bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

# Decoupling of heterosis and inbreeding variances is evidenced in yeast's life-history and proteomic traits

**M. Petrizzelli***, **D. de Vienne*** **and C. Dillmann***,[1]

*Génétique Quantitative et Evolution – Le Moulon, INRA, Université Paris-Saclay, Université Paris-Sud, CNRS, AgroParisTech, 91190 Gif-sur-Yvette, France

**ABSTRACT** Heterosis (hybrid vigor) and inbreeding depression, commonly considered as corollary phenomena, could nevertheless be decoupled under certain assumptions according to theoretical population genetics works. In order to explore this issue on real data, we analyzed the components of genetic variation in a population derived from a half-diallel cross between strains from *Saccharomyces cerevisiae* and *S. uvarum*, two related yeast species involved in alcoholic fermentation. A large number of phenotypic traits, either molecular (coming from quantitative proteomics) or related to fermentation and life-history, were measured during alcoholic fermentation. Because parental strains were included in the design, we were able to distinguish between inbreeding effects, which measures phenotypic differences between inbred and hybrids, and heterosis, which measures phenotypic differences between a specific hybrid and the other hybrids sharing a common parent. The sources of phenotypic variation differed depending on the temperature, indicating the predominance of genotype by environment interactions. Decomposing the total genetic variance into additive (intra- and inter-specific), inbreeding, and heterotic (intra- and inter-specific) variances, we showed that the distribution of variance components defined clear-cut groups of proteins and traits. Moreover, it was possible to cluster fermentation and life-history traits into most proteomic groups. Within groups, we observed positive, negative or null correlation between inbreeding and heterotic variances. To our knowledge, such a decoupling had never been experimentally demonstrated. This result suggests that, despite a common evolutionary history of individuals within a species, the different types of traits have been subject to different selective pressures.

**KEYWORDS** Hybrid vigor; inbreeding depression; diallel crossing; mixed effect genetic model

Heterosis, or hybrid vigor, refers to the common superiority of hybrids over their parents for quantitative traits. This phenomenon has been observed for virtually any quantitative trait, from mRNA abundances to fitness, and in a large diversity of species, including microorganisms. For decades it has been extensively studied and exploited for plant and animal breeding, since it affects traits of high economical interest such as biomass, fertility, growth rate, disease resistance etc. (Gowen 1952; Schnable and Springer 2013).

There are three classical, non exclusive genetic models to account for hybrid vigor: dominance, overdominance and epistasis. In the dominance model, the hybrid superiority results from the masking of the deleterious alleles of one parent by the non deleterious ones of the other parent (Davenport 1908). In the overdominance model, the hybrid superiority is due to the advantage *per se* of the heterozygous state at a given locus (Hull 1946). Actually, more common is pseudo-overdominance, which is due to dominance at two loci linked in repulsion, *e.g.* in maize (Graham *et al.* 1997; Lariepe *et al.* 2012) or yeast (Martì-Raga *et al.* 2017). Lastly, the epistasis model postulates favorable intergenic interactions created in the hybrids (Powers 1944). In particular, "less-than-additive" (antagonistic) epistasis, which is quite common in plant and animal species (Redden 1991; Shao *et al.* 2008) can account for best-parent heterosis (Fievet *et al.* 2010). In this last paper, it is theoretically shown that epistasis can result in best-parent heterosis even if there is no dominance at any locus. The respective parts of the various genetics effects in heterosis depends on the trait, the species and the genetic material (Xiao *et al.* 1995; Huang *et al.* 2016; Seymour *et al.* 2016). Altogether, heterosis appears to be a pervasive phenomenon, accounted for by the common non-linearity of the genotype-phenotype map

doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Wednesday 4th July, 2018
[1]To whom correspondence should be addressed: UMR Génétique Quantitative et Évolution - Le Moulon, Ferme du Moulon, 91190 Gif-sur-Yvette, France. E-mail: christine.dillmann@inra.fr

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

(Wright 1934; Omholt *et al.* 2000; Fiévet *et al.* 2018).

Because heterosis is associated with heterozygosity, heterosis for life-history traits is associated with genetic load: the average population fitness can never exceed the maximum fitness. Genetic load drives the evolution of sexual reproduction, of mating systems as well as the fate of small populations. Indeed, high levels of homozygosity in outcrossing species is generally associated with decreased growth rate, survival or fertility (discussed in Charlesworth and Willis (2009)). In population genetics, inbreeding depression is defined as the fitness of self-fertilized progenies as compared with fitness of outcrossing progenies. In sexual species, the balance between selfing and outcrossing is driven by the genetic load due to inbreeding depression relative to the cost of sexual reproduction (twice as expensive as clonal reproduction): selfing can evolve whenever inbreeding depression is less costly than the sexual reproduction, or after purging deleterious mutations as can arise in small populations (Lande and Schemske 1985). However, heterosis due to less-than-additive epistasis could explain the large number of predominantly (but not fully) selfing species exhibiting a persistent amount of inbreeding depression and heterosis (Charlesworth *et al.* 1991). Considering a metapopulation, Roze and Rousset (2004) defined inbreeding depression as the fitness reduction of selfed progeny relative to outcrossed progeny within populations, and heterosis as the difference between the fitness of the outcrossed progeny within population and the outcrossed progeny over the whole metapopulation. They showed that while selfing reduced both inbreeding depression and heterosis, inbreeding depression decreased and heterosis increased with the degree of subdivision of the metapopulation. Hence, from a population genetics point of view, heterosis is expected even in predominantly selfing species.

In a breeding perspective, the pioneer work of Shull (1908) in maize predicted that given the large amounts of heterosis within the species, the best way to maximize yield was to create inbreds from existing population varieties in order to seek for the best hybrid combinations. Diallel designs were popularized as the most comprehensive designs for estimating genetic effects, predicting hybrid values and generating breeding populations to be used as basis for selection and development of elite varieties (i.e. Hallauer and Filho (1988)). The simplest and most popular analytic decomposition of genetic effects in diallel designs is that of Griffing (1956), in which the mean phenotypic value, $y_{ij}$, of the cross between lines $i$ and $j$ is modeled as:

$$y_{ij} = \mu + GCA_i + GCA_j + SCA_{ij}, \quad (1)$$

where $\mu$ is the mean phenotypic value of the population, $GCA_i$ (resp. $GCA_j$) is the *General Combining Ability* of line $i$ (resp. $j$), *i.e.* the average performance of line $i$ (resp. $j$) in hybrid combinations expressed as a deviation from the mean value of all crosses, and $SCA_{ij}$ is the *Specific Combining Ability* of hybrid $i \times j$. It is defined as the difference between the mean phenotypic value of the progeny and the sum of the combining abilities of the parental lines (Sprague and Tatum 1942). Therefore, superior individuals can be selected from their GCA and/or SCA. Numerous extensions of the Griffing's model have been proposed to extract other effects, such as maternal and paternal effects or sex-linked variations (Cockerham and Weir 1977; Bulmer 1980; Zhu and Weir 1996; Greenberg *et al.* 2010). In many crop species, combining ability groups have been identified, with lines from the same group characterized by high specific combining ability with other groups (Hallauer *et al.* 1988). Generally, combining ability groups are redundant with population structure within a species (Melchinger and Gumber 1998; Ramya *et al.* 2018), which is consistent with the population genetics predictions of Roze and Rousset (2004).

When parental lines are included in the analysis, $GCA$ and $SCA$ effects can be decomposed in more suitable genetic effects. Indeed, the value of a particular hybrid can be compared either to the average value of its inbred parents, or to the average value of the other hybrids sharing either parent. Heterosis can be split into *average heterosis* (average difference between inbreds and outbreds), *variety heterosis* (average difference between one inbred parent and all crosses sharing the same parents), and *specific heterosis* (difference between the hybrid and all hybrids sharing at least one parent) (Eberhart and Gardner 1966). A modern version of this model have been proposed by Lenarcic *et al.* (2012) along with a Bayesian framework to estimate the genetic effects.

In this work, we study a half-diallel design with diagonal constructed from the crosses between 11 yeast strains belonging to two close species, *Saccharomyces cerevisiae* and *S. uvarum*. The design included both intra- and inter-specific crosses. Two categories of phenotypic traits were considered: (i) protein abundances measured at one time point of alcoholic fermentation (Blein-Nicolas *et al.* 2013, 2015); (ii) a set of fermentation traits measured during and/or at the end of fermentation, which were divided into kinetic parameters, basic enological parameters, aromas and life-history traits (da Silva *et al.* 2015). All traits were independently measured at two temperatures.

We propose a decomposition of the genetic effects based on Lenarcic *et al.* (2012) that takes into account the presence of two species in the diallel design and that distinguishes inbreeding and heterosis effects. We could characterize every trait by the set of its variance components and we could clearly cluster the traits from this criterion, which suggests that traits sharing a similar pattern of variance components could share common life-history. We were able to assign each fermentation trait to one group of protein traits, which shows that integrated phenotypes and proteins can share similar life-history. Finally, our results show a poor correlation between inbreeding and heterosis variances within groups. This confirms the importance of epistatic interactions in determining the components of phenotypic variation both within and between close species. Altogether, our results suggest that despite a common demographic history of individuals within a species, the genetic variance components of the traits can be used to trace back other trait-specific evolutionary pressures, like selection.

## Materials and Methods

### *Materials*

The genetic material of the experimental design consisted in 7 strains of *S. cerevisiae* and 4 strains of *S. uvarum*. The strains were diploid and fully homozygous, and derived from strains associated to various food processes or isolated from natural environment. These strains, called W1, D1, D2, E2, E3, E4, E5 for *S. cerevisiae* and U1, U2, U3, U4 for *S. uvarum* were the parental strains of a half-diallel design with diagonal, *i.e.* including the inbred lines (da Silva *et al.* 2015). All parental lines were selfed and pairwise crossed, which resulted in a total of 66 strains: 11 inbred lines, 27 intra-specific hybrids (21 for *S. cerevisiae*, noted *S. c.*, and 6 for *S. uvarum*, noted *S. u.*) and 28 inter-specific (noted *S. u.* × *S. c*).

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The 66 strains were grown in triplicate in fermentors at two temperatures, 26° and 18°, in a medium close to enological conditions (Sauvignon blanc grape juice) (da Silva *et al.* 2015). From a total of 396 alcoholic fermentations (66 strains × 2 temperatures × 3 replicas), 31 failed due to poor fermenting abilities of some strains. The design was implemented considering a block as two sets of 27 fermentations (26 plus a control without yeast to check for contamination), one carried out at 26° and the other at 18°. The distribution of the strains in the block design was randomized to minimize the residual variance of the estimators of the strain and temperature effects, as described in Albertin *et al.* (2013b).

For each alcoholic fermentation, two types of phenotypic traits were measured or estimated from sophisticated data adjustment models: 35 fermentation traits and 615 protein abundances.

The fermentation traits were classified into four categories (da Silva *et al.* 2015):

- *Kinetics parameters*, computed from the $CO_2$ release curve modeled as a Weibull function fitted on $CO_2$ release quantification monitored by weight loss of bioreactors: the fermentation lag-phase, *t-lag* (h); the time to reach the inflection point out of the fermentation lag-phase, *t-$V_{max}$* (h); the fermentation time at which 45 $gL^{-1}$ and 75 $gL^{-1}$ of $CO_2$ was released, out of the fermentation lag-phase, *t-45* (h) and *t-75* (h) respectively; the time between *t-lag* and the time at which the $CO_2$ emission rate became less than, or equal to, $0.05gL^{-1}h^{-1}$, *AFtime* (h); the maximum $CO_2$ release rate, $V_{\max}$ ($gL^{-1}h^{-1}$); and the total amount of $CO_2$ released at the end of the fermentation, $CO_{2\max}$ ($gL^{-1}$).
- *Life history traits*, estimated and computed from the cell concentration curves over time, modeled from population growth, cell size and viability quantified by flow cytometry analysis: the growth lag-phase, *t-$N_0(h)$*; the carrying capacity, *K* (log[cells/mL]); the time at which the carrying capacity was reached, *t-$N_{max}$* (h); the intrinsic growth rate, *r* (log[cell division/mL/h]); the maximum value of the estimated $CO_2$ production rate divided by the estimated cell concentration, $J_{max}$ ($gh^{-1}10^{-8}cell^{-1}$); the average cell size at *t-$N_{max}$*, *Size-t-$N_{max}$*($\mu$m); the percentage of living cells at *t-$N_{max}$*, *Viability-t-$N_{max}$* (%); and the percentage of living cells at *t-75*, *Viability-t-75* (%).
- *Basic enological parameters*, quantified at the end of fermentation: *Residual Sugar* ($gL^{-1}$); *Ethanol* (%vol); the ratio between the amount of metabolized sugar and the amount of released ethanol, *Sugar.Ethanol.Yield* ($gL^{-1}$%$vol^{-1}$); *Acetic acid* ($gL^{-1}$ of $H_2SO_4$); *Total $SO_2$* ($mgL^{-1}$) and *Free $SO_2$* ($mgL^{-1}$).
- *Aromatic traits*, mainly volatile compounds measured at the end of alcoholic fermentation by GC-MS: two higher alcohols (*Phenyl-2-ethanol* and *Hexanol*, $mgL^{-1}$); seven esters (*Phenyl-2-ethanol acetate*, *Isoamyl acetate*, *Ethyl-propanoate*, *Ethyl-butanoate*, *Ethyl-hexanoate*, *Ethyl-octanoate* and *Ethyl-decanoate*, $mgL^{-1}$); three medium chain fatty acids (*Hexanoic acid*, *Octanoic acid* and *Decanoic acid*, $mgL^{-1}$); one thiol *4-methyl-4-mercaptopentan-2-one*, *X4MMP*($mgL^{-1}$) and the acetylation rate of higher alcohols, *Acetate ratio*.

For proteomic analyses the samples were harvested at 40 % of $CO_2$ release, corresponding to the maximum rate of $CO_2$ release. Protein abundances were measured by LC-MS/MS techniques from both shared and proteotypic peptides relying on original Bayesian developments (Blein-Nicolas *et al.* 2012). A total of 615 proteins were quantified in more than 122 strains × temperature combinations as explained in details in Blein-Nicolas *et al.* (2015).

Cross-referencing MIPS micro-organism protein classification (Ruepp *et al.* 2004), KEGG pathway classification (Kanehisa and Goto 2000; Kanehisa *et al.* 2016, 2017) and Saccharomyces Genome database (Cherry *et al.* 2012), we attributed each protein to a single functional category based on our expert knowledge (Table ST1). Considering the genes encoding the proteins, we also assigned to each protein a number of putative transcription factors (TFs). A total of 313 TFs with a consensus DNA-binding sequence were retrieved from the Yeastrack database (Teixeira *et al.* 2014; Abdulrehman *et al.* 2011; Monteiro *et al.* 2008; Teixeira *et al.* 2006).

### *Statistical Methods*

In order to estimate the genetic variance components for the different phenotypic traits, we adapted the model described in Lenarcic *et al.* (2012) to our particular half-diallel design that includes the diagonal with parental inbred strains from two species. Thus we included in our model intra- and inter-specific additive effects, inbreeding effects and intra- and inter-specific heterosis effects.

Formally, let $y_{ijk}$ be the observed phenotype for the cross between parents $i$ and $j$ in replica $k$. Our model reads:

$$\begin{aligned} y_{ijk} = \mu + I_{s(i)=s(j)}(A_{w_i} + A_{w_j}) + I_{s(i)\neq s(j)}(A_{b_i} + A_{b_j}) + \\ + I_{i\neq j}(I_{s(i)=s(j)}H_{w_{ij}} + I_{s(i)\neq s(j)}H_{b_{ij}}) + \\ + I_{i=j}(\beta_{s(i)} + B_i) + \epsilon_{ijk}, \end{aligned} \quad (2)$$

where:

- $\mu$ is the overall mean;
- $s(i)$ associates to each parental strain $i$ the specie it belongs to:
$$s(i) \in \{S.\ cerevisiae, S.\ uvarum\}$$
- $A_{w_i}$ and $A_{b_i}$ denote, respectively, the additive contributions of strain $i$ in intra-specific (within species, *i.e.* $s(i) = s(j)$), and inter-specific (between species, *i.e.* $s(i) \neq s(j)$) crosses;
- $H_{w_{ij}}$ and $H_{b_{ij}}$ denote the interaction effect between parents $(i, j)$ in intra-specific (within species) and inter-specific (between species) crosses, respectively. Due to our half-diallel design (no reciprocal crosses), they are assumed to be symmetric, *i.e.* $H_{w_{ij}} = H_{w_{ji}}$ and $H_{b_{ij}} = H_{b_{ji}}$. Hereafter we will refer to these effects as intra- and inter-specific heterosis effects, respectively;
- $\beta_{s(i)}$ and $B_i$ are, respectively, the deviation from the fixed overall effect for the species $s(i)$ and the associated strain-specific contribution of strain $i$ in the case of inbred lines. Hereafter we will refer to $B_i$ as inbreeding effect;
- $\epsilon_{ijk}$ is the residual, the specific deviation of individual $ijk$;
- $I_{condition}$ is an indicator variable. Its value is equal to 1 if the condition is satisfied and 0 otherwise.

Therefore, for the parental lines we have:

$$y^{p}_{iik} = \mu + 2A_{w_i} + \beta_{s(i)} + B_i + \epsilon_{iik}, \quad (3)$$

for the intra-specific hybrids:

$$y^{intra}_{ijk} = \mu + A_{w_i} + A_{w_j} + H_{w_{ij}} + \epsilon_{ijk}, \quad (4)$$

and for the inter-specific hybrids:

$$y^{inter}_{ijk} = \mu + A_{b_i} + A_{b_j} + H_{b_{ij}} + \epsilon_{ijk}. \quad (5)$$

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

All genetic effects were considered as random variables drawn from a normal distribution. Formally, letting $\boldsymbol{q} \in \{\boldsymbol{A_w}, \boldsymbol{A_b}, \boldsymbol{B}, \boldsymbol{H_w}, \boldsymbol{H_b}\}$ denote the genetic effect under consideration:

$$\forall i \quad q_i \sim \mathcal{N}(0, \sigma^2_{\boldsymbol{q}}). \tag{6}$$

The full mixed-effect genetic model is thus defined by three fixed effects (the intercept $\mu$ and the inbreeding effects $\beta_{Su}$ and $\beta_{Sc}$) and five genetic random effect variances ($\sigma^2_{\boldsymbol{A_w}}$, $\sigma^2_{\boldsymbol{A_b}}$, $\sigma^2_{\boldsymbol{B}}$, $\sigma^2_{\boldsymbol{H_w}}$, $\sigma^2_{\boldsymbol{H_b}}$).

