Title: Strength of selection potentiates distinct adaptive responses in an evolution experiment with outcrossing yeast

Authors: Mark A. Phillips, Rupinderjit K. Briar, Marcus Scaffo, Shenghao Zhou, Molly K. Burke

Abstract

Combining experimental evolution with whole-genome sequencing is now a well-established method for studying the genetics of adaptation and complex traits. In this type of work that features sexually-reproducing populations, studies consistently find that adaptation is highly polygenic and fueled by standing genetic variation. Less consistency is observed with respect to general evolutionary dynamics however; for example, investigators remain ambivalent about whether selection produces repeatable versus idiosyncratic responses, or whether small shifts in allele frequencies at many loci drive adaptation, versus selective sweeps at fewer loci. Resolving these open questions is a crucial next step as we move toward extrapolating findings from laboratory evolution experiments to populations inhabiting natural environments. We propose that subtle differences in experimental parameters between studies can influence evolutionary dynamics in meaningful ways, and here we empirically assess how one of those parameters – selection intensity – shapes these dynamics. We subject populations of outcrossing *Saccharomyces cerevisiae* to zero, moderate, and high ethanol stress for ~200 generations and ask: 1) does stronger selection intensity lead to greater changes in allele frequencies, and a higher likelihood of selective sweeps at sites driving adaptation; and (2) do targets of selection vary with selection intensity? We find some evidence for positive correlations between selection intensity and allele frequency change, but no evidence for more sweep-like patterns at high intensity. While we do find genomic regions that suggest some shared genetic architecture across treatments, we also identify distinct adaptive responses in each selection treatment. Combined with evidence of phenotypic trade-offs between treatments, our findings support the hypothesis that selection intensity might influence evolutionary outcomes due to pleiotropic and epistatic interactions. As such, we conclude that details of a selection regime should be a major point of consideration when attempting to generalize inferences across experimental evolution studies. Finally, our results demonstrate the importance of clearly defined controls when associating genomic changes with adaptation to specific selective pressures. Despite working with a presumably lab-adapted model system, without this element we would not have been able to distinguish between genomic responses to ethanol stress and laboratory conditions.

Introduction

Evolve and resequence ("E&R") experiments are now commonly used to study the genetics of adaptation and complex traits (Long et al. 2015; Schlötterer et al. 2015). In these studies, experimental populations are subjected to carefully designed selective regimes under controlled conditions, and sampled for DNA or RNA sequencing over many generations. This experimental framework allows for powerful statistical associations between observed, presumably adaptive, phenotypic responses with shifts in genetic variation or changes in gene expression. In addition to linking genotypes to phenotypes, the ability to control population size, level of starting genetic variation, and environmental conditions make these experiments a powerful tool for studying broad adaptive dynamics. It is now well understood that adaptation in E&R studies featuring outcrossing systems is fueled by standing genetic variation (Long et al. 2015; Barghi et al. 2020). However, the specific evolutionary dynamics observed across these studies vary considerably. We believe achieving a better understanding of what causes these differences is a key step towards extending general findings from E&R to real populations (Phillips and Burke 2021).

Major points of discrepancy in the literature today include topics like evolutionary repeatability, the role of contingency in the genetics of adaptation, and what some term "sweeps versus shifts" (Barghi et al. 2020). Using evolutionary repeatability as an example, while many studies report high levels of evolutionary repeatability, as evidenced by parallel responses to selection across replicate populations (e.g. Linder et al. 2021), others find more idiosyncratic and replicate-specific responses (e.g. Barghi et al. 2019. We suggest that this and other observed discrepancies are likely due to differences in experimental design among studies, as well as the genetic variation segregating in ancestral populations. In the case of evolutionary repeatability, differences in starting allele frequency distributions and effective population size stand out as potentially relevant factors (i.e., these factors directly influence the probability that selection will favor a specific beneficial allele in a given replicate). More broadly, we believe these discrepancies point to a need for deeper understanding of how experimental parameters influence adaptive processes before attempting to extend inferences from E&R studies to populations outside the lab. Here we focus on how a specific experimental choice – selection intensity – impacts evolutionary outcomes in large outcrossing populations with abundant standing genetic variation.

How a given selective pressure manifests (e.g. intensity, duration, constant versus dynamic, etc.) has clear ecological relevance that likely shapes evolutionary outcomes in both natural populations and laboratory experiments. While this topic has been explored across a number of theoretical studies (e.g. Kessner and Novembre 2015; Stetter et al. 2018; Christodoulaki et al. 2019; Vlachos and Kofler 2019; Hayward and Sella 2021), few E&R studies have sought to address it empirically. Here we aim to address this gap by directly testing the effects of selection intensity on the genetic architecture of adaptation using populations of outcrossing *S. cerevisiae*. Specifically, we compare the genomic response to moderate and high ethanol stress in replicate populations all derived from a single ancestral population. We are

primarily interested in how selection intensity shapes signatures of selection and whether treatments featuring different selection intensities show distinct genomic responses to selection.

Christodoulaki et al. (2019) explored the idea that selection intensity influences signatures of selection using simulations designed to mimic laboratory experimental evolution. They find that when a new trait optimum is distant (i.e. when a strong selective pressure is applied), the genomic response to selection often involves substantial increases in allele frequencies that are consistent with selective sweeps. However, for close trait optima (weaker selection pressure), adaptation proceeds through smaller shifts in allele frequencies. The "sweeps versus shifts" question – in other words, the question of whether population-genetic models focused on identifying few loci of large effect are better at explaining trait evolution than quantitative-genetic models that assume extensive polygenicity – looms beyond the experimental evolution literature (Höllinger et al. 2019), and we believe that selection intensity stands out as a major factor key to resolving this question. A major goal of this study is therefore to test to prediction that high selection intensity results in larger changes in allele frequencies at and around target sites.

