Title
Genetic architecture of highly complex chemical resistance traits across four yeast strains.

Permalink
https://escholarship.org/uc/item/08g2v6ck

Journal
PLoS genetics, 8(3)

ISSN
1553-7390

Authors
Ehrenreich, Ian M
Bloom, Joshua
Torabi, Noorossadat
et al.

Publication Date
2012

DOI
10.1371/journal.pgen.1002570

Peer reviewed
Genetic Architecture of Highly Complex Chemical Resistance Traits across Four Yeast Strains

Ian M. Ehrenreich1,2,3,4*, Joshua Bloom1,5*, Noorossadat Torabi1,5*, Xin Wang1,5*, Yue Jia1,3, Leonid Kruglyak1,2,3*

1 Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey, United States of America, 2 Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, United States of America, 3 Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey, United States of America, 4 Molecular and Computational Biology Section, University of Southern California, Los Angeles, California, United States of America, 5 Department of Molecular Biology, Princeton University, Princeton, New Jersey, United States of America

Abstract

Many questions about the genetic basis of complex traits remain unanswered. This is in part due to the low statistical power of traditional genetic mapping studies. We used a statistically powerful approach, extreme QTL mapping (X-QTL), to identify the genetic basis of resistance to 13 chemicals in all 6 pairwise crosses of four ecologically and genetically diverse yeast strains, and we detected a total of more than 800 loci. We found that the number of loci detected in each experiment was primarily a function of the trait (explaining 46% of the variance) rather than the cross (11%), suggesting that the level of genetic complexity is a consistent property of a trait across different genetic backgrounds. Further, we observed that most loci had trait-specific effects, although a small number of loci with effects in many conditions were identified. We used the patterns of resistance and susceptibility alleles in the four parent strains to make inferences about the allele frequency spectrum of functional variants. We also observed evidence of more complex allelic series at a number of loci, as well as strain-specific signatures of selection. These results improve our understanding of complex traits in yeast and have implications for study design in other organisms.

Introduction

Most traits of agricultural, evolutionary, and medical significance are genetically complex, involving multiple genes that interact with one another and the environment [1]. Despite decades of effort, our understanding of how such traits are specified at the genetic level remains incomplete [2]. Studies in model organisms can provide fundamental insights into the genetic basis of complex traits that are applicable to other species, including humans [3]. However, such studies typically detect only a small fraction of the loci that contribute to a trait due to low statistical power [4].

To improve genetic mapping of complex traits in *Saccharomyces cerevisiae*, we recently developed extreme QTL mapping (X-QTL), which is a bulk segregant mapping technique that employs millions of cross progeny [5]. X-QTL involves three key steps: generation of very large segregating populations, isolation of cross progeny with extreme trait values, and quantitative measurement of pooled allele frequencies across the genome in these phenotypically extreme individuals [5]. To make the pools of segregants that are the starting point for X-QTL, we use selectable markers to obtain an effectively unlimited number of progeny from a cross of two strains. We then employ selection-based phenotyping to isolate large numbers of segregants with extreme trait values from populations that contain millions of cross progeny. DNA is extracted from pools of phenotypically extreme segregants, and the allele frequencies of markers throughout these individuals’ genomes are determined using custom microarrays or next generation sequencing. In an X-QTL experiment, a locus that influences a trait is expected to show an allele frequency skew in the direction of the parental allele that contributes to a more extreme trait value.

By applying X-QTL to a number of chemical resistance phenotypes in a single cross of the lab strain BY4716 and the vineyard strain RM11-1a (hereafter, BY and RM, respectively), we were able to show that large numbers of loci can underlie quantitative trait variation between *S. cerevisiae* isolates [5]. Following our publication, another group observed similar results in a different cross [6], suggesting that high genetic complexity may be a common feature of heritable trait variation among yeast strains.