We did not declare mitochondrial effects because many genes encoding mitochondrial proteins are repressed under fermentation conditions, and because inter-specific hybrids harbor similar fermentation features for most fermentation kinetics and enological parameters whatever their mitochondrial genotype (Albertin *et al.* 2013a). In addition, we did not know the mitochondrial inheritance for most of the intra-specific crosses (table ST2).

### *The fitting algorithm*

Fixed effects, variance components of the genetic effects as well as their Best Linear Unbiased Predictors (*BLUPs*) were estimated using the *hglm* package in R (Ronnegard *et al.* 2010) that implements the estimation algorithm for hierarchical generalized linear models and allows fitting correlated random effects as well as random regression models by explicitly specifying the design matrices both for the fixed and random effects. The model, based on a maximum likelihood estimation, is deemed to produce unbiased statistics (Gumedze and Dunne 2011).

A separate analysis was conducted for each trait at each temperature, considering the vector of observations for the trait/temperature combination of interest, $\mathbf{y}$, and re-writing model (eq. (2)) in matrix form:

$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{u} + \boldsymbol{\epsilon}, \tag{7}$$

where $X$ is the design matrix for the fixed effects, $Z$ the design matrix for the random effects, $\boldsymbol{\beta} = (\mu, \beta_{S.u.}, \beta_{S.c.})$ and $\boldsymbol{u} = (\boldsymbol{A_w}, \boldsymbol{A_b}, \boldsymbol{B}, \boldsymbol{H_w}, \boldsymbol{H_b})$ are respectively the vectors of fixed effect parameters and random effect parameters, and $\boldsymbol{\epsilon}$ is the vector of residual errors. With this notation, the construction of the model is straightforward from the data (for details see The fitting algorithm in Supplementary Materials).

Whenever the full model (eq. 2) failed to converge, we considered the subsequent model obtained by removing one effect at a time following the hierarchy imposed by the order of the fitting algorithm, *i.e.* first heterosis, second inbreeding effects and finally additive effects. The full model converged for all proteomic data. For the fermentation traits, the model did not converge for most of the *Ethyl* esters (*Ethyl-propanoate*, *Ethyl-butanoate*, *Ethyl-hexanoate*, *Ethyl-octanoate* and *Ethyl-decanoate*), as well as for *Acetate Ratio* and for *Acetic acid* that were removed from the analysis. For all other fermentation traits, the full model converged, except for *t.lag* at 18°, for which the additive model applied. For this trait, other genetic variance components were set to zero.

In order to test the robustness of the results, a bootstrap analysis was performed by sampling the 55 hybrids with replacement, conditionally to the 11 parental strains. Each bootstrap sample was submitted to the same analysis as described above. For each variance component, we checked that the estimations in the experimental sample were close to the median of the estimations in the bootstrap samples.

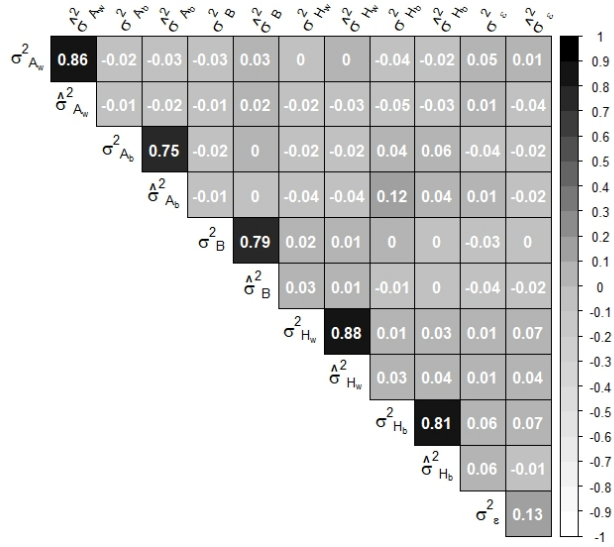



**Figure 1** Correlation between estimated variance components and their true value. Variances have been estimated on a simulated half-diallel between 11 parental strain (seven belonging to a specie, four to the other). Phenotypic values have been computed as detailed in section Testing for the identifiability of the model.

### *Testing for the identifiability of the model*

Computer simulations were performed to test the statistical power of the *hglm* algorithm in predicting the values of the observables while producing unbiased estimations of the model parameters. We simulated a half-diallel between 11 strains, seven belonging to a species, four to the other. We computed the phenotypic values of each simulated cross by first drawing $\mu$, $\beta_{specie1}$, $\beta_{specie2}$, $\sigma^2_{\boldsymbol{A_w}}$, $\sigma^2_{\boldsymbol{A_b}}$, $\sigma^2_{\boldsymbol{B}}$, $\sigma^2_{\boldsymbol{H_w}}$, $\sigma^2_{\boldsymbol{H_b}}$ and $\sigma^2_{\boldsymbol{\epsilon}}$ from a Gamma distribution fitted from the values estimated by the model on our dataset (see fig. SF1). Second, for each random effect $\boldsymbol{q} \in \{\boldsymbol{A_w}, \boldsymbol{A_b}, \boldsymbol{B}, \boldsymbol{H_w}, \boldsymbol{H_b}, \boldsymbol{\epsilon}\}$ we drew

$$\forall i \quad q_i \sim \mathcal{N}(0, \sigma^2_{\boldsymbol{q}}) \tag{8}$$

and computed the phenotypic values as in eq. 2, generating three replicas per cross.

We repeated the simulation 1000 times. We fitted the model and checked that the estimation of the random effects, the predicted phenotypic values as well as their variance components were close enough to the true values (fig. 1) and we noticed that inbreeding parameters were the most variable (fig. SF2 in *Supplementary figures*).

In addition, since we were interested in the correlation structure between the variance components of the genetic effects, we checked that possible correlations between random effects were not a statistical artifact of the model. Therefore, we simulated uncorrelated variances of random effects and we checked that no correlation structure was found between the estimated variance components, as can be seen in fig. 1. Simulations performed with different numbers of parental lines led to similar results (not shown).

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

#### Fermentation traits

Before fitting our model, we updated eq. 2 in order to account for a block effect:

$$y_{ijkl} = y_{ijk} + block_l + \epsilon_{ijkl}, \tag{9}$$

assuming that

$$\forall l \quad block_l \sim \mathcal{N}(0, \sigma^2_{\boldsymbol{block}}). \tag{10}$$

Many fermentation traits, mostly aromatic, were *log*-transformed in order to deal with the variable mean of the residuals. So as to handle the null values in the observations, we chose to consider the following transformation:

$$y_{ijk} = log(max(y_{ijk}, \delta)) \tag{11}$$

where $\delta \sim \mathcal{U}(0, min(\boldsymbol{y}))$. In this situation, as we introduced a random term in our analysis, which may skew parameter estimation, we decided to: (i) perform the *log*-transformation, (ii) compute the fitting algorithm, (iii) record the parameter's estimation, then after having computed it a hundred times, (iv) consider the median of the estimators in order to achieve a more robust statistics.

#### Protein abundances

For each cross, protein abundances have been quantified on average. Yet, to perform a diallel analysis at the proteomic level, replicas are critical for quantifying genetic variation. Therefore, we generated pseudo replicas using the residual variance estimated when quantifying protein abundances (Blein-Nicolas *et al.* 2013). Formally, let $y_{ij}$ be the average protein abundance of the cross between parents $i$ and $j$. We generated three replicas as follows:

$$y_{ijk} = y_{ij} + \epsilon_k \tag{12}$$

$$\epsilon_k \sim \mathcal{N}(0, \hat{\sigma}^2_\epsilon) \text{ for } k = 1, 2, 3 \tag{13}$$

where $\hat{\sigma}^2_\epsilon$ is the residual variance. Simulations of pseudo replicas and parameter estimations were performed 100 times. The final value of the parameters was the median of its estimation.

#### Variance component analysis

For each trait, our mixed model generates a vector of variance components

$$\boldsymbol{v} = (\hat{\sigma}^2_{\boldsymbol{A_w}}, \hat{\sigma}^2_{\boldsymbol{A_b}}, \hat{\sigma}^2_{\boldsymbol{B}}, \hat{\sigma}^2_{\boldsymbol{H_w}}, \hat{\sigma}^2_{\boldsymbol{H_b}}) \tag{14}$$

and the results were summarized in a matrix with rows being the different trait by temperature combinations, and columns the relative contribution of each component to the total genetic variance of the trait. We chose to perform unsupervised classification to compare the distributions of variance components between traits. Following the recommendations of Kurtz *et al.* (2015), percentages of variance components were transformed into real numbers using the following *clr*-transformation:

$$clr(\hat{\sigma}^2_{\boldsymbol{q}}) = log(\frac{\hat{\sigma}^2_{\boldsymbol{q}}}{(\prod_{k \in \boldsymbol{Q}} \hat{\sigma}^2_{\boldsymbol{k}})^{1/N_q}}) \tag{15}$$

where $N_q$ is the total number of random effects and $\boldsymbol{Q}$ is the set of random variables fitted by the model. For fermentation traits, $N_q = 7$ (accounting for block and residual variances, eq. 9), while $N_q = 6$ for proteomic traits (eq. 2). We chose the *clr*-transformation because it satisfies *scale invariance*, *subcompositional dominance* and *perturbation invariance* properties (Tsagris *et al.* 2011). Therefore the distance relationship between the original profiles is preserved by the selected sub-vectors thanks to the sub-compositional dominance property of the *clr*-transformation (see section Subcompositional dominance and distances in Supplementary Materials). The *clr*-transformation allowed us to test finite Gaussian mixture models using model-based clustering proposed in the *Mclust* package in R (Scrucca *et al.* 2016). Percentage of good assignments were computed by separating the data into training and validation sets.

This procedure was first applied separately for proteomic and fermentation traits (see Structuration of genetic variability at the fermentation trait level in Supplementary Materials). Protein groups were tested for enrichment in either Kegg pathways, transcription factors and heterotic proteins. Fermentation traits were tested for enrichment in the different trait categories (kinetic parameters, life-history, basic enological parameters, aromatic traits). For each cluster, Pearson's chi-square test of enrichment was computed on protein functional category frequencies taking as prior probability the expected categorical frequency found in the MIPS database.

Further, fermentation traits were assigned to clusters identified on protein abundances profiles based on their membership probability computed through Gaussian finite mixture models.

#### Data Availability

The data that support the findings of the current study are available from the corresponding author on reasonable request. Supplementary materials contain:

- Demonstration of the relationship between the *subcompositional dominance* property and distances in the Euclidean space;
- Detailed description of the fitting algorithm;
- Description of the construction of the simulated values on a half-diallel design based on the genetic models supposed to explain heterosis and inbreeding;
- Demonstration of the perfect correlation between inbreeding and heterosis variances in three parents half-diallel designs with no maternal effects;
- Clustering analysis for the fermentation and life-history traits;
- Strains characterization based on the estimated BLUP of their genetic effects;
- Table ST1: Protein functional category classification (available on https://figshare.com/s/97a08273de59d1a9cb8b );
- Table ST2: Mitochondrial inheritance of the phenotyped crosses of our study;
- Table ST3: Table of results from the Pearson's chi-square test of cluster enrichment in proteins with a particular functional category;
- Figure SF1: Density distribution of the genetic variances estimated by the model;
- Figure SF2: Predicted BLUPs and phenotypic values versus their prior value used to compute the values of simulated diallels;
- Figure SF3: Clustering profiles of fermentation and life-history traits;
- Figure SF4: Global correlations of the genetic variance components for both protein abundances and the more integrated traits;
- Figure SF5: Representation of the standardized Pearson's chi-square residuals of each cluster computed at 18° versus those at 26° estimated for the analysis of cluster enrichment in proteins with a particular functional category;

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

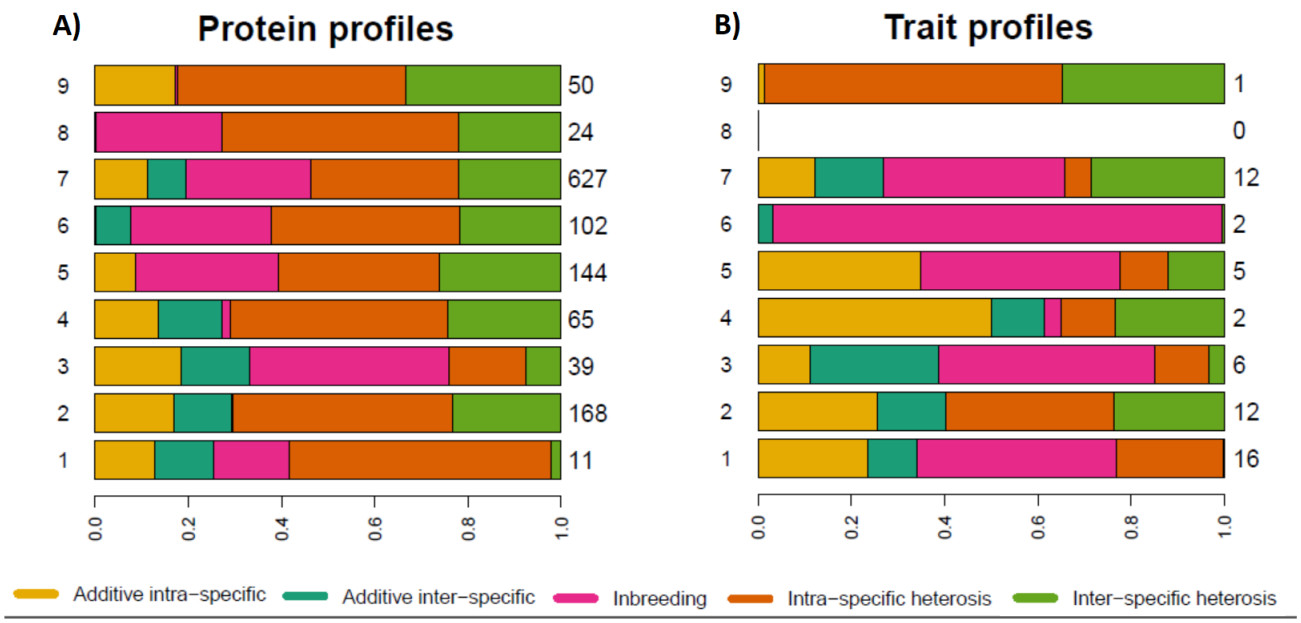



**Figure 2** Clustering profiles of genetic variance components for protein abundances (A) against profiles of fermentation traits (B) predicted in each cluster. Cluster numbers are reported on the left, on the right the number of proteins or traits found in each cluster.

- Figure SF6: Correlation plot between genetic effects of fermentation and life-history trait profiles;
- Figure SF7: Intra-cluster correlations of variance components profiles for fermentation and life-history traits;
- Figure SF8: Variance components of fermentation and life-history traits at the two temperatures;
- Figure SF9: Summary example of the density distribution of a genetic variance estimation through bootstrap analysis;
- Figure SF10: Representation of the relationship between inbreeding and heterosis variances simulated through different genetic models;
- Figure SF11: For each trait and for each genetic effect are shown the strains with highest and lowest contribution at both temperatures;
- Figure SF12: For each trait are shown the estimated BLUPs for each genetic parameter.

## Results

In order to estimate genetic variance components from a diallel cross involving two yeast species, we proposed a decomposition of genetic effects based on the model of Lenarcic *et al.* (2012) that allowed to split the classical General (GCA) and Specific (SCA) Combining Abilities into intra- and inter-specific additive and heterosis effects, and to take into account inbreeding effects, defined as the difference between the inbred line value and the average value of all the crosses that have this inbred as parent.

Simulations showed that despite the small number of parents in the diallel, our model led to unbiased estimations of variance components, and that correlations between variance components did not arise from unidentifiability of some model's parameter (fig. 1). Significance of variance components was assessed by bootstrap sampling. We found that whenever the fitting algorithm converged, variance component estimations were significant. For some traits and some variance components, the bootstrap distributions of the estimated variances were bimodal, suggesting a strong influence of a particular hybrid combination. However, the estimates were globally closed to the median of the bootstrap distribution (see example fig. SF9). Therefore, we are confident with our estimations, conditionally to the parents of the diallel.

Because temperature has a major effect on many traits and because, in previous work, numerous strain × temperature effects have been detected (da Silva *et al.* 2015; Blein-Nicolas *et al.* 2015), the model was applied to each trait separately at the two temperatures. We obtained estimations of fixed and random effect parameters, their corresponding variances, residuals and residual variances. For each trait, normality of residuals and homogeneity of variances was checked. Broad sense heritability was measured as the ratio of the sum of genetic variance components to the total phenotypic variance. It varied between 0.05 to 0.98 for protein abundances and between 0.04 to 0.95 for fermentation traits. Altogether, protein abundance measurements were highly repeatable (median heritability of 0.53), while fermentation traits were more variable. Median broad sense heritability was 0.77 for fermentation kinetic trait, 0.49 for life-history traits, 0.36 for basic enological products and 0.32 for aromatic traits. Whatever the amount of residual variance, all genetic variance components were significant for all traits, except for *t.lag* at 18°, for which only additive variance components were significant. We found that variances associated to each genetic effect differ in a large extent between the two temperatures (shown for fermentation traits in fig. SF12).

Because of their potential interest for wine-making, BLUPs of fermentation traits are presented in section Strain characterization of Supplementary Materials. In the following, we focus on genetic variance components.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

**Table 1** Pearson's chi-square test for count data: comparison between the number of heterotic proteins in each cluster and group membership probability. The statistics clearly highlight clusters enriched of heterotic proteins (p-value<0.05). In yellow (respectively pink) are highlighted the clusters significantly enriched (resp. depleted) in heterotic proteins.

| Cluster | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Number of proteins | 11 | 168 | 39 | 65 | 144 | 102 | 627 | 24 | 50 |
| Number of heterotic proteins | 7 | 35 | 3 | 22 | 13 | 13 | 72 | 5 | 2 |
| Proportion of heterotic proteins | 0.64 | 0.21 | 0.08 | 0.34 | 0.09 | 0.13 | 0.11 | 0.21 | 0.04 |
| Chi-square standardized residuals | 4.42 | 2.56 | –1.07 | 4.40 | –1.69 | –0.35 | –2.39 | 0.91 | –1.93 |

### *Structuration of genetic variance components at the proteomic level*

A Gaussian mixture model was used to classify the proteins according to their genetic variance components. The best model clearly identified nine clusters, each characterized by a particular profile of genetic variance components (fig. 2). Cluster 1 (88.4% of good assignments) consists of 11 proteins that have high intra-specific heterosis variance and the smallest inter-specific heterosis variance. Clusters 2, 4 and 9 have very small inbreeding variance components. Clusters 2 and 4 differ from cluster 9 by their significant additive inter-specific variance component. 6.4% of proteins from cluster 2 (composed of 168 proteins with 93.2% of good assignments) can be attributed to cluster 4 and 10.4% of proteins from cluster 4 (65 proteins, 80.5% good assignments) to cluster 2. Proteins from clusters 3 (80.5% of good assignments) and 7 (93.3% of good assignments) have similar profiles.