We are also interested in exploring if the same targets of selection respond to both high and moderate selection intensities. In the simplest scenario, there is some set of "ethanol resistance" alleles segregating in our ancestral population, and these should increase in frequency in both selection intensity treatments. However, if the genetics of complex traits is characterized by widespread pleiotropy as some suggest (Visscher and Yang 2016; Boyle et al. 2017) and/or epistasis, this might not be the case. For example, due to pleotropic effects and resulting trade-offs with other characters, alleles that are adaptive under strong ethanol stress conditions might not be favored under weaker ethanol stress conditions and vice versa. If this prediction holds true, it would suggest a clear need to incorporate this possibility as a factor when constructing models of polygenic adaptation. Lastly, alleles that are beneficial at late but not early stages of adaptation have been previously identified in evolving outcrossing *S. cerevisiae* populations, which could indicate epistatic interactions (Phillips et al. 2020). If this is the case and generally true, it could also drive treatment-specific responses in combination with pleiotropy; for example, an allele that is beneficial at moderate levels of ethanol stress may not confer those same benefits at higher ethanol stress due to negative interactions with alleles unambiguously beneficial under high ethanol stress.

In summary, with this work we aim to address the following questions: (i) do populations experiencing stronger selection intensities display larger changes in allele frequencies at and around targets of selection compared to populations experiencing weaker selection intensities; and (ii) do targets of selection vary with selection intensity? We address these questions using genomic data from outcrossing *S. cerevisiae* populations subjected to zero, moderate, and high ethanol stress for ~200 generations.

Materials and Methods

Experimental populations and selection regimes

All experimental populations used in this study were derived from the "S12" population described in detail in Phillips et al. (2021). Briefly, this population was created by combining 12 haploid strains from the barcoded SGRP yeast strain collection (Cubillos et al. 2009). The specific strains used are: Y12, YSP128, SK1, DBVPG6044, UWOPS05_217_3, L_1528, L_1374, DBVPG6765, YJM981, YJM975, BC187, and N273614. As described in Linder et al. (2020), these strains have been modified to enable easy crossing and diploid recovery – these modified strains were kindly provided by Dr. Anthony Long. The MATa and MATα strains contain *ho* deletions to prevent mating-type switching, but each contain a different drug-resistance marker inserted into a pseudogene (YCR043C) tightly linked to the mating-type locus (*MATa, hoΔ, ura3::KanMX-barcode, ycr043C::NatMX* and *MATα, hoΔ, ura3::KanMX-barcode, ycr043C::hphMX*). These genotypes enable haploids of each mating type to be recovered using media supplemented with either hygromycin B or nourseothricin sulfate, and they enable newly mated a/α diploids to be recovered in media supplemented with both drugs.

To create the S12 recombinant outbred population from these haploid strains, we used a semi-round-robin design of initial crossing, followed by 12 iterations of what we call our "weekly outcrossing protocol" to maximize genetic diversity and allow for some domestication to laboratory handling conditions (detailed in Burke et al. 2020; Phillips et al. 2021). The experimental populations featured in this study were then derived from samples of this base population: 20 control replicates (C_{1-20}), 20 moderate ethanol stress replicates (M_{1-20}), and 20 high ethanol stress replicates (H_{1-20}). Levels of ethanol used in the two stress treatments were chosen based on the results of exploratory growth rate assays (see next section for general assay methods). We found that 10% ethanol was close to the limit of what could support population growth (at least for a 48-hour period), and we thus chose this as the high stress treatment. We chose the moderate treatment to include 6% ethanol as this resulted in a doubling time approximately mid-way between the high stress and control conditions (Supplementary Figure 1).

The weekly outcrossing protocol described by Burke et al. (2020) was also used to maintain sexual reproduction in these 60 replicate populations, with minor modifications to increase throughput (i.e. smaller volumes). This protocol involves batch culture of diploids in 1mL of liquid medium in alternating wells of 24-well plates (Corning); every other well contained sterile YPD which was monitored for growth (which would indicate potential cross-contamination) throughout the experiment. After batch culture of diploids, the entire 1mL of culture was washed and resuspended in 1mL minimal sporulation media (1% potassium acetate), and incubated with shaking for 72 hours (30°C/200 rpm). After sporulation, a modified random spore isolation protocol was implemented to disrupt asci and isolate spores. This protocol involves resuspending sporulated cultures in 1mL Y-PER Yeast Protein Extraction Reagent (Thermo), followed by incubation at 50°C for 15 minutes to kill vegetative diploid cells. Cultures were then resuspended in a 1% zymolyase (Zymo Research) solution to weaken ascus

walls, and vortexed at maximum speed with 0.5 mm silica beads (BioSpec) to mechanically agitate the asci. Following these steps, spores were transferred to YPD agar plates supplemented with nourseothricin sulfate (100mg/L), hygromycin B (300mg/L) and G418 (200mg/L) and incubated at 30°C for 48 hours, during which time spores mated (due to close proximity on the plate) and diploids germinated. The resulting lawns of new diploid cells were scraped off plates using sterile glass slides and transferred to 10mL of sterile YPD media (1/10th of this culture was preserved in 15% glycerol and archived at -80°C for DNA extraction and sequencing). From these super-saturated cultures, 10 μL were transferred into 1mL of treatment-specific medium and incubated for 48 hours (30C/200rpm), with a 1/100 dilution halfway through to increase generational turnover. For the control treatment, standard YPD (2% yeast extract, 1% peptone, 2% dextrose) was used as the culture medium. For the moderate and high ethanol stress treatments, YPD was supplemented with either 6% or 10% ethanol (by volume), respectively. After this period of batch culture, the next weekly outcrossing iteration would commence (Supplementary Figure 2). All replicates were handled in parallel for a total of 15 weeks, with one cycle of outcrossing per week.

