A systems approach to elucidate heterosis of protein abundances in yeast


To cite this version:


HAL Id: hal-02636704
https://hal.inrae.fr/hal-02636704
Submitted on 8 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
A Systems Approach to Elucidate Heterosis of Protein Abundances in Yeast*\textsuperscript{S}

Mélisande Blein-Nicolas\textsuperscript{A}, Warren Albertin\textsuperscript{B,C}, Telma da Silva\textsuperscript{A,D}, Benoît Valot\textsuperscript{E}, Thierry Balliau\textsuperscript{A}, Isabelle Masneuf-Pomarède\textsuperscript{C,F}, Marina Bely\textsuperscript{C}, Philippe Marullo\textsuperscript{C,G}, Delphine Sicard\textsuperscript{H,I}, Christine Dillmann\textsuperscript{H}, Dominique de Vienne\textsuperscript{H}, and Michel Zivy\textsuperscript{J,K}

Heterosis is a universal phenomenon that has major implications in evolution and is of tremendous agro-economic value. To study the molecular manifestations of heterosis and to find factors that maximize its strength, we implemented a large-scale proteomic experiment in yeast. We analyzed the inheritance of 1,396 proteins in 55 interspecific hybrids obtained from \textit{Saccharomyces cerevisiae} and \textit{S. uvarum} that were grown in grape juice at two temperatures. We showed that the proportion of heterotic proteins was highly variable depending on the parental strain and on the temperature considered. For interspecific hybrids, this proportion was higher at non-optimal temperature. Unexpectedly, heterosis for protein abundance was strongly biased toward positive values in interspecific hybrids but not in intraspecific hybrids. Computer modeling showed that this observation could be accounted for by assuming concave relationships between protein abundances and their controlling factors, in line with the metabolic model of heterosis. These results point to nonlinear processes that could play a central role in heterosis. \textit{Molecular & Cellular Proteomics} 14: 10.1074/mcp.M115.048058, 2056–2071, 2015.

Nonadditive inheritance in hybrids, whereby the phenotype of offspring is not the average of the parental phenotypes, is commonly observed in all species. For monogenic traits, the departure from additivity is called dominance (1) or overdominance if the phenotypic value of the hybrid is outside the range defined by the parental values (2). For polygenic traits, it is called heterosis. Heterosis is commonly associated to macroscopic traits, but it also applies to less integrated traits such as metabolite abundances (3, 4), fluxes and enzyme activities (5–7), mRNA, and protein amounts (8).

The concept of heterosis is not universally shared and depends on the scientific communities. Strictly speaking, heterosis is defined as the superiority of the hybrid over the mean parental value (mid-parent heterosis, MPH\textsuperscript{1}) or over its parent exhibiting the highest value (best-parent heterosis, BPH). This definition and the associated terminology are historical and come from the fact that heterosis was commonly associated to traits such as growth rate, biomass, size, yield, or fertility, for which higher values are beneficial. However, lower values can also occur (e.g. (9)). The definition of heterosis has therefore been broadened to include also the cases where the hybrid is below the mean parental value (negative MPH) or below its parent exhibiting the lowest value (worst-parent heterosis, WPH). In this paper, we will adopt this second definition, motivated by its lack of presumption about whether the changes observed in hybrids are beneficial or detrimental.

Heterosis has fascinated scientists and breeders for more than 100 years. It has major implications in evolution and domestication of crop plants (10, 11), and it has been exploited since the 1930s in plant breeding to produce hybrids of high agronomic value (12). In this context, heterosis has proven to efficiently accelerate the process of selection for various crops (reviewed in (13)). Heterosis is opposite to inbreeding depression, that is supposed to be predominantly caused by the homozygosity of deleterious recessive alleles (14). Heterosis can provide a heterozygote advantage by buffering against these alleles and confers genetic plasticity to adapt to environmental changes (11).

\textsuperscript{1} The abbreviations used are: ANOVA: analysis of variance; MPH: mid-parent heterosis; BPH: best-parent heterosis; CFU, colony-forming unit; LC-MS/MS, liquid chromatography-tandem mass spectrometry; RPM, round per minute; TF, transcription factor; WPH, worst-parent heterosis; YPD, yeast extract peptone dextrose; ZALS, zwitterionic acid labile surfactant.
Given the importance of heterosis for agriculture and because it is an intriguing phenomenon, many studies have focused on the understanding of its genetic and molecular bases (11, 15–25). Three nonexclusive hypotheses based on genetic effects are classically put forward to explain heterosis. First, the dominance hypothesis attributes heterosis to complementation: In the hybrid, the recessive alleles are masked by dominant and generally favorable alleles (26, 27). Second, the overdominance hypothesis assumes that heterosis is inherent to the heterozygous state (2, 28). Third, the epistasis hypothesis proposes that heterosis is due to intergenic interactions created in the hybrid (29, 30). Scientists have long sought a unifying theory to account for heterosis, but it is now commonly admitted that this phenomenon likely arises from the combination of several genetic mechanisms, the relative effects of which vary according to the trait, the cross, or the species (23, 25).

These genetic effects are consistent with the factors known to maximize the occurrence of heterosis. When compiling the results obtained so far across numerous studies, it appears that heterosis is of greatest magnitude for highly integrated, and hence polygenic, traits such as crop yield (23, 31); it is larger in allogamous than in autogamous species (31); it requires genetic divergence between parents; and interspecific crosses commonly produce higher levels of heterosis (21, 24, 32). However, these general trends are not sufficient for a reliable prediction of heterosis, which is a major challenge for plant and animal breeding and for biotechnology. Future strategies for heterosis prediction will have to rely both on an accurate description of its manifestations and on the detailed knowledge of the factors that maximize its strength. To address these issues, we performed a large-scale study of heterosis by analyzing the inheritance of the abundance of a high number of proteins in an unprecedented number of yeast hybrids grown in two conditions. The proteomic level is particularly relevant to the large-scale study of heterosis because protein abundances are polygenic molecular traits (33) that can be measured by high-throughput quantitative proteomics (34, 35).