Here, we examined how genetic complexity varies among strains and crosses. We used X-QTL to identify the genetic basis of resistance to 13 diverse chemicals in all 6 pairwise crosses of strains BY, RM, YJM789, and YPS163. YJM789 (hereafter, YJM) is derived from a clinical isolate, and YPS163 (hereafter, YPS) is an...
Author Summary

Most heritable traits of agricultural, evolutionary, and medical significance are specified by multiple genetic loci. Despite decades of research, we have only a limited understanding of the genetic basis of such complex traits. Studies in model organisms have the potential to provide fundamental insights into this research area, but most genetic mapping studies in these species have had low statistical power to detect multiple loci with small effects. Using a technique in which we employed millions of cross progeny in genetic mapping, we previously showed that resistance to chemicals has a highly complex genetic basis in a cross of a lab strain and a wine strain of the budding yeast Saccharomyces cerevisiae. Because we only examined a single cross, it was unclear how general our findings were. Here, we expand our work to all six possible crosses of four strains—the two isolates we used in our last study, as well as an isolate from an immunocompromised human being and an isolate from the sap of an oak tree. Our results based on these four ecologically and genetically distinct S. cerevisiae strains suggest that resistance to chemicals commonly exhibits a highly complex genetic basis among yeast isolates.

Results/Discussion

We previously noted that levels of genetic complexity underlying heritable variation in growth differed among chemical conditions in a single cross [5]. Here, we sought to determine the generality of our previous finding by examining additional crosses. We first generated the strains and microarrays to conduct X-QTL in all 6 pairwise crosses of the BY, RM, YJM, and YPS strains (Materials and Methods). Because the statistical power gained by using very large mapping populations, we detected approximately an order of magnitude more loci than did previous studies involving multiple crosses of yeast strains [15,17,20], allowing us to gain deeper insights into the genetic architecture and evolution of complex traits in S. cerevisiae.

Using a technique in which we employed millions of cross progeny in genetic mapping, we previously showed that resistance to chemicals has a highly complex genetic basis in a cross of a lab strain and a wine strain of the budding yeast Saccharomyces cerevisiae. Because we only examined a single cross, it was unclear how general our findings were. Here, we expand our work to all six possible crosses of four strains—the two isolates we used in our last study, as well as an isolate from an immunocompromised human being and an isolate from the sap of an oak tree. Our results based on these four ecologically and genetically distinct S. cerevisiae strains suggest that resistance to chemicals commonly exhibits a highly complex genetic basis among yeast isolates.

oak strain. These 4 strains are highly diverged at the sequence level [7,8,9,10,11] and exhibit a wide range of heritable phenotypic differences [12,13,14,15,16,17,18,19]. Because of the statistical power gained by using very large mapping populations, we detected approximately an order of magnitude more loci than did previous studies involving multiple crosses of yeast strains [15,17,20], allowing us to gain deeper insights into the genetic architecture and evolution of complex traits in S. cerevisiae.
affinity glucose transporter [21], suggesting that variability in glucose uptake may contribute to growth differences among the strains. The effect of RED1 on tunicamycin resistance is less clear, as this gene is thought to be involved in chromosome segregation [21], and tunicamycin affects the unfolded protein response. We also constructed allele replacement strains for two other genes: NUP157, which lies within a copper sulfate resistance locus with the resistance allele coming from BY, and PTK1, which lies within a paraquat resistance locus with the resistance allele coming from YPS. However, we obtained inconsistent results for NUP157 and PTK1: the allele replacements produced effects on resistance that were in the opposite direction from those seen in the X-QTL selections, and also caused growth defects on standard rich medium, suggesting that we did not identify the right candidate genes for these loci.

In addition to the simple bi-allelic patterns, we observed other more complex patterns of peaks (Figure 2A). Some of these are consistent with the presence of allelic series, in which either three or four alleles with different phenotypic effects are present among the four strains; we observed 29 examples involving at least 3 alleles and 9 examples that can only be explained by the presence of 4 different alleles (Table S2). The other 210 loci (51% of all loci) showed patterns of peaks that were not easily interpretable in terms of specific allelic classes. This probably reflects a mixture of false negatives in which a peak was present but not detected in a given cross, and cross-specific effects due to non-additive interactions and linkage between loci.