By estimating cell density at various steps of the outcrossing protocol (via OD_{600} of liquid cultures, and/or colony counts of dilution series), we can approximate the number of asexual generations that occurred in each treatment during the experiment. We estimate that ~10 cell doublings occurred during the 48-hour period of batch culture – this phase of competitive growth in treatment-specific medium is most relevant to our major questions relating to selection strength. We infer that an additional ~4 cell doublings occur during the non-competitive phases of the protocol (i.e. growth on agar medium). We observed that growth was slower (and cell densities lower) in the high ethanol stress treatment compared to the others, but this did not lead to dramatically different estimates of cell doublings among the treatments. Ultimately, we project that over 200 asexual generations elapsed in all treatments over 15 weeks. This generational turnover is difficult to precisely estimate for several reasons, chiefly because some aspects of population biology were not routinely quantified throughout the experiment. For example, our protocol development (*cf.* Burke et al. 2020) suggests that sporulation and mating efficiencies are high in the base population, and that the mortality induced by the weekly outcrossing protocol is low. But it was not possible to quantify these metrics in all 60 experimental replicates every week. If sporulation efficiency, mating efficiency, and/or spore viability was low in a particular replicate during a given week, this would reduce the number of viable cells in the population (compared to the baseline expectation) and lead to a higher estimate of cell doublings for the week. Thus, we feel that our ballpark estimates can be considered rather conservative.

<u>Growth rate assays</u>

Growth in liquid medium was the primary phenotype we tracked in this experiment. We compared growth between evolved populations and the ancestor in each media type to detect

evidence of phenotypic adaptation in a given treatment. We also compared growth of evolved populations across all media types to detect evidence of potential phenotypic trade-offs.

For each population assayed, a culture would first be grown overnight in 10 mL YPD in a shaking incubator (30 °C/200 rpm). Cultures were then quantified via absorbance at OD_{600} (values of 1/100 dilutions typically ranged from 0.10 to 0.15 across all treatments and timepoints) and diluted to a standard starting concentration of OD_{600} = 0.1 using the appropriate media type. These dilutions were then transferred to individual wells of a 96-well plate (Corning) in 200 μL volumes. A Tecan Spark multimodal microplate reader measured absorbance in each well every 30 minutes over a 48-hour period at 30°C with no shaking. Measurements were taken at four positions within each well and the average values were used for subsequent data analysis. To avoid edge effects, wells on the outer perimeter of the plates were not used for assays and simply filled with sterile YPD. In a given plate reader assay, we included 9 technical replicates of the ancestral population and 10 randomly selected replicates from each treatment. Multiple plate reader assays were run under control, moderate ethanol stress, and high ethanol stress conditions. Growth curves were generated using the raw readings, and the R (R CoreTeam 2021) package "Growthcurver" (Sprouffske and Wagner 2016) used to estimate doubling times and carrying capacity. These estimates are obtained by fitting data to the following logistic equation that gives the number of cells N_t (as measured by absorbance) at time t:

$$N_t = \frac{K}{1 + \left(\frac{K - N_0}{N_0}\right) e^{-rt}}$$

where starting population size is represented by N_0 carrying capacity by K. Here carrying capacity is simply defined as the maximum population size in a particular environment. Lastly, r represents the growth rate that would occur if there were no limits on total population size. This value is also used calculate the doubling time which is defined as $\frac{Ln2}{R}$. Figures shown in the manuscript are representative of patterns seen across multiple assays. However, we note that while patterns were consistent across assays, some specific values varied between plate reader runs (e.g. carrying capacity estimates from cycle 15 samples varied between runs under high ethanol conditions, but the observation that carrying capacity in H replicates was lower than replicates from the other treatments remained consistent).

<u>DNA extraction, sequencing, and read mapping</u>

The 20 replicates of each treatment were sampled for pooled-population genome sequencing at three distinct timepoints: after a single outcrossing cycle, after 7 outcrossing cycles, and after 15 outcrossing cycles. To sample populations for sequencing, 1 mL of freezer stocks were revived on YPD agar plates. After 48 hours of growth at 30ºC, the resulting lawns were broadly sampled by wooden applicator (to capture as much genetic diversity as possible) and these cells were inoculated into 10mL of liquid YPD culture, and growth overnight in the

shaking incubator. Genomic DNA was extracted from samples with Qiagen's Yeast/Bact Kit following the manufacterer's protocol. After checking DNA quantity, sequencing libraries were prepared for Illumina sequencing with the Nextera Kit DNA Sample Preparation Kit, implementing some routine modifications to increase throughput (e.g. Baym et al. 2015). Libraries were pooled into groups of 48 and run on at least one PE150 lane of the HiSeq3000 housed at OSU's Center for Quantitative Life Sciences (CQLS); samples with lower-than-average coverages were re-quantified, re-pooled and re-sequenced such that high (>50X average genome-wide) coverages were achieved across all experimental replicates.