Yeast has only rarely been used to study heterosis (9, 36–41). Yet, it is amenable to large-scale laboratory experiments, and it is of major industrial interest for wine making. Hybrids with exceptional performances were reported in Saccharomyces cerevisiae and S. uvarum Beijerinck or S. kudriavzevii isolated in wine and natural environments showed important biotechnological potential, such as a better robustness than their parents (44–47). In addition, several wine strains empirically selected for their biotechnological properties proved to be interspecific hybrids (48, 49). For all these reasons, we chose to study heterosis for protein abundance in yeast strains from S. cerevisiae and S. uvarum, which are the two main species associated with grape juice fermentation (50).

By using shotgun proteomics, we analyzed more than 1,300 proteins in an experimental design, including 11 parental strains of S. cerevisiae and S. uvarum and their 55 intra- and interspecific hybrids, which were grown at two temperatures to take into account adaptation differences between parental species (18 °C and 26 °C optimal for S. uvarum and S. cerevisiae, respectively (44, 51, 52)). We showed that heterosis for protein abundance was strongly biased toward positive values in interspecific hybrids but not in intraspecific hybrids, which, to our knowledge, has never been reported. We also showed that our experimental results were consistent with results obtained from modeling approaches assuming nonlinear relationships between protein abundances and their controlling factors.

**EXPERIMENTAL PROCEDURES**

**Yeast Strains**—Four diploid S. uvarum strains, seven diploid S. cerevisiae strains, and their 55 hybrids produced from a half diallel design (53) were analyzed in this study. Parental strains were derived from strains isolated from different geographical locations and from either natural or food-processing origins (Table I): the S. cerevisiae strains were isolated from diverse media (distillery, enology, oak exudate) to maximize the genetic diversity within this species (54); the S. uvarum strains, originating from grape must or cider fermentation, were chosen to cover a wide part of the genetic diversity of the S. uvarum species (Masneuf-Pomarède, I., personal communication). For each original strain, one meiospore was isolated with a micromanipulator (Singer MSM Manual, Singer Instrument, Somerset, UK). All the original strains but Alcotec 24 were homothallic (HO/HO); therefore, fully homozygous diploid strains were spontaneously obtained by fusion of opposite mating type cells. For A24 (ho/ho), one isolated haploid meiospore was diploidized via transient expression of the HO endonuclease (55). All strains were grown at 30 °C in YPD medium containing 1% yeast extract (Difco Laboratories, Detroit, MI), 1% bactopeptone (Difco), and 2% glucose, supplemented or not with 2% agar. When necessary, antibiotics were added at the following concentrations: 100 μg/ml for G418 (Sigma, L’Isle d’Abeau, France), and nourseothricin (Werner bioagent, Jena, Germany) and 300 μg/ml for hygromycin B (Sigma).

**Construction of the Half Diallel Design**—Hybrid construction was performed as described in Albertin et al. (53). Briefly, the 11 diploid parental strains were transformed with a cassette containing the HO allele disrupted by a gene of resistance to either G418 (ho::KanR), hygromycin B (ho::HypR), or nourseothricin (ho::NatR). Strain transformation allowed conversion to heterothallism for the homothallic strains. Then the mating-type (MATa/MATα) of antibiotic-resistant monosporic clones was determined using testers of well-known mating type. For each cross, parental strains of opposite mating type were put in contact 2 to 6 h in YPD medium at room temperature and plated on YPD-agar containing the appropriate antibiotics. Ten independent hybrids per cross were recovered. After recurrent cultures on YPD-agar corresponding to ~ 80 generations, the nuclear chromosomal stability of the hybrids was controlled by pulsed field electrophoresis (CHEF-DR3, Bio-Rad, Marnes-La-Coquette, France) as well as homoplasmy (only one parental mitochondrial genome). One hybrid per cross was finally retained for further experiments.

**Yeast Strain Characterization**—Two polymorphic microsatellites specific to S. cerevisiae (Sc-YFR038 and Sc-YML091 (56)) and two
specific to \textit{S. uvarum} (focus 4 and 9 (57)) were used to discriminate rapidly the hybrids from the parental strains. These four markers were amplified in a multiplex PCR reaction (95 °C for 5 min for initial denaturation step; 95 °C for 30 s, 55 °C for 90 s, and 72 °C for 60 s repeated 35 times; a final elongation step of 30 min at 60 °C). The PCR products were analyzed on an ABI3730 apparatus (Applied Biosystem, Villebon-sur-Yvette, France), and microsatellite lengths were analyzed using the Peak Scanner tool (Applied Biosystem).

**Alcoholic Fermentation in Grape Must**—All the 66 strains (11 parents and 55 hybrids) were grown in the same batch of white grape must obtained from Sauvignon grapes harvested in vineyards in the Bordeaux area (2009 vintage). Tartaric acid precipitation was stabilized, and turbidity was adjusted to 100 NTU (nephelometric turbidity unit) before storage at −20 °C. The nitrogen content was 242 mg.l$^{-1}$, and the pH was 3.3. 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 strain were run in half-diluted must filtered through a 0.45 µm nitrate-cellulose membrane (24 °C, 150 rpm (rounds per minute)) during 24 h, after what one million cells per ml were sampled and added to a final volume of 125 ml of Sauvignon must. Then, fermentations were run into 125 ml glass reactors at two temperatures (18 °C and 26 °C, 300 rpm) and repeated three times independently. In total, 396 alcoholic fermentations were performed with at least two successful fermentations were kept for further conditions, Nieuwkoop, Netherlands) at 7.5 µl.min$^{-1}$ and desalted with 0.1% formic acid and 2% ACN. After 3-10 times in 50 mM of ammonium bicarbonate, proteins were reduced 1 h in 100 mM DTT, alkylated 1 h in 40 mM iodoacetamide, and digested overnight at 37 °C with 1/50 (w/w) trypsin (Promega, Charbonnière, France). Digestion was stopped by adding 0.4% of trifluoroacetic acid (TFA). Peptides were purified on solid phase extraction using polymeric C18 column (Phenomenex, Le Pecq, France) with a washing solution containing 0.06% acetic acid and 3% acetonitrile (ACN). After elution with 0.06% acetic acid and 70% ACN, peptides were speedvac-dried and suspended in 2% ACN and 0.08% TFA.