The allele frequency spectrum of causal loci is critical for the design of genetic mapping studies and for understanding sources of missing heritability in natural populations, including humans. As discussed above, we were able to distinguish and enumerate two simple allelic classes—singletons and doubletons. We estimated the peak detection rate to be 51%, with a 95% confidence interval of 39%–62%, and the ratio of allelic singletons to doubletons to be 3.03, with a 95% confidence interval of 1.7–5.3 (Figure 3A; Figure S3). This result suggests that despite the high statistical power of X-QTL, a substantial fraction of loci with weaker effects likely still go undetected in any one cross. Interestingly, the estimate of the ratio of allelic singletons to doubletons is similar to that observed for nonsynonymous polymorphisms in the genomes of the parent strains (2.97), and is shifted toward singletons relative to both the neutral expectation of 2.67 and the observed ratio of 2.57 for 109,585 SNPs genome-wide (Figure 3A). Thus, the frequency spectrum of variants that contribute to complex trait variation in yeast appears to be mildly shifted toward lower frequencies by purifying selection, but, given the wide confidence interval for the estimated ratio of allelic singletons to doubletons, we cannot rule out that the variant frequencies follow the neutral spectrum.

Several lines of evidence suggest that lineage-specific selection or demography has shaped variation among the four strains. We observed an excess of allelic singletons at detected loci for BY and RM, and a deficit for YJM and YPS, relative to the numbers of singleton SNPs in the parent genomes ($\chi^2 = 35.98$; d.f. = 3, $p<0.0001$; Figure 3B). The laboratory strain BY also exhibits other signatures of selection for both general and chemical-specific resistance. For instance, BY carries a marginally significant excess of allelic singletons that confer resistance relative to the other three strains (Fisher’s exact test, Bonferroni-corrected $p = 0.06$; Figure 3C; Table 1). In addition, trait-specific sign tests [22] identified one significant result: an excess of copper sulfate resistance alleles contributed by BY in the BYxRM cross (18 loci with BY carrying the resistance allele and 2 loci with RM carrying the resistance allele; binomial test, Bonferroni-corrected $p = 0.031$; Figure 3D). Interestingly, BY is among the most copper-resistant S. cerevisiae strains [23,24], and our data suggest that this resistance in
BY may be the result of selection, possibly due to the use of high levels of copper or another chemical with similar effects in standard growth media. However, the BYxYJM and BYxYPS crosses do not show significant excess of BY alleles, and RM is also among the more highly copper-resistant strains [23], making the excess of BY resistance alleles in the BYxRM cross difficult to explain. Overall, our results are consistent with previous analyses that have shown lab strains isogenic to BY exhibit high evolutionary rates relative to other yeast isolates [25], probably due to both relaxed purifying selection [26] and adaptation [26,27].

We have shown that variation in chemical resistance among yeast strains is typically due to a large number of underlying loci. The level of genetic complexity, as measured by the number of loci detected, is largely a property of each resistance trait, although it is also affected to a lesser extent by the choice of parent strains. The total number of distinct loci detected for a trait in these crosses among four strains ranged from 8 to 57, and these numbers substantially exceeded those seen in any one cross. These observations suggest that the total number of loci affecting certain resistance traits in _S. cerevisiae_ can be very large, since many of them
Table 1. Patterns used to identify allelic singletons and allelic doubletons in the X-QTL data, and the number of loci detected with these patterns.

