Materials and Methods

Biological material and synthetic fermentative media

In a previous report, we studied the genetic diversity of 26 *Saccharomyces cerevisiae* food-processing strains used in enology, brewery, distillery and bakery (1). A sub-panel of nine representative strains was formed (Supplementary Table 1), excluding bakery strains that were mostly autotetraploids (1). For each of these nine strains, a fully homozygous meiosporic derivate was constructed as detailed in previous work (2). The resulting biological material was constituted by four enology strains (E1 to E4), two brewery strains (B1 and B2) and three distillery strains (D1 to D3) (Supplementary Table 1). All nine strains were grown three times independently in three synthetic fermentative media, so that 81 fermentations were run.

Three synthetic fermentative media were used, differing by their amount of sugar, nitrogen, pH, osmotic pressure and anaerobic growth factors in order to reflect main changes of fermentation medium between brewery, bakery and enology contexts (Supplementary Table 2). The composition of bakery, brewery and winery media (designated as BAM, BREM and WIM respectively) is detailed in Albertin *et al.* (2). The media were filtered through a 0.45 µm nitrate-cellulose membrane before inoculation. Pre-cultures were run in diluted half-synthetic medium 20 h at 24°C with orbital agitation (150 rpm) and population size were measured using a particle counter (Z2 Coulter Counter, Beckman Coulter, Villepinte, France). The fermentative media were inoculated at 10^6 cells per mL, and fermentations were run in 1.2 L glass-reactors, locked to maintain anaerobiosis, with permanent stirring at 22°C.

Metabolic traits: CO_2 specific flux and dosage of alcoholic fermentation products

For each of the 81 fermentations (9 strains x 3 media x 3 repetitions), the amount of CO_2 released was determined by automatic measurement of glass-reactor weight loss every 20 min (3). The CO_2 production rate ($dCO_2\ dt^{-1}$, g L^{-1} h^{-1}) was calculated using a local polynomial regression fitting (loess function, R program (4)). The population growth was monitored regularly using a particle counter (Z2 Coulter Counter, Beckman Coulter, Villepinte, France): more than 20 samples per fermentation were taken from the inoculation time until the carrying capacity (*K* or maximum population size) was reached. The experimental points were fitted with a logistic model that allowed estimating *K* (cells *per* mL) (Albertin *et al*. (2) for details). The CO_2 specific flux (the CO_2 production rate *per* cell, g h^{-1} $cell^{-1}$) was calculated by dividing the CO_2 production rate by the number of cells at the stage of cell harvesting for proteomics.

At the end of the alcoholic fermentation, several dosages were made: ethanol concentration (% vol.) was determined by infrared reflectance (Infra-Analyzer 450,Technicon, Plaisir, France) and acetic acid production (g L^{-1}) were measured by colorimetry (A_{460nm}) in continuous flux (Sanimat, Montauban, France) in the supernatant. External glycerol (g L^{-1}) was assayed by enzymatic method (Boehringer kits 10 148 270 035 and 11 112 732 035, R-Biopharm, Darmstadt, Germany). Biomass dry weights (g L^{-1}) were measured from 200 mL of final fermentation medium using a desiccator (SMO 01, Scaltec Instruments GmbH, Göttingen, Germany). Ethanol, acetate, glycerol and biomass contents were then divided by the number of cells to express them on a per-cell basis.

Two-dimensional gel electrophoresis (2-DE)

Five to ten mL of fermentative medium were harvested regularly (every two to three hours) until the maximal CO_2 production rate, *Vmax* (g L^{-1} h^{-1}), was reached. CO_2 production rate drop reflects CO_2 production rate per cell, which is associated with nutriments depletion (such as nitrogen starvation) in the medium and with several physiological and metabolic changes (5). Thus, one sample *per* fermentation was chosen *a posteriori*, to be as close as possible of the *Vmax* but just before it is reached so that all studied samples display comparable physiological stage (maximal CO_2 production rate before nutriments starvation). After centrifugation (5 min, 2750 g), the pellets were rinsed two times with 5 to 10 mL of H_2O, frozen in liquid nitrogen and conserved at –80°C. Cells were lysed using a FastPrep-24 instrument (MP Biomedicals, Illkirch, France): 500 µL of glass beads (acid-washed, 425-600 µm, Sigma, Lyon, France) and 500 µL of cold TCA solution (see below) were added to the frozen pellets, and two 20-second runs were made. Lysed samples were cooled with ice, and proteins were extracted using standard TCA-acetone precipitation method (6). Briefly, the proteins were precipitated 1 hr at – 20°C in 1.2 mL TCA solution (10% TCA and 0.07% 2-mercaptoethanol in acetone). After centrifugation (10 min, maximum speed), the pellets were washed two times with cold acetone containing 0.07% 2-mercaptoethanol, and residual acetone was removed by vacuum drying. Proteins were then solubilized in R2D2 buffer (7) and quantified using the PlusOne 2-D Quant kit (Amersham Biosciences, Arlington Heights, IL). Isoelectrofocusing was carried out using 24-cm long, pH 3–10 NL Immobiline DryStrips (Amersham Biosciences) rehydrated in R2D2 solubilization buffer to which 250 mg of protein extract was added, while two samples upon 81 were discarded from the analysis due to weak concentrations. Full focusing was achieved after application of 84,000 V h at 20°C in a Protean IEF Cell (Bio-Rad, Hercules, CA). Strips were equilibrated as previously