**LC-MS/MS Analysis**—LC-MS/MS analyses were performed using a NanoLC-Ultra System (nano2DUltira, Eksigent, Les Ulis, France) connected to a Q-Exactive mass spectrometer (Thermo Electron, Waltham, MA, USA). A 700 ng of protein digest were loaded onto a PepMap C18 precolumn (0.3 × 5 mm, 100 Å, 5 µm; NanoSeparations, Nieuwkoop, Netherlands) at 7.5 µl.min$^{-1}$ and desalted with 0.1% formic acid and 2% ACN. After 3 min, the precolumn was connected to a PepMap C18 nanocolumn (0.075 × 150 mm, 100 Å, 3 µm). Buffers were 0.1% formic acid in water (A) and 0.1% formic acid and 100% ACN (B). Peptides were separated using a linear gradient from 5 to 35% buffer B for 40 min at 300 n.l.min$^{-1}$. One run took 60 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, Woburn, MA, USA). Peptide ions were analyzed using Xcalibur 2.2 (Thermo Electron) with the following data-dependent acquisition steps: (1) MS scan (mass-to-charge ratio (m/z) 400 to 1,400, 70 000 resolution, profile mode), (2) MS/MS (17,500 resolution, collision energy = 30%, profile mode). Step 2 was repeated for the eight major ions detected in step 1. Dynamic exclusion was set to 40 s. Xcalibur raw data files were transformed to mzXML open source format using msconvert software in the ProteoWizard 3.0.3706 package (59). During conversion, MS and MS/MS data were centroided.

**MS Data Availability**—The raw MS output files were deposited online using PROTeINDB database (60–62) at the following URL: http://moulon.inra.fr/protic/heteroys3t.
**Protein Identification**—Protein identification was performed using the custom database described in Blein-Nicolas et al. (58). This 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 October 5, 2010 and December 15, 2003, respectively). The proteins of *S. cerevisiae* and of *S. uvarum* encoded by orthologous genes were attributed unique labels. 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 juice was also interrogated. The decoy database comprised the reverse protein sequences of the custom database. Database search was performed with X!Tandem (version 2011.12.01.1; http://www.thegpm.org/TANDEM/) with the following settings. 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 1 ppm. Fragment mass tolerance was 0.02 Th. A refinement search was added with the same settings, except that protein N-ter acetylations were also searched. Only peptides with an E-value smaller than 0.05 were reported. Identified proteins were filtered and sorted by using X!Tandem-Pipeline (version 3.3.0, http://pappso.inra.fr/biinfo/xtandempipeline/). Criteria used for protein identification were (i) at least two different peptides identified with an E-value smaller than 0.05 and (ii) 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 of 0.12% and 1.15% for peptide and protein identification, respectively.

**Peptide Quantification and Processing Intensity Data**—Peptides were quantified based on extracted ion chromatograms using MassChroQ software version 1.2.2 (63) with the parameters given in File S1. The detection threshold on min and max were set at 30,000 and 50,000, respectively. Due to progressive fouling of the quadrupole, sensitivity losses were observed over time, leading to a global decrease of measured intensities, particularly for hydrophobic peptides. To take these sensitivity losses into account, samples were classified according to their running order and divided into five blocks representing homogeneous global intensities. For each peptide, the block effect was retrieved and subtracted from intensity measures by using an analysis of variance (ANOVA). Then, normalization was performed to take into account possible global quantitative variations between LC-MS runs. 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.

Raw data (containing intensity measures of 25,060 peptides) were then filtered to remove (i) dubious peptides for which standard deviation of retention time was superior to 60 s, (ii) peptide × strain × temperature combinations quantified in only one replicate, and (iii) peptides shared by several proteins, representing less than 5% of all the quantified peptides. To avoid bias on the estimation of total protein abundances in hybrids, we removed parent-specific peptides by using peptides presenting presence/absence variation among parental strains as a proxy. However, parent-specific peptides were confounded with species-specific peptides, which represented nearly 65% of the valid peptides. To exploit as far as possible the data available for intraspecies crosses, we thus split the dataset into three subsets: one contained *S. cerevisiae* triplets (hybrid and its parents), another contained *S. uvarum* triplets, and the last one contained interspecific triplets. Parent-specific peptides were removed separately in the three subsets. To finish, in order to estimate the peptide effect properly, peptides quantified in less than four strains × temperature combinations in a given subset of data were removed.

**Detection of Protein Abundance Changes**—Protein abundances were estimated independently in the three subsets of data by using

$$\log(I_{i,pt}) = \theta_{i,pt} + D_i + B_t + C_{str} + \varepsilon_{i,pt}$$

where $I_{i,pt}$ is the normalized intensity value for peptide $i$ in strain $s$, temperature $t$, and replicate $r$, $\theta_{i,pt}$ is the natural logarithm of the abundance of protein $k$ in strain $s$ and temperature $t$, $D_i$ is the normal distribution, and $B_t$ is the biological variation of replicate, $C_{str}$ is an error due to the technical variation of sample $str$, $D_i \sim N(0,\sigma_D^2)$ is an error due to the biological variation of replicate, and $\varepsilon_{i,pt} \sim N(0,\sigma^2)$ is the residual error.