Indeed, 19.5% of the proteins from cluster 3 can be attributed to cluster 7 and 4% of the proteins from cluster 7 can be attributed to cluster 3. Cluster 3 consists in 39 proteins with relatively higher additive and inbreeding variance components. Cluster 7 has 627 proteins with higher heterosis variance components. Proteins from cluster 5 (144 proteins, 96% of good assignments) have significant intra-specific additive variance but null inter-specific additive variance, high inbreeding and heterosis variances. On the contrary, cluster 6 (102 proteins, 96.2% of good assignments) has null intra-specific additive variance, but a small additive inter-specific, and high inbreeding and heterosis variances. Cluster 8 (96.9% of good assignments) consists of 24 proteins that have null additive variances with high inbreeding and high heterosis variances. Finally, the 50 proteins in cluster 9 (95.4% of good assignments) are characterized by a null additive inter-specific and inbreeding variance and high heterosis components. Overall the same protein is generally found in two different clusters at the two temperatures (only 37% of proteins belong to the same cluster at the two temperatures).

The nine clusters were also clearly distinguishable from each other from their pattern of correlation between variance components (fig. 3). Globally, all variance components are negatively correlated, except for the heterosis variances, $\sigma^2_{H_w}$ and $\sigma^2_{H_b}$, that are positively correlated ($r = 0.47$, fig. SF4).

Therefore, we can state that the 615 proteins at 18° and 26° form highly structured and well defined clusters according to their genetic variance component profiles.

### *Proteins sharing a similar variance component structure share functional properties*

In each protein cluster we tested for enrichment in functional categories at the two temperatures separately. Clusters were split into two groups of proteins, those measured at 18° and those measured at 26°, and the enrichment analysis was performed for each group. The statistical tests were significant for each cluster, except for cluster 1 at 18° and cluster 6 at 26° (tab. ST3). Even though one protein generally falls into two different clusters at two different temperatures, functional enrichments were globally the same at the two temperatures. Indeed, we found a high correlation between Pearson's chi-squared residuals at both temperatures, except for clusters 3 and 9 (fig. SF5). Whenever a functional category was enriched/depleted at one temperature, it also tended to be enriched/depleted at the other temperature.

Cluster 1 is enriched with proteins quantified at 26° linked to response to stress, mating and transcription, and depleted with proteins related to cell fate and protein synthesis. Cluster 3 is enriched with proteins measured at 18° linked to amino-acid and nucleotide metabolism, and at 26° to cell fate and response to stress. Cluster 6 is enriched with proteins quantified at 18° linked to protein synthesis and nucleotide metabolism, and depleted in proteins linked to metabolism, other than amino acid, nucleotide and carbon metabolism. Cluster 9 is enriched in proteins linked to transcription at both temperatures, it is enriched in proteins measured at 18° linked to response to stress and mating, and depleted in proteins linked to protein synthesis and cell fate; at 26° it is enriched in proteins linked to nucleotide metabolism and transport. The other protein clusters have the same profile at both temperatures. Cluster 2 is enriched with proteins linked to amino-acids and carbon metabolism, cell fate and response to stress, and depleted in proteins linked to transport and mating. Cluster 4 is enriched in proteins linked to amino-acid metabolism, and to stress response at 26°. Cluster 5 is enriched in proteins linked to protein synthesis, amino-acid, nucleotide and other but not carbon metabolism, and depleted in proteins linked to transcription. Cluster 7 is enriched in proteins linked to amino-acids and carbon metabolism, and depleted in proteins linked to transcription, transport and signal. Cluster 8 is enriched in proteins linked to cell fate, stress response, nucleotide metabolism and mating, and depleted in proteins linked to other metabolisms, transport and protein synthesis. Hence, genetic variance components tend to cluster proteins having similar functions at both temperatures.

Concerning the number of transcription factors, we found no correlation between the number of transcription factors and the components of genetic variation of protein abundances.

Finally, Pearson's chi-square test have been performed in order to investigate if there were differences between clusters regarding the proportion of heterotic proteins quantified in Blein-Nicolas *et al.* (2015). Results are shown in tab. 1: cluster 1, 2, 4 are enriched with heterotic proteins while in clusters 5, 7, 9 heterotic proteins are scarce ($\chi^2 = 54.29$, p-value<0.05). Hence, heterotic

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

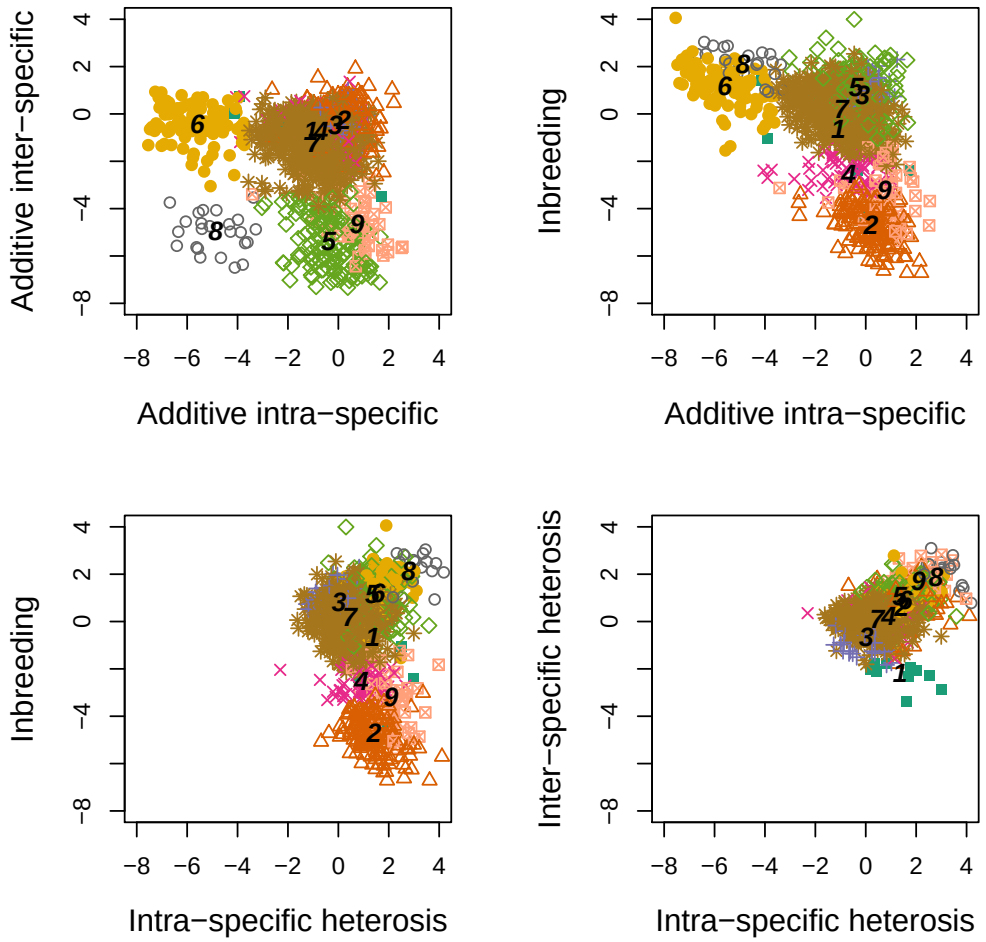



**Figure 3** Patterns of correlations between genetic variance components of protein abundances. Points correspond to proteins, type and color combinations identify the clusters obtained by their classification based on a Gaussian Mixture model. Numbers from 1 to 9 identify class centers for each cluster.

proteins are preferably found in clusters characterized by low inbreeding variance and high heterosis variance.

Briefly, despite poor correlations between variance components measured for the same protein at two temperatures, the nine clusters of proteins identified from the distribution of variance components group together proteins of similar function, based on their functional annotation. Heterotic proteins that show non-additive inheritance between parents and hybrids are mostly found in protein clusters with high heterosis and low inbreeding variance.

#### *Variance components of fermentation traits fall into the proteomic landscape*

Using for the fermentation/life history traits the same clustering approach as for the proteins, we clearly identified three profiles of genetic variance components (fig. SF3; see description in the section Structuration of genetic variability at the fermentation trait level of Supplementary Materials).

In order to compare the patterns of genetic variation of protein abundances and fermentation traits, we tried to assign fermentation traits to proteomic clusters based on the Gaussian Mixture model fitted on protein abundances profiles, as explained in section Variance component analysis of Materials and Methods. We chose for each fermentation trait the cluster of maximal membership probability. Most traits were assigned to a single protein cluster with a probability higher than 80%. The exceptions were *Sugar/EthanolYield* (26°), *X4MPP* (26°), *t.75* (26°), *t.lag* (26°) and *t.lag* at both temperatures. Average variance components for each cluster are represented in fig. 2. Altogether, the 56 fermentation traits fall into eight proteomic clusters, most of them being assigned to clusters 1 (16 traits), 2 (12 traits), 7 (12 traits), 3 (6 traits), 5 (5 traits). Note that no trait was assigned to cluster 8, which corresponds to the cluster with the lowest additive variance components. Despite similarities with protein abundance traits, fermentation traits are characterized by higher additive and inbreeding variances and globally higher contrasts in genetic variance components (fig. 4). Overall, 8 traits were attributed to the same cluster at the two temperatures: $J_{max}$, *r*, *t-*$N_{max}$, *Viability-t-75*, *X4MMP*, *Hexanoic acid*, *Hexanol*, *Ethanol*.

In addition, we investigated, for each temperature, the link between protein category in each cluster and type of fermen-

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

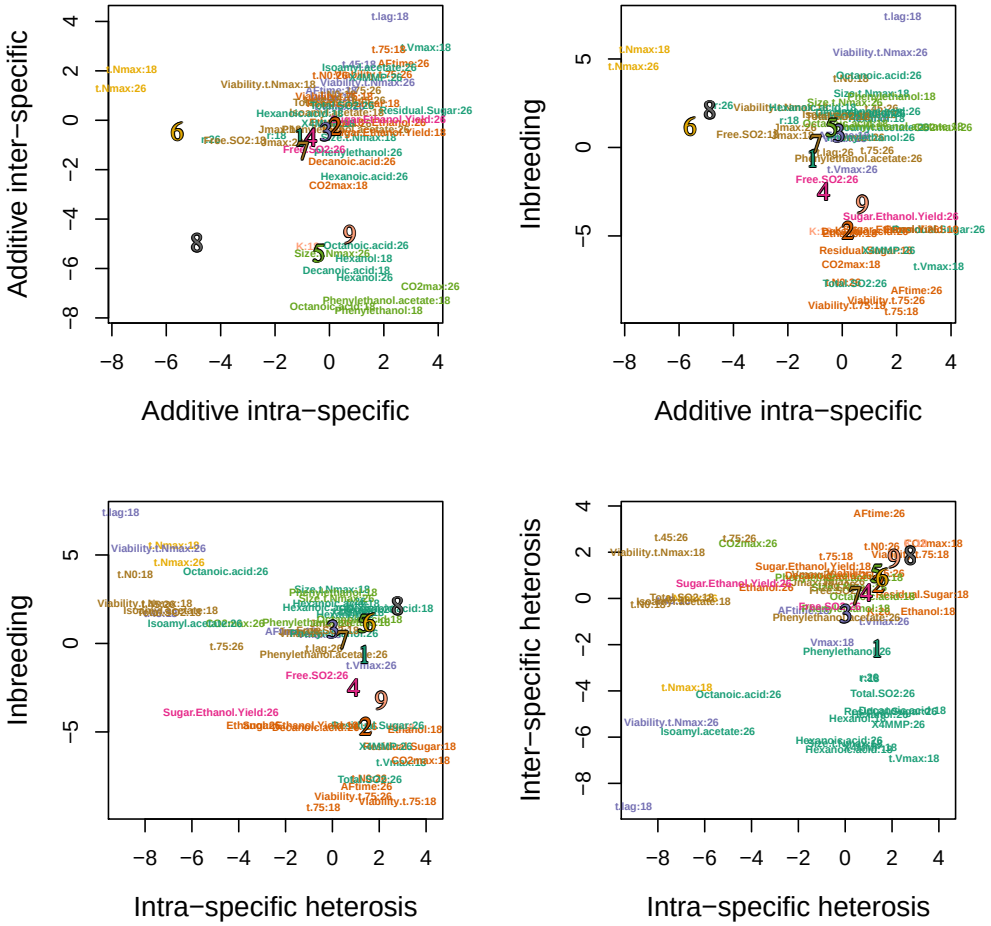



**Figure 4** Variance components of fermentation traits. Fermentation traits are assigned to clusters identified at the proteomic level based on their membership probability computed through Gaussian finite mixture models. They are identified by their label and the color of the cluster to which they are assigned. Numbers 1 to 9 identify class centers for each protein cluster.

tation trait. We see that at 18°, most Basic Enological Parameters (BEP) fall in cluster 2 where we found proteins involved in metabolism and stress response. Life History Traits fall in cluster 7 (amino-acid and carbon metabolism) and carrying capacity $K$ falls in cluster 9 (cell growth) while $t$-$N_{max}$ is found in cluster 6 (nucleotide metabolism and protein synthesis). At 26°, most Aromatic Traits fall in cluster 1 (cell fate, stress response), most Fermentation Kinetics traits are found in cluster 7 (amino-acid and carbon metabolism), and BEP are in cluster 4 (stress response).

In conclusion, traits are generally attributed to different clusters at the two temperatures, based on the underlying components of genetic variation. Those clusters are characterized by the enrichment in proteins with a certain functional category, that may vary between temperatures. Interestingly, we found an association between traits linked to different metabolic processes and proteins involved in such processes just by taking into account their genetic variance decomposition.

### *Intra-cluster correlations between variance components*

Pearson's correlation coefficients were computed for each pair of variance components within each cluster of proteins. Results clearly show different correlation structures between groups, particularly concerning correlation between inbreeding and heterosis variances (fig. 5). In cluster 1, additive variances strongly and negatively correlate with each other. In cluster 3, there is a slightly negative correlation between $\sigma^2_{A_w}$ and the heterotic variances, and there is a strong correlation between $\sigma^2_{A_b}$ and inbreeding variance. Cluster 4 is characterized by a weak negative correlation between $\sigma^2_{A_w}$, $\sigma^2_{A_b}$, $\sigma^2_{H_w}$ variances, and between $\sigma^2_{A_b}$, inbreeding and heterosis variances. Clusters 5 and 7 preserve the global correlation structure. In cluster 2, inbreeding and intra-specific heterosis variances are negatively correlated, in cluster 6 inbreeding and heterotic variances are positively correlated, in cluster 8 inbreeding and inter-specific heterosis variances are positively correlated, and in cluster 9 inbreeding and heterosis variances are negatively correlated. Altogether, when a statistical significant correlation between additive, inbreeding and heterotic variances is found, it is negative.

Additive variances tend to be negatively correlated to the heterosis and inbreeding variances, and there is no straightforward relationship between inbreeding and heterotic variances: $\sigma^2_B$ can be either negatively (cluster 9) or positively (cluster 6) correlated to both $\sigma^2_{H_b}$ and $\sigma^2_{H_w}$, negatively correlated to $\sigma^2_{H_w}$ (cluster 2), positively correlated to $\sigma^2_{H_b}$ (cluster 8). However, $\sigma^2_B$ can also be

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

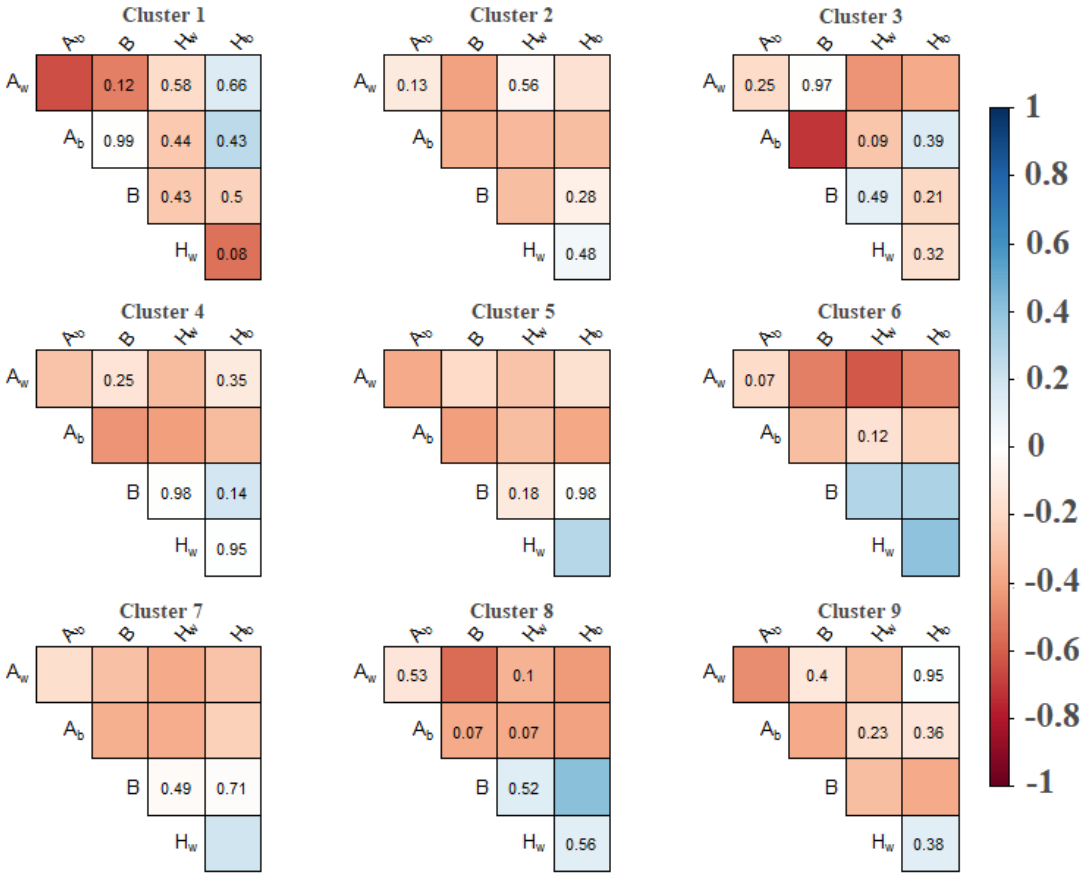



**Figure 5** Pearson's correlation test performed to investigate the intra-cluster correlations on proteomic data. For each cluster, correlation between variances of the genetic effects are indicated by a color-code. Warm colors stand for negative correlations and cold colors for positive correlations. P-value are indicated when they are not significant (p-value>0.05).

independent from either $\sigma^2_{H_w}$ or $\sigma^2_{H_b}$ (clusters 1, 2, 3, 4, 5, 7, 8).