The SNP calling pipeline we routinely use has been described previously (e.g. Phillips et al. 2020. We used GATK v4.0 (Van der Auwera & O'Connor 2020) to align raw data to the *S cerevisiae* S288C reference genome (R64-2-1) and create a single VCF file for all variants identified across all populations, using standard best practices workflows and default filtering options. This VCF file was converted into a "raw" SNP frequency table by extracting the AD (allele depth) and DP (unfiltered depth) fields for all SNPs passing quality filters; the former field was used as the minor allele count observed at a presumed SNP, and the latter was used as the total coverage observed at that site. The VCF file was also used as an input for SnpEff v4.3 (Cingolani et al. 2012) to extract potential functional effects of individual SNPs. SnpEff annotates each variant in a VCF file (e.g. by tagging whether it occurs within a protein-coding sequence) and calculates the effect(s) each produces on known genes (e.g. amino acid changes).

We applied several filtering and quality control steps to the raw SNP table generated from the steps above ensure only high-confidence sites were used in subsequent analysis. Mean genome-wide coverage across the 180 samples sequenced for this study ranged from ~50X to ~300X with a median value of ~80X (Supplementary Table 1). We began our filtering process by only considering sites where sequence coverage exceeded 20X in all populations. Next, we removed all sites that were not expected to be polymorphic based on the previously described sequences of the founder strains used to create the ancestral population (*cf.* Phillips et al. 2021). We also removed sites that were not polymorphic in ancestral population itself, as for this study our objective is to track the evolution of standing genetic variants, and not *de novo* mutations. Finally, we only considered sites where the minor allele frequencies between 0.02 and 0.98 across the entire dataset. This removes sites where the alternate nucleotide is fixed across all populations, and sites where errors in sequencing and/or variant calling may create the appearance of polymorphism. These steps ultimately resulted in a SNP table with 61,281 SNPs.

<u>Identifying candidate sites and selection response categories</u>

In pursuing the goal of associating particular candidate SNPs with adaptation, we found it useful to assign these to one of five distinct categories: (i) general adaptation to laboratory handling, (ii) general adaptation to ethanol stress, (iii) specific adaptation to the high ethanol treatment, (iv) specific adaptation to the moderate ethanol treatment, and (v) specific adaptation to the control treatment. Comparing the magnitude of change at sites across these categories allows us to assess whether greater selection intensity results in greater shifts in allele

frequencies, and comparing the location of sites observed across these categories allows us to assess whether or not different selection intensities involve different genomic targets. Notably, the decision to include category (v) above was made *after* observing evidence of continued and unique adaptation in our control treatment; this was an unanticipated outcome that has thematically influenced our interpretation of all other results. To identify and differentiate between these types of responses we used a combination of within- and between-treatment comparisons. For within-treatment comparisons (e.g. comparing SNP frequencies between cycle 1 and cycle 15 for the $C_{1\text{-}20}$ populations), we relied on the Cochran-Mantel-Haenszel (CMH) tests. For between-treatment comparisons (e.g. comparing SNP frequencies between $C_{1\text{-}20}$ and $H_{1\text{-}20}$), we used a generalized linear mixed model (GLMM) approach. Patterns of overlap between these different comparisons were then used to isolate the different types of responses we were interested in (see Table 1 for details on how we define each type of response). As coverage variation can impact statistical results when using nucleotide counts, as both these methods do we scaled coverage across the genome to a uniform level of 50X before performing these tests (Wiberg et al. 2017). Importantly, we chose approaches best-suited to the detection of parallel responses to selection across replicates, as we believe such sites are most likely to be true targets of selection.

The CMH tests for within-treatment comparisons are commonly used in E&R studies, and simulations suggest this is one of the most accurate methods available for identifying responses to selection based on sampling populations over time (Vlachos et al. 2019). Tests were performed using the "poolSeq" package (Taus et al. 2017) in R. For a given within-treatment comparison, the data were paired appropriately (e.g. C_1 cycle 1 with C_1 cycle 15, C_2 cycle 1 with C_2 cycle 15, and so on for the control treatment) and tests were performed at all 61,281 K SNPs. Significant sites show a consistent change in frequency from cycle 1 to cycle 15, across all replicate populations. As we sequenced populations at an intermediate timepoint (cycle 7) we also had the option to use more sophisticated time-series approaches for this task. However, given that simulations (Vlachos et al. 2019) and empirical comparisons (Phillips et al. 2020) have shown that these approaches perform similarly to the CMH test using starting and ending points in an E&R study, we favored the latter approach.