**Estimation of the parameters of the model** was performed as described in Blein-Nicolas et al. (64). Protein abundance changes were detected by multiple test procedure across four different contrasts: (i) hybrid-mean of parents, (ii) hybrid–parent 1, (iii) hybrid–parent 2, (iv) parent 1–parent 2. Since several couples of strains × temperature combinations and several proteins were tested, $p$-values were adjusted for multiple testing by a Benjamini–Hochberg procedure (65). Of note, the statistical power was reduced in the subset of data containing interspecific hybrids compared with the two other subsets since intensity data were more drastically filtered (on average, there were 6.2 peptides per protein in the subset containing interspecific hybrids against 8.9 and 8.2 in the subsets containing *S. cerevisiae* hybrids and *S. uvarum* hybrids, respectively).

**Data Analysis**—Protein abundances estimated in different subsets of data were not directly comparable. To overcome this drawback, the subset of data containing interspecific hybrids (further named B for between) was taken as a reference, and the following linear regression was performed for each protein in the subsets of data containing interspecific hybrids (referred to as W for within):

$$\theta_{i,pt}^W = a + b \theta_{i,pt}^B + \epsilon_{i,pt}$$

where $\theta_{i,pt}^B$ and $\theta_{i,pt}^W$ are the abundances estimated in parental strain $p$ at temperature $t$ in the subsets of data $W$ and $B$, respectively, $a$ and $b$ are the parameters of intercept and slope, respectively, and $\epsilon_{i,pt}$ is the residual error.

The median of the coefficient of determination $R^2$ was 0.83, indicating that the protein abundances estimated separately in different subsets of data were globally well correlated. For proteins with $b$ significantly different from 0 (adjusted $p < 0.05$), estimators of $a$ and $b$ were used to correct the abundances estimated for intraspecies hybrids:

$$\theta_{i,pt}^W = (\theta_{i,pt}^W - \hat{a})/\hat{b}$$

where $\theta_{i,pt}^W$ is the abundance estimated in hybrid $h$ at temperature $t$ in the subset $W$. Then, protein abundances in the subset $B$ were gathered with the $\theta_{i,pt}^W$ computed in the subset $W$.

A total of 615 proteins quantified in more than 122 strains × temperature combinations were kept for data representation as heat map and principal component analysis. Missing data were imputed from a uniform distribution with minimum = 0 and maximum = 10^6 under the hypothesis that they corresponded to low abundance values.

All data analyses and graphical representations were performed using R version 3.0.2 (66). Appropriate statistical tests were used for each kind of data: $\chi^2$ tests were used to compare distributions; Student and Mann–Whitney tests were used to compare means in the
Heterosis for Protein Abundance in Yeast

case of normally distributed and nonparametric data, respectively; Pearson and Kendall correlation tests were used to analyze associations between normally distributed data and counting data, respectively; and analyses of variance were performed by using a linear model for normally distributed data or generalized linear model with Poisson distribution for counting data. Residuals were examined for normality and independence.

**In Silico Simulation of Heterosis**—We wrote an R program to simulate heterosis in the framework of a nonlinear genotype–phenotype relationship (File S2). We assumed that the protein abundances were controlled by 10 factors \( j = 1, \ldots, 10 \), the inheritance of which was either additive or nonadditive. We varied the number of polymorphic factors from 1 to 10. Each factor was defined by its value \( E_j \) (concentration or activity of factor \( j \)) and its contribution to the abundance of the controlled protein, \( a_j \). The \( E_j \) values were drawn in a gamma distribution (mean = 6, coefficient of variation = 0.3; –\( \Gamma(11.11, 0.54) \)) and the \( a_j \) in a uniform distribution \(-U(1,10)\). These parameters were chosen to get distributions of protein abundances similar to the distributions observed for the most abundant proteins, which are right skewed and have coefficients of variation around 0.2–0.3. For a given protein, the \( a_j \)'s were the same for the parents and their hybrid. Homozygous parents were created by randomly attributing an \( E_j \) value (allelic value) to each factor. The Euclidean distance between two parents \( j \) and \( j' \) was computed as follows:

\[
D_{jj'} = \sqrt{\frac{1}{10} \sum_{i=1}^{10} (E_{ij} - E_{ij'})^2}
\]

Protein abundances were computed assuming a concave relationship between the factors and the abundances. To this end, we used a simple hyperbolic function derived from that of the metabolic control theory (67):

\[
A_j = \frac{X}{\sum_{i=1}^{10} (\frac{1}{a_i E_i})}
\]

where \( A_j \) is the abundance of the protein in parent \( j \) and \( X \) is a constant.

To compute protein abundance in the hybrids, we took into account an index of inheritance \( X_{pp} \) drawn in a normal distribution \( N(0.5, 0.15) \). If \( X_{pp} = 0.5 \), the factor was additively inherited in the hybrid between parents \( j \) and \( j' \); otherwise there was positive or negative deviation from additivity. If \( X_{pp} = 0 \) (respectively \( X_{pp} = 1 \)), there was strict dominance of parent \( j \) (respectively \( j' \)) over parent \( j' \) (respectively \( j \)). Therefore, the abundance of a protein in a hybrid is written:

\[
A_{jj'} = \frac{X}{\sum_{i=1}^{10} (\frac{1}{(X_{pp} n_i - 1) a_i E_i})}
\]

where \( A_{jj'} \) is the abundance of the protein in the hybrid between parents \( j \) and \( j' \), and \( n_i = E_{i j'} / E_{i j} \).

The simulations were performed with 20,000 proteins (Figs. 8C and 8E and Fig. S6) or 400,000 proteins (Figs. 8B and 8D)

**RESULTS**

**Protein Quantification by LC-MS/MS**—A total of 396 alcoholic fermentations (66 strains \( \times \) 2 temperatures \( \times \) 3 replicates) were performed, of which 31 failed due to the poor fermenting abilities of some strains (Table S1). Yeast samples taken from the 365 successful fermentations were analyzed by shotgun label-free quantitative proteomics. Detailed information on all the peptides and proteins identified in all LC-MS/MS runs are shown in Table S2 and S3, respectively.