| Allele          | Effect   | BYxRM | BYxYJM | BYxYPS | RMxYJM | RMxYPS | YJMxYPS | Subclass Total | Class Total |
|-----------------|----------|-------|--------|--------|--------|--------|---------|---------------|-------------|
| BY singleton Resistant | BY       | BY    | BY     |        |        |        |         | 26            | 42          |
| Susceptible     | RM       | YJM   | YPS    |        |        |        |         | 16            |             |
| RM singleton Resistant | RM       |        |        | RM     | RM     |        |         | 19            | 49          |
| Susceptible     | BY       |        |        |        | YJM    | YPS    |         | 30            |             |
| YJM singleton Resistant | YJM      | YJM   | YJM    |        | YJM    |        |         | 18            |             |
| Susceptible     | .        | BY    |        |        |        | YPS    |         | 11            |             |
| YPS singleton Resistant | YPS      |        |        | YPS    | YPS    |        |         | 9             | 26          |
| Susceptible     | .        |        |        | RM     | YJM    |        |         | 17            |             |
| BY = RM doubleton BY & RM resistant | BY    | BY    | RM    | RM     |        |        |         | 7             | 12          |
| BY & RM susceptible |        | YJM   | YPS    | YJM    | YPS    |        |         | 5             |             |
| BY = YJM doubleton BY & YJM resistant | YJM    | YJM   | YJM    | YJM    | YJM    |        |         | 3             | 6           |
| BY & YJM susceptible | RM     | YPS   |        | RM     | YPS    |        |         | 3             |             |
| BY = YPS doubleton BY & YPS resistant | YPS    |        |        | YPS    | YPS    |        |         | 7             | 10          |
| BY & YPS susceptible | RM     | YJM   |        | RM     | YJM    |        |         | 3             |             |

The strain mentioned under a cross indicates which allele should have been selected in that cross for the given pattern to hold. Both exact match patterns and patterns that allow for one undetected peak were used to generate this table. A full listing of the exact match and "one off" patterns is described in Table S1.

doi:10.1371/journal.pgen.1002570.t001

will have escaped detection because they don’t vary among the four parent strains examined here, have effect sizes that are too small, or are too closely linked to be resolved as separate loci by our mapping technique. Our results suggest that the functional variants underlying complex traits are broadly distributed across the frequency spectrum from rare to common alleles, and that many loci harbor more than two allelic variants. These findings provide multiple non-exclusive explanations for the sources of the “missing heritability” of complex traits, and illustrate the power of a simple model system for probing genetic complexity.

Materials and Methods

Construction and use of segregating pools for X-QTL

The Synthetic Genetic Array marker system [28] was used to generate MATa haploid pools as previously reported [5], with the exception that thialysine and the dominant sensitive LIP1/lyp1A marker system were not employed. All six pairwise crosses of BY, RM, YJM, and YPS were made, with one strain in a cross having the genotype MATa can1::STE2pr-SpHIS5 his3D and the other having the genotype MATa his3D. In notation describing crosses (e.g., BYxRM), we first list the MATa and then the MATa parent. The selection experiments used for X-QTL were conducted as previously described [5]. The drug doses used in the selections, which were determined by plating millions of cells across a range of drug doses and finding a concentration at which 300 to 1,000 colonies could be resolved, are given in File S1. Each experiment was conducted once, as we previously found that biological replicates conducted on the same day produced highly similar results [5].

Microarray design and use

Microarrays were designed from the BY genome sequence obtained from the Saccharomyces Genome Resequencing Project [10]. Note that the YPS606 genome was used to design the YPS array, as YPS606 is isogenic to YPS163. We aligned the genomes chromosome-by-chromosome using Fast Statistical Alignment (FSA) [29]. These multiple sequence alignments were filtered for SNPs using the following criteria: i) all 4 strains had to have been sequenced at a position and ii) all 4 strains had to have a specific base called (i.e. A, C, G, or T) at the position. These SNPs were then used for microarray design, as well as for downstream population-genetic analyses. Cross-specific microarrays were designed using only bi-allelic SNPs. Probes were chosen to have a length between 21 and 27 nucleotides and a melting temperature between 54 and 56°C as described previously [5,30]. One probe was designed for each allele of a SNP, and the two probes for a SNP were randomly positioned on the microarray. Probes were targeted to regions where only one SNP would be covered by the probes. Markers were chosen to provide near-uniform coverage of the genome. The arrays were tested using control DNA from both parents and the heterozygous diploid to ensure that they could discriminate the two alleles of a SNP. All hybridizations and processing was done as previously described [5]. All microarray data is available in the Princeton University MicroArray database (http://puma.princeton.edu/). The processed log10 hybridization intensities are included in Files S2, S3, S4, S5, S6, S7.