For between-treatment comparisons, as there was no meaningful way to pair replicate populations (e.g. pairing C_1 with H_1 is no more meaningful than C_1 with H_2), the CMH test was not appropriate. We instead opted to use the GLMM approach, which emphasizes parallel changes across replicate populations, and is becoming popular in E&R work (e.g. Jha et al., 2015, Kawecki et al., 2021). Three pairwise comparisons were performed using this approach: $C_{1\text{-}20}$ versus $H_{1\text{-}20}$, $C_{1\text{-}20}$ versus $M_{1\text{-}20}$, and $H_{1\text{-}20}$ versus $M_{1\text{-}20}$ (note: these comparisons only use cycle 15 data). For each comparison, the "lme4" package (Bates et al. 2015) was used to fit a GLMM with binomial error structure to nucleotide counts at a given polymorphic site. The following model statement was used:

$$y_i = Treatment + Replicate(Treatment) + e_i$$

where y_i is the frequency of the *i*th SNP, $Treatment$ is a fixed effect, and $Replicate$ is a random factor nested within $Treatment$, and e_i refers to the binomial error term. To test for significance, the "anova" function in R was used to compare the full model to a simplified model without the fixed effect. Using this approach, we were able to identify SNPs that were consistently differentiated between treatments at the end of the experiment (cycle 15).

To correct for multiple comparisons and establish significance thresholds, we used the same general permutation approach described in Burke et al. (2014) and Graves et al. (2017). Using $C_{1\text{-}20}$ cycle 1 versus $C_{1\text{-}20}$ cycle 15 as an example, the relevant replicate populations were randomly assigned to one of two groups, the CMH tests was performed at each SNP in the shuffled data set, and the smallest p-value generated was recorded. This was then repeated 999 times for a total of 1000 permutations. The "quantile" function in R was then applied to the resulting list of 1000 p-values to define a genome-wide false-positive rate, per site, at 0.5%. That is to say, a Type I error rate of 0.005 is required for a site to be considered significantly differentiated. We used a 0.5% cutoff instead of a 5% cutoff seen in previous efforts based on recent work making the case for 0.005 thresholds instead of 0.05 when interpreting p-values (Benjamin et al. 2018). This same general procedure was used for between-treatment comparisons using the GLMM approach. However, for GLMM comparisons, we only ran 200 permutations instead of 1000 as this method is more computationally intensive than the CMH. We did perform a trial using 1000 permutations for one comparison and found that it produced a very similar threshold.

Once SNPs belonging to each category of selection response listed in Table 1 were identified, we compared these for potential differential enrichment of gene ontology (GO) terms. These lists were created using the SnpEff output: for each response type, we created a list of genes associated with at least one candidate SNP in that category (note: here we consider all effect types as valid associations). Text files with candidate list and associated SnpEff annotation are available through Dryad for interested individuals (see Data Availability statement for details). GO term enrichment analyses for each gene list were then performed using Metascape (Zhou et al. 2019) with default settings.

<u>Sweeps versus shifts</u>

To assess the prediction that more intense selection should be associated with greater changes in allele frequencies, we simply compared the mean changes in SNP frequency at candidate sites between treatments using ANOVAs. If this core idea is correct, among sites associated with general ethanol stress, we would expect to see significantly greater changes in SNP frequency in the high ethanol stress treatment compared to the moderate ethanol stress treatment. Next, to evaluate the prediction that more intense selection should involve signatures of selection more consistent with selective sweeps (larger allele frequency changes at fewer loci) rather than polygenic shifts (smaller allele frequency changes at more loci), we generated genome-wide heterozygosity plots for replicates in each treatment to see if there were signs of

the types of depressions in heterozygosity characteristic of selective sweeps (Burke 2012). If it is the case that more intense selection is associated with more sweep-like dynamics, this should be evident when comparing patterns of genetic variation in the high ethanol stress replicates to the moderate ethanol stress and control replicates.

We defined heterozygosity as *2pq* where *p* and *q* refer to the reference (major) and alternate (minor) nucleotide frequencies. To assess genome-wide patterns, we generated sliding-window plots showing estimates across each chromosome. To do this we first calculated heterozygosity at each polymorphic site in our dataset in each replicate population, then averaged across replicates within a given treatment on a per-site basis for a given timepoint. To define windows, we used the GenWin package (Beissinger et al. 2015) package in R. This package uses a smoothing spline technique to define windows biased on breakpoints in the data. As a "smoothness" parameter is user defined, we tried a range of values before settling on "smoothness = 3000". Our final selection was made based on what most clearly represented the data. Scripts and data to generate plots with more or less smoothness are available for interested parties through Dryad and GitHub (see "Data availability" statement for details).

Results

Adaptation and growth rates

To characterize the long-term consequences of each experimental treatment, we assayed the growth of evolved (cycle 15) replicates in all three media types (control, moderate ethanol, high ethanol) and compared these to the growth of the ancestral population (Figure 1). To assess statistical differences, we used Kruskal-Wallis tests to compare mean values across all groups, and pairwise Wilcoxon rank sum tests to determine which treatment pairs were significantly different (note: the Benjamini-Hochberg procedure was used to correct for multiple comparisons). We find that in plain YPD, differences in doubling time and carrying capacity are small (Figure 1A and B, Supplementary Figure 3A), but significant for both doubling time (Kruskal-Wallis test p=0.0001) and carrying capacity (p= 0.001). In the case of doubling time, there are no significant differences between the C, M, and H replicates. However, the ancestral population has a significantly longer doubling time than evolved populations (mean of 1.3 hours for the ancestor versus 1.25 to 1.26 for C, M, and H populations); this implies that replicates in all treatments adapted to laboratory conditions in ways that improved growth in YPD. For carrying capacity, the H replicates appear to stop doubling at a lower cell concentration compared to the ancestral population, and the C and M treatments (mean of 0.9 in the H replicates versus ~1 for the others, note: carrying capacity is represented as log(OD) as shown in Supplementary Figure 3). While we do find some statistically significant differences here, they are clearly quite small. As such, it appears that adaptation to ethanol stress in the moderate and high ethanol stress has not had a large negative impact on the ability of replicates in these groups to grow in YPD (i.e. no obvious suggestion of an adaptive trade-off).