In total, 1,396 proteins were quantified in at least one strain \( \times \) temperature combination (Table S5). Of them, 1,396 proteins were quantified both in a hybrid and its parents at the same temperature. These 1,396 proteins belonged to 16 functional categories following the MIPS Functional Catalogue Database (68) (Fig. S1, Table S6). Metabolism was the most represented category, with 534 proteins (31.1% coverage; Fig. S1).

Representation of protein abundances as a heat map showed that the strain \( \times \) temperature combinations were separated in three main clusters corresponding globally to \( S. uvarum \) strains, interspecific hybrids, and \( S. cerevisiae \) strains (Fig. 1, clusters A, B, and C, respectively). Interspecific hybrids differed from all the other strains by a cluster of proteins that were globally more abundant than in the other strains (Fig. 1, cluster II). \( S. uvarum \) strains and \( S. cerevisiae \) strains differed by two clusters of proteins: one containing proteins that were more abundant in \( S. cerevisiae \) (Fig. 1, cluster I) and one containing proteins that were more abundant in \( S. uvarum \) (Fig. 1, cluster III). Except for a particular group containing the parental strain D2 and all its descendants including interspecific hybrids (Fig. 1, cluster D), the strains \( \times \) temperature combinations within the clusters A, B, and C were grouped by temperature.

**Protein Inheritance Patterns**—To analyze the inheritance of protein abundances at a given temperature, we considered the triplets (formed by one hybrid and its parents) where at least two successful fermentations were obtained for each member. This was the case for 53 triplets at 18 °C and for 44 triplets at 26 °C (Table S1). For each protein \( \times \) hybrid \( \times \) temperature combination, we computed the deviation from additivity (\( d \)) as the difference between hybrid abundance and mid-parental abundance. A protein was considered as heterotic whenever \( d \) was significantly different from zero (Wald test, adjusted \( p < .05 \), Table S7). A total of 97,360 protein \( \times \) hybrid \( \times \) temperature combinations were examined. For 65.2% (63,469) of them, no significant abundance variation was detected neither between a hybrid and its parent nor between parents (invariant proteins). The remaining 33,891 protein \( \times \) hybrid \( \times \) temperature combinations were classified depending on their inheritance pattern (Tables S7 and S8, Fig. 2): 66.8% (22,634) displayed additivity; 11.7% (3,965) displayed negative or positive MPH, meaning that the protein abundance in the hybrid was within the parental range; 11.0% (3,746) displayed BPH or WPH, meaning that the protein abundance in the hybrid fell outside the parental range; and 10.5% (3,546) corresponded to cases of unresolved heterosis.
because statistical tests did not allow us to distinguish between mid-parent and best/worst-parent heterosis.

The proportion of heterotic proteins per hybrid × temperature combination (invariant proteins omitted) was highly variable, ranging from 8.4 to 61.2% with a median at 31.4% (Table II). Globally, hybrids having at least one \( S. \text{cerevisiae} \) strain as a parent showed more heterotic proteins at 18 °C than at 26 °C (Fig. 3). On the contrary, \( S. \text{uvarum} \) intraspecific hybrids showed slightly more heterotic proteins at 26 °C than at 18 °C (Fig. 3).

**Interspecific Hybrids Exhibit Specific Characteristics Regarding Protein Abundance Inheritance**—We further analyzed heterosis for protein abundance in inter- versus intraspecific hybrids. By examining the distribution of relative additivity deviation (computed as \( d/m \), where \( m \) is the parental mean), we showed that \( d/m \) was globally higher in inter- than in intraspecific hybrids (Fig. 4A, Fig. S2). In addition, the proportion of heterotic proteins with positive \( d \) values was, on average, much higher in inter- than intraspecific hybrids (78.8%, 52.3%, and 42.6% in interspecific, \( S. \text{cerevisiae} \) and \( S. \text{uvarum} \) hybrids, respectively; Fig. 4B). This indicates a strong bias toward positive heterosis in interspecific hybrids.

---

**Fig. 1.** *Heat map representation of the estimated protein abundances.* Each line corresponds to a protein and each column to a strain × temperature combination. A total of 615 proteins that were quantified in at least 122 strains × temperature combinations are presented (see experimental procedures for details). For each protein, abundance values were scaled and represented by a color code as indicated by the color-key bar: blue for low abundances and red for high abundances. Hierarchical clusterings of the strains (top) and of the proteins (left) were built by using Euclidean distances and Ward aggregation method. Letters on the top indicate clusters of strain × temperature combinations presenting similar proteomes. Roman numerals on the left indicate clusters of proteins exhibiting similar abundance patterns. Membership of a protein to the set \( H^+ \) (see results and Fig. 6) is shown in brown on the right. The type of strain and the growth temperature is indicated in brown at the bottom.

**Fig. 2.** *Inheritance pattern of the proteins exhibiting abundance variation between hybrid and parental strains.* P1, parent 1; H, hybrid; P2, parent 2.
We next looked whether the temperature affected protein inheritance similarly in interspecific hybrids compared with intraspecific hybrids. For the majority of the protein × hybrid combinations (82.5%), the protein was heterotic at only one temperature. This indicates that heterosis for protein abundance is generally dependent on the temperature. For the remaining 17.5%, four scenarios were possible depending on the sign of $d$: positive at both 18 °C and 26 °C (+/+), negative at both 18 °C and 26 °C (−/−), positive at 18 °C and negative at 26 °C (±), negative at 18 °C and positive at 26 °C (−/+).

We next looked whether the temperature affected protein inheritance similarly in interspecific hybrids compared with intraspecific hybrids. For the majority of the protein × hybrid combinations (82.5%), the protein was heterotic at only one temperature. This indicates that heterosis for protein abundance is generally dependent on the temperature. For the remaining 17.5%, four scenarios were possible depending on the sign of $d$: positive at both 18 °C and 26 °C (+/+), negative at both 18 °C and 26 °C (−/−), positive at 18 °C and negative at 26 °C (±), negative at 18 °C and positive at 26 °C (−/+).