## Discussion

In this paper, we focused on the comparative analysis of genetic variance components estimated through the decomposition of traits value quantified in a half-diallel cross during or at the end of alcoholic fermentation. The cross design involved 11 yeast strains from two related species naturally associated with wine fermentations, *S. cerevisiae* and *S. uvarum*, and the set of traits quantified spanned from protein abundances to fermentation and life-history.

Genetic variances have been estimated through a comprehensive genetic model that allowed us to decompose the phenotypic value of a cross, including the parental inbred strains, in terms of additive and interaction effects within (intra-specific) or between (inter-specific) the two species. This decomposition can be described in the following way. The parental inbred lines have two identical haploid genomes, while the hybrids have two different haploid genomes, each inherited by one parent. Additive effects refer to the average value conferred by a single haploid genome with respect to any other haploid genome from the same specie (intra-specific additive effect) or from another species (inter-specific additive effect), and interaction effects refer to the non-additive effect of a particular genotype computed as the difference between the particular diploid value and the average additive effect of its haploid genomes. The presence of the parental inbreds in the experimental design permits a decomposition of those effects into inbreeding and heterosis effects. Inbreeding effect is defined as the difference between the value of the inbred strain (with the same haploid genome twice) and the average of all the crosses having at least one copy of the haploid parental genome. Heterosis effect is defined as the difference between a single pairwise genome combination and the average value of hybrids having one or the other haploid genome. We made also the difference between intra-specific heterosis and inter-specific heterosis.

This general model could be adapted to consider mitochondrial effects, which we did not declare for biological and technical reasons given in Materials and Methods. If such effects do exist in our genetic material they are expected to be weak and confounded with other effects.

The variance components of the genetic effects defined above have been estimated using the linear mixed model (*LMM*) described in eq. 2. Whenever a variance component was significant, it meant that genetic differences were found between strains. We checked the ability of the *LMM* to estimate genetic parameters by means of computer simulations and the robustness of the estimations through bootstrap analysis. In the simulations, despite residual variances that were not well correlated to their true

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

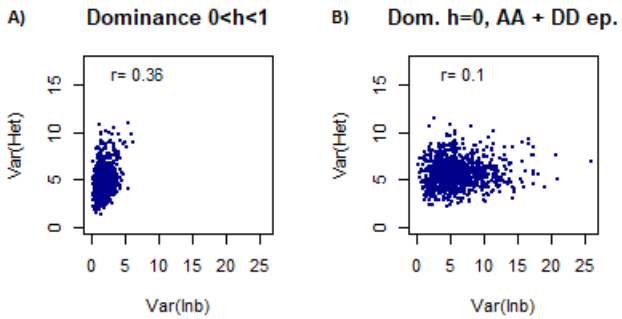



**Figure 6** Correlation between inbreeding and heterosis variance for: A) Additive model with symmetrical dominance (no epistasis), B) Additive model with dominance of the strongest allele, additive × additive and dominance × dominance epistasis. The simulated half-diallel consisted of 11 parental lines. Phenotypic values were supposed to depend on 10 loci, and the number of alleles per loci was imposed to 11. Allele values were drawn from a gamma distribution ($k$=10, $\theta$=20) and epistatic effects from a normal distribution ($\mathcal{N}(0,3)$).

value, estimated genetic variances were found to highly correlate with their true value (fig. 1). However, residuals quantified on the proteomic data highly correlate with their true value (see section Protein abundances). Bootstrap analysis, performed by sampling the 55 hybrids with replacement, conditionally to the 11 parental strains, revealed that for each variance component the estimations in the experimental sample were close to the median of the estimations in the bootstrap samples. For some traits and some variance components, the distribution of the bootstrap estimated variances were bimodal, suggesting a strong influence from a particular hybrid combination. However, it was never flat or smooth, in agreement with the non arbitrary choice of the parameters. Therefore, we are confident about the estimations of the genetic variances, conditionally to the parents of the diallel.

We were able to characterize the 615 proteins and the 28 fermentation and life-history traits quantified at 18° and 26° by a particular profile of genetic variance components despite the small number of parental inbred strains from which the half-diallel was built. Each trait has been treated at each temperature separately, considering trait × temperature as independent characters. Indeed, genotype-by-environment interactions affect very commonly phenotypic variation. In particular, it is well documented that the genetic architecture of a trait is not stable under varying environments, highlighting the fact that evolutionary processes may depend largely upon ecological conditions (Falconer 1960; Lynch and Walsh 1998; Hermisson and Wagner 2004; Robinson *et al.* 2009; Malosetti *et al.* 2013). Accordingly we found a weak correlation between genetic variances at the two temperatures.

The molecular phenotypes (protein abundances) reflect the underlying genetic factors involved in the cellular processes regulating the most integrated traits. So we investigated the distribution of the components of genetic variation of protein abundances in relation to fermentation and life-history trait variance components. We found nine clear-cut clusters of protein variance components, and we were able to assign traits to these clusters based on their genetic variance components. Overall, the profiles of the fermentation and life-history traits associated to each cluster were close to that of the proteomic level, but they were characterized by higher additive variances; further, we could not assign any trait to cluster 8, which has null additive variance components, *i.e.* which is the group with the less heritable proteins. Altogether these results reveal that the most integrated traits have a higher evolutionary potential compared to protein abundances.

We tested for cluster enrichment in protein functions, based on the functional annotation of the proteins. Clusters were found to group together proteins of similar functions. Despite the fact that 63% of the proteins were found in different clusters at the two temperatures, the metabolic functions were preserved. This suggests temperature-specific regulatory changes that achieve the maintenance of cell functions. At the trait level, 16 over 28 fermentation/life-history traits (57%) fell into the same cluster at the two temperatures (fig. SF8). For the 12 remaining traits, changes in the distribution of variance components between the two temperatures can be explained by $G \times E$ interactions.

Beside, we have shown that the clusters were characterized by a particular profile of genetic variance components, which suggests that traits that group together share a similar evolutionary history. If all traits were neutral, they would have shown the same equilibrium level of total genetic variance of approximately $2NV_m$ ($N$ the effective population size and $V_m$ the mutational variance (Lynch and Hill 1986)) with a similar partition of genetic variance components. The existence of different profiles of variance components probably reflects that the different types of traits have been subject to particular selective pressures.

Beyond, the nine clusters were clearly distinguishable from each other from their pattern of correlation between variance components. Overall, additive genetic variances were negatively correlated to inbreeding and heterosis variances. This may reveal differences in the patterns of allele frequencies at the underlying loci. In a biallelic case, additive genetic variance is always maximum for intermediate allele frequencies, while dominance and epistatic variances (which are components of inbreeding and heterosis variances) are maximum for more extreme allele frequencies (Hill *et al.* (2008)). A trait with a high additive variance is therefore expected to have lower dominance or epistatic variances. Conversely, a trait with low additive variance may exhibit high dominance and epistatic variances.

In the common view, heterosis and inbreeding are corollary effects. However, we have shown that inbreeding and heterosis variances could be negatively, positively or not correlated to each other. For a better understanding of such a decoupling, we simulated a half-diallel design between $N$ parental strains (for details see section Half-diallel simulation construction in Supplementary Materials). We computed the phenotypic values of the parental lines and hybrids starting with a simple additive model (neither dominance at any locus nor epistasis), then we added dominance and/or epistasis effects. We considered different degrees of dominance for each couple of alleles (including dominance of the strongest allele, h=0) and *additive* × *additive* and *dominance* × *dominance* epistasis, and we let the number of alleles per locus to vary. We considered all possible combinations of these effects. Finally we decomposed the values of the simulated traits into additive, inbreeding and heterosis effects.

Not surprisingly, inbreeding and heterosis variances are both null when there is neither dominance nor epistasis. If there is *additive* × *additive* epistasis with no dominance, inbreeding and heterosis variances are strictly correlated, with very low heterosis variance. In the other conditions, the results depend on the number of parental lines. With three parents, heterosis and

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

inbreeding variances are strictly equal, as it can be shown analytically (Inbreeding depression and heterosis variances are equal in three parents diallel of Supplementary Materials). Otherwise the correlation between inbreeding and heterosis variances varies in function of the number of loci affecting the trait of interest, on the frequency of alleles in the population and on the presence of dominance and epistatic effects. In general, the correlation between heterosis and inbreeding variances tends to become null when the number of parental lines, the number of alleles per locus and the number of loci increase. Given these parameters, whether there is dominance or not, and whatever the type of dominance, the lowest correlations between inbreeding and heterosis variances are observed when there are both types of epistasis together (fig. 6 and fig. SF10). However in no case we get negative correlations between the two variances. Further, we decided to consider the data obtained on all the different cases together and we run as previously a Gaussian Mixture Model to cluster genetic variances components. We computed intra-cluster correlations varying the number of alleles per locus, the number of loci and the distribution in which we drew allele values. Those correlations did not show profiles similar to those obtained with real data (correlations between genetic effects are commonly positive or null).

Classical genetic studies and modern molecular evolutionary approaches now suggest that inbreeding effects and heterosis are predominantly caused by the presence of recessive deleterious mutations in the population (Charlesworth and Charlesworth 1999; Charlesworth and Willis 2009). Therefore understanding the effects of selection against deleterious alleles is crucial. Population structure also plays a key role in this framework. Indeed, population subdivision increases homozygosity through inbreeding, an effective process for purging deleterious alleles, but it also decreases selection efficiency by decreasing the genetic diversity. Allele frequency changes also modify the genetic variance components (Hill *et al.* 2008; Barton 2017). A more complex model, which takes into account selection, allele frequency, population structure and the presence of deleterious mutations is thus needed to explain our observations. Glémin *et al.* (2003) have discussed about the patterns of correlation between inbreeding effects and heterosis in a structured population assuming low frequencies of deleterious mutations, only present in the heterozygous state. They defined within- and between-demes inbreeding depression as the decline in mean fitness of selfed individuals relative to out-crossed individuals within the demes and as the decline in mean fitness of selfed individuals relative to out-crossed individuals between demes, respectively; and heterosis as the excess in mean fitness of individuals produced by out-crosses between demes relative to mean fitness of individuals produced by out-crosses within the demes. They stated that population structure decreases within-demes inbreeding depression while it increases between-deme inbreeding depression, and that increasing the inbreeding coefficient reduces within- and between-deme inbreeding depression and heterosis. A similar result was obtained by Roze and Rousset (2004) who considered a diffusion model in a population of partially selfing individuals subdivided according to an island model, with a large but finite number of demes. They found that generally within-deme inbreeding depression and heterosis are positively correlated upon selfing and, when the degree of population subdivision is high, inbreeding depression and heterosis are negatively correlated. To our knowledge, the present study reports the first experimental example of such a decoupling.

In conclusion, our findings have special relevance in three main directions: *(i) Detection of Quantitative Trait Loci (QTL):* Additive genetic variances are crucial for the detection of genes that have a significant gene substitution effect, and heterosis/inbreeding variances for the detection of gene-gene interactions when the part of genetic variance they explain is large; *(ii) Integration of proteomic data into Genome Scale Metabolic (GSM) model:* we assigned fermentation traits to clusters obtained on the components of genetic variation of protein abundances. Traits associated to a metabolic process were linked to proteins involved to such process, therefore we are confident that integrating proteins related to the most integrated traits into a GSM could improve their prediction, with particular attention to the prediction of heterosis; *(iii) Model inbreeding and heterosis variation:* we have highlighted various patterns of variation between inbreeding and heterosis variances that cannot be explained with simple quantitative genetics models. It would be interesting to construct *in silico* experiments to search for the key parameters that drive these patterns.

## Acknowledgments

We thank very much Dr. Arnaud Le Rouzic for exciting discussions and its material to purse the preliminary analysis of the diallel design. We thank very much Dr. Monique Bolotin for helping in the functional annotation of the proteins. This work was supported by a public PhD grant of the French National research Agency (ANR) as part of the "Investissement d'Avenir" program, through the "IDI 2015" project funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02.

## Literature Cited

Abdulrehman, D., P. T. Monteiro, M. C. Teixeira, N. P. Mira, A. B. Lourenço, *et al.*, 2011 Yeastract: providing a programmatic access to curated transcriptional regulatory associations in saccharomyces cerevisiae through a web services interface. Nucleic Acids Research **39**: D136–D140.

Albertin, W., T. da Silva, M. Rigoulet, B. Salin, I. Masneuf-Pomarede, *et al.*, 2013a The mitochondrial genome impacts respiration but not fermentation in interspecific saccharomyces hybrids. PLOS ONE **8**: 1–14.

Albertin, W., P. Marullo, M. Bely, M. Aigle, A. Bourgais, *et al.*, 2013b Linking Post-Translational Modifications and Variation of Phenotypic Traits. Molecular & Cellular Proteomics **12**: 720–735.

Barton, N. H., 2017 How does epistasis influence the response to selection? Heredity (Edinb) **118**.

Blein-Nicolas, M., W. Albertin, T. da Silva, B. Valot, T. Balliau, *et al.*, 2015 A systems approach to elucidate heterosis of protein abundances in yeast. Mol Cell Proteomics **14**: 2056–71.

Blein-Nicolas, M., W. Albertin, B. Valot, P. Marullo, D. Sicard, *et al.*, 2013 Yeast proteome variations reveal different adaptive responses to grape must fermentation. Molecular Biology and Evolution **30**: 1368.

Blein-Nicolas, M., H. Xu, D. de Vienne, C. Giraud, S. Huet, *et al.*, 2012 Including shared peptides for estimating protein abundances: A significant improvement for quantitative proteomics. PROTEOMICS **12**: 2797–2801.

Bulmer, M. G., 1980 *The mathematical theory of quantitative genetics /M.G. Bulmer*. Clarendon Press ; New York : Oxford University Press Oxford.

Charlesworth, B. and D. Charlesworth, 1999 The genetic basis of inbreeding depression. Genetical Research **74**: 329–340.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1991 Multilocus models of inbreeding depression with synergistic selection and partial self-fertilization. Genetics Research **57**: 177–194.

Charlesworth, D. and J. H. Willis, 2009 The genetics of inbreeding depression. Nature Reviews Genetics **10**: 783—-796.

Cherry, J. M., E. L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, *et al.*, 2012 Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Research **40**: D700–705.

Cockerham, C. C. and B. S. Weir, 1977 Quadratic analyses of reciprocal crosses. Biometrics **33**: 187–203.

da Silva, T., W. Albertin, C. Dillmann, M. Bely, S. la Guerche, *et al.*, 2015 Hybridization within saccharomyces genus results in homoeostasis and phenotypic novelty in winemaking conditions. PLOS ONE **10**: 1–24.

Davenport, C. B., 1908 Degeneration, albinism and inbreeding. Science **28**: 454–455, WOS:000201859500057.

Eberhart, S. A. and C. O. Gardner, 1966 A General Model for Genetic Effects. Biometrics **22**: 864–881.

Falconer, D. S., 1960 *Introduction to quantitative genetics*. New York,: Ronald Press Co.

Fievet, J. B., C. Dillmann, and D. de Vienne, 2010 Systemic properties of metabolic networks lead to an epistasis-based model for heterosis. Theoretical and Applied Genetics **120**: 463–473, WOS:000272803700025.

Fiévet, J. B., T. Nidelet, C. Dillmann, and D. de Vienne, 2018 Heterosis is a systemic property emerging from nonlinear genotype-phenotype relationships: evidence from in vitro genetics and computer simulations. Frontiers in Genetics **9**.

Glémin, S., J. Ronfort, and T. Bataillon, 2003 Patterns of inbreeding depression and architecture of the load in subdivided populations. Genetics **165**: 2193–2212.

Gowen, J. W., 1952 *Heterosis*. Iowa state pr edition.

Graham, G. I., D. W. Wolff, and C. W. Stuber, 1997 Characterization of a yield quantitative trait locus on chromosome five of maize by fine mapping. Crop Sci. **37**: 1601–1610, WOS:A1997XZ35500033.

Greenberg, A. J., S. R. Hackett, L. G. Harshman, and A. G. Clark, 2010 A hierarchical bayesian model for a novel sparse partial diallel crossing design. Genetics .

Griffing, B., 1956 Concept of general and specific combining ability in relation to diallel crossing systems. Australian Journal of Biological Sciences **9**: 463–493.

Gumedze, F. and T. Dunne, 2011 Parameter estimation and inference in the linear mixed model. Linear Algebra and its Applications **435**: 1920 – 1944.

Hallauer, A., W. Russell, and K. LAMKEY, 1988 Corn breeding: 463-564. Sprague, GF and. IW Dudley: Corn and corn improvement. Agron. Monogr.(third edition). ASA, CSSA and SSSA Madison, WI .

Hallauer, A. R. and J. B. M. Filho, 1988 *Quantitative Genetics in Maize Breeding*. Iowa State University Press.

Hermisson, J. and G. P. Wagner, 2004 The Population Genetic Theory of Hidden Variation and Genetic Robustness. Genetics **168**: 2271–2284.

Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLOS Genetics **4**: 1–10.

Huang, X., S. Yang, J. Gong, Q. Zhao, Q. Feng, *et al.*, 2016 Genomic architecture of heterosis for yield traits in rice. Nature **537**: 629–633.

Hull, F., 1946 Overdominance and Corn Breeding Where Hybrid Seed Is Not Feasible. Journal of the American Society of Agronomy **38**: 1100–1103, WOS:A1946UC58500007.

Kanehisa, M., M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, 2017 KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research **45**: D353–D361.

Kanehisa, M. and S. Goto, 2000 KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research **28**: 27–30.

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, 2016 KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research **44**: D457–462.

Kurtz, Z. D., C. L. Müller, E. R. Miraldi, D. R. Littman, M. J. Blaser, *et al.*, 2015 Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLOS Computational Biology **11**: e1004226.

Lande, R. and D. W. Schemske, 1985 The evolution of self-fertilization and inbreeding depression in plants. i. genetic models. Evolution **39**: 24–40.