Peak detection

For a given SNP, the difference in the log10 ratios of the intensities of the MATa and MATa parent-specific probes on a single array was computed (subsequently referred to as a ‘log10 intensity difference’), and this metric was used in downstream analyses. Background allele frequency changes that occur during
pool construction were removed from the data for each X-QTL selection. This was done separately for each SNP by subtracting the average log10 intensity difference obtained in seven cross-specific control experiments from the log10 intensity difference observed in an X-QTL selection. A peak detection algorithm was then employed that used a Savitzky-Golay filter to smooth the data within sliding windows of 100 probes. This smoothing approach was used to preserve local maxima in the data. Loci were called at a 1% FDR threshold, where the number of false discoveries was determined by running the peak caller on the control data using a range of thresholds, and the total number of discoveries was determined by running the peak caller on the selection data at the same thresholds used to analyze the controls. Thresholds were set by examining the quantiles of log10 intensity differences observed for every 100 SNP genomic window on an array, and taking the median interquantile range between the x and 1-x quantiles, where x ranged from 0.005 to 0.45. We found that setting x as 0.045 resulted in a 1% FDR. Peak calling and all other statistical analyses were conducted in R (http://www.r-project.org/). The peak caller and an associated function library are included in Files S8 and S9. The detected peaks are listed in File S10.

Testing for effects of chemical and genetic background on the number of peaks detected in a cross

The test for cross effect was conducted using the model $y = \text{chemical} + \text{cross}$, while the test for strain effect was conducted using the model $y = \text{chemical} + \text{strain}1 + \text{strain}2 + \text{strain}3$. Implementing the second test required specifying the design matrix for the strain effect. Each row in the design matrix represented a single X-QTL experiment from a particular combination of chemical and cross. Entries in the design matrix were parameterized as follows: a strain had a value of 2, 1, or 0 if it was the MATa parent, the MATα parent, or not a parent in a particular experiment, respectively. Only three strains were included in the test, because the information for the fourth could be obtained from the other three. To ensure that results were not dependent on the three included strains, we conducted the test with all four possible combinations of the three strains and...
Testing for disproportionate contributions of particular strains to the genetic complexity of traits

We first conducted \( \chi^2 \) tests in which single strains were examined. This test has two categories – one that is the sum of the peaks detected in the three crosses involving the query strain and another that is the sum of the peaks detected in the other three crosses. The expectation is that each of these classes will contain half of the peaks detected for a trait. We then conducted \( \chi^2 \) tests in which two strains were examined. The first category here is the sum of the peaks detected in the four crosses involving the two strains, while the second is sum of the peaks detected in the other two crosses. Here, the expectation is that the first category will contain two-thirds of the peaks, while the second will contain one-third of the peaks.

Identification of distinct loci for a trait

Peaks identified across the six crosses for a single trait were grouped into distinct loci. We started with the most strongly selected peak on each chromosome and grouped with it all peaks that occurred within a 200-kilobase window surrounding it. This window size accommodated the grouping of peaks that exhibited weak but significant allele frequency changes, and may result in the underestimation of the total number of loci due to the overgrouping of peaks. Remaining peaks were grouped into distinct loci using additional iterations of the procedure until all peaks identified for a trait were members of a group.

Analysis of distinct loci across traits

We divided the genome into equally sized bins ranging from 20 to 100 kb and counted the number of distinct loci that fell into each bin. A bin was considered to have an excess of distinct loci if the number present in it exceeded the number expected by chance from a Poisson distribution, given the number of distinct loci divided by the total number of bins and a multiple testing correction for the number of bins. With the 50 kb bin size reported in the text, 8 or more distinct loci were required to be present in a bin for the bin to be considered significant.

Identification of allelic series, doubletons, and series

The distinct loci identified for each trait were used to classify singletons and doubletons. The specific patterns used to identify the allelic classes are described in Table S1. We focused on exact pattern matches and on patterns that were missing an expected peak at a given locus in one cross. A number of distinct loci had peaks detected in four or more crosses, but did not conform to the patterns expected for allelic doubletons. We considered these loci as allelic series, and for each of these putative series we determined the possible logical relationships of the parent alleles to each other. These relationships are reported in Table S2.

