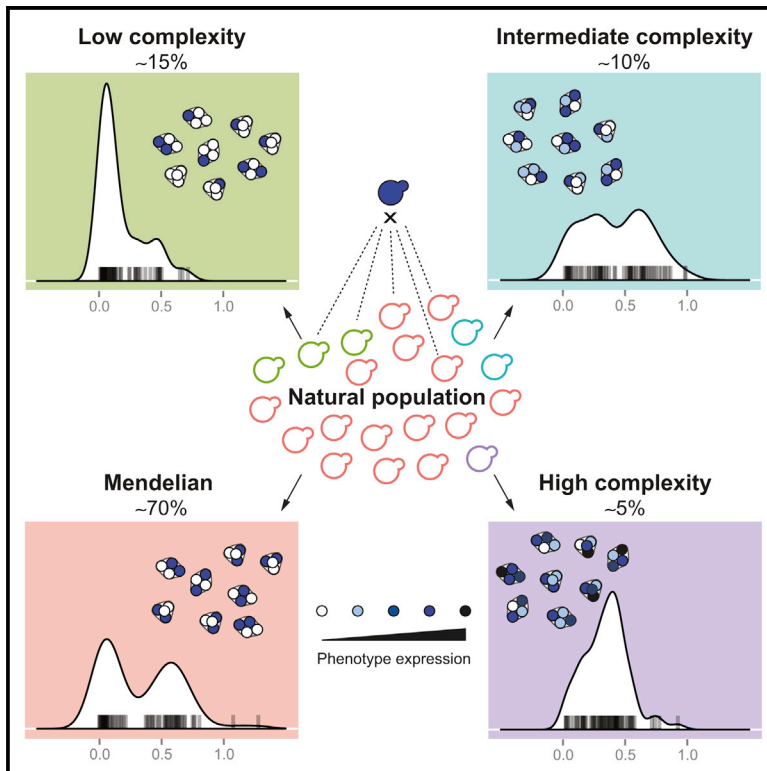


## The Hidden Complexity of Mendelian Traits across Natural Yeast Populations

### Graphical Abstract



### Authors

Jing Hou, Anastasie Sigwalt, Téo Fournier, ..., Jacky de Montigny, Maitreya J. Dunham, Joseph Schacherer

### Correspondence

[schacherer@unistra.fr](mailto:schacherer@unistra.fr)

### In Brief

Mendelian traits are considered to be the least complex of all heritable phenotypes. By systematically identifying causal Mendelian variants and following their effects and inheritances across the *S. cerevisiae* species, Hou et al. uncovered the continuous complexity spectrum of monogenic mutations, where genotype is hardly predictive of a phenotype.

### Highlights

- In a natural population, at least ~9% of traits follow a Mendelian inheritance
- Non-random distribution of causal Mendelian loci leads to co-segregating phenotypes
- Monogenic variants may have a continuous expressivity spectrum in natural population
- Background could lead to the transition from Mendelian to complex traits



# The Hidden Complexity of Mendelian Traits across Natural Yeast Populations

Jing Hou,<sup>1,3</sup> Anastasie Sigwalt,<sup>1,3</sup> Téo Fournier,<sup>1</sup> David Pflieger,<sup>1</sup> Jackson Peter,<sup>1</sup> Jacky de Montigny,<sup>1</sup> Maitreya J. Dunham,<sup>2</sup> and Joseph Schacherer<sup>1,\*</sup>

<sup>1</sup>Department of Genetics, Genomics and Microbiology, University of Strasbourg/CNRS UMR 7156, 67000 Strasbourg, France

<sup>2</sup>Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA

<sup>3</sup>Co-first author

\*Correspondence: [schacherer@unistra.fr](mailto:schacherer@unistra.fr)

<http://dx.doi.org/10.1016/j.celrep.2016.06.048>

## SUMMARY

Mendelian traits are considered to be at the lower end of the complexity spectrum of heritable phenotypes. However, more than a century after the re-discovery of Mendel's law, the global landscape of monogenic variants, as well as their effects and inheritance patterns within natural populations, is still not well understood. Using the yeast *Saccharomyces cerevisiae*, we performed a species-wide survey of Mendelian traits across a large population of isolates. We generated offspring from 41 unique parental pairs and analyzed 1,105 cross/trait combinations. We found that 8.9% of the cases were Mendelian. Further tracing of causal variants revealed background-specific expressivity and modified inheritances, gradually transitioning from Mendelian to complex traits in 30% of the cases. In fact, when taking into account the natural population diversity, the hidden complexity of traits could be substantial, confounding phenotypic predictability even for simple Mendelian traits.