Lariepe, A., B. Mangin, S. Jasson, V. Combes, F. Dumas, *et al.*, 2012 The Genetic Basis of Heterosis: Multiparental Quantitative Trait Loci Mapping Reveals Contrasted Levels of Apparent Overdominance Among Traits of Agronomical Interest in Maize (Zea mays L.). Genetics **190**: 795–U835, WOS:000300621200037.

Lenarcic, A. B., K. L. Svenson, G. A. Churchill, and W. Valdar, 2012 A general bayesian approach to analyzing diallel crosses of inbred strains. Genetics **190**: 413–435.

Lynch, M. and W. G. Hill, 1986 PHENOTYPIC EVOLUTION BY NEUTRAL MUTATION. Evolution; International Journal of Organic Evolution **40**: 915–935.

Lynch, M. and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer, Google-Books-ID: UhCCQgAACAAJ.

Malosetti, M., J.-M. Ribaut, and F. A. van Eeuwijk, 2013 The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. Frontiers in Physiology **4**.

Martì-Raga, M., E. Peltier, A. Mas, G. Beltran, and P. Marullo, 2017 Genetic Causes of Phenotypic Adaptation to the Second Fermentation of Sparkling Wines in Saccharomyces cerevisiae. G3: Genes, Genomes, Genetics **7**: 399–412.

Melchinger, A. E. and R. K. Gumber, 1998 Overview of Heterosis and Heterotic Groups in Agronomic Crops. In *Concepts and Breeding of Heterosis in Crop Plants*, edited by K. R. Larnkey and J. E. Staub, pp. 29–44, Crop Science Society of America, USA.

Monteiro, P. T., N. D. Mendes, M. C. Teixeira, S. d'Orey, S. Tenreiro, *et al.*, 2008 Yeastract-discoverer: new tools to improve the analysis of transcriptional regulatory associations in saccharomyces cerevisiae. Nucleic Acids Research **36**: D132–D136.

Omholt, S. W., E. Plahte, L. Øyehaug, and K. Xiang, 2000 Gene Regulatory Networks Generating the Phenomena of Additivity, Dominance and Epistasis. Genetics **155**: 969–980.

Powers, L., 1944 An expansion of Jones's theory for the explanation of heterosis. The American Naturalist **78**: 275–280.

Ramya, A. R., L. Ahamed M, C. T. Satyavathi, A. Rathore, P. Katiyar, *et al.*, 2018 Towards Defining Heterotic Gene Pools in Pearl Millet [Pennisetum glaucum (L.) R. Br.]. Frontiers in Plant Science **8**.

Redden, R., 1991 The effect of epistasis on chromosome mapping of quantitative characters in wheat. I. Time to spike emergence. Australian Journal of Agricultural Research **42**: 1.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Robinson, M. R., A. J. Wilson, J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, *et al.*, 2009 The Impact of Environmental Heterogeneity on Genetic Architecture in a Wild Population of Soay Sheep. Genetics **181**: 1639–1648.

Ronnegard, L., X. Shen, and M. Alam, 2010 hglm: A package for fitting hierarchical generalized linear models. The R Journal **2**: 20–28.

Roze, D. and F. Rousset, 2004 Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. Genetics **167**: 1001–1015.

Ruepp, A., A. Zollner, D. Maier, K. Albermann, J. Hani, *et al.*, 2004 The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. Nucleic Acids Research **32**: 5539–5545.

Schnable, P. S. and N. M. Springer, 2013 Progress toward understanding heterosis in crop plants. Annu Rev Plant Biol **64**: 71–88.

Scrucca, L., M. Fop, T. B. Murphy, and A. E. Raftery, 2016 mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. The R Journal **8**: 205–233.

Seymour, D. K., E. Chae, D. G. Grimm, C. Martín Pizarro, A. Habring-Müller, *et al.*, 2016 Genetic architecture of non-additive inheritance in Arabidopsis thaliana hybrids. Proc. Natl. Acad. Sci. U.S.A. **113**: E7317–E7326.

Shao, H., L. C. Burrage, D. S. Sinasac, A. E. Hill, S. R. Ernest, *et al.*, 2008 Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis. Proceedings of the National Academy of Sciences **105**: 19910–19914.

Shull, G. H., 1908 The Composition of a Field of Maize. Journal of Heredity **os-4**: 296–301.

Sprague, G. F. and E. L. Tatum, 1942 General vs. specific combining ability in single crosses of corn. PROTEOMICS **34**: 923–932.

Teixeira, M. C., P. Monteiro, P. Jain, S. Tenreiro, A. R. Fernandes, *et al.*, 2006 The yeastract database: a tool for the analysis of transcription regulatory associations in saccharomyces cerevisiae. Nucleic Acids Research **34**: D446–D451.

Teixeira, M. C., P. T. Monteiro, J. F. Guerreiro, J. P. Gonçalves, N. P. Mira, *et al.*, 2014 The yeastract database: an upgraded information system for the analysis of gene and genomic transcription regulation in saccharomyces cerevisiae. Nucleic Acids Research **42**: D161–D166.

Tsagris, M. T., S. Preston, and A. T. A. Wood, 2011 A data-based power transformation for compositional data. ArXiv e-prints .

Wright, S., 1934 Physiological and evolutionary theories of dominance. American Naturalist **68**: 24–53, WOS:000200907000002.

Xiao, J., J. Li, L. Yuan, and S. Tanksley, 1995 Dominance Is the Major Genetic-Basis of Heterosis in Rice as Revealed by Qtl Analysis Using Molecular Markers. Genetics **140**: 745–754, WOS:A1995RA36600028.

Zhu, J. and B. S. Weir, 1996 Mixed model approaches for diallel analysis based on a bio-model. Genetical Research **68**: 233–240.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

## Supplementary Materials

### Subcompositional dominance and distances

We consider the central log-ratio transformation in order to pursue our analysis without considering both the block effect and residuals for the more integrated traits and residuals for protein abundances. We are aloud to do so since the clr-transformation satisfies the subcompositional dominance property, *i.e.*, for each couple of vectors, $\boldsymbol{x}$ and $\boldsymbol{y}$, and for each pair of subvectors $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, obtained by selecting the same set of components, the distance between the subvectors is always less than or equal to the distance between the original vectors, *i.e.*

$$d(\boldsymbol{x},\boldsymbol{y}) \geq d(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}) \tag{S1}$$

Therefore, for each $\boldsymbol{z}$ such that $d(\boldsymbol{x},\boldsymbol{y}) \geq d(\boldsymbol{x},\boldsymbol{z})$, we have that, dividing eq.(S1) by $\frac{d(\boldsymbol{x},\boldsymbol{z})}{d(\hat{\boldsymbol{x}},\hat{\boldsymbol{z}})} \geq 1$

$$\alpha d(\hat{\boldsymbol{x}},\hat{\boldsymbol{z}}) \geq k d(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}) \tag{S2}$$

where $\alpha = \frac{d(\boldsymbol{x},\boldsymbol{y})}{d(\boldsymbol{x},\boldsymbol{z})} \geq 1$ and $k = \frac{d(\hat{\boldsymbol{x}},\hat{\boldsymbol{z}})}{d(\boldsymbol{x},\boldsymbol{z})} \leq 1$. So, since $k/\alpha \leq 1$

$$d(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}}) \geq d(\hat{\boldsymbol{x}},\hat{\boldsymbol{z}}) \tag{S3}$$

As a consequence, distance relationship between the original vectors is preserved by selected subvectors.

### The fitting algorithm

The hglm package implements the estimation algorithm for hierarchical generalized linear models. It fits generalized linear models with random effects, where the random effect may come from a conjugate exponential-family distribution (Gaussian, Gamma, Beta or inverse-Gamma) and it is possible to explicitly specify the design matrices both for the fixed and random effects, which allows fitting correlated random effects as well as random regression models.
In order to perform the diallel analysis, we considered $\mathbf{y}$, the vector of observations for the trait of interest, and we re-wrote the model (eq.(2)) in matrix a form:

$$\boldsymbol{y} = X\boldsymbol{\beta} + Z\boldsymbol{u} + \boldsymbol{\epsilon} \tag{S4}$$

where $X$ is the design matrix for the fixed effects, $Z$ the design matrix for the random effects, $\boldsymbol{\beta} = (\mu, \beta_{S.uvarum}, \beta_{S.cerevisiae})$ and $\boldsymbol{u} = (\boldsymbol{A_w}, \boldsymbol{A_b}, \boldsymbol{B}, \boldsymbol{H_w}, \boldsymbol{H_b})$ are respectively the vectors of fixed effects parameters and random effects parameters, and $\boldsymbol{\epsilon}$ is the vector of random errors. With this notation, the construction of the model is straight forward since we just have to construct the design matrices for both fixed and random effects.

Let $n$ be the number of observations, $J$ the total number of parental strains, $N_{intra}$ (resp. $N_{inter}$) the number of intra-specific (resp. inter-specific) crosses, and $K$ the total number of random effects parameters. $X$ is a $n \times 3$ matrix, with, by construction, the first column equal to $(1, 1, ..., 1)$, while the elements of the second and third columns (for respectively *S. uvarum* and *S. cerevisiae*) are 1 or 0 depending on whether the strain is inbred and or not.
$Z$ will be a $n \times K$ matrix and, more precisely, it can be thought as the following block matrix:

$$Z = \left[ \begin{array}{c:c:c:c:c} Z_{\boldsymbol{A_w}} & Z_{\boldsymbol{A_b}} & Z_{\boldsymbol{B}} & Z_{\boldsymbol{H_w}} & Z_{\boldsymbol{H_w}} \end{array} \right] \tag{S5}$$

where $Z_{\boldsymbol{A_w}}$, $Z_{\boldsymbol{A_b}}$, $Z_{\boldsymbol{B}}$, $Z_{\boldsymbol{H_w}}$, $Z_{\boldsymbol{H_b}}$ denote the design matrices, respectively, of the random effect parameters $\boldsymbol{A_w}$, $\boldsymbol{A_b}$, $\boldsymbol{B}$, $\boldsymbol{H_w}$ and $\boldsymbol{H_b}$. In particular, $Z_{\boldsymbol{A_w}}$, $Z_{\boldsymbol{A_b}}$, $Z_{\boldsymbol{B}}$ are $n \times J$ matrices, $Z_{\boldsymbol{H_w}}$ is a $n \times N_{intra}$ matrix and $Z_{\boldsymbol{H_b}}$ is a $n \times N_{inter}$ matrix with entries:

$$z_{\boldsymbol{A_{w_{ij}}}} = \begin{cases} 2 & \text{If the } i\text{-observation belongs to a parental strain, the } j\text{-th;} \\ 1 & \text{If the } i\text{-observation belongs to an hybrid achieved through an intra- specific cross} \\ & \text{in which the parental strain } j \text{ is involved;} \\ 0 & \text{otherwise;} \end{cases} \tag{S6}$$

$$z_{\boldsymbol{A_{b_{ij}}}} = \begin{cases} 1 & \text{If the } i\text{-observation belongs to an hybrid achieved through an inter- specific cross} \\ & \text{in which the parental strain } j \text{ is involved;} \\ 0 & \text{otherwise;} \end{cases} \tag{S7}$$

$$z_{\boldsymbol{B_{ij}}} = \begin{cases} 1 & \text{If the } i\text{-observation belongs to a parental strain, the } j\text{-th;} \\ 0 & \text{otherwise;} \end{cases} \tag{S8}$$

and, enumerating the intra-specific/inter-specific hybrid strains with $k_{intra}/k_{inter}$ from 1 to $N_{intra}/N_{inter}$, respectively,

$$z_{\boldsymbol{H_{w_{ik_{intra}}}}} = \begin{cases} 1 & \text{If the } i\text{-observation belongs to the } k_{intra}\text{- hybrid strain;} \\ 0 & \text{otherwise;} \end{cases} \tag{S9}$$

$$z_{\boldsymbol{H_{b_{ik_{inter}}}}} = \begin{cases} 1 & \text{If the } i\text{-observation belongs to the } k_{inter}\text{- hybrid strain;} \\ 0 & \text{otherwise;} \end{cases} \tag{S10}$$

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

### Half-diallel simulation construction

In order to elucidate our findings about the decoupling of inbreeding and heterotic variances, we simulated a half-diallel between $N$ parental strains. We supposed the phenotypic values of each trait to depend on a fixed number of loci, $L$, and we considered all the possible combinations of genetic effects, namely presence/absence of dominance, of additive $\times$ additive epistasis and of additive $\times$ additive epistasis.
We let the number of alleles at each locus to vary between 1 and $N$ and we drew values for allele $a$ at locus $i$ ($a_i$) from a Gamma distribution ($\Gamma(k,\theta)$), for additive $\times$ additive epistatic effect between $a_i$ and $a_j$ ($aa^{ij}$) and for dominance $\times$ dominance epistatic effect ($dd^{ij}$) from a Gaussian distribution ($\mathcal{N}(0,\sigma^2)$). The dominance effect between alleles $a$ and $b$ at locus $i$ ($d^i_{ab}$) are drawn from an uniform distribution $\mathcal{U}(0,m)$ with $m = 0.5$ for dominance of the strongest allele, and $m = 1$ for symmetrical dominance. Therefore, the phenotypic value of the parental lines $P_k$ and of the hybrid, $H_{lk}$, between parents $P_k$ and $P_l$ are given by:

**1) Additive model**

$$y_{P_k} = 2\sum_i k_i\,, \quad y_{H_{lk}} = \sum_i k_i + \sum_i l_i \tag{S11}$$

**2) Additive model plus dominance**

$$y_{P_k} = 2\sum_i k_i\,, \quad y_{H_{lk}} = \sum_i k_i + \sum_i l_i + \sum_i d^i_{kl} \tag{S12}$$

**3) Additive model plus *additive* $\times$ *additive* effect**

$$y_{P_k} = 2\sum_i k_i + \sum_{ij} aa^{ij}\,, \quad y_{H_{lk}} = \sum_i k_i + \sum_i l_i \tag{S13}$$

**4) Additive model plus *dominance* $\times$ *dominance* effect**

$$y_{P_k} = 2\sum_i k_i\,, \quad y_{H_{lk}} = \sum_i k_i + \sum_i l_i + \sum_{ij} dd^{ij} \tag{S14}$$

**5) Additive model plus *additive* $\times$ *additive* and *dominance* $\times$ *dominance* effect**

$$y_{P_k} = 2\sum_i k_i + \sum_{ij} aa^{ij}\,, \quad y_{H_{lk}} = \sum_i k_i + \sum_i l_i + \sum_{ij} dd^{ij} \tag{S15}$$

**6) Additive model plus dominance and *additive* $\times$ *additive* effect**

$$y_{P_k} = 2\sum_i k_i + \sum_{ij} aa^{ij}\,, \quad y_{H_{lk}} = \sum_i k_i + \sum_i l_i + \sum_i d^i_{kl} \tag{S16}$$

**7) Additive model plus dominance and *dominance* $\times$ *dominance* effect**

$$y_{P_k} = 2\sum_i k_i\,, \quad y_{H_{lk}} = \sum_i k_i + \sum_i l_i + \sum_i d^i_{kl} + \sum_{ij} dd^{ij} \tag{S17}$$

**8) Additive model plus dominance, *additive* $\times$ *additive* and *dominance* $\times$ *dominance* effect**

$$y_{P_k} = 2\sum_i k_i + \sum_{ij} aa^{ij}\,, \quad y_{H_{lk}} = \sum_i k_i + \sum_i l_i + \sum_i d^i_{kl} + \sum_{ij} dd^{ij} \tag{S18}$$

### Inbreeding depression and heterosis variances are equal in three parents diallel

Inbreeding and heterosis variances are equal in the particular case of a three-parent diallel when no maternal effect is present. It can be easily seen by the direct computation of their value.
In order to do that we decompose the phenotypic values of the $i-$parent, $P_i$, as

$$P^d_i = \mu + 2A_i \tag{S19}$$

and of the $i \times j$ hybrid, $H_{ij}$, as

$$H^d_{ij} = \mu + A_i + A_j \tag{S20}$$

where $\mu = \frac{1}{6}(P_1 + P_2 + P_3 + H_{12} + H_{13} + H_{23})$ is the mean phenotypic value of the population and

$$A_i = \frac{1}{3}(P_i + \sum_{j\neq i} H_{ij}) - \mu \tag{S21}$$

the $GCA$ of strain $i$. Therefore, we can express the inbreeding depression variance as the deviation of the decomposed phenotypic value of the parents, $P^d$, and their true value $P$

$$Var(inbreeding) = \frac{1}{3}\sum_i (P^d_i - P_i - (\overline{P^d - P}))^2 \tag{S22}$$

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and the heterosis variance analogously

$$Var(heterosis) = \frac{1}{3}\sum_{i<j}(H^d_{ij} - H_{ij} - (\overline{H^d - H}))^2 \tag{S23}$$

In which we have used the fact that $H_{ij} = H_{ji}$, since no maternal effects are present.