described (8). Second-dimension electrophoresis was performed at 14°C (16 h; 30 mA/gel) on a 24 x 24-cm gel (11% acrylamide, 2.9% of PDA crosslinker) in a Protean Plus Dodeca cell (Bio-Rad). 2-DE gels were fixed in 2% phosphoric acid–50% ethanol, washed in 2% phosphoric acid, and stained 3 days in 2% phosphoric acid–15% ammonium sulfate–17% ethanol–0.1% Coomassie Brilliant Blue G-250. Stained gels were scanned using the PowerLook III scanner (Umax) and LabScan software (Amersham Biosciences). The 79 resulting gels were submitted to 2-DE image analysis using Progenesis software (Nonlinear Dynamics, Newcastle, UK). A co-electrophoresis gel (*i.e.* made from an equimolar mixture of all 27 *medium x strain* combinations) was used as master gel (Figure 1). Detection of spots of interest (*i.e.* spots involved in glycolysis, ethanol, acetate and glycerol pathways, see mass spectrometry analyses below) was checked, corrected manually if necessary, and automatic "Progenesis" background subtraction was applied. The spots of interest represented a large proportion of the total spot volume *per* gel (≈ 35%), so that the usual spot quantification through normalization in percent of total spot volume for each 2-DE gel was not adapted. Thus, we decided to use 107 additional random spots (designed as normalization spots), common to all 2-DE gels and displaying intermediate abundance. Spots were thus quantified in relative abundance (% of normalization spots for each 2-DE gel).

Protein identification through mass spectrometry analyses

Spots of interest, *i.e.* involved in glycolysis, ethanol, acetate and glycerol pathways, were localized by comparison between our 2-DE gels and *S. cerevisiae* reference maps available in the literature (9-12), and were submitted to mass spectrometry (MS) analyses to verify protein identity.

In-gel trypsin digestion: About one hundred spots were excised, and in-gel digestion was performed with a Progest system (Genomic Solution, Huntingdon, UK) using a standard trypsin protocol as described previously (13). Briefly, gel plugs were first washed twice with 10% (v/v) acetic acid, 40% (v/v) ethanol in water, and then with acetonitrile. They were further washed with 25 mM NH_4CO_3 and dehydrated in acetonitrile (two alternating cycles). Following reduction (10 mM DTT, 1 h at 57°C) and alkylation (55mM iodoacétamide, 45 min at 20°C), digestion was performed for 6 h at 37°C with 125 ng of modified trypsin (Promega) dissolved in 20% (v/v) methanol in 20 mM NH_4CO_3. Tryptic peptides were first extracted with 50% (v/v) acetonitrile, 0.5% (v/v) trifluoroacetic acid in water, and then with pure acetonitrile. Both peptide extracts were pooled, dried in a vacuum speed concentrator and suspended in 25 µL of 2% (v/v) acetonitrile, 0.08% (v/v) trifluoroacetic acid in water.

Maldi-TOF identification: The samples were used for matrix-assisted laser desorption ionization time flight (MALDI-TOF). One microliter of sample was spotted directly onto the MALDI plate. The sample was then dried at room temperature before adding 1 µL aliquot of the matrix solution. This dried-droplet sampling method was employed using a matrix solution prepared fresh daily composed of 3 mg/mL α-cyano-4-hydroxycinnamic acid (α-CHCA) in 50% (v/v) acetonitrile and 0.1% (v/v) trifluororacetic acid. Mass spectra were acquired on MALDI-TOF mass spectrometer (Voyager-DE-STR, Applied Biosystems, Framingham, MA) equipped with a nitrogen laser (Laser Science, Franklin, MA). Spectra were then calibrated using tryptic autodigestion ion peaks, $(M + H)^+$ = 2211.104 and 842.509 Da, as an internal calibration. The peptide mass lists were used for identification of proteins using *S. cerevisiae* amino acid sequence database (UniprotKB, 09.09.2010, 29988 entries), using the MS-Fit v 3.2.1 software. To release online MALDI-TOF data in PROTICdb (see below), MASCOT identifications (14), that were all congruent with MS-Fit software, were used and

exported to PROTICdb (MSDB_database, 20060831 release, 16477795 sequences). When protein identification was unsuccessful with MALDI-TOF, nanoHPLC-MS/MS was used.