## INTRODUCTION

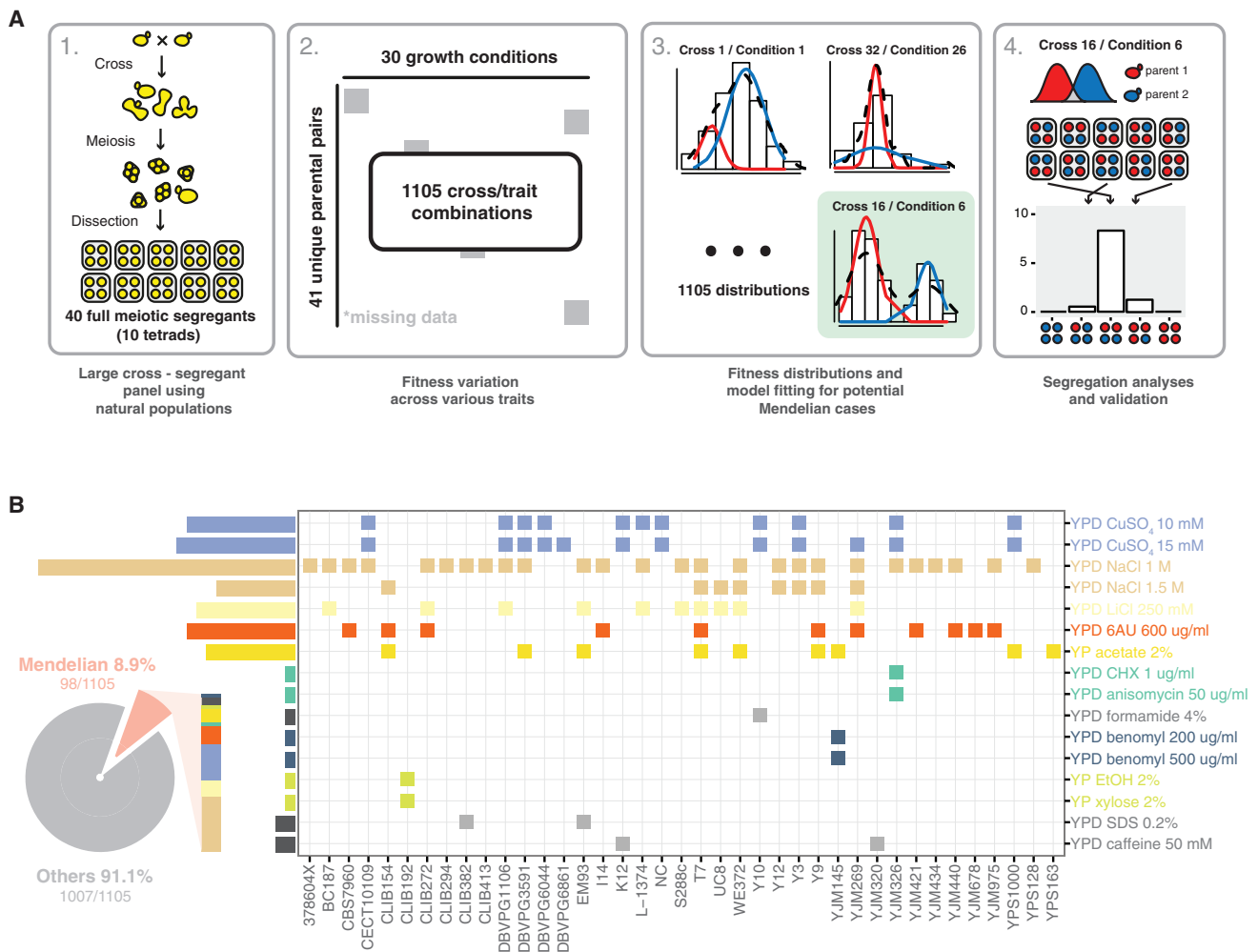
Elucidating the genetic causes of the astonishing phenotypic diversity observed in natural populations is a major challenge in biology. Within a population, individuals display phenotypic variations in terms of morphology, growth, physiology, behavior, and disease susceptibility. The inheritance patterns of phenotypic traits can be classified as either monogenic or complex. While many traits are complex, resulting from variation within multiple genes, their interaction, and environmental factors (Mackay et al., 2009), some traits are primarily monogenic and conform to a simple Mendelian inheritance (Antonarakis and Beckmann, 2006). Nevertheless, while useful, this overly simplistic dichotomic view could potentially mask the continuous level of the underlying genetic complexity (Antonarakis et al., 2010; Badano and Katsanis, 2002; Dipple and McCabe, 2000). More than a century after the rediscovery of Mendel's law, we still lack a global overview of the spectrum of

genetic complexity of phenotypic variation within any natural population.

Complex traits can be predominantly controlled by variation in a single gene (Dipple and McCabe, 2000). Similarly, monogenic traits can be influenced by multiple genes in specific genetic backgrounds (Badano and Katsanis, 2002; Cooper et al., 2013; Dorfman, 2012; Nadeau, 2001; Thein, 2011). In fact, it is increasingly evident that monogenic mutations do not always strictly adhere to Mendelian inheritance (Cooper et al., 2013; Dorfman, 2012; Nadeau, 2001). For example, many human monogenic disorders, including sickle cell anemia and cystic fibrosis, could display significant clinical heterogeneity such as incomplete penetrance and variable levels of severity due to allelic interactions and background-specific modifiers (Cooper et al., 2013). Recent genome-scale surveys of loss-of-function mutations have revealed considerable background effects in various model systems (Dowell et al., 2010; Hamilton and Yu, 2012; Paaby et al., 2015; Vu et al., 2015) and human cell lines (Blomen et al., 2015; Hart et al., 2015; Wang et al., 2015), where the mutant phenotypes could be highly variable even between closely related individuals.

However, although background effects on "monogenic" loss-of-function mutations are readily seen, such specific mutation type does not reflect the overall genetic diversity and complexity observed in natural populations (Auton et al., 2015; Cao et al., 2011; Strobe et al., 2015). Specifically, the global landscape of natural genetic variants leading to Mendelian traits has never been thoroughly explored in any species, and their phenotypic effects and inheritance patterns within natural populations have been largely unknown.

Here, we carried out a first species-wide identification of causal variants of Mendelian traits in the yeast *S. cerevisiae* to characterize in depth their phenotypic effects and transmission patterns across various genetic backgrounds. We generated a large number of crosses using natural isolates and analyzed the fitness distribution and segregation patterns in the offspring for more than 1,100 cross/trait combinations. We found that 8.9% of the cases were Mendelian, among which most were caused by common variants and showed stable inheritances across the *S. cerevisiae* species. Interestingly, global phenotypic distribution patterns of multiple Mendelian traits across an extremely large population (~1,000 isolates) were not necessarily correlated with patterns observed



**Figure 1. Comprehensive Landscape of Mendelian Traits in *S. cerevisiae***

(A) Workflow of the detection of Mendelian traits. The workflow was defined as four steps, consisting with offspring generation, fitness measurements, model fitting, and segregation analysis, as indicated.

(B) Distribution of all identified Mendelian traits spanning different crosses (x axis) on conditions tested (y axis). Each square represents any single Mendelian case, and colors indicate different conditions. Pie chart represent the fraction of Mendelian cases relative to the entire dataset. Method summary and genomic mapping results for selected Mendelian cases are included in the [Supplemental Information](#).

See also [Figures S1](#) and [S2](#).

in the offspring from individual crosses. We further characterized a causal variant related to drug resistance and traced its effects across multiple genetic backgrounds. Significant deviations from the Mendelian expectation were observed with variable genetic complexities, illustrating the hidden complexity of a monogenic mutation across a yeast natural population.

## RESULTS

### Global Landscape of Mendelian Traits in *S. cerevisiae*

To obtain an overview of natural genetic variants leading to Mendelian traits in the *S. cerevisiae* species, we selected 41 diverse natural isolates spanning a wide range of ecological sources (tree exudates, *Drosophila*, fruits, various fermentation and clin-

ical isolates) and geographical sources (Europe, America, Africa, and Asia) and performed systematic crosses with one strain  $\Sigma$ 1278b ([Supplemental Experimental Procedures](#)). For each cross, we generated 40 offspring representing 10 individual meiosis (full tetrads), summing up to a panel of 1,640 full meiotic segregants from diverse parental origins ([Figure 1A](#), panel 1). All segregants, as well as the respective parental isolates, were tested for 30 stress-responsive traits related to various physiological and cellular processes, including different carbon sources, membrane and protein stability, signal transduction, sterol biosynthesis, transcription, and translation, as well as osmotic and oxidative stress ([Supplemental Experimental Procedures](#)). In total, we tested 1,105 cross/trait combinations and analyzed the offspring fitness distribution patterns for each combination ([Figure 1A](#), panel 2).