We do observe differences in growth patterns among treatments in 6% ethanol (Figure 1C), with respect to both doubling time (Figure 1D, Kruskal-Wallis test $p = 1.75 \times 10^{-6}$) and

carrying capacity (Supplementary Figure 3B, Kruskal-Wallis test $p=8.05x10^{-7}$). Both the M and C treatments grow significantly faster than the ancestral population, evidencing adaptation. By contrast, doubling time in the H treatment could not be significantly distinguished from the ancestor, and was correspondingly slower than the C and M treatments. Carrying capacity is significantly different in all pairwise comparisons except for the ancestror vs. C treatment. Carrying capacity is lowest in the H treatment (mean=0.98) followed by the M treatment (mean = 1.07), and the C treatment (mean=1.1) resembled that of the ancestor (mean=1.1).

In populations assayed in 10% ethanol (Figure 1E), we also find significant differences in both doubling time (Figure 1F, Kruskal-Wallis test $p=1.16x10^{-5}$) and carrying capacity (Supplementary Figure 3C, Kruskal-Wallis test $p=5.46x10^{-7}$) among treatments. With respect to doubling time, we find significant differences between all pairwise comparisons except for the M replicates versus the H replicates. Doubling time is slowest in the ancestral population (mean = 3.64 hours), followed by the C replicates (mean = 3.26 hours), followed by the M and H replicates (mean = 2.64 and 2.57 hours). For carrying capacity, all pairwise comparisons are significantly different except for the ancestral population versus the C replicates. The ancestral population has the highest carrying capacity (mean = 1.54, same for the C replicates), followed by the M replicates (1.11) and H replicates (mean = 0.78). So, while populations adapted to ethanol stress treatments generally have faster growth rates under high ethanol stress conditions, they also stop doubling sooner. And adaptation to high ethanol stress specifically is associated with the lowest carrying capacity.

While these growth rate assays are useful for ranking phenotypic differences among ancestral and evolved populations (and they are internally consistent), we do not believe the raw metrics reported necessarily reflect what occurred during the actual selection experiment. For instance, in the plate reader assays, populations in 10% ethanol appear to still be in lag phase after nearly 24 hours of growth (Figure 1E). While we observed that replicates of the H treatment grew more slowly than the other treatments throughout the experiment, routine OD_{600} readings indicate that all populations were in log phase at 24 hours. So, whenever cultures were transferred to fresh liquid media during the experiment, cells were in log phase in all treatments; we felt that it was important to avoid sampling cells at different stages as this could introduce unwanted variation among treatments. There are many reasons why the absolute values of growth parameters estimated during the phenotype assays differ from what was observed in the selection experiment, but we suspect the main drivers are volume differences (200 μL in 96 well plates versus 1mL in 24 well plates), and agitation (no shaking in the plate reader versus 200rpm in the shaking incubator).

<u>Evidence for different response categories among candidate SNPs</u>

To assess the genomic response to selection, we first compared SNP frequencies between cycles 1 and 15 for each treatment using the CMH test (Supplementary Figure 4). Here we find clear responses to selection as evidenced by consistent changes in SNP frequencies in each treatment, including the controls. As the number of candidate regions overlap across both ethanol

stress treatments and controls, it appears that adaptation to general laboratory handling (i.e., our weekly outcrossing protocol) is a major feature of this study. To better parse these responses to selection, we used a GLMM approach to identify differentiated SNPs between treatments during cycle 15 (Supplementary Figure 5). Based on patterns of overlap between all of these comparisons, we identified 5 major response types: (i) general adaptation to laboratory handling, (ii) general adaptation to ethanol stress, (iii) specific adaptation to the high ethanol treatment, (iv) specific adaptation to the moderate ethanol treatment, and (v) specific adaptation to the control treatment. (Table 1). As seen in Figure 2 where these categories are overlaid on the CMH results comparing cycle 1 and 15 for each treatment, almost all significant SNPs can be assigned to one of these 5 categories.

Considering the major peaks in Figure 2, we do observe some mixing of categories within certain peaks, but most are either made up of a single response type or have a clear majority type. And while there are a number of regions that correspond to our three ethanol response types (general, high, and moderate ethanol responses), our strongest signals are observed in regions associated with general laboratory selection. So, while ethanol exposure clearly imposes stress, laboratory conditions appear to be major selective pressures on their own. Here it should be noted that the ancestral population sampled to establish all experimental replicates had already experienced 12 “domestication” cycles with outcrossing prior to this experiment. As such, there was some expectation that these populations would already be adapted to routine culture and outcrossing protocols. A principal component analysis (PCA) of SNP frequencies indicates general differentiation between treatments, and within treatments over time, based on clustering patterns (Figure 3).

The existence of treatment-specific responses addresses our hypothesis that complex genetic architectures (e.g. extensive pleiotropy and/or epistasis) could lead to different targets of selection despite the same core selective pressure. As shown in Figure 2, treatment specific responses can be observed in each group. For instance, there are regions of the genome showing significant responses to selection in the H replicates (Figure 2A) and not the M replicates (Figure 2B) and vice versa despite both being exposed to ethanol stress. We also find cases where significant regions are found in the C replicates (Figure 3A) but not in either the H or M replicates despite all three treatments adapting to the same general laboratory conditions and maintenance protocols.