NanoHPLC-MS/MS: HPLC was performed with the Ultimate LC system combined with Famos autosample and Switchos II microcolumn switching for preconcentration (LC Packings, Amsterdam). The sample was loaded on the column (PEPMAP C18, 5 mm, 75 mm internal diameter, 15 cm; LC Packing), using a preconcentration step on a microprecolumn cartridge (300 mm internal diameter, 5 mm). Five microliters of sample were loaded on the precolumn at 5 mL/min. After 3 min, the precolumn was connected with the separating column and the gradient was started at 200 nL/min. Buffers were 0.1% HCOOH, 3% ACN (A) and 0.1% HCOOH, 95% ACN (B). A linear gradient from 5 to 30% (B) for 25 min was applied. Including the regeneration step, one run was 60 min length. The LCQ deca xp1 (Thermofinnigan, les Ulis, France) was used with a nanoelectrospray interface. Ionization (1.2–1.4 kV ionization potential) was performed with liquid junction and noncoated capillary probe (New Objective, Cambridge, MA). Peptide ions were analyzed by the Nth-dependent method as follows: (i) full MS scan (m/z 500–1500), (ii) ZoomScan (scan of the two major ions with higher resolution), and (iii) MS/MS of these two ions. The raw mass data were first converted to mzXML format with the ReAdW software (http://tools.proteomecenter.org/software.php). Protein identification was performed querying MS/MS data against a *S. cerevisiae* amino acid sequence database (Integr8, 20071211 version, 5815 protein entries, http://www.ebi.ac.uk/), together with an in-house contaminant database, using the X!Tandem software (X!Tandem Tornado 2008.02.01.3, http://www.thegpm.org) with the following parameters: one trypsin missed cleavage allowed, alkylation of cysteine and conditional oxidation of methionine, precursor and fragment ion set at 10 ppm and 0.5 Da, respectively. A refined search was added with similar

parameters except that semitryptic peptides and possible N-terminal acetylation of proteins were searched. All peptides matched with an E-value lower than 0.05 were parsed with an in-house program (X!Tandem pipeline 3.1.2, http://pappso.inra.fr/bioinfo/xtandempipeline). Proteins identified with at least two unique peptides and a log (E-value) lower than 1 10^{e-8} were validated.

The detail of enzymes and metabolites abbreviations is Pgi: Phosphoglucoisomerase, Pfk: Phosphofructokinase, Fba: Fructose-biphosphatase aldolase, Tpi: Triose-phosphate isomerase, Tdh: Triose-phosphate dehydrogenase, Pgk: 3-Phosphoglycerate kinase, Gpm: Glycerate phosphomutase, Eno: Enolase, Pyk: Pyruvate kinase, Pdc: Pyruvate decarboxylase, Adh: Alcohol dehydrogenase, Gpd: Glycerol-3-phosphate dehydrogenase, Hor: Hyperosmolarity-responsive (DL-glycerol-3-phosphatase), Rhr: Related to HOR2 (DL-glycerol-3-phosphatase), Ald: Aldehyde dehydrogenase, Glucose-6P: glucose-6-phosphate, F6P: fructose-6-phosphate, FBP: Fructose-1,6-biphosphate, DHAP: dihydroxyacetone phosphate, G3P: Glycerol-3-phosphate, GA3P: glyceraldehyde-3-phosphate, BPG: glycerate-1,3-biphosphate, 3PG: glycerate-3-phosphate, 2PG: glycerate-2-phosphate, PEP: phosphoenol-pyruvate.

Some of the identified spots were strain-specific and corresponded to allelic variants: for Pfk1p, Pgk1p, Eno1p and Eno2p, Tdh3p, three spots were identified for each strain, yet shifted for E1 B2, E2, E1 and B2/D2 strains respectively (Figure 1B). These allelic variations were confirmed by sequence analysis (Supplementary Information Dataset 1). For example, Eno1p had a predicted isoelectric point of 6.16 for all strains except E2 (6.28), in accordance with the basic shift of Eno1p spots on 2-DE gels (Figure 1). Tdh1p and Pfk2p were also associated with theoretical isoelectric shift, yet not observed on 2-DE gels: this may be

explained by the fact that both Tdh1p and Pfk2p are located in 2-DE gel border, for which slight variation may be difficult to discriminate. For functional analyses (see below), allelic variants were re-aligned: for example, three isoforms of Eno1p (basic, intermediary and acidic isoforms) were taken into account, corresponding respectively to spots 4913, 2067 and 4923 for E2 strain, and to spots 2067, 4923 and 2024 for the others strains (Figure 1B). For the enzymes having paralogous counterparts (pyruvate decarboxylase, enolase, etc.), we verified that at least one discriminant peptide allows accurate paralogous gene assignation (Supplementary Information Dataset 3).

Almost all enzymes involved in glycolysis and ethanol pathways were identified, or at least the major/most abundant isozymes in case of paralogous genes. Pyk2p, Pdc5p and Pdc6p, Adh3p, Adh4p, Adh5p are minor pyruvate kinase, pyruvate decarboxylase and alcohol dehydrogenase isozymes respectively (15-17). For acetate and glycerol pathways also, the major/most abundant enzymes (Ald6p, Rhr2p and Hor2p) were also identified: Ald2p, Ald3p, Ald4p, and Ald5p are minor aldehyde dehydrogenase isozymes (18), Gpd1p is a poorly abundant glycerol-3-phosphate dehydrogenase isozyme (see the Yeast Protein Map: http://www.ibgc.u-bordeaux2.fr/YPM/), and Gpd2p has never been identified to date in 2-DE experiments, suggesting low abundance (19).