We next looked whether the temperature affected protein inheritance similarly in interspecific hybrids compared with intraspecific hybrids. For the majority of the protein × hybrid combinations (82.5%), the protein was heterotic at only one temperature. This indicates that heterosis for protein abundance is generally dependent on the temperature. For the remaining 17.5%, four scenarios were possible depending on the sign of $d$: positive at both 18 °C and 26 °C (+/+), negative at both 18 °C and 26 °C (−/−), positive at 18 °C and negative at 26 °C (±), negative at 18 °C and positive at 26 °C (−/+).

We next looked whether the temperature affected protein inheritance similarly in interspecific hybrids compared with intraspecific hybrids. For the majority of the protein × hybrid combinations (82.5%), the protein was heterotic at only one temperature. This indicates that heterosis for protein abundance is generally dependent on the temperature. For the remaining 17.5%, four scenarios were possible depending on the sign of $d$: positive at both 18 °C and 26 °C (+/+), negative at both 18 °C and 26 °C (−/−), positive at 18 °C and negative at 26 °C (±), negative at 18 °C and positive at 26 °C (−/+).
Globally, interspecific hybrids presented an excess of $+/+$ scenarios (451 over 656, $\chi^2$ test, $p = 7.6 \times 10^{-19}$; Fig. 4C). This result holds true for nearly all interspecific hybrids (Fig. S3A). Regarding intraspecific hybrids, $S.\ cerevisiae$ hybrids presented an excess of $\pm$ scenarios (252 over 719, $\chi^2$ test, $p = 1.1 \times 10^{-15}$; Fig. S4A), while $S.\ uvarum$ hybrids lacked $\pm$ scenarios (6 over 79, $\chi^2$ test, $p = 4.8 \times 10^{-14}$; Fig. S4B). However, this result largely depended on the hybrid considered (Fig. S3B).

The Remodeling of the Proteome of Interspecific Hybrids Predominantly Affects Particular Categories of Proteins—Principal component analysis based on the estimated protein abundances was performed in order to visualize the effects of the strains and of the temperature on the proteome (Fig. 5). The first axis (PC1, 15% of the total variance) separated the parental and hybrid strains of $S.\ cerevisiae$ from those of $S.\ uvarum$, with interspecific hybrids located between the two species. Interestingly, within each type of hybrid ($S.\ cerevi-
Heterosis for Protein Abundance in Yeast

Fig. 5. Principal component analysis based on the estimated abundances of 615 proteins quantified in at least 122 strains x temperature combinations (see experimental procedures for details). Parental strains are written in upright (18 °C) or italics (26 °C) characters. Plain and dotted lines represent the limits that contain 99.9% of the distribution of the PC1 and PC2 coordinates of strain x temperature combinations for each group. They were obtained by simulating the kernel densities from group’s means and variances assuming bivariate normal distributions and using the R package MASS.

Fig. 6. Characteristics of the proteins in the set H+. The proteins correlated with $r > 0.5$ to the second axis of the PCA shown Fig. 5 (Pearson correlation test, set H+) were compared with those that were not correlated (set NH) for: (A) the mean abundance in parental strains (Student test); (B) the proportion of proteins encoded by essential genes ($\chi^2$ test); (C) the proportion of proteins involved in protein metabolism ($\chi^2$ test). Symbols: $6 \times 10^{-2} \times P$; $5 \times 10^{-2} \times P$; $5 \times 10^{-4} \times P$.

majority of the proteins in the set H+ (i.e. the set H without the protein negatively correlated to PC2) were included in cluster II. These proteins contributed poorly to PC1, which is consistent with the fact that they displayed little abundance variation between the parents of interspecific hybrids and between temperatures (Fig. S5). Compared with the proteins that were not correlated to PC2 (set NH, 253 proteins; Table S9), these proteins exhibited other specific characteristics: They were more abundant than other proteins (average abundance in parental strains: $2.2 \times 10^7$ in NH versus $3.1 \times 10^7$ in H+; Student test, $p = 2.5 \times 10^{-64}$, Fig. 6A); they were significantly enriched in proteins encoded by essential genes, i.e. genes that are required for viability of S. cerevisiae under standard laboratory conditions (69, 70) (22.6% in NH versus 55.3% in H+; $\chi^2$ test, $p = 1.6 \times 10^{-7}$, Fig. 6B); and they were slightly enriched in proteins involved in protein metabolism (protein synthesis and protein fate; 32.0% in NH versus 49.0% in H+; $\chi^2$ test, adjusted $p = .029$; Fig. 6C).

Altogether, these results show that interspecific hybridization caused BPH for a defined portion of the proteome that...
contains proteins characterized by the stability of their abundances toward genetic and environmental changes, by their high abundances, and by their importance for the cell viability.

Heterosis for Protein Abundance Is Partly Related to the Complexity of Transcriptional Regulation—To determine the extent to which the factors controlling protein abundances could be involved in heterosis for protein abundance, we focused on the transcription factors (TFs) possibly involved in the regulation of the genes encoding the proteins quantified in our study. A total of 162 TFs with a consensus DNA-binding sequence were retrieved from the Yeastract database (www.yeastract.com; 71–74). On average, the genes encoding proteins that were heterotic in at least one hybrid × temperature combination were putative targets of a higher number of TFs than the genes encoding non-heterotic proteins (27.7 versus 21.4; Mann–Whitney test p = 1.96 × 10^{-15}; Fig. 7A). In addition, a significant correlation was found between the number of putative TFs of a gene and the proportion of hybrids × temperature combinations in which the encoded protein was heterotic (Kendall correlation test, r = 0.18, p < 2.2 × 10^{-16}, Fig. S6).