To address our hypothesis that more intense selection would lead to greater changes in allele frequencies, we first compared the mean change in candidate SNP frequencies per replicate per treatment between cycle 1 and 15 for sites that fall into the general ethanol selection category (Figure 4A). As expected, there is a greater shift in SNP frequencies in the ethanol treatments compared to the controls at these sites. However, we do not find clear evidence of greater changes in SNP frequencies in the high ethanol stress replicates compared to the moderate ethanol stress replicates. We also repeated this approach for the general laboratory selection category. Here we find that while changes in candidate SNP frequencies are not significantly different between the control and moderate ethanol stress selection treatments, shifts are smaller

in the high ethanol stress treatment compared to both. This perhaps suggests some interaction between adaptation for high ethanol stress conditions and general laboratory adaptation.

Observed changes in SNP frequencies between cycle 1 and 15 for treatment-specific responses also suggest complex underlying dynamics. Starting with control-specific responses, we find that these sites show significantly more change in the control treatment versus both ethanol stress treatments (Figure 5A). And, the magnitude of change is much smaller in the high ethanol stress treatment compared to the moderate ethanol stress treatment. Again, we suggest this may evidence a possible interaction between adaptation to ethanol stress and laboratory conditions. We observe the opposite pattern when we focus on the high ethanol stress-specific candidate SNPs (Figure 5B). Here as expected, we see that the change in frequencies in the high ethanol stress SNPs is greater than the other two. These sites appear to shift more and more consistently in the moderate ethanol stress treatment versus the controls as well, perhaps suggesting some weak selection in their favor under moderate ethanol stress. Finally, among the moderate ethanol stress-specific SNPs, we observe greater changes in the moderate ethanol stress sites compared to the other treatments as expected (Figure 5C). However, this change is not significantly different from what was observed in the control treatment, suggesting perhaps that this category also involves weekly selected laboratory selection responses.

<u>SNP frequency trajectories</u>

Given the complex genomic responses to selection observed in each experimental treatment, we examined individual SNP trajectories in each response type in an attempt to deepen our understanding of how selection is acting in our system. For some of the most significant markers in peaks corresponding to our different response categories, we plotted raw allele frequency in each replicate over the three sequenced timepoints. In representative sites associated with general laboratory stress and general ethanol stress, we observe the expected pattern of a clear and consistent increase in frequency in all three treatments for the former but only the two ethanol stress treatments for the latter (Figure 6 and Supplementary Figure 6). However, patterns are more complicated when looking at treatment specific response types.

In representative high ethanol stress-specific sites, we observe the greatest consistent increase in the high ethanol stress replicates (Figure 7A and Supplementary Figure 7A). However, the trajectories in the moderate ethanol stress treatment trend consistently upward to a greater degree than what we see in the control treatment. This again suggests that some high ethanol stress response sites are being weakly selected for in the moderate ethanol treatment. We do not find such clear patterns looking at our representative moderate ethanol stress-specific responses. Here we see the greatest consistent change in the moderate ethanol stress treatment, but changes in the other two treatments show an upward trend as well (Figure 7B and Supplementary Figure 7B). Here one interpretation is that this “moderate ethanol treatment” category is capturing sites that are also beneficial in both or one of the other treatments, but more so under moderate ethanol stress.

Finally, in representative control-specific sites, we again observe evidence of strong directional selection in the control treatment and perhaps weaker selection in the moderate ethanol stress treatment (Figure 7C and Supplementary Figure 7C). However, we find that in the high ethanol stress treatment, SNP frequency changes very little over the course of the experiment and this observation is quite consistent across replicates and independent peaks. This is also consistent with the small changes observed across all sites in the category as shown in Figure 5A (red boxplot). Our interpretation of this result is that these sites might provide significant benefits in terms of adaptation to general laboratory conditions and outcrossing protocols, but these benefits may diminish (or incur costs) as greater ethanol stress is introduced. However, there does not appear to be strong enough selection for them to be completely eliminated, at least on the time scale of this experiment.

Sweeps versus shifts in response to selection

The assess whether more intense selection was associated selective sweeps, we simply visualized patterns of heterozygosity across the genome and looked for evidence of dramatic reductions in genetic diversity (Burke 2012). Specifically, we compared genome-wide levels of variation within each treatment at cycles 1 and 15 to evaluate the effect of 15 cycles of selection on diversity, in a way that was consistent among replicates. Mean heterozygosity across the C, M, and H treatments during cycle 15 does not provide any clear indication of more sweep-like regions with depressed heterozygosity in the ethanol stress treatments versus the controls, or in the high ethanol stress treatment versus the moderate (Supplementary Figure 8B, D, and F). While we certainly observe genomic regions where heterozygosity is low at the end of the experiment, it is important to contextualize this observation relative to the heterozygosity observed at the start of the experiment; in most of these regions heterozygosity was low to begin with (Supplementary Figure 8A vs 8B; 8C vs 8D; 8E vs 8F). This can be visualized more clearly in Supplementary Figure 9 where we plot the absolute difference in mean heterozygosity over time in each treatment. We do find that some of our most significant peaks from CMH comparisons between cycles 1 and 15 are associated with pronounced reductions in heterozygosity (e.g. the regions of chromosome 7 and 11 where heterozygosity dramatically decreased correspond to most significant peaks in Supplementary Figure 4). However, the depth of these depressions does not appear to be associated with intensity of selection. For instance, on chromosome 3 there is a relatively isolated region we believe is responding to ethanol stress in the both M and H but not the C populations (Figure 2). Here we find there is also a large change in heterozygosity seen in the ethanol treatments that is not present in the controls (Supplementary Figure 9). However, it is not the case this region looks like more of a complete sweep in the high ethanol stress treatment than in the moderate ethanol stress treatment (Supplementary Figure 9).