For a Mendelian trait, contrasting phenotypes between the parental isolates were controlled by a single locus; therefore, half of the offspring would inherit the causal allele and display a 2:2 segregation in any given tetrad. Consequently, the global offspring fitness distribution would follow a bimodal pattern with equal partitioning of segregants in either parental phenotype cluster. To detect such cases, we applied a bimodal distribution model with random latent variables for the observed fitness distributions for each cross/trait combination, using an expectation maximization (EM) algorithm (Figure 1A, panel 3; Figure S1; Supplemental Experimental Procedures). The principle of this method relies on unsupervised iterations of latent variables to fit any given set of data with maximum a posteriori (MAP) probability to a predefined model; in our case, the presence of two normal distributions for the observed fitness values where the mean of each assigned cluster corresponds to either parental fitness value.

For each fitness distribution observed in a given cross/trait combination, the posterior probability that an individual belongs to either fitness cluster was computed (Experimental Procedures), and the general features of the fitted bimodal model, such as the means and SDs for both clusters as well as their relative proportions, were extracted (Experimental Procedures). To determine the cutoff values that allow for high confidence calling of bimodal cases and subsequent cluster assignments, we generated a simulated dataset of 1,000 fitness distributions with the same general features compared to the real data and reapplied the model-fitting procedure. Using the simulated data as a training set, we determined that a cutoff of posterior probability > 0.8 for cluster assignment while allowing less than 10% overlapping between the clusters (i.e., the percentage of individuals cannot be assigned at a given posterior probability cutoff) was the best parameter to maintain a high detection performance (area under the receiver operating characteristic [ROC] area under the curve [AUC] = 0.824; Figure S1A) while minimizing case loss (Figure S1B).

By applying these parameters, 318 cross/trait combinations were detected as bimodal, with the parental isolates belonging to distinct clusters. All detected bimodal cases were robust against experimental noise, where the mean trait differences between the assigned clusters exceeded at least 2.5 times the SD estimated using technical replicates of the common parental strain  $\Sigma 1278b$  ( $N \geq 72$ ) (Figure S1C). Considering that each segregant tested for a given cross/trait combination was genotypically distinct, bimodal cases detected using our method, while robust against noise, were rather conservative and, therefore, likely represented a lower-bound estimation.

For all bimodal cases, we further analyzed the phenotypic segregation patterns in the tetrads and identified 98 as Mendelian, displaying the characteristic 2:2 segregation (Figure 1A, panel 4). In total, identified Mendelian cases represented 8.9% (98/1105) across our sample and were interspersed among various conditions, including a large number of instances related to NaCl (28 crosses), CuSO<sub>4</sub> (13 crosses), 6-azauracil (11 crosses), and acetate (9 crosses) (Figure 1B). Other low-frequency cases were found in conditions related to signal transduction (caffeine) and carbon sources (ethanol and xylose); various other conditions (formamide, benomyl, and SDS); and

conditions related to the antifungal drugs cycloheximide (CHX) and anisomycin (Figure 1B). In addition, we observed co-segregation of unrelated traits (NaCl, acetate, and 6-azauracil; Figure 1B) where the fitness variation patterns in the segregants were highly correlated (Pearson's correlation  $\rho > 0.9$ ). We further characterized cases with co-segregations, high-frequency cases related to CuSO<sub>4</sub>, and the low-frequency case related to resistance to the drugs CHX and anisomycin in detail. For the selected cases, 80 additional full tetrads were tested, and the 2:2 phenotypic segregation patterns were confirmed.

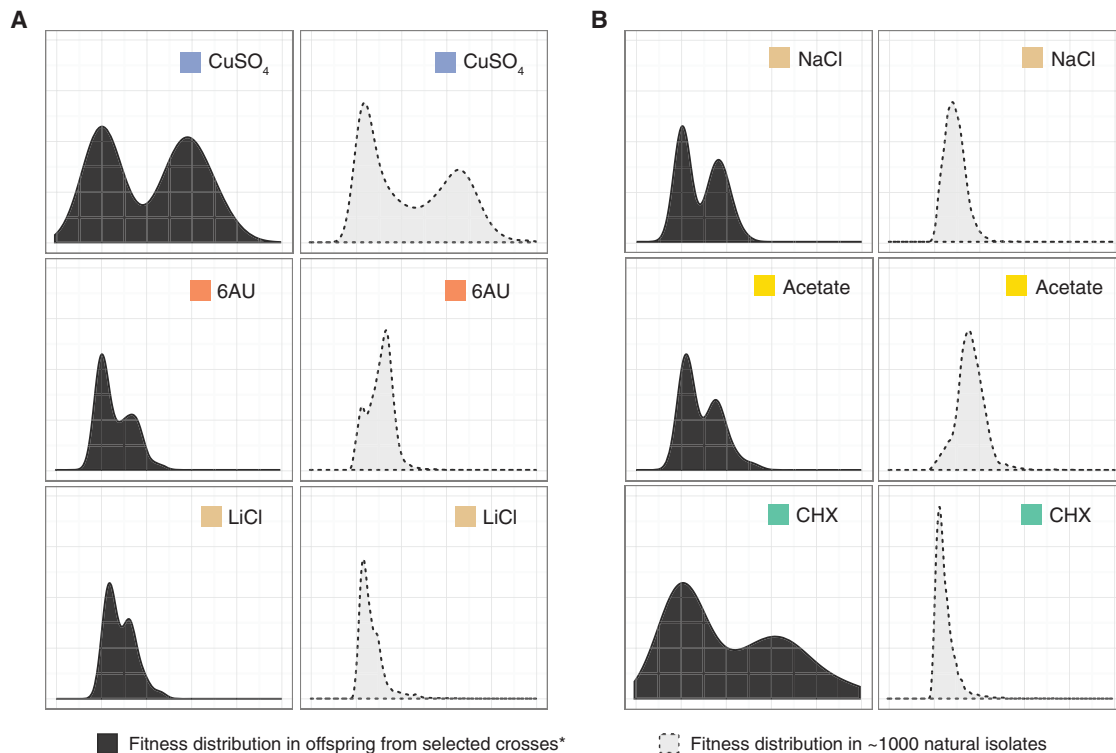
### Molecular Characterization of Identified Mendelian Traits

Using bulk segregant analysis followed by whole-genome sequencing, we identified one locus for each case as expected. For all crosses displaying co-segregation with NaCl, the same ~60-kb region (480,000–540,000) on chromosome IV was mapped, spanning the *ENA* genes encoding for sodium and/or lithium efflux pumps (Figure S2A). While variations of the *ENA* genes were known to lead to osmotic stress tolerance (Ruiz and Ariño, 2007), the phenotypic associations with other co-segregating traits (acetate and 6-azauracil) were previously unknown. Causal genes related to acetate and 6-azauracil were suspected to be in close genetic proximity with the *ENA* locus; however, the precise identities of these genes remained unclear. For cases related to CuSO<sub>4</sub>, we mapped a 40-kb region on chromosome VIII (190,000–230,000; Figure S2C). We identified the *CUP1* gene in this region, which encodes for a copper-binding metallothionein (Figure S2C). In this case, the common parental strain  $\Sigma 1278b$  was resistant to both concentrations of CuSO<sub>4</sub> tested, and the allelic version of *CUP1* in  $\Sigma 1278b$  led to stable Mendelian inheritance across multiple genetic backgrounds (Figure 1B).