Substituting S19 and S21 in S22, we get

$$\begin{aligned} Var(inbreeding) &= \frac{1}{3}\sum_{i=1}^{3}(\mu + 2A_i - P_i - \frac{1}{3}\sum_{k=1}^{3}(\mu + 2A_k - P_k))^2 = \\ &= \frac{1}{3}\sum_{i=1}^{3}(\mu + \frac{2}{3}(P_i + \sum_{j\neq i} H_{ij}) - 2\mu - P_i - \frac{1}{3}\sum_{k=1}^{3}(\mu + \frac{2}{3}(P_k + \sum_{j\neq k} H_{kj}) - 2\mu - P_k))^2 = \\ &= \frac{1}{243}\sum_{i=1}^{3}(6P_i + 6\sum_{j\neq i} H_{ij} - 9\mu - 9P_i - \sum_{k=1}^{3}(2P_k + 2\sum_{j\neq k} H_{kj} - 3\mu - 3P_k))^2 = \\ &= \frac{1}{243}\sum_{i=1}^{3}(6\sum_{j\neq i} H_{ij} - 9\mu - 3P_i + \sum_{k=1}^{3} P_k - 4\sum_{j<k} H_{kj} + 9\mu)^2 = \\ &= \frac{1}{243}\sum_{i=1}^{3}(-2P_i + P_j + P_k + 2H_{ik} + 2H_{ij} - 4H_{kj})^2 \end{aligned} \tag{S24}$$

where $i \neq j \neq k$. In the same way, substituting S20 and S21 in S23, we get

$$\begin{aligned} Var(heterosis) &= \frac{1}{3}\sum_{i<j}(\mu + A_i + A_j - H_{ij} - \frac{1}{3}\sum_{k<m}(\mu + A_k + A_m - H_{km}))^2 = \\ &= \frac{1}{3}\sum_{i<j}(\frac{1}{3}(P_i + P_j + \sum_{k\neq i} H_{ik} + \sum_{k\neq j} H_{jk}) - \mu - H_{ij} - \frac{1}{3}\sum_{k<m}(\frac{1}{3}(P_k + P_m + \sum_{l\neq k} H_{kl} + \sum_{l\neq m} H_{ml}) - \mu - H_{km}))^2 = \\ &= \frac{1}{243}\sum_{i<j}(3(P_i + P_j + \sum_{k\neq i} H_{ik} + \sum_{k\neq j} H_{jk}) - 9\mu - 9H_{ij} - (2\sum_{k} P_k + \sum_{k<m} H_{km} - 9\mu))^2 = \\ &= \frac{1}{243}\sum_{i<j}(3(P_i + P_j + \sum_{k\neq i} H_{ik} + \sum_{k\neq j} H_{jk}) - 9H_{ij} - 2\sum_{k} P_k - \sum_{k<m} H_{km})^2 = \\ &= \frac{1}{243}\sum_{i<j}(P_i + P_j - 2P_k - 4H_{ij} + 2H_{ik} + 2H_{jk})^2 \end{aligned} \tag{S25}$$

where again $i \neq k \neq j$. Therefore,

$$\begin{aligned} Var(inbreeding) &= \frac{1}{243}\sum_{i=1}^{3}(-2P_i + P_j + P_k + 2H_{ik} + 2H_{ij} - 4H_{kj})^2 = \\ &= \frac{1}{243}((-2P_1 + P_2 + P_3 + 2H_{12} + 2H_{13} - 4H_{23})^2 + (-2P_2 + P_1 + P_3 + 2H_{12} + 2H_{23} - 4H_{13})^2 + \\ &+ (-2P_3 + P_1 + P_2 + 2H_{13} + 2H_{23} - 4H_{12})^2 = \frac{1}{243}\sum_{i<j}(-2P_k + P_i + P_j + 2H_{ik} + 2H_{jk} - 4H_{ij})^2 = Var(heterosis) \end{aligned} \tag{S26}$$

### Structuration of genetic variability at the fermentation trait level

A Gaussian mixture model is run to classify life-history and fermentation traits according to their genetic variance components.

The best model clearly identify three clusters (fig.SF3 and fig.SF6). Cluster 1 (99.9% of good assignments) is composed by 9 traits, characterized by having null inter-specific additive variance component, relatively low inter-specific heterosis variance and high intra-specific additive and inbreeding components. In this cluster we can find most volatile compounds such as *Octanoic acid* and *Hexanol* at both temperatures, *Phenyl-2-ethanol*, *Phenyl-2-ethanol acetate* and *Decanoic acid* at 18°C, the kinetic parameter $CO_{2max}$ and the life-history trait *Size-t-*$N_{max}$ at 26°C. Cluster 2 (98.9% of good assignments) consists of 28 traits that are characterized by high inter-specific additive and inbreeding components ($\sigma^2_{\boldsymbol{A_b}}$ and $\sigma^2_{\boldsymbol{B}}$), relatively low heterosis ($\sigma^2_{\boldsymbol{H_w}}$ and $\sigma^2_{\boldsymbol{H_b}}$) and intra-specific additive variances ($\sigma^2_{\boldsymbol{A_w}}$). Most kinetic parameters and life-history traits belongs to this cluster: *t-lag*, $V_{max}$, *t-45*, *r*, *t-*$N_{max}$, $J_{max}$ and *Viability-t-*$N_{max}$ at both temperatures; *t-*$V_{max}$ and *t-75* at 26°C; *AFtime*, *t-*$N_0$, *Size-t-*$N_{max}$ at 18°C. We can also find some basic enological parameters and aromatic traits - *Isoamyl acetate* and *Hexanoic acid* at both temperatures; *Phenyl-2-ethanol* and *Phenyl-2-ethanol acetate* at 26°C; *X4MMP*, *Free* $SO_2$ and *Total* $SO_2$ at 18°C. Traits attributed to cluster 3 (19 traits, 97.3% of good assignments) have high additive and heterotic variances and null inbreeding variance. The rest of the basic enological parameters and aromatic traits along with some kinetics parameters and life-history traits belongs to it.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

As for protein abundances, we choose to consider life-history and fermentation traits at two temperatures (18°C and 26°C) as different traits. Indeed, after computation of genetic variance components for each trait, correlations between temperatures are not found to be significant except for 6 traits (*t-*$V_{max}$, *t-45*, *r*, *t-*$N_{max}$, *Viability-t-*$N_{max}$ and *Hexanol*) that are highly and positively correlated. All of them fall in the same cluster at the two temperatures, except *t-*$V_{max}$. Overall, we find that 79% of traits do not belong to the same cluster at the two temperatures. Further, Pearson's correlation tests are performed to investigate the correlation between genetic effects at the two temperatures. They were not significant except for the additive inter-specific component ($cor = 0.74$, p-value<0.05). Therefore, at the fermentation trait level, genotype by environment interactions predominate.

Globally, correlations between variance components, when present, are found to be negative (fig.SF4). However, the pattern changes when considering intra-group correlations. Indeed, in cluster 2, even if inbreeding is negatively correlated to the heterotic variances, it is positively correlated to the additive inter-specific variance, and in cluster 3, additive genetic variances are positively correlated to each other. In cluster 1, there is no statistical significant correlation between genetic effects (fig.SF7).

Therefore, we can state that three well defined groups of traits can be differentiated according to their genetic variance profiles and we show that the part of phenotypic variation explained by the model's parameters depends on trait's category and temperature: in cluster 1, we can find mostly aromatic traits; in cluster 2 kinetics parameters and life-history traits and in cluster 3 most enological parameters. Further, closely related phenotypes show similar profiles in terms of variance components, such as $CO_{2max}$, *Ethanol* and *Residual Sugar* that clusters together at 18°C; *Total* $SO_2$ and *Free* $SO_2$ are found in cluster 2 at 18°C and in cluster 3 at 26°C; *t-*$N_0$ and *t-lag* in cluster 2 at 18°C. We finally see that inbreeding variance can be either negatively, or not correlated to heterotic effects.

## Strain characterization

We characterized the strains based on their genetic contribution to the total phenotypic value of a trait at a certain temperature (fig. SF11). Strain D1 is found to be the strain with the lowest additive contribution for *Phenyl-2-ethanol* at both temperatures and for *Sugar.Ethanol.Yield* (except in inter-specific crosses at 18°C), with the highest additive intra-specific contribution for *Decanoic* and *Octanoic acid*, while displaying the highest heterosis contribution for *Octanoic acid* when crossed with E2 at 18°C, with E5 and U1 at 26°C, and for *Decanoic acid* when crossed with E4 at 26°C and U2 at 18°C. D2 and E2 strains have the highest or lowest additive contributions across almost all traits, mostly fermentation kinetics parameters and life history traits. In particular, D2 strain shows the highest intra- and inter-specific additive effects, and inbreeding values for *t.45*, *t.75* and *AFtime* at both temperatures, where the highest heterosis effect is achieved when crossed with E2, U1 for *t*.45 at 18°C, with E5 and U1 for *t*.75 with the first at both temperatures and the latter at 18°C. Similarly, the additive intra-specific effect of U4 is the highest or the lowest for almost all aromatic traits at 18°C (higher for *Phenyl-2-ethanol, Hexanol* and *Hexanoic acid*; lowest for *Decanoic acid* and *Octanoic acid*). Strain U1 shows the highest additive inter-specific effect in aromatic traits at 26°C (*Phenyl-2-ethanol, Phenyl-2-ethanol acetate, Hexanol, Hexanoic acid* and *Octanoic acid*). In particular, the heterosis effect in the inter-specific cross with strain D2 is the highest for *Hexanol* and with strain E2 for *Phenyl-2-ethanol*. For all traits, E5 produces intermediate heterosis values when crossed with E2, E3, E4, W1, U1 and U4 at 18°C, but its cross with E4 results in the highest heterosis value for *t.*$N_{max}$, and the lowest for *Decanoic acid* with E3 and for *Total* $SO_2$ with W1 at 26°C. In the same way, crosses between E3 and U1, U2 or U3, between E4 and U1 or W1 never show extreme heterosis values for any trait.

## Supplementary tables

**Table ST2** Diallel table representing the mitochondrial inheritance for each phenotyped cross: the data clearly shows too many *unknown*s to enter a mitochondrial effect in the model. Backslashes indicate the not phenotyped reciprocals.

| P1\ P2 | D1 | D2 | E2 | E3 | E4 | E5 | W1 | U1 | U2 | U3 | U4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **D1** | D1 | D2 | unknown | \ | unknown | \ | unknown | \ | U2 | U3 | U4 |
| **D2** | \ | D2 | E2 | \ | E4 | \ | W1 | \ | \ | \ | \ |
| **E2** | \ | \ | E2 | unknown | unknown | E5 | unknown | \ | U2 | U3 | U4 |
| **E3** | D1 | D2 | \ | E3 | unknown | \ | W1 | \ | U2 | U3 | \ |
| **E4** | \ | \ | \ | \ | E4 | E5 | W1 | U1 | U2 | U3 | U4 |
| **E5** | D1 | unknown | \ | unknown | \ | E5 | unknown | U1 | U2 | U3 | \ |
| **W1** | \ | \ | \ | \ | \ | \ | W1 | \ | U2 | U3 | \ |
| **U1** | D1 | D2 | E2 | E3 | \ | \ | CW1 | U1 | \ | \ | \ |
| **U2** | \ | D2 | \ | \ | \ | \ | \ | unknown | U2 | \ | \ |
| **U3** | \ | D2 | \ | \ | \ | \ | \ | unknown | unknown | U3 | \ |
| **U4** | \ | unknown | \ | E3 | \ | E5 | W1 | unknown | unknown | unknown | U4 |

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

**Table ST3 Pearson's chi-square.** For each cluster and at each temperature (18° and 26°) we tested for enrichment in proteins belonging to a certain functional category using as prior probability the frequency of proteins functional category based on MIPS database. In yellow (resp. pink) are highlighted the functional category enhanced (resp. depleted) for each cluster and at each temperature when the statistical test was significant.

| Cluster | T | P-val | Metabolism (other) | Amino acid metabolism | Nucleotide metabolism | Carbon metabolism | Replication | Transcription | Protein synthesis | Protein fate | Transport | Stress | Mating | Signal | Unknown |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 18 | 0.873 | -0.092 | 0.592 | -0.580 | 0.296 | -0.637 | -0.845 | 0.859 | 1.091 | 1.214 | -0.591 | -0.880 | -0.456 | -1.021 |
| 1 | 26 | 0.000 | -1.853 | -0.053 | -0.230 | -1.990 | -1.102 | 2.009 | -2.702 | -2.022 | -1.458 | 10.183 | 5.900 | 0.446 | -2.336 |
| 2 | 18 | 0.000 | -0.468 | 3.799 | 0.519 | 3.424 | -1.970 | -1.784 | -1.269 | 3.335 | -2.024 | 2.721 | -2.322 | -1.409 | -2.094 |
| 2 | 26 | 0.000 | -0.431 | 3.850 | 0.543 | 3.477 | -1.425 | -1.764 | -1.235 | 2.697 | -2.000 | 2.186 | -2.303 | 0.061 | -2.425 |
| 3 | 18 | 0.042 | -0.062 | 2.012 | 2.550 | 1.332 | -1.239 | -0.327 | -1.193 | -0.387 | -1.379 | 1.564 | 0.829 | -0.886 | -1.421 |
| 3 | 26 | 0.001 | 0.428 | -0.053 | 1.100 | -0.199 | -1.102 | -1.319 | -0.466 | 3.951 | -1.458 | 2.989 | -1.447 | 0.446 | -1.929 |
| 4 | 18 | 0.026 | -0.532 | 3.519 | 0.755 | -1.109 | -1.220 | -0.283 | 0.424 | 1.345 | -1.341 | 1.622 | -0.397 | -0.873 | -1.383 |
| 4 | 26 | 0.001 | -0.206 | 2.549 | 0.301 | -1.490 | -0.691 | -0.175 | -0.319 | 1.483 | -1.747 | 4.180 | -0.864 | 0.989 | -1.793 |
| 5 | 18 | 0.001 | 2.239 | 2.060 | 3.638 | 0.219 | -1.393 | -1.262 | 2.081 | -1.803 | -1.685 | -0.488 | -1.359 | 0.031 | -1.731 |
| 5 | 26 | 0.001 | 2.063 | 2.300 | 2.324 | -0.165 | -1.645 | -2.067 | 2.744 | -0.319 | -0.712 | -0.931 | -1.107 | -0.856 | -2.440 |
| 6 | 18 | 0.008 | -2.145 | 0.044 | 2.562 | 1.287 | -1.048 | -1.243 | 3.483 | 0.101 | -0.943 | -0.882 | -0.900 | 0.515 | -0.596 |
| 6 | 26 | 0.118 | -1.933 | 0.855 | 2.056 | 1.939 | -0.617 | -0.642 | 1.237 | -0.284 | -0.114 | 0.350 | -1.330 | 1.094 | -1.193 |
| 7 | 18 | 0.000 | 0.891 | 6.649 | -1.135 | 4.290 | -1.673 | -3.194 | 1.885 | -0.304 | -2.198 | 1.672 | -1.634 | -2.007 | -3.807 |
| 7 | 26 | 0.000 | 1.379 | 7.162 | -0.865 | 5.001 | -1.397 | -3.095 | 1.398 | -1.032 | -2.895 | 1.776 | -1.220 | -2.237 | -4.176 |
| 8 | 18 | 0.000 | -3.121 | -2.200 | 2.324 | -2.679 | -1.645 | 0.603 | -2.935 | 7.348 | -2.367 | 7.463 | 3.682 | -0.186 | -3.093 |
| 8 | 26 | 0.000 | -2.899 | -2.046 | 2.651 | -2.506 | -1.510 | 0.529 | -2.693 | 5.751 | -2.143 | 7.516 | 4.186 | -0.743 | -2.902 |
| 9 | 18 | 0.000 | -1.884 | -0.085 | -0.252 | -1.568 | -1.120 | 1.959 | -2.728 | -2.050 | -1.488 | 10.081 | 5.819 | 0.424 | -2.360 |
| 9 | 26 | 0.013 | -1.032 | 1.433 | 2.002 | -0.882 | -0.643 | 2.830 | -1.584 | -0.824 | 2.372 | -0.488 | -0.230 | 0.031 | -1.731 |

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

## Supplementary figures

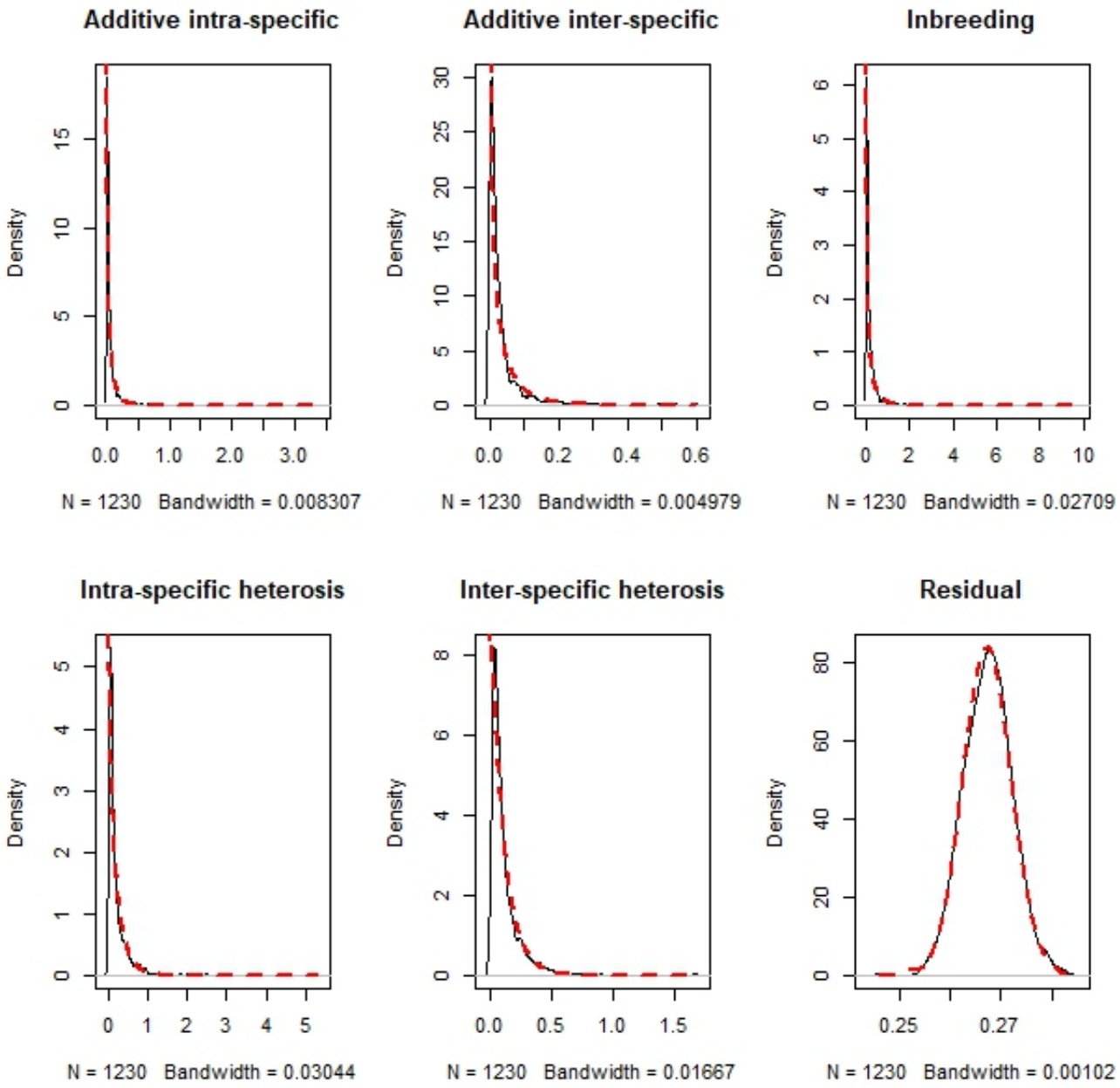



**Figure SF1** Density of the variance components estimated by the *hglm* algorithm for the 1230 proteins. Red dashed lines represent the fitted distributions used to simulate and test parameter inference of the proposed model.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