Identification of the post-translational modification through MS

Isoforms of Pdc1p, Adh1p and Tdh1p were identified as significantly related to metabolic or life-history traits. To identify the underlying post-translational modification(s), we performed additional mass spectrometry analyses using a high-resolution mass spectrometer (QExactive, Thermo Scientific). Fresh 2-DE gels were used, spots were picked on two gels

corresponding to two strains in different medium (BAM-D1 and WIM-D2) and in-gel trypsin digestion was performed as described above. The mass spectrometry (MS) data were used to search specifically for phosphorylation and N-terminal acetylation using X!Tandem software (X!Tandem Tornado 2008.02.01.3, http://www.thegpm.org), which are very common post-translational modifications in yeast and were previously described for Pdc1p, Tdh1p and Adh1p (http://www.ibgc.u-bordeaux2.fr/YPM/, http://www.phosphogrid.org).

For Pdc1p, we were unable to identify the corresponding post-translational modifications. For Adh1p, X!Tandem software identified an N-terminal acetylation (after methionine excision) harbored by Adh1p-4808 (*SIPETQK), while Adh1p-4799, the only remaining identified isoform, corresponded to the native protein: more than 99% of Adh1p-4808 N-terminal peptide was acetylated, while around 79% of the Adh1p-4799 N-terminal peptide was non-acetylated for both D2 and D3 strains. In theory, we would expect 100% versus 0% of acetylated residue for Adh1p-4808 and the inverse for Adh1p-4799, but Adh1p is an abundant protein and cross-contamination inevitably occurs between the acetylated and non-acetylated isoforms.

For Tdh1p-2824, we identified a phosphorylated serine (position 201, TA*SGNIIPSSTGAAK) that discriminated Tdh1p-2824 and Tdhp-2757 (that may correspond to the native protein): Tdh1p-2824 displayed a 10-fold enrichment in the phosphorylated residue (TAS*GNIIPSSTGAAK) in comparison with Tdh1p-2757 for both D2 and D3 strains, confirming the occurrence of Tdh1p-phosphorylation. The post-translational modifications associated with the two other isoforms of Tdh1p (2729 and 2775) were not identified using X!Tandem software, but multiple combination of post-translational modifications are possible, rendering hazardous their identification by mass spectrometry.

PROTICdb release

2-DE gels were released online using PROTICdb database at the following URL: http://moulon.inra.fr/protic/adaptalevure (login: adaptalevure, password: review). PROTICdb is a web-based application designed for large-scale proteomic programs to store and query data related to protein separation by 2-DE and protein identification by MS (20, 21). Initially designed for plant proteomics, PROTICdb can also be applied to animal and microorganism proteomics. PROTICdb is fully integrated in the World-2DPAGE Portal (http://www.expasy.org/world-2dpage/), allowing to perform queries simultaneously on all federated proteomics databases. In our case, one representative 2-DE gel for each 27 medium x strain combinations were released online. Annotated 2-DE gel images can be displayed through the "Gel Browser," and identified spots are tagged by red crosses. When mouse pointer is moved over a spot, identification data appear. By clicking on a spot, a full report about gel image analysis (quantification, *etc*.) and MS experiment is generated , including peptides sequences and scores either from MALDI-TOF identification (MASCOT software) or NanoHPLC-MS/MS (SEQUEST software). A second tool, the "GelComparator" is useful for visual comparison of up to four gel images simultaneously and offers zooming and image display synchronization functionalities (20, 21). Full spot report is also available from this tool by clicking on a spot. The whole data set, including phenotypic and proteomic data, is also available online (http://moulon.inra.fr/protic/adaptalevure, login: adaptalevure, password: review).

Statistical analyses

The variation of each isoforms or enzyme (in the later case all the isoforms of the enzyme were summed) was investigated using the lme4 package (R program), through the following mixed model of ANOVA:

$Z = \mu + medium_i + strain_j + block_k + position_l + batch_m + medium * strain_{ij} + \varepsilon_{ijklm}$

where Z is the variable, *medium* is the medium effect (i = 1,2,3), *strain* is the strain effect (j = 1, …, 9), *block* is the random block effect (effect of each weekly experimental repetition, k = 1, …, 11), *position* is the random position effect (bioreactor position, l = 1, …, 15), *batch* is the random 2-DE batch effect (m = 1, …, 6), *medium * strain* is the interaction effect between medium and strain factors, and ε is the residual error. For some enzymes/isoforms, data transformation was necessary to obtain normally distributed residues (log, inverse or square-root transformation, see Table 1). For further analyses (hierarchical clustering, PCA, LDA, regression analysis, etc.), we used the data predicted by the ANOVA model, *i.e.* data corrected for the random effects (*block*, *position* and *batch* effects). In addition, the inverse transformation (if any) was applied so that all enzymes/isoforms display comparable abundance range. The final dataset is available as Supplementary Information (Dataset 4) and online (http://moulon.inra.fr/protic/adaptalevure, login: adaptalevure, password: review). Note that the mean coefficient of variation for all considered isoforms is 18.4%, which is particularly accurate for proteomic studies and allows pertinent subsequent analyses.