Finally, the last characterized case involved the two anti-fungal drugs CHX and anisomycin, which was found in the cross between a clinical isolate YJM326 and  $\Sigma 1278b$  (Figure 1B). Pooled segregants belonging to the higher fitness cluster showed allele frequency enrichment for the YJM326 parent across an ~100-kb region on chromosome VII (420,000–520,000; Figure S2B). Further analyses yielded *PDR1* as the potential candidate, which encodes for a transcription factor involved in multidrug resistance. Using reciprocal hemizygosity analysis (Figure S3A), as well as a plasmid-based complementation test (Figure S3B), we showed that the *PDR1*<sup>YJM326</sup> allele was necessary and sufficient for the observed resistance.

### Fitness Distribution of Identified Mendelian Traits across Large Natural Populations

Although Mendelian traits could exhibit distinctive offspring distribution and segregation patterns in individual crosses, the general phenotypic distribution of such traits within a population was unclear. We measured the fitness distribution of an extremely large collection of ~1,000 natural isolates of *S. cerevisiae* (the 1002 Yeast Genomes Project; <http://1002genomes.u-strasbg.fr/>) on selected conditions related to identification of Mendelian traits, including resistance to NaCl, LiCl, acetate, 6-azauracil, CuSO<sub>4</sub>, and CHX (Figure 2). Interestingly, while some traits followed the same bimodal distribution



**Figure 2. Fitness Distribution Patterns of Identified Mendelian Traits within a Large Natural Population**

Comparisons of the fitness distribution on six selected conditions in individual crosses (left panel,  $n = 40$ ) and across ~1,000 natural isolates of *S. cerevisiae* (right panel,  $n = 960$ ) are shown. Conditions tested are color coded.

(A) Bimodal distribution patterns observed both in crosses and at the population level. 6AU, 6-azauracil.

(B) Bimodal distributions observed only in crosses but not within a population.

model across the population as was observed in offspring from single crosses (Figure 2A), other traits with a clear Mendelian inheritance pattern in crosses appeared to vary continuously at the population level (Figure 2B). This observation suggested that the phenotypic distribution within the population might not necessarily reflect the underlying genetic complexity of traits. Instead, the inheritance pattern for any given trait might largely be determined by specific combinations of parental genetic backgrounds.

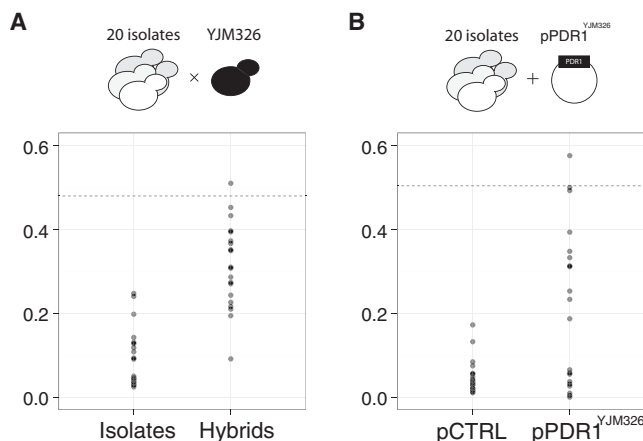
### Hidden Complexity of a Rare Mendelian Variant across Different Genetic Backgrounds

While focusing on highly frequent cases such as  $\text{CuSO}_4$  and NaCl provided indications about the transmission stability of common Mendelian variants and revealed previously unknown co-segregations, we were particularly interested in rare cases where the phenotypic effects and the general inheritance patterns across different genetic backgrounds were unknown. The identified Mendelian case related to the anti-fungal drugs CHX and anisomycin could be considered as such. Across our panel, the parent YJM326 was the only highly fit isolate, and few isolates showed similar resistance levels within the whole species (Figure 2B). To test the effect of the  $PDR1^{\text{YJM326}}$  allele in different backgrounds, we crossed the resistant isolate YJM326 with 20 diverse sensitive isolates. Counterintuitively, the resulting

hybrids displayed continuous variation of the resistance in the presence of CHX (Figure 3A). To test whether the resistance variation in the hybrids was due to allelic interactions at the  $PDR1$  locus in different backgrounds, we introduced a plasmid carrying the  $PDR1^{\text{YJM326}}$  allele ( $pPDR1^{\text{YJM326}}$ ) into the same set of isolates and quantified their fitness in the presence of CHX (Figure 3B). Across all isolates tested, about half (11/20) expressed the resistant phenotype to various degrees (Figure 3B; Figure S3B). However, fitness between haploid isolates carrying  $pPDR1^{\text{YJM326}}$  and the corresponding hybrids was only weakly correlated (Pearson's correlation  $\rho = 0.434$ ), indicating that allelic interactions at the  $PDR1$  locus only partly accounted for the observed variation (Figure S3B).

The lack of correlation between hybrids and isolates carrying the plasmid with the  $PDR1^{\text{YJM326}}$  allele led us hypothesize the presence of potential modifiers in various hybrid backgrounds. To test this hypothesis, we evaluated the fitness distributions of the drug resistance in the offspring across the 20 hybrids generated previously. For each hybrid, 20 complete tetrads were tested in the presence of CHX, and the fitness distributions as well as the segregation patterns were assessed in the offspring (Figure S4A). In the absence of modifiers, haploid segregants are expected to have complete phenotypic penetrance, as the effects of intralocus interaction were eliminated. In this scenario, all crosses between any sensitive parental isolate





**Figure 3. Effects of the  $PDR1^{YJM326}$  Allele in Different Genetic Backgrounds**

(A) Fitness variation of 20 isolates (left panel) in comparison with the same set of strains hybridized with YJM326 in the presence of drug. Fitness values (y axis) correspond to the ratio between the growth in the presence of cycloheximide (YPD CHX 1  $\mu$ g/ml) and control media YPD. Dashed line indicates the fitness of the resistant strain YJM326.

(B) Fitness variation of 20 isolates carrying empty control plasmid (pCTRL, left panel) or plasmid containing the  $PDR1^{YJM326}$  allele under its native promoter (pPDR1<sup>YJM326</sup>, right panel). Fitness values were measured in the presence of cycloheximide (YPD CHX 1  $\mu$ g/ml) with hygromycin to maintain plasmid stability. Dashed line indicates the fitness value of YJM326 carrying the plasmid pPDR1<sup>YJM326</sup>. The Supplemental Information includes a detailed comparison for the effect of hybrid and plasmid in individual genetic backgrounds.

See also Figure S3.

and YJM326 should display a bimodal distribution in the offspring, with a 2:2 segregation of the phenotype.

