 co-authors). 2004. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res.* 32:5539–5545.

Salvadó Z, Arroyo-López FN, Barrio E, Querol A, Guillamón JM. 2011. Quantifying the individual effects of ethanol and temperature on the fitness advantage of *Saccharomyces cerevisiae*. *Food Microbiol.* 28:1155–1161.

Sampaio JP, Gonçalves P. 2008. Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and S. paradoxus. *Appl Environ Microbiol.* 74: 2144–2152.

Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, Rine J, Johnston M, Hittinger CT. 2011. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces* sensu stricto. *G3 (Bethesda)* 1:11–25.

Schade B, Jansen G, Whiteway M, Entian KD, Thomas DY. 2004. Cold adaptation in budding yeast. *Mol Biol Cell.* 15:5492–5502.

Schrimpf SP, Weiss M, Reiter L, et al. (13 co-authors). 2009. Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol.* 7:e48.

Sicard D, Legras J-L. 2011. Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C R Biol.* 334:229–236.

Swiegers JH, Pretorius IS. 2007. Modulation of volatile sulfur compounds by wine yeast. *Appl Microbiol Biotech.* 74:954–960.

Tehlivets O, Scheuringer K, Kohlwein SD. 2007. Fatty acid synthesis and elongation in yeast. *Biochim Biophys Acta.* 1771:255–270.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015

sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882.

Thomson JM, Gaucher EA, Burgan MF, De Kee DW, Li T, Aris JP, Benner SA. 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet.* 37:630–635.

Torriani S, Zapparoli G, Suzzi G. 1999. Genetic and phenotypic diversity of *Saccharomyces sensu stricto* strains isolated from Amarone wine. *Antonie van Leeuwenhoek* 75:207–215.

Tosi E, Azzolini M, Guzzo F, Zapparoli G. 2009. Evidence of different fermentation behaviours of two indigenous strains of *Saccharomyces cerevisiae* and *Saccharomyces uvarum* isolated from Amarone wine. *J Appl Microbiol.* 107:210–218.

Vuralhan Z, Morais MA, Tai S-L, Piper MDW, Pronk JT. 2003. Identification and characterization of phenylpyruvate decarboxylase genes in *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 69:4534–4541.

Wang S, Spor A, Nidelet T, Montalent P, Dillmann C, De Vienne D, Sicard D. 2011. Switch between life history strategies due to changes in glycolytic enzyme gene dosage in *Saccharomyces cerevisiae*. *Appl Environ Microbiol.* 77:452–459.

Weiss M, Schrimpf S, Hengartner MO, Lercher MJ, Von Mering C. 2010. Shotgun proteomics data from multiple organisms reveals remarkable quantitative conservation of the eukaryotic core proteome. *Proteomics* 10:1297–1306.

Wolfe KH, Shields DC. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387:708–713.

Wynne C, Edwards NJ, Fenselau C. 2010. Phyloproteomic classification of unsequenced organisms by top-down identification of bacterial proteins using cap LC-MS/MS on an Orbitrap. *Proteomics* 10: 3631–3643.

Zeder MA, Emshwiller E, Smith BD, Bradley DG. 2006. Documenting domestication: the intersection of genetics and archaeology. *Trends Genet.* 22:139–155.

Zuzuarregui A, Del Olmo M. 2004a. Analyses of stress resistance under laboratory conditions constitute a suitable criterion for wine yeast selection. *Antonie van Leeuwenhoek* 85:271–280.

Zuzuarregui A, Del Olmo M. 2004b. Expression of stress response genes in wine strains with different fermentative behavior. *FEMS Yeast Res.* 4:699–710.

Downloaded from http://mbe.oxfordjournals.org/ by guest on January 13, 2015