Since classical significance tests are controversial for fixed effects in mixed models, we used an alternative method based on Markov Chain Monte Carlo (MCMC) sampling by means of R's *language R* package version 2.14.2 (4, 22). *P*-values were then adjusted for multiple testing using Benjamini-Hochberg methods. For each variable, the homogeneity of the

variance was assessed using a Levene test by means of R's *car* package version 2.14.2 (4), as well as the normality of residual distribution using a Shapiro test (4).

Hierarchical clustering: R's pvclust package (23) was used to make hierarchical clustering (ward's clustering method and Euclidean distances) and to compute p-values for each cluster by multi-scale bootstrap resampling. For Figure 3A, the proteomic data included the relative abundance (% of normalization spots) of all the isoforms involved in glycolysis, ethanol, acetate and glycerol pathway (mean of three replicate for each strain * medium combination, data mean-centered and scaled). For Figure 3B, the genetic data included the genotyping of the nine food-processing strains using 8 microsatellites markers, as detailed previously (1). Genetic hierarchical clustering was based on presence/absence of the different alleles, using mean-centered and scaled data.

Proteomic-trait relationship: to analyze the putative links between metabolic and life-history traits (CO_2 specific flux, metabolite contents, K and cell size) and proteomic data, we conducted multiple linear regression using R's *lm* function (4) to find enzymes/isoforms whose abundance was significantly related to phenotypic traits. More precisely, stepwise regression in both directions was used to select a subset of enzymes/isoforms explaining best the phenotypic variation (taking into account the partial correlation existing between enzymes/isoforms), on the basis of Akaike Information Criterion 'AIC' (24). The best model (*i.e.* lowering AIC criterion) was submitted to bootstrap (1000 runs, lmg metrics) to estimate the relative importance of each enzyme/isoform as well as confidence intervals using R's package *relaimpo* (25). For such functional analysis, invariant enzymes/isoforms (as shown by ANOVA) were not included.

The different equations were as follow:

CO_2 flux = 2.31 10^{e-11} + 7.21 10^{e-11} x $A_{TDH1-2757}$ - 1.43 10^{e-10} x $A_{PDC1-1589}$ + 1.70 10^{e-10} x $A_{TDH2-4872}$ - 2.30 10^{e-10} x $A_{TDH1-2824}$ - 2.46 10^{e-10} x $A_{RHR2-3003}$ - 2.10 10^{e-10} x $A_{HOR2-3129}$ + 2.50 10^{e-10} x $A_{TDH1-2729}$ + 3.13 10^{e-10} x $A_{ALD6-1565}$ + 1.16 10^{e-10} x $A_{ADH1-4808}$ - 1.21 10^{e-10} x $A_{PGK1-Acidic}$ + 5.94 10^{e-11} x $A_{PGK1-interm}$ + 7.17 10^{e-11} x $A_{ENO2-bAsic}$ - 1.60 10^{e-10} x $A_{ADH1-4799}$ + 6.33 10^{e-11} x $A_{ENO2-interm}$ - 8.94 10^{e-11} x $A_{TDH2-4740}$ - 2.1 10^{e-10} x $A_{PGK1-bAsic}$ + 1.39 10^{e-10} x $A_{TDH2-4732}$ + 3.08 10^{e-11} x $A_{PYK1-1310}$

Ethanol = 1.32 10^{e-8} + 1.40 10^{e-7} x $_{A}PDC1-4854$ - 5.20 10^{e-8} x $A_{TDH1-2775}$ - 1.34 10^{e-7} x $A_{TDH2-4740}$ + 1.72 10^{e-7} x $A_{PGK1-Acidic}$ + 1.36 10^{e-7} x $A_{RHR2-3003}$ - 3.28 10^{e-7} x $A_{TDH2-4872}$ + 1.09 10^{e-7} x $A_{ADH1-4808}$ - 3.08 10^{e-8} x $A_{TDH1-2757}$ - 3.01 10^{e-8} x $A_{ENO2-bAsic}$

Glycerol = 1.06 10^{e-9} + 1.44 10^{e-8} x $A_{PDC1-4854}$ - 1.26 10^{e-8} x $A_{ADH1-4799}$ - 4.60 10^{e-9} x $A_{TDH1-2775}$ + 1.67 10^{e-8} x $A_{ADH1-4808}$ + 1.48 10^{e-8} x $A_{TDH2-4732}$ - 9.39 10^{e-9} x $A_{TDH2-4740}$ - 2.66 10^{e-8} x $A_{TDH2-4872}$ - 2.97 10^{e-9} x $A_{TDH1-2757}$ - 2.63 10^{e-9} x $A_{ENO2-bAsic}$