Interestingly, while most of the tested crosses (14/20) displayed Mendelian segregation, as was observed in the cross between YJM326 and  $\Sigma$ 1278b, several crosses showed clear deviation of the expected phenotypic distribution (Figure 4; Figure S4). In addition to Mendelian cases (Figure 4A), three other types of distribution were observed (Figures 4B–4D). In total, such cases represent ~30% of all crosses tested. Of these crosses, 15% (3/20, between YJM320, Y3, Y9, and YJM326) showed incomplete penetrance, indicating possible suppressors of the  $PDR1^{YJM326}$  allele (Figure 4B). We observed a 1:4:1 ratio between tetrads containing 2, 1, and 0 resistant segregants, respectively, possibly indicating that two independent loci, including  $PDR1$ , were involved (Figure 4B; Figure S4). Furthermore, 10% of the crosses (2/20, between S288c, YJM440, and YJM326) showed enriched high-fitness offspring, with an intermediate peak between the sensitive and resistant clusters. This observation suggests the presence of epistatic interactions from these specific genetic backgrounds, resulting as a transitional resistant phenotype cluster with higher genetic complexity (Figure 4C). The levels of genetic complexity in these cases are suspected to be low, but the precise number of genes involved remains unclear.

In addition to cases with low levels of deviation from Mendelian expectations, we also found one cross (between YJM653 and YJM326) with largely biased offspring fitness distribution,

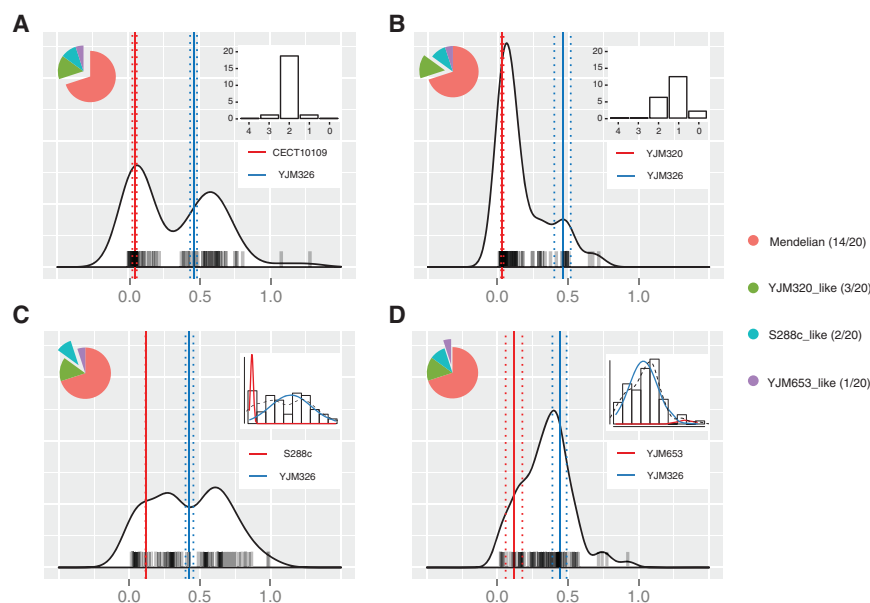
for which a bimodal distribution with distinctive parental clusters was rejected by our model. In this case, the resistant phenotype was no longer caused by a single Mendelian factor, and the underlying genetic determinants were undoubtedly complex (Figure 4D). For selected crosses in each type of biased distributions, additional offspring (160 from 40 complete tetrads) were generated, and the fitness distribution patterns were further confirmed (Figure S4B). Contrasting with other identified Mendelian traits with a stable inheritance patterns across the population, the  $PDR1$  case represented a perfect example illustrating the hidden complexity of a simple Mendelian trait within the natural population of the yeast *S. cerevisiae*.

## DISCUSSION

By performing a species-wide survey of monogenic variants in *S. cerevisiae*, we obtained a first lower-bound estimation of the proportion of Mendelian traits within a natural population. We showed that genes and alleles underlying the onset of Mendelian traits are variable in terms of their type, frequency, and genomic distribution at the population level. Remarkably, by tracing the effect of one causal Mendelian variant  $PDR1^{YJM326}$  across the population, we demonstrated that the genetic complexity of traits could be dynamic, transitioning from clear Mendelian to diverse complex inheritance patterns, depending on various genetic backgrounds.

Yeasts—and, more particularly, *S. cerevisiae*—have been extensively used as a model for dissecting many complex traits that were of medical, industrial, and evolutionary interests (Bloom et al., 2013; Ehrenreich et al., 2012; Mukherjee et al., 2014; Steinmetz et al., 2002; Treusch et al., 2015). A trend emerging from studying complex traits in this species was that causal variants do not distribute randomly across the genome, and several hotspots have been identified (Fay, 2013). As a result, a low number of loci were found to be involved in high numbers of unrelated phenotypes, despite the fact that underlying causal genes could be different. Interestingly, causal variants in Mendelian traits seemed to follow the same trend, as supported by our data. In fact, we observed phenotypic co-segregation of unrelated conditions such as resistance to acetate, 6-azauracil, and osmotic stress and showed that only a single region on chromosome IV was involved (Figure S2). In addition, the observed co-segregations showed relatively high population frequencies, with more than 15% of the crosses co-segregating on at least two different conditions (Figure 1B). This effect of linkage could possibly lead to biased phenotype assortments across the population, although the underlying evolutionary origin is unknown.

In general, Mendelian traits were considered as rare, especially in human disorders; however, no direct estimation of the proportion of Mendelian traits relative to complex traits was available at the population level, and the types of genes that were more susceptible to cause Mendelian inheritance were unknown. Our data showed that, across a yeast natural population, causal alleles involved in direct response to stress, such as transporters (*ENA*) or metal-binding genes (*CUP1*), were more likely to follow Mendelian inheritance. In fact, a large number of Mendelian traits identified in our sample were related to these



**Figure 4. Post-Mendelian Inheritance Patterns of Drug Resistance in Different Hybrid Contexts**

Offspring fitness distribution patterns observed in hybrids originated from 20 sensitive isolates and YJM326 in the presence of cycloheximide (YPD CHX, 1  $\mu$ g/ml). 80 offspring were tested for each case, and examples of Mendelian (A) and non-Mendelian (C and D) inheritance patterns are shown.

(A and B) Phenotypic segregation is indicated at the upper right side for cases with biased bimodal distributions.

(C and D) For cases with more complex patterns, maximum likelihood fittings of a bimodal model were shown instead. For non-bimodal cases the model-fitting results were shown instead. Parental origins for each cross are shown, and the fitness values of the sensitive (red) or resistant (blue) parental strains are presented as vertical bars, with dotted bars corresponding to  $\pm$  standard variation ( $n = 4$ ). The [Supplemental Information](#) includes offspring fitness distributions for each of the 20 crosses and additional offspring distributions for selected crosses with biased Mendelian patterns.