The number of putative TFs of a gene depended significantly on the functional category of the gene (generalized linear model, ANOVA p < 2.2 × 10^{-16}). As a consequence, the frequency at which a protein was heterotic was also dependent on its functional category. For example, the genes involved in metabolism, energy and cell rescue, defense, and virulence had, on average, more putative TFs, and their proteins were more frequently heterotic than those involved in cell differentiation (Fig. 7B). Note that the protein synthesis category appeared as an outlier, containing proteins that were heterotic in a high proportion of hybrids but not presenting a very high number of putative TFs.

Altogether, these results suggest that the number of factors involved in transcriptional regulation may have an influence on heterosis for protein abundance, which may also explain why some functional categories are more prone to heterosis than others.

Predicting Protein Inheritance According to a Nonlinear Model—A general property of metabolic systems is the nonlinear response of the fluxes to genetic variations of enzyme concentrations and/or activity parameters (75). This relationship allowed Kacser and Burns (76) to propose a metabolic basis for dominance. In addition, Fiévet et al. (7) showed that when two or more enzymes are variable, the concave relationship between a flux and its parameters necessarily results in positive MPH or in BPH for the flux when the enzyme parameters, i.e. any genetic parameter that determines the enzymatic activity, are additively inherited (Fig. 8A).

Interestingly, the protein synthesis rates seem also to be a concave function of various factors, such as mRNA amount (77), translation factor abundance (78), ribosomal initiation rate, and elongation rate (79). Therefore, the basis of heterosis put forward for metabolic fluxes could apply for protein abundances, even though the relationship is mathematically different. In order to test this hypothesis and interpret our results, we used a simple nonlinear function for modeling and simulating the consequences of concavity on protein heterosis.
We analyzed the relationships between heterosis for protein abundance and (i) the type of inheritance of the genetic factors controlling abundance; (ii) the number of polymorphic factors controlling a protein, which is an indicator of the complexity of its genetic control; (iii) the Euclidean distance between parents computed from the values of the factors controlling protein abundances; and (iv) the phenotypic distance, i.e. the difference in protein abundance between parents.

When there is only one polymorphic factor, only positive MPH can be observed if there is additivity of the factor and positive and negative MPH in case of nonadditivity (Fig. 8B).
When two or more factors are polymorphic, positive MPH and BPH are possible if there is additivity of the factor, and if there is nonadditivity, the four types of heterosis are possible. When the number of polymorphic factors increases, BPH proportion increases at the expense of the other types of heterosis (Fig. 8E).

The Euclidean distance between parents was positively correlated with d/m (Fig. 8C), which is consistent with the well-known relationship between genetic distance and heterosis. More interestingly, the proportions of the different types of heterosis depended on the distance. For the smallest distances, all the types of heterosis were observed and all d/m values were small, while for the largest distances almost exclusively BPH and positive MPH were observed and their d/m values were high (Figs. 8C and 8D). This observation was valid whatever the number of polymorphic factors (Fig. S7). As expected from the concavity of the function, the distribution of the four types of heterosis tightly depended also on the phenotypic distance between parents. For the closest parents, there was a majority of BPH cases and few WPH, while the hybrids between distant parents displayed mainly positive MPH and to a lesser extent negative MPH (Figs. 8A and 8E and Fig. S7).

**DISCUSSION**

We used label-free quantitative proteomics in yeast to perform a large-scale study of heterosis for protein abundance. In agreement with previous results (58), we confirmed that the proteomes of S. cerevisiae and S. uvarum were highly differentiated. Interestingly, this differentiation is partly related to the adaptation of these species to their optimal temperatures (18 °C for S. uvarum and 26 °C for S. cerevisiae), as evidenced by the fact that lower temperatures drive S. cerevisiae's proteome close to that of S. uvarum, while the higher temperature drives S. uvarum's proteome close to that of S. cerevisiae.

**Heterosis for Protein Abundance Is Subject to Genotype × Environment Interactions**—Heterotic proteins were detected in every hybrid × temperature combinations analyzed. This is in line with previous results showing that heterosis for gene expression and protein abundance is a common occurrence, regardless of the species or genotypes considered (reviewed in (11)). The proportion of heterotic proteins varied from 8.4 to 61.2% depending on the hybrid × temperature combination considered. Comparatively, Khan et al. (80) found 85.9% of heterotic proteins (342 out of 398) in one S. cerevisiae × S. uvarum cross. However, these authors used an arbitrary threshold without statistical test to decide on the inheritance of the proteins, which may explain the discrepancy with our results. In any case, our study is much more representative of both the proteome and the genetic diversity of S. cerevisiae and S. uvarum since we examined 1,396 proteins quantified in 55 crosses and at two temperatures. This allowed us to show that there were genotype × environment interactions for heterosis since the temperature did not affect protein inheritance similarly in the different types of hybrid examined. Indeed, the proportion of heterotic proteins was higher at 18 °C for S. cerevisiae and interspecific hybrids and at 26 °C for S. uvarum hybrids. Note that in the case of intraspecific hybrids, these temperatures were nonoptimal, suggesting that there may be a relationship between the proportion of heterotic proteins and stressful growth conditions. In addition, the sign of d was little affected by temperature in interspecific hybrids, which was not the case in intraspecific hybrids.

**Heterosis for Protein Abundance Primarily Affects Highly Regulated Proteins**—Our results suggest that the number of putative TFs of a gene is related to the heterosis for the abundance of the encoded protein. Regulation of transcription is complex, involving a combination of several TFs individually acting as activator and/or repressor (81). Previous studies have shown that genetic polymorphism in cis and trans regulators can influence the inheritance pattern of gene expression level, polymorphism of trans regulators being preferentially associated to heterotic patterns (39, 82–84). If the number of polymorphic TFs increases with the number of TFs, the relationship between the number of putative TFs of a gene and the frequency at which the encoded protein is heterotic is consistent. Conceptually, our results are similar to what has been observed for agronomic traits in plants. Indeed, the results obtained from previous studies show that highly complex, polygenic traits such as yield are more prone to heterosis (23, 31). By analyzing a very high number of traits, we show here that the relationship between genetic complexity and heterosis is robust, even for less-integrated traits such as protein abundance.