Acetate = 2.76 10^{e-10} + 2.98 10^{e-9} x $A_{PDC1-4854}$ - 4.47 10^{e-9} x $A_{ADH1-4799}$ - 5.50 10^{e-9} x $A_{PGK1-bAsic}$ + 3.31 10^{e-9} x $A_{ADH1-4808}$ - 2.91 10^{e-9} x $A_{PYK1-1630}$ - 1.11 10^{e-9} x $A_{TDH1-2757}$ + 3.26 10^{e-9} x $A_{TDH1-2729}$

Biomass = 1.41 10^{e-9} - 9.84 10^{e-9} x $A_{ADH1-4799}$ - 1.49 10^{e-8} x $A_{ALD6-1565}$ - 7.01 10^{e-9} x $A_{PGK1-Acidic}$ - 2.66 10^{e-9} x $A_{PDC1-1605}$ + 2.55 10^{e-9} x $A_{ENO2-interm}$ + 2.051 10^{e-9} x $A_{ENO2-bAsic}$ + 5.42 10^{e-9} x $A_{TDH1-2729}$ + 2.00 10^{e-9} x $A_{TDH3-Acidic}$

K = 2.39 10e7 + 8.98 10e8 x $A_{ADH1-4799}$ – 7.91 10e8 x $A_{ADH1-4808}$ + 3.13 10e8 x $A_{PGK1-Acidic}$ + 1.01 10e9 x $A_{PGK1-bAsic}$ + 5.20 10e8 x $A_{TDH1-2824}$ + 6.43 10e8 x $A_{ALD6-1565}$ – 1.53 10e8 x $A_{PDC1-4854}$ + 1.26 10e8 x $A_{TDH2-4740}$ – 2.60 10e8 x $A_{TDH2-4732}$ – 1.68 10e8 x $A_{TDH1-2729}$

Cell size = 6.57 - 17.5 x $A_{ADH1-4799}$ + 9.01 x $A_{ADH1-4808}$ - 3.80 x $A_{PDC1-1589}$ + 3.74 x $A_{ENO2-bAsic}$ + 2.96 x $A_{ENO2-interm}$ - 8.78 x $A_{PGK1-Acidic}$ - 24.0 x $A_{ALD6-1565}$ + 6.56 x $A_{TDH2-4732}$ - 8.47 x $A_{TDH1-2824}$ + 9.42 x $A_{TDH1-2729}$ - 5.74 x $A_{PGK1-bAsic}$ + 1.87 x $A_{TDH3-Acidic}$ - 7.18 x $A_{RHR2-3003}$ + 4.83 x $A_{PYK1-1630}$ + 25.1 x $A_{PFK1-Acidic}$ - 1.32 x $A_{PGK1-interm}$

with $A_{enzyme-isoform}$ designing the abundance of the corresponding isoform.

Impact of human domestication: a Linear Discriminant Analysis (LDA) was made to discriminate beer, distillery and wine strains on the basis of the relative spot abundance (% of normalization spots) using R's *mda* package (26). Discriminant isoforms were identified through stepwise variable selection and through the calculation of the « ability to separate » (AS) criterion using R's *klaR* package (27). Briefly, the matrix of posterior probabilities predicted by the LDA was used to include/exclude every single isoform and to estimate the AS criterion for all the isoforms retained by the model.

Supplementary Tables

Supplementary Table 1. Origins of *Saccharomyces cerevisiae* strains.

Food-processing parental strain	Monosporic derivate	Collection / supplier[a]	Food origin	Area of origin	Reproductive mode	Ploidy level
CLIB-382	B1	CIRM-Levures	Brewery	Japan	Homothallic	Diploid
NRRL-Y-7327	B2	NRRL	Brewery	Tibet	Heterothallic	Diploid
CLIB-294	D1	CIRM-Levures	Distillery	France	Homothallic	Diploid
Alcotec 24	D2	Hambleton Bard	Distillery	UK	Heterothallic	Diploid
963	D3	Confidential	Distillery	Confidential	Heterothallic	not available
CLIB-328	E1	CIRM-Levures	Enology	UK	Homothallic	Diploid
BO213	E2[b]	LAFFORT Œnologie	Enology	France	Homothallic	Diploid
VL1	E3[b]	LAFFORT Œnologie	Enology	France	Homothallic	diploid
F10	E4[b]	LAFFORT Œnologie	Enology	France	Homothallic	diploid

[a] CIRM-Levures (http://www.inra.fr/internet/Produits/cirmlevures); NRRL (http://nrrl.ncaur.usda.gov); Hambleton Bard (http://www.hambletonbard.com); LAFFORT

Oenologie (http://www.laffort.com/); [b] E2 E3 and E4 were previously referenced as: SB, GN and G-4A respectively (28, 29).