See also [Figure S4](#).

two loci, and the inheritance patterns were extremely stable, displaying 2:2 segregations with little influence of the genetic backgrounds. A similar pattern was found in a Mendelian-trait-related ammonium resistance in natural isolates of *S. cerevisiae*, where a transporter gene *TRK1* was involved (Reisser et al., 2013). The stable inheritance patterns of traits caused by alleles with a direct phenotypic effect could potentially be due to the lack of regulatory complexity. As was supported by laboratory evolution experiments, amplifications of this type of genes were frequent, conferring to rapid acquisition of resistances in stress conditions such as salt (Anderson et al., 2010), copper (Fogel and Welch, 1982; Gerstein et al., 2015), sulfate (Gresham et al., 2008), and glucose limitations (Dunham et al., 2002).

By contrast, depending on the gene involved, a given Mendelian trait could lead to complex inheritance patterns across different genetic backgrounds, as evidenced by the causal allele *PDR1* related to resistance to CHX and anisomycin. By crossing the strain YJM326 carrying the resistant allele *PDR1*<sup>YJM326</sup> with diverse natural isolates, we showed that, although most crosses retained stable 2:2 segregations, the inheritance pattern of the resistance phenotype, in some cases, displayed various deviations from Mendelian expectation, including reduced penetrance (3/20), increased genetic complexity (2/20), and in one extreme case, transition from monogenic to complex trait. We propose that the observed post-Mendelian inheritance patterns are due to the functional nature of the *PDR1* gene. In fact, as *PDR1* encodes for a transcriptional factor with complex regulatory networks and impacts multiple downstream effector genes (Moye-Rowley, 2003), the resulting phenotypic expression would possibly be influenced by variations of a large number of genes that are involved in the same network in different genetic backgrounds.

Overall, our data provided a first comprehensive view of natural genetic variants that lead to the onset of Mendelian traits in a yeast population. We showed that monogenic mutations could exhibit post-Mendelian modifications such as pleiotropy, incomplete dominance, and variations in expressivity and penetrance due to differences in specific genetic backgrounds. Depending on the parental combination, the inheritance might display a Mendelian, intermediate, or complex pattern, showing the continuum of the complexity spectrum related to a monogenic mutation, as illustrated by the example of the drug resistance involving *PDR1*<sup>YJM</sup>. However, while Mendelian traits could be related to common or rare variants, we found that the overall fitness distribution patterns of such traits at the population level—for some instances, if not all—were not informative regarding their genetic complexity. Collectively, phenotypic prediction, even for simple Mendelian variants, may not be an easy task, in part due to the lack of prediction power using population data and the scarcity of large-scale family transmission information, such as the case for diseases in human. Future studies using pairwise crosses covering a larger panel of conditions in yeasts, or in other model organisms, may provide general trends and a more complete picture regarding the phenotypic predictability of monogenic traits.

## EXPERIMENTAL PROCEDURES

### Strains

Isolates from diverse ecological and geographical sources used in this study are detailed in the [Supplemental Experimental Procedures](#). All strains are stable haploids with deletion of the *HO* gene (Liti et al., 2009; Schacherer et al., 2009). Laboratory strains FY4, FY5 (isogenic to S288c), and  $\Sigma$ 1278b were used. Deletion mutants in the  $\Sigma$ 1278b background were obtained from the gene deletion collection kindly provided by Dr. Charles Boone (Dowell et al., 2010). YJM326  $\Delta$ *pdr1* strain was generated by insertion of the hygromycin-resistance cassette *HygMX* using homologous recombination.

### Media and Culture Conditions

Detailed media compositions for phenotyping of the segregant panel are listed in the [Supplemental Experimental Procedures](#). Growth and maintenance of the strains are carried on standard rich YPD media (1% yeast extract, 2% peptone and 2% glucose). A final concentration of 200  $\mu\text{g/ml}$  hygromycin (Euromedex) was supplemented to maintain the plasmids carrying the resistance marker gene *HygMX*. Sporulation was induced on potassium acetate plates (1% potassium acetate, 2% agar). All procedures were performed at 30°C unless otherwise indicated.

### Crosses and Generation of the Offspring

For the construction of the segregant panel, 41 diverse isolates (*MAT $\alpha$* ) were crossed with the lab strain  $\Sigma 1278b$  (*MATa*) on YPD ([Supplemental Experimental Procedures](#)). Resulting diploids were sporulated for 2–3 days on sporulation medium (10 g/l potassium acetate, 20 g/l agar) at 30°C. Tetrad dissections were performed using the MSM 400 dissection microscope (Singer Instruments) on YPD agar after digestion of the tetrad asci with zymolyase (MP Biomedicals MT ImmunO 20T). A total of ten tetrads containing four viable spores were retained per cross. The same protocol was used for the validation of Mendelian cases with 2:2 segregation, as well as the crosses between 20 isolates with the drug-resistant strain YJM326 ([Supplemental Experimental Procedures](#)); in these cases, 80 and 20 full tetrads were tested for each cross, respectively.

### High-Throughput Phenotyping and Growth Quantification

Quantitative phenotyping was performed using endpoint colony growth on solid media. Strains were pregrown in liquid YPD medium and pinned onto a solid YPD matrix plate to a 384 density format using a replicating robot RoTor (Singer Instruments). At least two replicates of each parental strain were present on the corresponding matrix, and 16 replicates were present for the common parent  $\Sigma 1278b$ . The matrix plates were incubated overnight to allow sufficient growth, which were then replicated in 30 media conditions, including YPD as a pinning control (see detailed compositions of the media in the [Supplemental Experimental Procedures](#)). The plates were incubated for 48 hr at 30°C and were scanned at the 24-, 40-, and 48-hr time points with a resolution of 600 dpi at 16-bit grayscale. Quantification of the colony size was performed using the Colony Area plugin in ImageJ, and the fitness of each strain on the corresponding condition was measured by calculating the normalized growth ratio between stress media and YPD using the software package ScreenMill ([Dittmar et al., 2010](#)). Phenotyping data related to crosses with YJM326 were analyzed using the R package Gitter ([Wagih and Parts, 2014](#)).