The number of putative TFs of a gene depended significantly on the functional category of the gene, explaining why the proteins from some functional categories showed more heterosis than others. Among the functional categories containing genes putatively regulated by a high number of TFs and showing frequently heterotic proteins, we found energy, metabolism, and cell rescue, defense, and virulence. This result is consistent with many studies in plants that showed that these categories were involved in heterosis for gene expression (reviewed in (22, 24)). In addition, since these categories are generally involved in response to environmental changes (85), they were expected to be highly regulated.

The proteins involved in protein synthesis appeared as outliers regarding the relationship between the number of putative TFs of a gene and heterosis for protein abundance, presenting frequencies of heterosis higher than expected based on the number of putative TFs of their encoding genes. To explain the peculiar behavior of these proteins, we assume that factors other than TFs are involved in heterosis for protein abundance as, for example, posttranslational modifications, that were recently shown to be related to the variations of phenotypic traits (86).

**Best-Parent Heterosis Is Related to Proteins Under Evolutionary Constraints**—BPH in interspecific hybrids was more
particulars related to a particular group of proteins (set H+) that were highly abundant and exhibited little abundance variations between temperatures and between S. cerevisiae and S. uvarum, yet two distantly related species (87). Observation of interspecific heterosis for these proteins necessarily implies that the two species are genetically contrasted at the loci controlling protein abundances (Protein Quantity Loci (33)). This is in agreement with dominance hypothesis, which attributes heterosis to allele complementation. In addition to these results, we also showed that the set H+ was enriched in proteins encoded by essential genes and in proteins involved in protein metabolism. Essential genes are thought to be under strong purifying selection since they are highly conserved across large evolutionary distances in yeasts and mammals (88, 89). Moreover, protein metabolism includes proteins of ribosomes and proteasome that are structurally and functionally conserved (90, 91). This suggests that the proteins of the set H+ are under evolutionary constraint. However, we have currently no hypothesis to establish a relationship between evolutionary constraint and heterosis.

**Heterosis for Protein Abundance Is Consistent with a Model of Nonlinear Genotype–Phenotype Relationship**—It has been observed from numerous experiments that heterosis is generally biased toward positive values (for example, (3, 9)). This bias is accounted for in the dominance hypothesis, where recessive deleterious alleles are complemented by dominant superior alleles (26, 27). In the context of metabolic systems, dominance of the high over the low allele is explained by the hyperbolic response of fluxes toward the variations of enzyme parameters (e.g., activity, concentration): Due to the concavity of the curve, the flux value in a hybrid obtained from a cross between two parents presenting contrasted enzyme parameters is systematically biased toward the highest parent, provided the value of the enzyme parameter is additively inherited (76, 92). Generalized to networks with several variable enzymes, this hyperbolic relationship generates heterosis for the metabolic flux (7, 93).

In this study, we analyzed a high number of traits in a large number of hybrids, which allowed us to examine the extent to which the bias toward positive heterosis was robust. Unexpectedly, we showed that heterosis for protein abundance was strongly biased toward positive values in interspecific hybrids but not in intraspecific hybrids and (ii) BPH in interspecific hybrids occurred preferentially for a special group of proteins assumed to be under evolutionary constraint. These results shed new light on heterosis by supporting a model where protein abundances would be related to transcriptional and translational parameters by concave relationships. In agreement with this hypothesis, we also showed that the complexity of transcriptional regulation, estimated through the number of putative TFs of a gene, is related to heterosis for protein abundance, which supports a general relationship between heterosis and trait complexity. Taken together, our results show the interest of high-throughput technologies to provide a more comprehensive view of complex biological phenomena such as heterosis.

By simulating heterosis for protein abundance using a nonlinear model of genotype–phenotype relationship, we obtained in silico results in agreement with those obtained from the experiments. First, we showed that negative heterosis can occur when there is nonadditive inheritance of the genetic factors, which is biologically realistic. Second, we showed that for small genetic distances positive and negative heterosis can be observed, while for large distances there is much more positive than negative heterosis. This is consistent with the bias we observed between intra- and interspecific hybrids. Third, we showed that the proportion of BPH was maximal for hybrids obtained from distant parents and for proteins displaying similar abundances in the parents. This is consistent with the frequent BPH observed in interspecific hybrids for the proteins of set H+. Finally, we showed that heterosis was related to the number of polymorphic factors controlling a protein. This is consistent with the observation that the proteins regulated by a high number of TFs were more prone to heterosis.

To conclude, we performed a large-scale study of heterosis, which allowed us to obtain original results: (i) heterosis was strongly biased toward positive values in interspecific hybrids but not in intraspecific hybrids and (ii) BPH in interspecific hybrids occurred preferentially for a special group of proteins assumed to be under evolutionary constraint. These results shed new light on heterosis by supporting a model where protein abundances would be related to transcriptional and translational parameters by concave relationships. In agreement with this hypothesis, we also showed that the complexity of transcriptional regulation, estimated through the number of putative TFs of a gene, is related to heterosis for protein abundance, which supports a general relationship between heterosis and trait complexity. Taken together, our results show the interest of high-throughput technologies to provide a more comprehensive view of complex biological phenomena such as heterosis.

* This work was supported by the Agence Nationale de la Recherche (ANR-08-ALIA-09 HeterosYeast).

This article contains supplemental material Tables S1 to S9 and Figs. S1 to S7.

To whom correspondence should be addressed: PAPPSO, Plateforme d’Analyse Proteomique de Paris Sud-Ouest, UMR de Genetique Vegetale, INRA-Univ Paris Sud-CNRS, Ferme du Moulon, F-91190 Gif-sur-Yvette, France. Fax: +33 1 69 33 23 40; E-mail: zivy@moulon.inra.fr.

**REFERENCES**


92. Wright, S. (1934) Molecular and evolutionary theories of dominance. Am. Nat. 68, 24–53