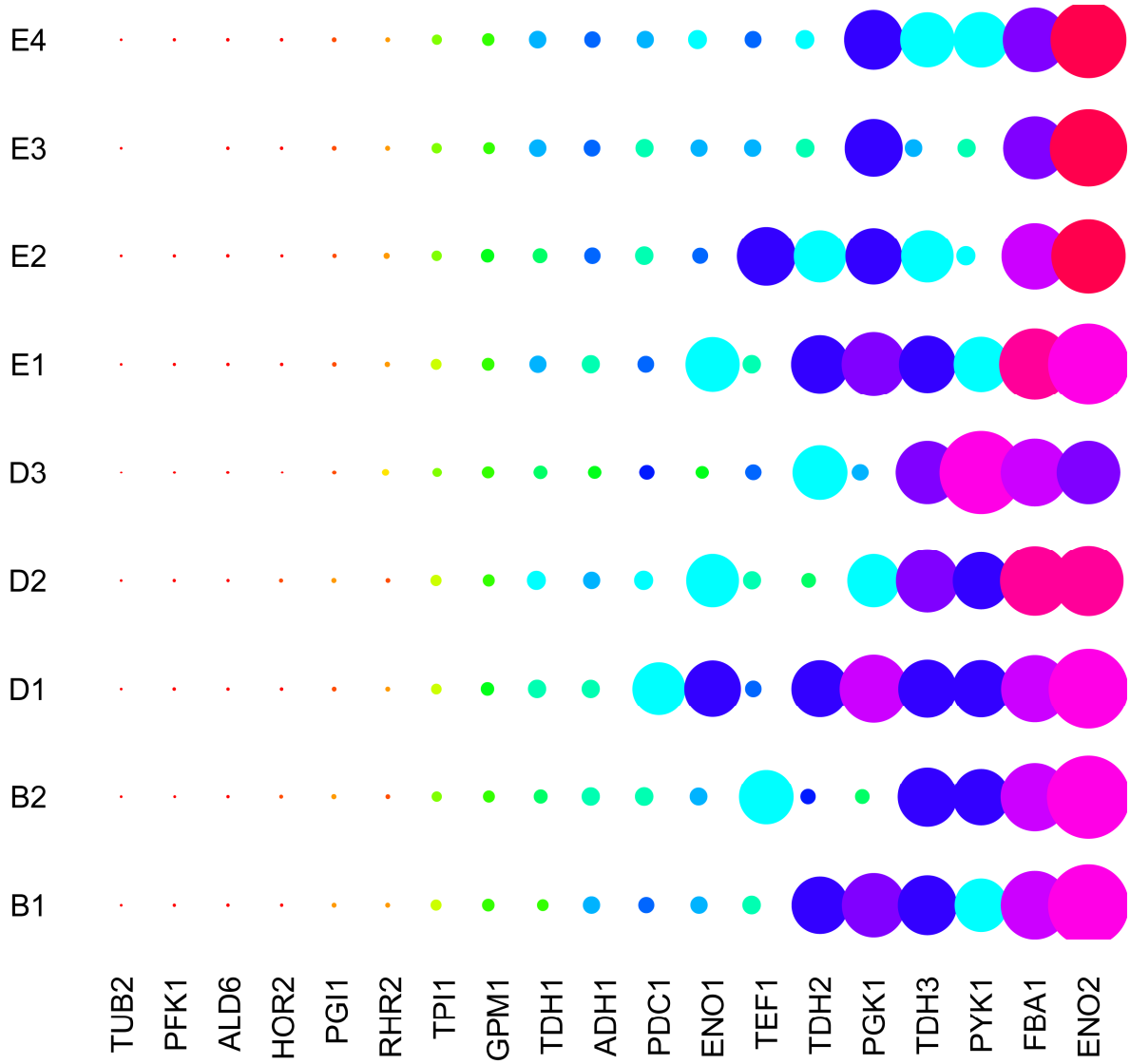
Supplementary Table 2. Main differences between the synthetic bakery, brewery and enology media.

Main components	BAM - Bakery medium	BREM - Brewery medium	WIM - Enology medium
Glucose	80 g L^{-1}	80 g L^{-1}	220 g L^{-1}
Assimilable nitrogen	0.4 g L^{-1}	0.4 g L^{-1}	0.2 g L^{-1}
pH	5.5	4.5	3.5
Sorbitol (for osmotic pressure)	150 g L^{-1}	No	No
Anaerobic growth factors	No	No	Yes

See Albertin at al. (2) for full composition.

Supplementary Figures

Supplementary Figure 1. Average protein abundances in each strain.



In order to perform comparisons between proteins, we summed the abundances of all the spots corresponding to the same proteins. An analysis of variance was performed with Protein, Medium and Strain effects, as well as all possible interactions. All the effects were found significant, but differences between proteins were by far the most important effect. Average protein abundances for each strain were compared using pairwise t-tests. As most of the comparisons were highly significant, a pvalue of 0.05 corresponded to a 0.005 False

Discovery Rate was chosen. Each color represents a different class of abundance: enzymes with different colors have significant different abundance means. Circle colors result from a hierarchical clustering of the means based on euclidian distances. The tree was cut at the height corresponding to the 95% quantile of the Student distribution with 982 degrees of freedom times the average standard deviation between two means, and resulted into 20 classes of abundances. Proteins are ranked from lower to higher average abundances. Strains are ranked according to their industrial origin. Circle size is proportional to the average abundance of the corresponding protein for each strain.