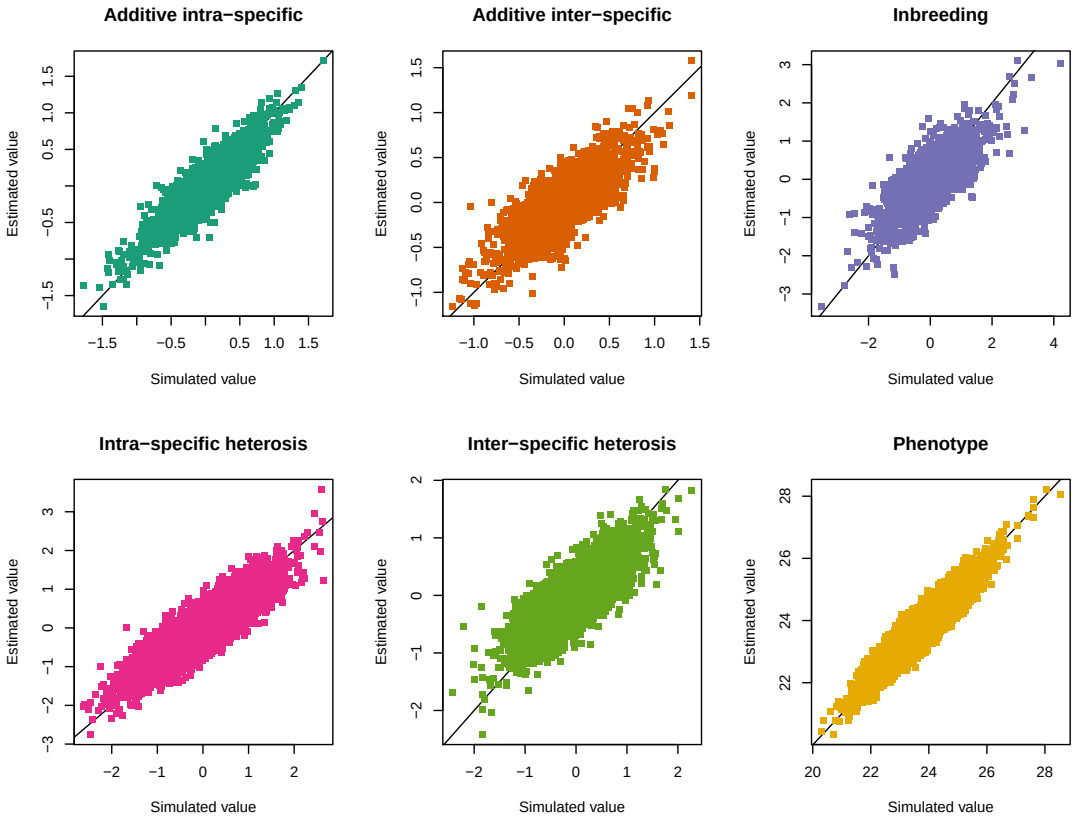



**Figure SF2** Fitted Best Linear Unbiased Predictors of the random effects parameters and predicted phenotypic value plotted against the simulated genetic parameters and the simulated phenotypic value. Fixed the number of parental strains and the number of individuals of each species, we performed the simulation 1000 times. Here, we show the case of eleven parents, with 7 belonging to one specie and 4 to the other.

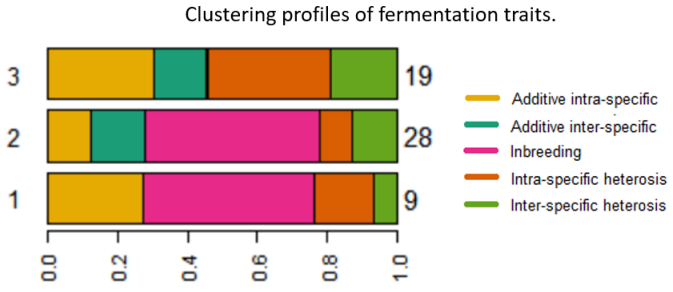



**Figure SF3** Clustering profiles of fermentation and life-history traits. Clusters number are reported on the left, on the right the number of traits found in each cluster.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

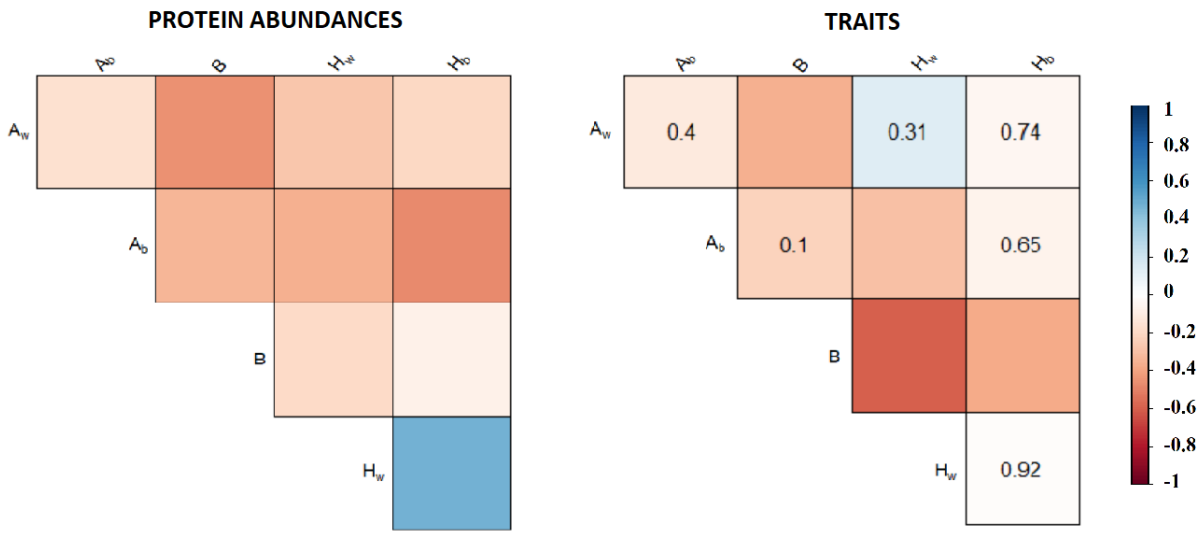



**Figure SF4** Global correlations between genetic variance components: on the left correlations at the proteomic level, on the right at the more integrated level. Crosses indicate point correlations that are not statistically significant (p-value>0.05).

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

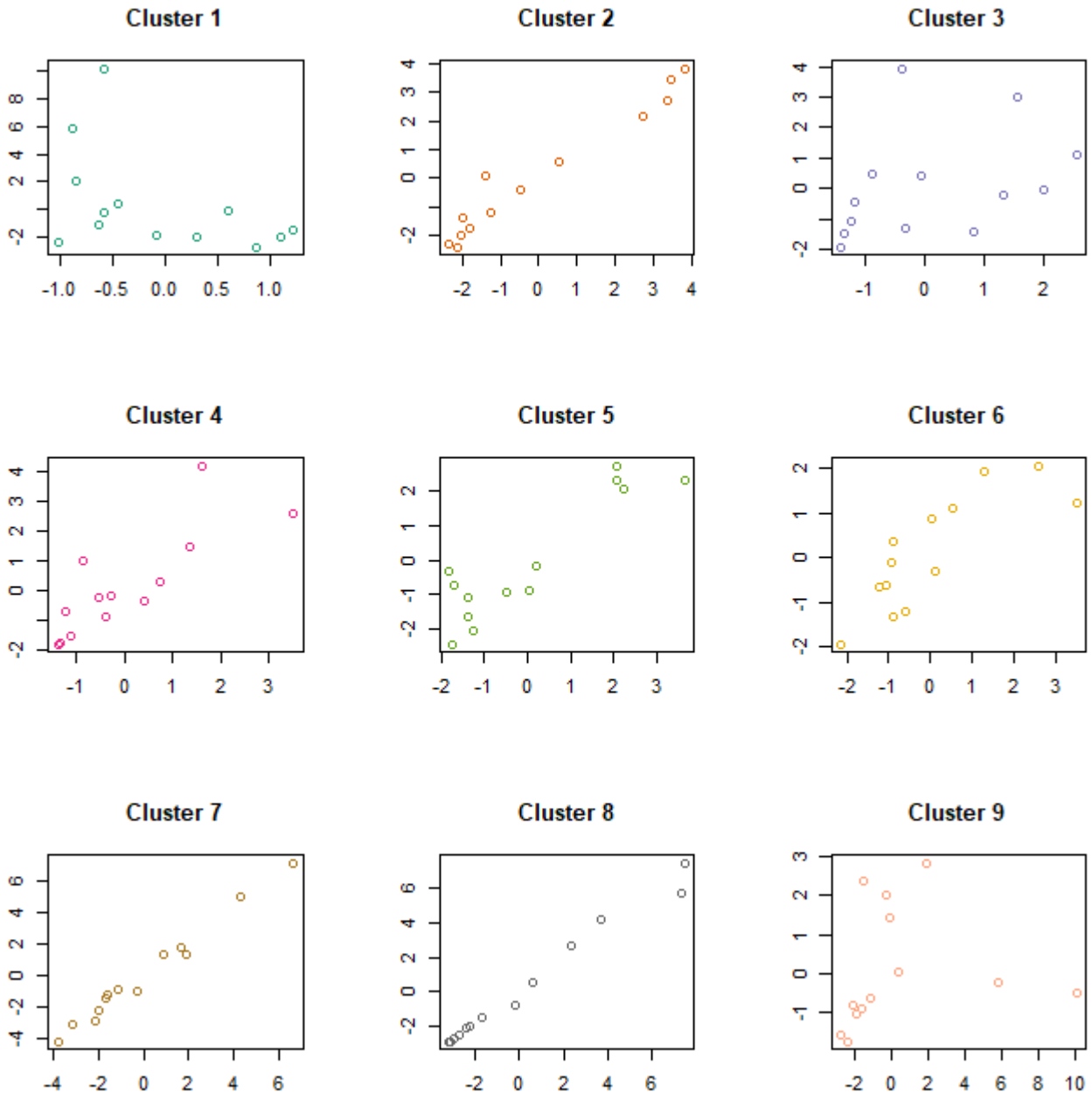



**Figure SF5** Pearson's chi-square test of enrichment: For each cluster are represented the chi-square standardized residuals at 18° (abscissa) and at 26° (ordinate).

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

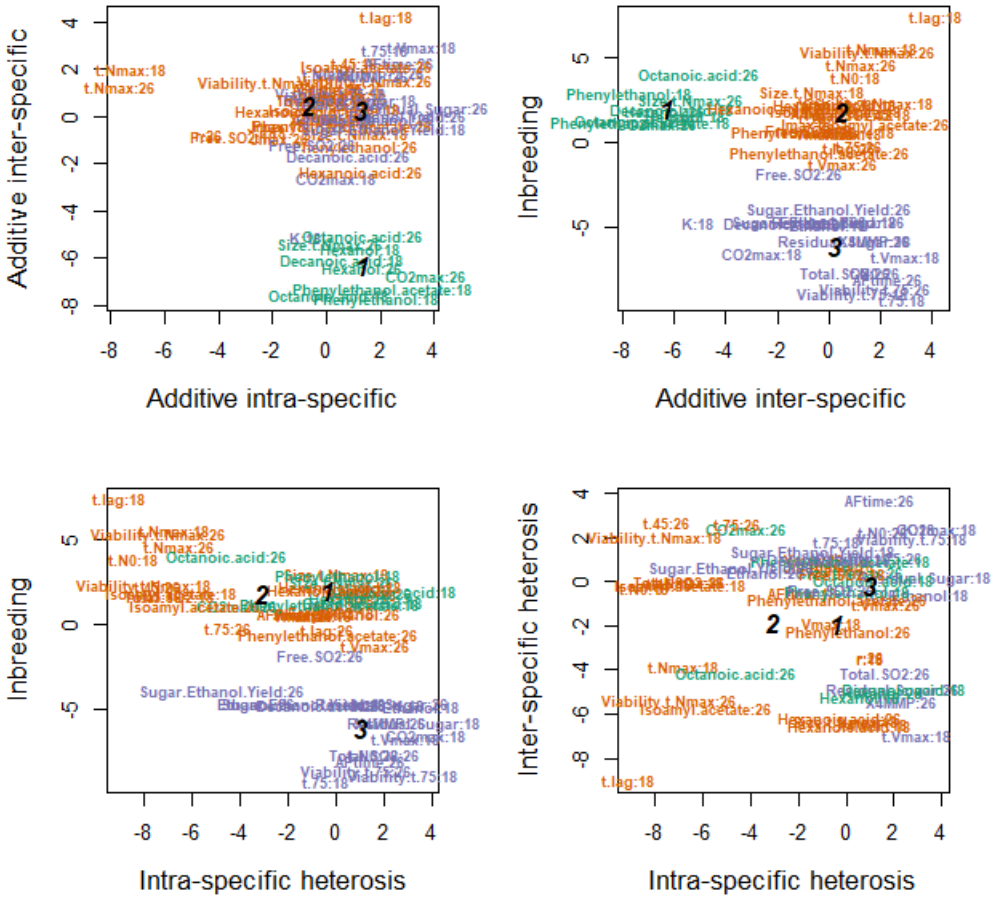



**Figure SF6** Life-history and fermentation traits profiles. Traits are identified by their label, color combinations identify the clusters obtained by their classification based on a Gaussian Mixture model.

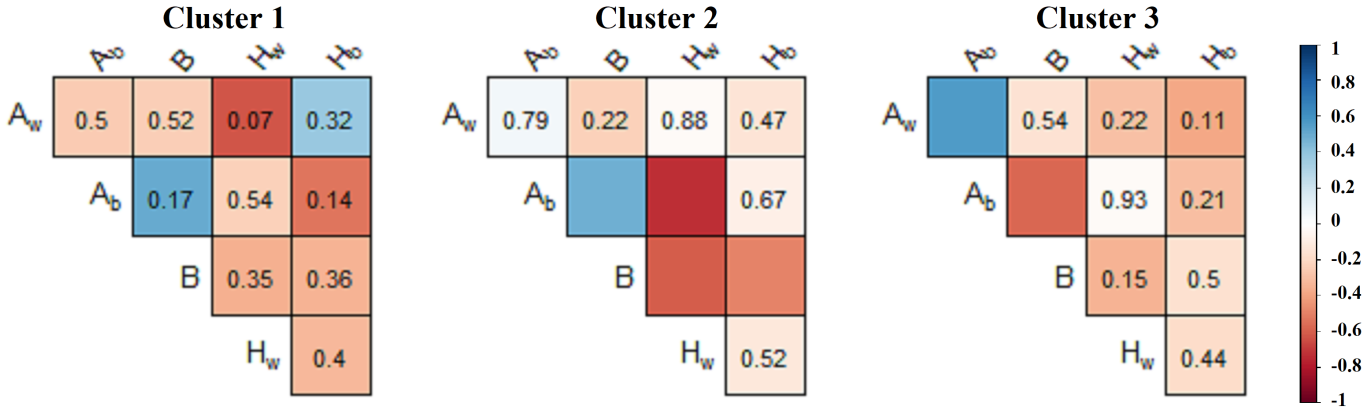



**Figure SF7** Pearson's correlation test performed to investigate the intra-cluster correlations at the trait level: for each cluster, the figure shows the correlation between variances of the genetic effects, and its p-value when the statistical test is not significant.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

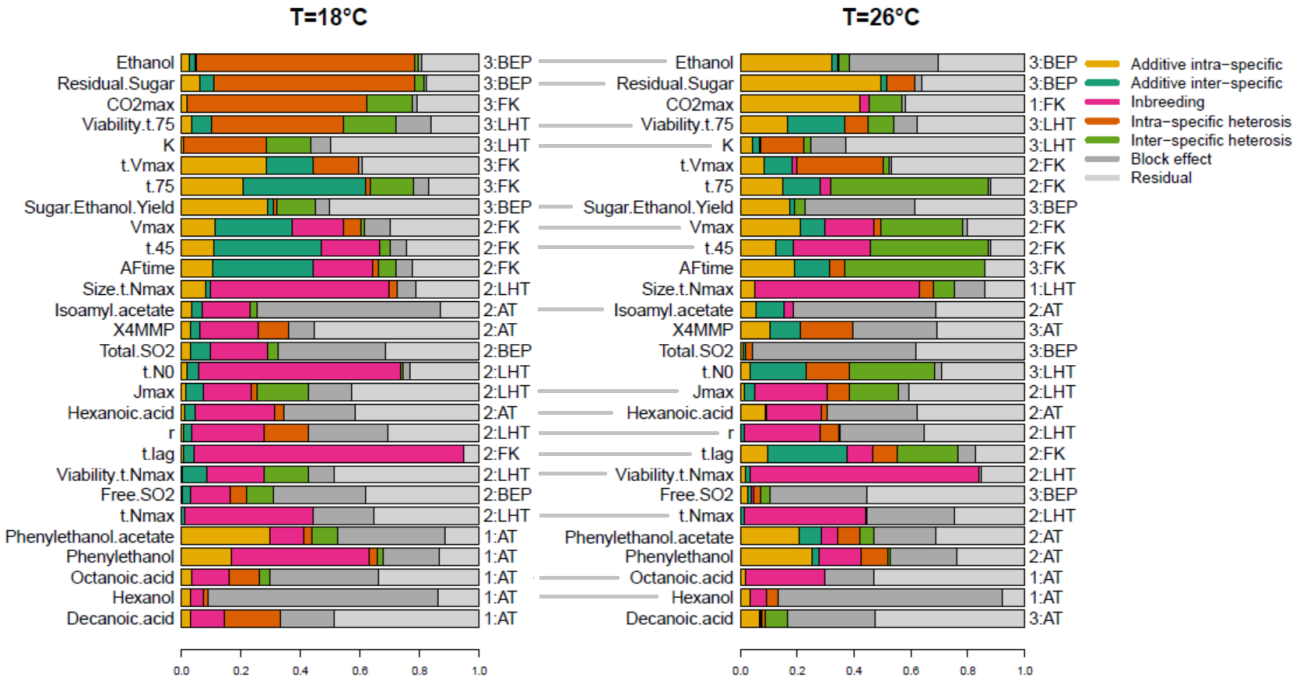



**Figure SF8 Variance components of fermentation traits**. **Left**: Traits measured at 18°C. **Right**: Traits measured at 26°C. Each variance component is attributed a different color. Traits are ranked according to their cluster number at 18°C. Trait category and cluster number is indicated on the right-hand-side of the plot.