### Model-Fitting Procedure and Detection of Traits with Mendelian Inheritance

For each cross/condition combination, a bimodal distribution was fitted using the R package mixtools (<https://cran.r-project.org/web/packages/mixtools/index.html>) with  $k = 2$  and  $\text{maxit} = 500$ . Mean ( $\mu$ ), SD ( $\sigma$ ), proportion of each cluster ( $\lambda$ ), and posterior probability of each cluster for each individual were extracted from the output file. To determine cutoff values of posterior probability for cluster assignment, a simulated dataset was generated by simulating two normal distributions with  $n \times \lambda$  and  $n \times (1 - \lambda)$  individuals for each cluster, respectively, with mean and SD randomly sampled from observations in real data. For each simulated set, the two normal distributions generated were combined, and the procedure was repeated 1,000 times to generate a training set with 1,000 distributions ([Supplemental Experimental Procedures](#)). The training set was then subjected to model fitting with the same parameters ([Supplemental Experimental Procedures](#)). The mean ( $\mu$ ), SD ( $\sigma$ ), proportion of each cluster ( $\lambda$ ), and posterior probability of each cluster for each individual were extracted again, and the training dataset was evaluated against the real data ([Supplemental Experimental Procedures](#)). In this case, as the prior probability of cluster assignment was known for each simulated individual, it was possible to test for the detection sensitivity and specificity (ROC) using varied cutoff parameters. A sequence starting from 0.5 to 0.95 (increment, 0.05) for posterior probability and a sequence from 0 to 0.9 (increment, 0.1) for percentage of non-overlapping individuals were tested. The ROC curves and AUC were calculated for each combination of cutoff parameters using the R package ROCR ([Sing et al., 2005](#)). Cutoffs of 0.8 for posterior probability and 0.9 for

percentage of non-overlapping were retained to ensure confident detections ([Figures S1A and S1B](#)). In this case, a maximum of 10% of individuals are allowed to have a posterior probability of less than 0.8, adjusting for the internal phenotypic stochasticity and noise. The defined parameters were applied to real data, and cases that passed the filter proceeded to cluster assignment. For bimodal cases with parental pairs that belong to either phenotype cluster, the segregation patterns were determined. As each cross/condition combination consists of 40 individual, a cutoff of 0.9 for percentage of non-overlapping would allow a maximum of four individuals that could not be assigned with confidence. In theory, four tetrads may be impacted. To adjust for this effect, a trait is considered as Mendelian when at least 7/10 of the tetrads display a 2:2 segregation with posterior probability  $> 0.8$ . All analyses were performed in R.

### Evaluation of Detection Robustness against Experimental Noise

The levels of experimental noise were estimated using technical replicates of the  $\Sigma 1278b$  parental strain. For each condition, at least 72 replicates were tested, and the noise level for a given condition was calculated as the SD. To evaluate the robustness of the detected bimodal distributions, we compared the mean fitness differences between the assigned clusters to the observed noise related to the corresponding condition ([Figure S1C](#)). For all 318 detected bimodal cases, the mean fitness differences between the assigned clusters were higher than at least 2.5 times the SD. Among which, the identified Mendelian cases displayed mean cluster fitness differences higher than three times standard deviation ( $Z$  scores =  $\pm 3$ ; [Figure S1C](#)). These observations suggest that the Mendelian cases identified here were of high confidence and were unlikely to be due to distribution stochasticity or experimental noise.

### Bulk Segregant Analysis

In total, six crosses representative of the identified Mendelian traits were subjected to bulk segregant analyses in order to identify the genomic regions involved. Crosses involved were between  $\Sigma 1278b$  and EM93, I14, YJM269, and CLIB272 in the presence of salt and a second stressor, and between  $\Sigma 1278b$  and YJM326 for antifungal drugs and copper sulfate. For each case, 50 independent viable spores tetrads exhibiting 2:2 segregation on the corresponding conditions were separately grown overnight at 30°C in liquid YPD and were pooled by equal optical density readings at 600 nm. Pooled segregants were subjected to whole-genome sequencing, and the genomic regions involved in each trait were determined by looking at allele frequency variation.

### Genotyping Strategy and Data Treatment

Genomic DNA from the pool was extracted using Genomic-tip 100/G columns and genomic DNA buffers (QIAGEN), as described previously ([Friedrich et al., 2015](#)). Sequencing of the samples was performed using the Illumina HiSeq 2000 sequencing system, except for the CHX pool, for which we used MiSeq technology. Reads were mapped to the  $\Sigma 1278b$  genome with the Burrows-Wheeler Aligner (BWA, version 0.7.4) allowing five mismatches and one gap ([Li et al., 2009](#)). The “-I” flag has been added for the MiSeq Pool because reads were encoded in Illumina 1.9 format. Single-nucleotide polymorphism (SNP) calling was performed using GATK v3.3-0 ([McKenna et al., 2010](#)), with default parameters. The allele frequency of  $\Sigma 1278b$  was calculated for each polymorphic position by adding the allele balance ratio, with the “VariantAnnotator” command of GATK.

### Reciprocal Hemizygosity Test

To perform a reciprocal hemizygosity test on drug resistance, the wild-type strains  $\Sigma 1278b$  and YJM326 were crossed with each other and with deletion mutant strains of *PDR1* in both genetic backgrounds. Individual zygotes were isolated using the MSM 400 dissection microscope (Singer Instruments) on a YPD plate, and the ploidy of the hybrids was checked on sporulation media after 2–3 days at 30°C. Phenotypic effects of the sensitive and resistance alleles were evaluated using a drop test on selective media YPD CHX 1  $\mu\text{g/ml}$  and YPD as a growth control. Plates were scanned after 48 hr of incubation at 30°C.

### Plasmid Construction and Phenotyping

Centromeric plasmid was constructed to test the allelic effect of the drug resistance allele *PDR1*<sup>YJM326</sup> in different genetic backgrounds using Gateway



cloning technology (Invitrogen). A fragment containing *PDR1* and its native promoter and terminator regions flanked by attB1/attB2 recombination sites was amplified from the genomic DNA of YJM326 and  $\Sigma 1278b$  and was cloned into an empty centromeric plasmid with the HphMX resistance marker (pCTRL), according to instructions (Treusch et al., 2015). The resulting plasmids, pPDR1<sup>YJM326</sup> and pPDR1 <sup>$\Sigma 1278b$</sup> , were verified using restriction enzymes and PCR amplification with internal primers of *PDR1* gene. 20 diverse natural isolates were transformed with pPDR1<sup>YJM326</sup> as well as the empty control plasmid pCTRL using the EZ-Yeast Transformation Kit (MP Biomedicals). Transformants were selected on YPD media containing 200  $\mu$ g/ml hygromycin. Growth quantification of isolates carrying the pPDR1<sup>YJM326</sup> or pCTRL plasmids in the presence of drug was performed as described previously, and 200  $\mu$ g/ml hygromycin was supplemented to all media to maintain the selection pressure during the phenotyping procedure.

## ACCESSION NUMBERS

The accession number for all Illumina sequencing reads generated in this study is European Nucleotide Archive: PRJEB11500.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2016.06.048>.

## AUTHOR CONTRIBUTIONS

J.H. and J.S. designed the experiments. J.H., A.S., T.F., and J.S. performed the experiments. J.H., D.P., P.J., and J.S. analyzed the data. J.d.M. and M.J.D. provided feedback. J.H. and J.S. wrote the paper.