References

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

2. Albertin, W., Marullo, P., Aigle, M., Dillmann, C., de Vienne, D., Bely, M., and Sicard, D. (2011) Population size drives the industrial yeast alcoholic fermentation and is under genetic control. *Appl Environ Microbiol* 77, 2772–2784.

3. El Haloui, N., Picque, D., and Corrieu, G. (1988) Alcoholic fermentation in winemaking: On-line measurement of density and carbon dioxide evolution. *Journal of Food Engineering* 8, 17-30.

4. R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

5. Schulze, U., Liden, G., Nielsen, J., and Villadsen, J. (1996) Physiological effects of nitrogen starvation in an anaerobic batch culture of Saccharomyces cerevisiae. *Microbiology* 142 (Pt 8), 2299-2310.

6. Damerval, C., de Vienne, D., Zivy, M., and Thiellement, H. (1986) Technical improvments in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. *Electrophoresis* 7, 52-54.

7. Mechin, V., Consoli, L., Le Guilloux, M., and Damerval, C. (2003) An efficient solubilization buffer for plant proteins focused in immobilized pH gradients. *Proteomics* 3, 1299-1302.

8. Görg, A., Postel, W., Weser, J., Günther, S., Strahler, J. R., Hanash, S. M., and Somerlot, L. (1987) Elimination of point breaking on silver stained two-dimensional gels by addition of iodoacetamine to the equilibration buffer. *Electrophoresis* 8, 122-124.

9. Kobi, D., Zugmeyer, S., Potier, S., and Jaquet-Gutfreund, L. (2004) Two-dimensional protein map of an "ale"-brewing yeast strain: proteome dynamics during fermentation. *FEMS Yeast Res* 5, 213-230.

10. Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuentes, R., and Payne, W. E. (1997) Proteome studies of Saccharomyces cerevisiae: identification and characterization of abundant proteins. *Electrophoresis* 18, 1347-1360.

11. Perrot, M., Sagliocco, F., Mini, T., Monribot, C., Schneider, U., Shevchenko, A., Mann, M., Jeno, P., and Boucherie, H. (1999) Two-dimensional gel protein database of Saccharomyces cerevisiae (update 1999). *Electrophoresis* 20, 2280-2298.

12. Fievet, J., Dillmann, C., Lagniel, G., Davanture, M., Negroni, L., Labarre, J., and de Vienne, D. (2004) Assessing factors for reliable quantitative proteomics based on two-dimensional gel electrophoresis. *Proteomics* 4, 1939-1949.

13. Mechin, V., Balliau, T., Chateau-Joubert, S., Davanture, M., Langella, O., Negroni, L., Prioul, J. L., Thevenot, C., Zivy, M., and Damerval, C. (2004) A two-dimensional proteome map of maize endosperm. *Phytochemistry* 65, 1609-1618.

14. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.

15. Boles, E., Schulte, F., Miosga, T., Freidel, K., Schluter, E., Zimmermann, F. K., Hollenberg, C. P., and Heinisch, J. J. (1997) Characterization of a glucose-repressed pyruvate kinase (Pyk2p) in Saccharomyces cerevisiae that is catalytically insensitive to fructose-1,6-bisphosphate. *J Bacteriol* 179, 2987-2993.

16. Pronk, J. T., Yde Steensma, H., and Van Dijken, J. P. (1996) Pyruvate metabolism in Saccharomyces cerevisiae. *Yeast* 12, 1607-1633.

17. Leskovac, V., Trivic, S., and Pericin, D. (2002) The three zinc-containing alcohol dehydrogenases from baker's yeast, Saccharomyces cerevisiae. *FEMS Yeast Res* 2, 481-494.

18. Saint-Prix, F., Bonquist, L., and 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.

19. Reinders, J., Zahedi, R. P., Pfanner, N., Meisinger, C., and Sickmann, A. (2006) Toward the complete yeast mitochondrial proteome: multidimensional separation techniques for mitochondrial proteomics. *J Proteome Res* 5, 1543-1554.

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

21. Ferry-Dumazet, H., Houel, G., Montalent, P., Moreau, L., Langella, O., Negroni, L., Vincent, D., Lalanne, C., de Daruvar, A., Plomion, C., Zivy, M., and Joets, J. (2005) PROTICdb: a web-based application to store, track, query, and compare plant proteome data. *Proteomics* 5, 2069-2081.

22. Baayen, R. H. (2009) languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics".

23. Suzuki, R., and Shimodaira, H. (2009) Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling.

24. Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716-723.

25. Groemping, U. (2006) Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software* 17.

26. Leisch, F., Hornik, K., and Ripley, B. D. (2009) mda: Mixture and flexible discriminant analysis.

27. Weihs, C., Ligges, U., Luebke, K., and Raabe, N. (2005) klaR Analyzing Business Cycles. In: Baier, D., Decker, R., and Schmidt-Thieme, L., eds. *Data Analysis and Decison Support*, pp. 335-343, Springer-Verlag, Berlin.

28. Marullo, P., Bely, M., Masneuf-Pomarede, I., Pons, M., Aigle, M., and 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.

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