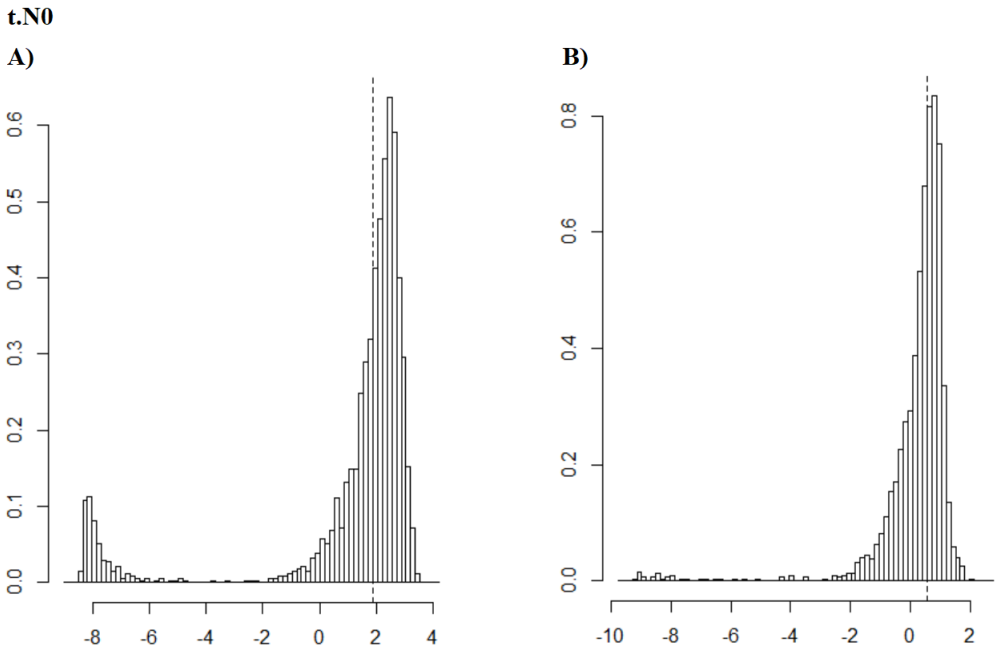



**Figure SF9** Bootstrap summary example: Distribution of intra-specific variance estimates for the growth lag-phase, *t.N0*, at **A)** 18° and **B)** 26°C.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

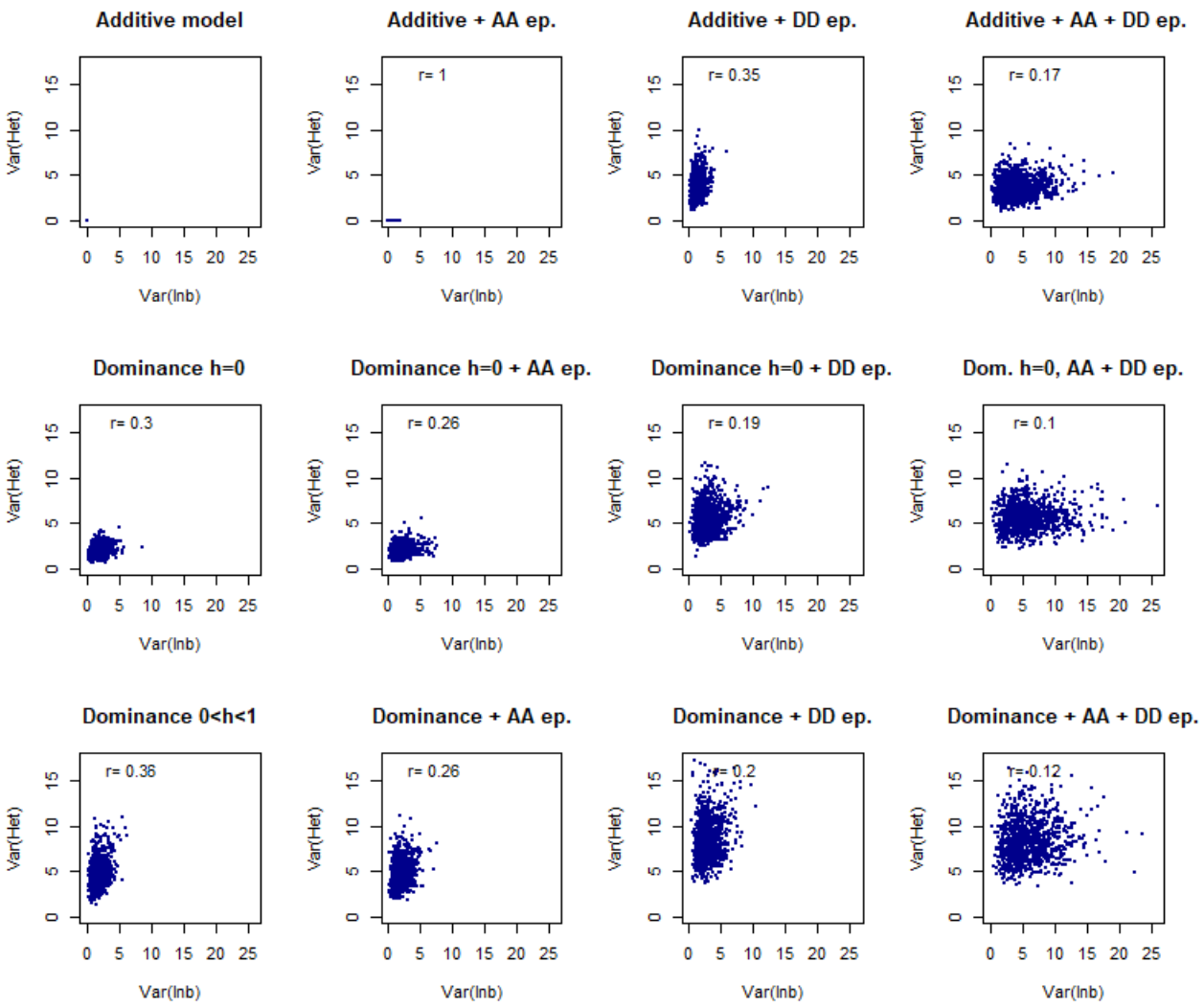



**Figure SF10** Simulations genetic models. Correlation between inbreeding and heterosis variance computed on the basis of a combination of multilocus genetic models: Additive model with/without additive by additive, dominance by dominance, or both epistatic effects, dominance of the strongest allele with/without additive by additive, dominance by dominance, or both epistatic effects; symmetrical dominance with/without additive by additive, dominance by dominance, or both epistatic effects. The simulated half-diallel consisted of 11 parental lines. Phenotypic values were supposed to depend on 10 loci, and the number of alleles per loci was imposed to 11. Alleles values were drawn from a gamma distribution ($k$=10, $\theta$=20) and epistatic effects from a normal distribution ($\mathcal{N}(0,3)$).

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

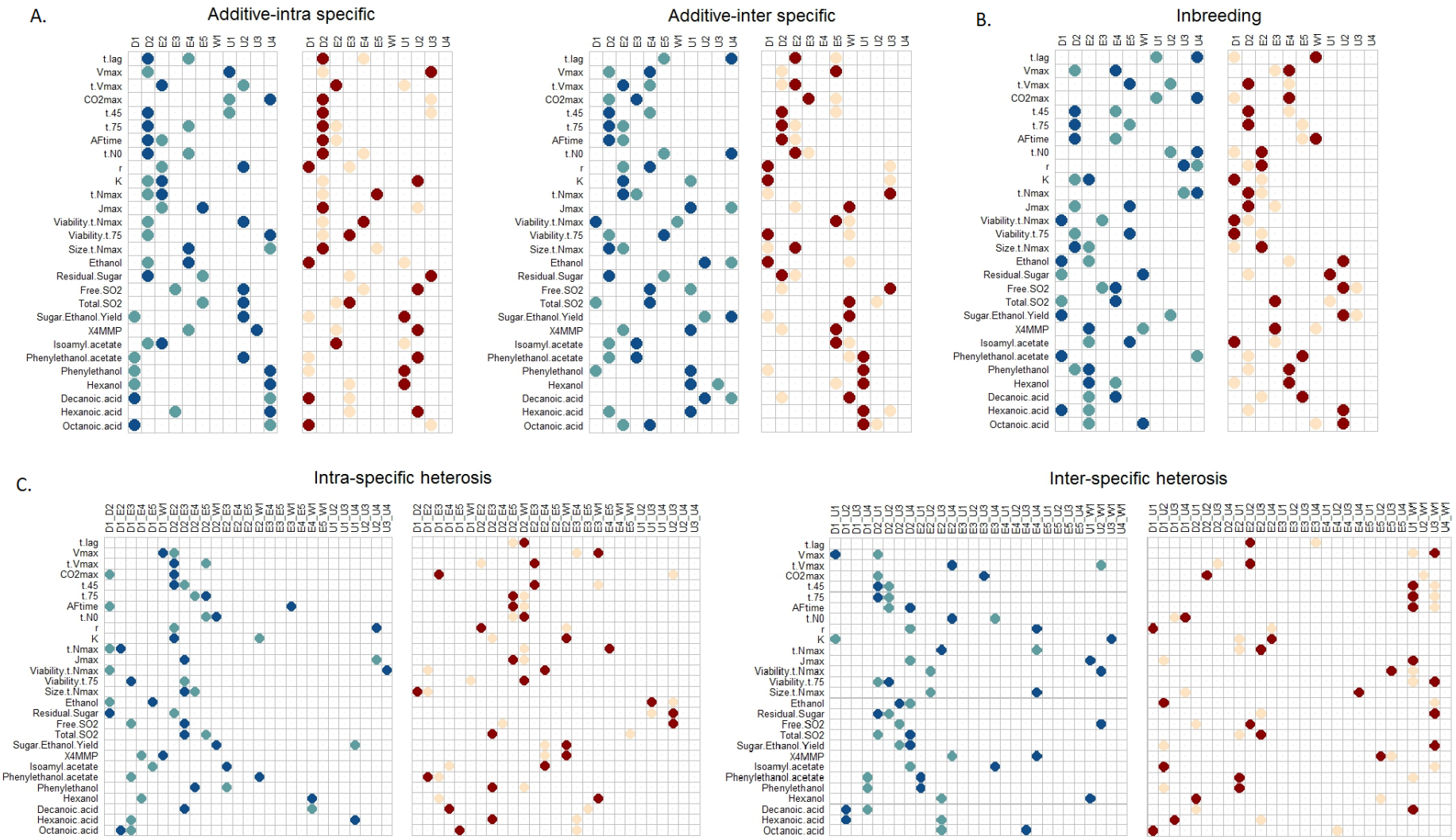
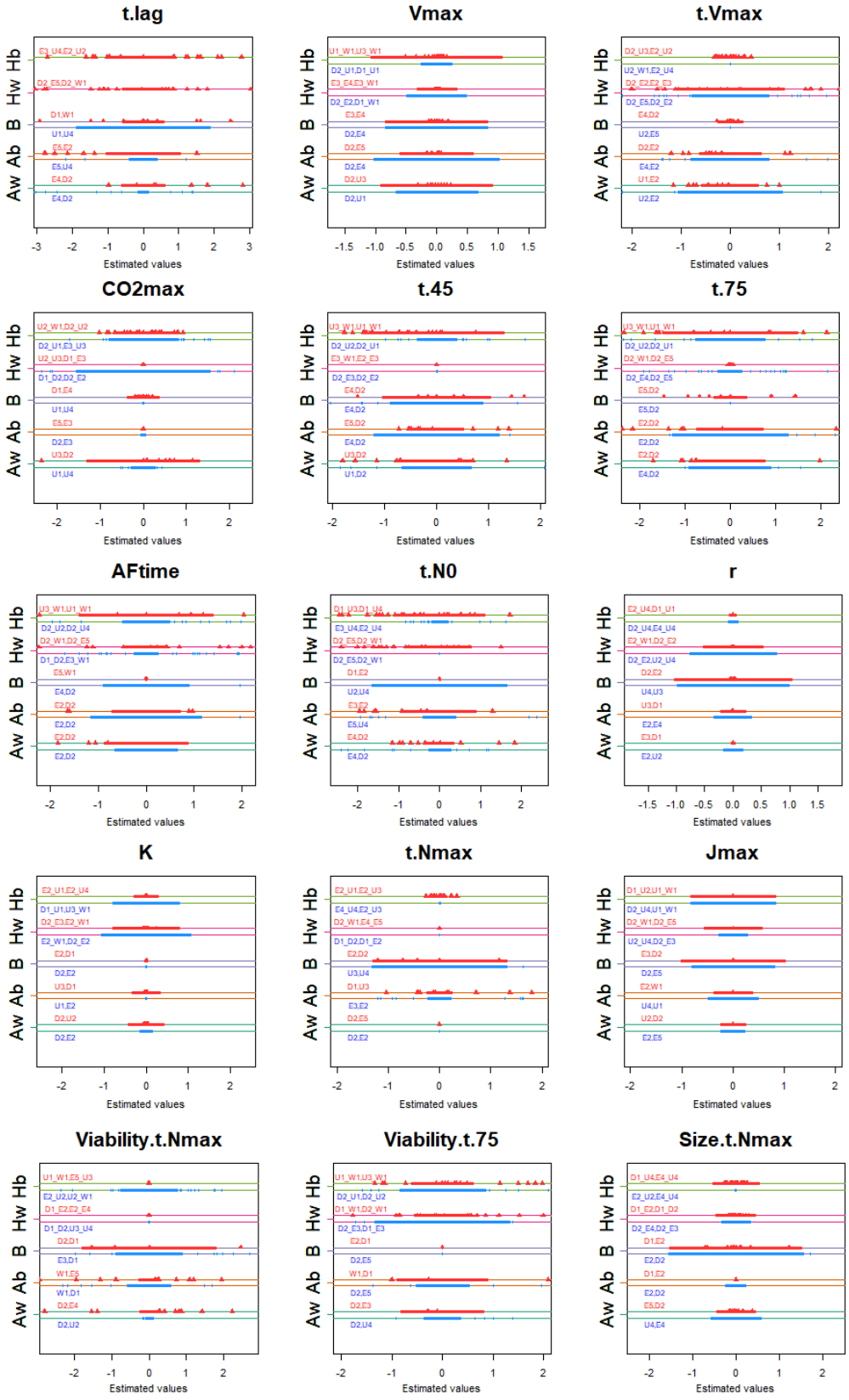



**Figure SF11** Dots show strains with highest and lowest genetic contribution per trait and temperature, in blue at 18°C and in red at 26°C. Dark and light colors report the strains with the highest and lowest additive (A), inbreeding (B) and heterosis (C) contributions, respectively. Strains are sorted by species and traits by category.

bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/362418; this version posted July 4, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

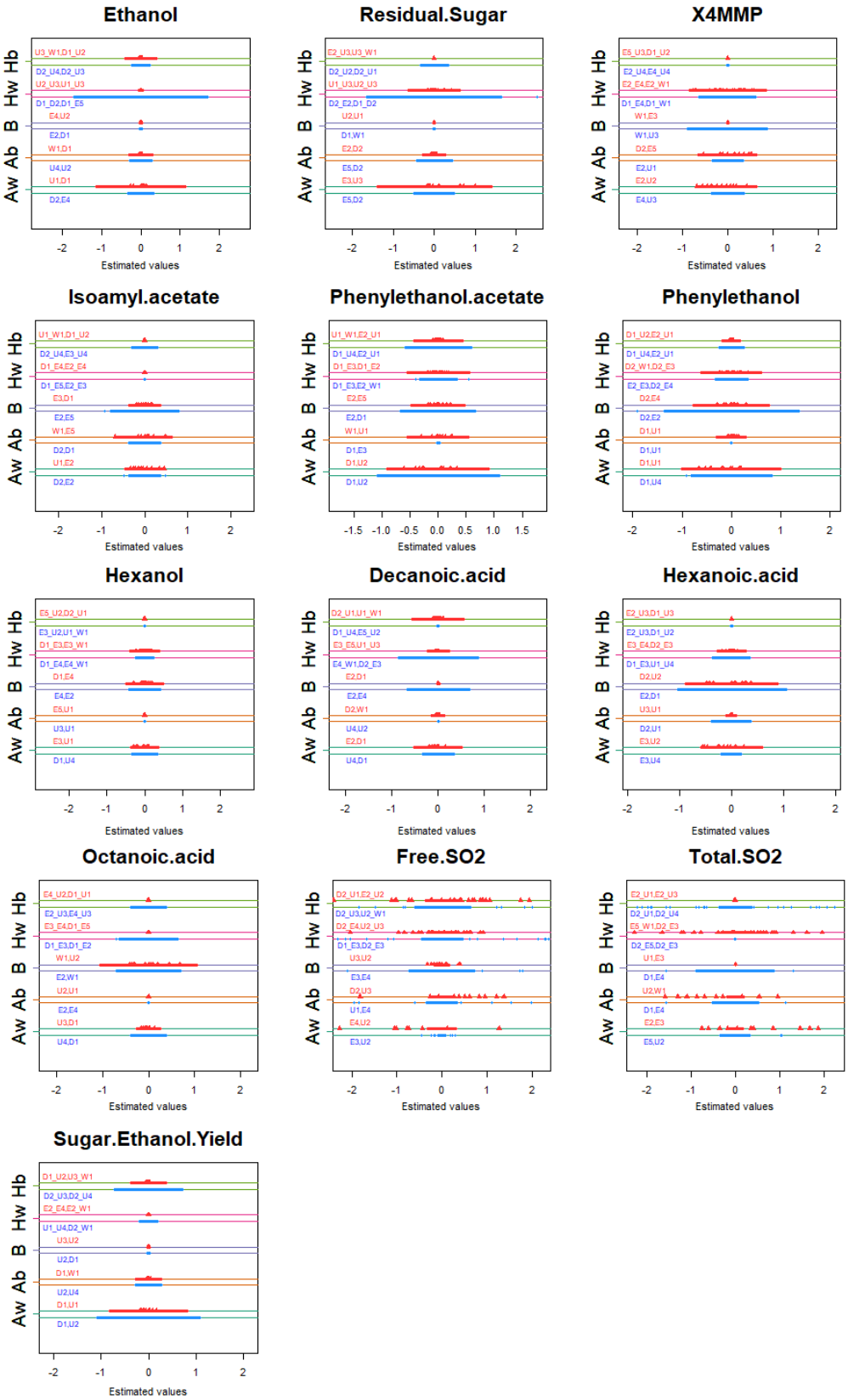



**Figure SF12** Interval plots. For each trait we plot random effect estimators, namely $\hat{A}_w$, $\hat{A}_b$, $\hat{B}$, $\hat{H}_w$, $\hat{H}_b$, at 18°C (blue) and 26°C (red). Horizontal bars show for each parameter the region of highest density that covers nearly 95% ($\sim \pm 2\hat{\sigma}_{\boldsymbol{q}}$) of the parameter density. On the left hand-side of each plot are listed, for each genetic effect, the strains which give the lowest and the greatest contribution.