## ACKNOWLEDGMENTS

We thank Matt Rockman, David Botstein, and Bernard Dujon for comments on the manuscript. The authors thank the NIH (grant R01 GM101091-01) and the Agence Nationale de la Recherche (ANR grant 2011-JSV6-004-01) for financial support. M.J.D. is a Rita Allen Foundation Scholar and a senior fellow in the Genetic Networks program at the Canadian Institute for Advanced Research. A.S. is supported by a grant from Région Alsace. J.H. is supported in part by a grant from the Ministère de l'Enseignement Supérieur et de la Recherche and in part by a fellowship from the medical association La Ligue contre le Cancer.

Received: October 24, 2015

Revised: March 15, 2016

Accepted: June 10, 2016

Published: July 7, 2016

## REFERENCES

Anderson, J.B., Funt, J., Thompson, D.A., Prabhu, S., Socha, A., Sirjusingh, C., Dettman, J.R., Parreiras, L., Guttman, D.S., Regev, A., and Kohn, L.M. (2010). Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. *Curr. Biol.* 20, 1383–1388.

Antonarakis, S.E., and Beckmann, J.S. (2006). Mendelian disorders deserve more attention. *Nat. Rev. Genet.* 7, 277–282.

Antonarakis, S.E., Chakravarti, A., Cohen, J.C., and Hardy, J. (2010). Mendelian disorders and multifactorial traits: the big divide or one for all? *Nat. Rev. Genet.* 11, 380–384.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74.

Badano, J.L., and Katsanis, N. (2002). Beyond Mendel: an evolving view of human genetic disease transmission. *Nat. Rev. Genet.* 3, 779–789.

Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science* 350, 1092–1096.

Bloom, J.S., Ehrenreich, I.M., Loo, W.T., Lite, T.L., and Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. *Nature* 494, 234–237.

Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., et al. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* 43, 956–963.

Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., and Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* 132, 1077–1130.

Dipple, K.M., and McCabe, E.R. (2000). Modifier genes convert “simple” Mendelian disorders to complex traits. *Mol. Genet. Metab.* 71, 43–50.

Dittmar, J.C., Reid, R.J., and Rothstein, R. (2010). ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC Bioinformatics* 11, 353.

Dorfman, R. (2012). Modifier gene studies to identify new therapeutic targets in cystic fibrosis. *Curr. Pharm. Des.* 18, 674–682.

Dowell, R.D., Ryan, O., Jansen, A., Cheung, D., Agarwala, S., Danford, T., Bernstein, D.A., Rolfe, P.A., Heisler, L.E., Chin, B., et al. (2010). Genotype to phenotype: a complex problem. *Science* 328, 469.

Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., and Botstein, D. (2002). Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 99, 16144–16149.

Ehrenreich, I.M., Bloom, J., Torabi, N., Wang, X., Jia, Y., and Kruglyak, L. (2012). Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet.* 8, e1002570.

Fay, J.C. (2013). The molecular basis of phenotypic variation in yeast. *Curr. Opin. Genet. Dev.* 23, 672–677.

Fogel, S., and Welch, J.W. (1982). Tandem gene amplification mediates copper resistance in yeast. *Proc. Natl. Acad. Sci. USA* 79, 5342–5346.

Friedrich, A., Jung, P., Reisser, C., Fischer, G., and Schacherer, J. (2015). Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Mol. Biol. Evol.* 32, 184–192.

Gerstein, A.C., Ono, J., Lo, D.S., Campbell, M.L., Kuzmin, A., and Otto, S.P. (2015). Too much of a good thing: the unique and repeated paths toward copper adaptation. *Genetics* 199, 555–571.

Gresham, D., Desai, M.M., Tucker, C.M., Jenq, H.T., Pai, D.A., Ward, A., DeSevo, C.G., Botstein, D., and Dunham, M.J. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* 4, e1000303.

Hamilton, B.A., and Yu, B.D. (2012). Modifier genes and the plasticity of genetic networks in mice. *PLoS Genet.* 8, e1002644.

Hart, T., Chandrasekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163, 1515–1526.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., et al. (2009). Population genomics of domestic and wild yeasts. *Nature* 458, 337–341.

Mackay, T.F., Stone, E.A., and Ayroles, J.F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010).

- The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Moye-Rowley, W.S. (2003). Transcriptional control of multidrug resistance in the yeast *Saccharomyces*. *Prog. Nucleic Acid Res. Mol. Biol.* 73, 251–279.
- Mukherjee, V., Steensels, J., Lievens, B., Van de Voorde, I., Verplaetse, A., Aerts, G., Willems, K.A., Thevelein, J.M., Verstrepen, K.J., and Ruyters, S. (2014). Phenotypic evaluation of natural and industrial *Saccharomyces* yeasts for different traits desirable in industrial bioethanol production. *Appl. Microbiol. Biotechnol.* 98, 9483–9498.
- Nadeau, J.H. (2001). Modifier genes in mice and humans. *Nat. Rev. Genet.* 2, 165–174.
- Paaby, A.B., White, A.G., Riccardi, D.D., Gunsalus, K.C., Piano, F., and Rockman, M.V. (2015). Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *eLife* 4, 09178.
- Reisser, C., Dick, C., Kruglyak, L., Botstein, D., Schacherer, J., and Hess, D. (2013). Genetic basis of ammonium toxicity resistance in a sake strain of yeast: a Mendelian case. *G3 (Bethesda)* 3, 733–740.
- Ruiz, A., and Ariño, J. (2007). Function and regulation of the *Saccharomyces cerevisiae* ENA sodium ATPase system. *Eukaryot. Cell* 6, 2175–2183.
- Schacherer, J., Shapiro, J.A., Ruderfer, D.M., and Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458, 342–345.
- Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCr: visualizing classifier performance in R. *Bioinformatics* 21, 3940–3941.
- Steinmetz, L.M., Sinha, H., Richards, D.R., Spiegelman, J.I., Oefner, P.J., McCusker, J.H., and Davis, R.W. (2002). Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 416, 326–330.
- Strope, P.K., Skelly, D.A., Kozmin, S.G., Mahadevan, G., Stone, E.A., Magwene, P.M., Dietrich, F.S., and McCusker, J.H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25, 762–774.
- Thein, S.L. (2011). Genetic modifiers of sickle cell disease. *Hemoglobin* 35, 589–606.
- Treusch, S., Albert, F.W., Bloom, J.S., Kotenko, I.E., and Kruglyak, L. (2015). Genetic mapping of MAPK-mediated complex traits Across *S. cerevisiae*. *PLoS Genet.* 11, e1004913.
- Vu, V., Verster, A.J., Schertzberg, M., Chuluunbaatar, T., Spensley, M., Pajkic, D., Hart, G.T., Moffat, J., and Fraser, A.G. (2015). Natural variation in gene expression modulates the severity of mutant phenotypes. *Cell* 162, 391–402.
- Wagih, O., and Parts, L. (2014). gitter: a robust and accurate method for quantification of colony sizes from plate images. *G3 (Bethesda)* 4, 547–552.
- Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. *Science* 350, 1096–1101.