Identification of candidate causal genes

For each bi-allelic locus, we evaluated a 30 kb interval centered on its estimated position for polymorphisms that segregated among the parent strains in the same pattern as the X-QTL peaks. Any gene that harbored a polymorphism in the coding region or in the immediate upstream and downstream regions was considered a candidate. The candidate genes are listed in File S11.

Allele replacement strategy

To generate the replacement strains, we used the allele replacement technique described by Storici et al. [31]. This method is a two-step process that involves knocking out a gene with a selectable marker cassette, and then replacing the selectable marker cassette with a different allele of the gene. We made each allele replacement strain once in one parental background, and then compared the phenotypes of the strains to their progenitors. For the two genes that exhibited the expected phenotypic effect, we made a second version of the allele replacement strain to validate the presence of functional variation in the gene.

Maximum likelihood estimation of the ratio of singletons to doubletons

The observed counts of exact-match allelic singletons and doubletons and near-exact-match allelic singletons and doubletons were modelled using two parameters: the detection rate of peaks (\( \alpha \)) and the ratio of singletons to doubletons (\( \beta \)). The formulae underlying this computation are provided in Text S1. The likelihood of each combination of parameter values was examined across a two-dimensional grid of parameter values using \( \chi^2 \) tests with 3 degrees of freedom. The likelihood reached a maximum at \( \alpha = 0.31 \) and \( \beta = 3.03 \). We obtained 95% confidence intervals for \( \alpha \) and \( \beta \) by using the \( \chi^2 \) distribution with 3 degrees of freedom and identifying the \( \chi^2 \) value for the 95% quantile. We then identified parameter combinations that produced an \( \chi^2 \) value below this threshold (7.81), and determined the minimum and maximum values of \( \alpha \) and \( \beta \) that satisfied this condition.

Supporting Information

Figure S1 Plots of X-QTL mapping results. The data for each trait is plotted as the difference between the MAT\( \alpha \) and MAT\( \alpha \) allele-specific probes on the selection array minus the average of the differences between the MAT\( \alpha \) and MAT\( \alpha \) allele-specific probes from seven control arrays. The red vertical lines indicate positions that were called as peaks at a 1% FDR.

Figure S2 Cloning of genes. (A–C) show the steps taken to clone \( HXT6 \), while (D–E) show the steps taken to clone \( RED1 \). In both cases, a locus was identified in all three crosses sharing one parent—the crosses involving RM in control conditions for \( HXT6 \) (A) and the crosses involving BY on tunicamycin for \( RED1 \) (B). The regions underlying the detected peaks were surveyed for polymorphisms that segregated across the parent strains in the same pattern as the detected locus. Both \( HXT6 \) and \( RED1 \) were chosen because they carry a number of nonsynonymous polymorphisms relative to other genes in their genomic regions (B and E). Allele replacement strains were made in the RM background using the BY strain as a template. Each strain was independently constructed twice and phenotyped using serially diluted colony growth assays (C and F). Overnight cultures were grown for each strain and then pinned onto agar plates using the Singer RoToR. The \( HXT6 \) strains were measured after 24 hours of growth at 30°C, while the \( RED \) strains were measured after 65 hours of growth at 30°C. RM grows better on standard medium when it carries its own allele of \( HXT6 \), while the BY allele of \( RED1 \) confers a growth advantage on tunicamycin. In B, dubious ORFs are colored in blue. In C and F, cultures were grown undiluted (abbreviated “Und.”) and at two successive tenfold dilutions. Each dilution of a strain was pinned in a square of four technical replicates.

Figure S3 Likelihood surface for the estimates of the ratio of allelic singletons to doubletons and the detection rate. P\( \chi^2 \)Data|Model is shown, with the correspondence of colors to probabilities.
given in the key. This was generated using the model described in Text S1, and evaluating the model across a wide, two-dimensional range of detection rates and singleton to doubleton ratios. In addition, this likelihood surface was used to generate the confidence intervals described in the main text.

**References**

1. Falconer DS, Mackay TF (1996) Introduction to quantitative genetics (4th edition). Harlow, England: Pearson Education Limited.
2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753.
3. Mackay TF, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10: 565–577.
4. Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A 102: 1572–1577.
5. Ehrenreich IM, Torabi N, Jia Y, Kent J, Maris S, et al. (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464: 1039–1042.
6. Parts L, Cubillos FA, Warringer J, Jain K, Salinas F, et al. (2011) Revealing the genetic structure of a trait by sequencing a population under selection. Genome Res 21: 1131–1138.
7. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, et al. (2008) A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet 4: e1000103. doi:10.1371/journal.pgen.1000103.
8. Wei W, McCusker JH, Hyman RW, Jones T, Ning Y, et al. (2007) Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. Proc Natl Acad Sci U S A 104: 12825–12830.
9. Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L (2008) Population genomic analysis of outcrossing and recombination in yeast. Nat Genet 39: 1077–1081.
10. Li T, Garter DM, Moses AM, Warringer J, Parts L, et al. (2009) Population genomics of domestic and wild yeasts. Nature 458: 337–341.
11. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. Nature 458: 342–345.
12. Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296: 752–755.

**File S1** Candidate loci for bi-allelic loci.

**File S11** Candidate genes for bi-allelic loci.

**Table S1** All patterns used to identify allelic singletons and allelic doubletons in the X-QTL data, and the number of loci detected with these patterns. The strain mentioned under a cross indicates which allele should have been selected in that cross for the given pattern to hold. We show all exact patterns used to identify singletons and doubletons, as well as each of the patterns that indicate the presence of a singleton or doubleton if one undetected peak is allowed. Both the exact match and “one off” patterns were used in the counts of bi-allelic loci described throughout the paper.

**Table S2** Allelic series inferred from the data. The most parsimonious relationship of alleles to each other is indicated. Greater than and equal signs indicate the effects of the alleles relative to each other, with “A>B” meaning that allele A confers higher resistance than allele B and “A=B” meaning that the effects of allele A and allele B are not distinguishable. In some cases, there are two equally parsimonious relationships that can explain the data.

**Acknowledgments**

The authors are grateful to members of the Kruglyak lab, Audrey Gasch, and three anonymous reviewers for comments that helped improve the manuscript. We thank David Botstein and Harmen Bussemaker for helpful discussions and Joshua Shapiro for input regarding the maximum likelihood analysis of allele frequencies.

**Author Contributions**

Conceived and designed the experiments: IME LK. Performed the experiments: IME NT XW. Analyzed the data: IME JB. Wrote the paper: IME LK. Constructed strains: IME XW YJ. Designed microarrays: IME.
24. Kvitek DJ, Will JL, Gasch AP (2008) Variations in stress sensitivity and genomic expression in diverse S. cerevisiae isolates. PLoS Genet 4: e1000223. doi:10.1371/journal.pgen.1000223.

25. Warringer J, Zorgo E, Cubillos FA, Zia A, Gjuvsland A, et al. (2011) Trait variation in yeast is defined by population history. PLoS Genet 7: e1002111. doi:10.1371/journal.pgen.1002111.

26. Gu Z, David L, Petrov D, Jones T, Davis RW, et al. (2005) Elevated evolutionary rates in the laboratory strain of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 102: 1092–1097.

27. Fraser HB, Moses AM, Schadt EE (2010) Evidence for widespread adaptive evolution of gene expression in budding yeast. Proc Natl Acad Sci U S A 107: 2977–2982.

28. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294: 2364–2368.

29. Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, et al. (2009) Fast statistical alignment. PLoS Comput Biol 5: e1000392. doi:10.1371/journal.pcbi.1000392.

30. Gresham D, Curry B, Ward A, Gordon DB, Brizuela I, et al. (2010) Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. Proc Natl Acad Sci U S A 107: 1482–1487.

31. Storici F, Lewis LK, Resnick MA (2001) In vivo site-directed mutagenesis using oligonucleotides. Nat Biotechnology 19: 773–776.