Candidate genes associated with different response categories

To compare potential differences in genes and pathways associated with our major response types (Table 1), we performed gene enrichment analyses using the candidate genes

associated with each category. As shown in Supplementary Table 2, we find do find enrichment terms like “signaling” and “growth” that are shared across categories. However, we also find distinct enriched terms within categories (both broadly and narrowly defined). This lends some creditability to the idea that selection is targeting different biological mechanisms across treatments. For instance, considering terms enriched in the general ethanol selection category, we find “alcohol catabolic process” and “response to organic substance” – terms that intuitively might underlie adaptation to long-term ethanol stress – that are not enriched in the general laboratory selection category. We also find distinct enrichment patterns with respect to treatment-specific categories. The most enriched terms in the high ethanol stress-specific response category are associated with cytokinesis and asexual reproduction, while the most enriched terms in the moderate ethanol stress-specific response category are more associated with pseudohyphal and filamentous growth, as well as metabolic function (e.g. lipid metabolism and sugar homeostasis). Lastly, the most enriched terms in the control-specific response category are related to autophagy and immune function.

Discussion

Using ethanol stress as the focal selective pressure, our experiment was designed to test if stronger selection was associated with greater changes in allele frequencies at target sites, and if targets of selection varied among treatments involving different selection intensities due to complex trait architectures. However, our genomic analysis makes it clear that despite the use of an ancestral population that was previously subjected to a significant period of laboratory domestication (involving a dozen cycles of outcrossing and hundreds of asexual generations), adaptation to laboratory conditions is a major feature of our results (Figure 2 and Supplementary Figure 4C). While this complicates our interpretations, we believe our study still has the power to address our core questions, as the control treatment allows us to distinguish between adaptive responses associated with laboratory domestication and those associated with ethanol stress (Table 1). And while we cannot definitively say ethanol exposure is exerting a greater selective pressure than general laboratory handling, it stands to reason that total selection intensity is highest in the high ethanol stress treatment, followed by the moderate ethanol stress treatment, and lowest in the control treatment.

Patterns of growth in evolved populations imply different avenues of adaptation

The observed growth-related phenotypes in our experimentally-evolved populations suggest complex evolutionary dynamics underlying adaptation to ethanol stress. In YPD (the control medium), replicates of the C, M, and H treatments all have similar doubling times after 15 cycles and all grow slightly faster than the ancestral population (Figure 1). This is perhaps expected given that our genomic results indicate all populations are continuing to adapt to laboratory conditions. However, under moderate and high ethanol stress conditions, patterns are more complicated. Under moderate ethanol stress, replicates of the H treatment have slower growth rates and lower carrying capacities (Figure 1C-D, and Supplementary Figure 3B) than all

other groups assayed. And while they do have higher doubling times than controls and the ancestral population in high ethanol stress conditions (Figure 1E-5), they again have the lowest carrying capacities (Supplementary Figure 3C). Replicates of the M treatment have similar doubling times under high ethanol conditions, but do not display the same level of reduction in carrying capacity (Supplementary Figure 3C). Therefore, we conclude that adaptation to ethanol stress involves mechanisms not directly related to growth rate. Differences between the moderate and high ethanol stress populations across media types also implicate potential trade-offs between growth rate and final sustainable population size.

These observations of potential trade-offs support the idea that pleiotropy and genetic interactions are general features of complex trait architecture. In the case of stress resistance, there is already support for this idea in the literature based on the work of Kawecki et al. (2021). Here the authors find that even though the same genes underlie starvation resistance in in experimentally-evolved *D. melanogaster*, directionally of allele frequency shifts change depending on whether selection is imposed at larval or adult life stages due to underlying trade-offs. More broadly, we do not find it is simply the case that increased selection intensity results in some clear and proportional phenotypic response. As such, selection intensity should be carefully considered when using experimental evolution to predict how specific traits might change in response to a given selective pressure.

<u>The magnitude of allele frequency change at target sites is context-dependent</u>

If increased selection intensity does result in greater changes in SNP frequency at and around targets of selection as simulation work by Christodoulaki et al. (2019) suggest, we would expect that at candidate sites underlying the general response to general stress, changes should be greater in the high ethanol stress treatment compared to the moderate ethanol treatment. In principle, this should also be true for candidate sites associated with general laboratory adaptation; we would predict that changes in the high and moderate ethanol stress treatments should be greater than those observed in the control treatment. With this in mind, our results are mixed when looking across the different response types outlined in Table 1. In the general ethanol selection category, the magnitude of change in candidate SNPs is similar in the high and moderate ethanol stress treatment (Figure 4A). And for the general laboratory selection category, SNP changes in the high ethanol stress treatment are actually smaller than both the control and moderate ethanol stress treatments (Figure 4B). These observations refute the expectation that stronger selection intensity should result in larger frequency changes at target sites. However, our other response types suggest there is perhaps still some merit to this idea.

Among sites we have identified as control-specific and high ethanol-specific, we do observe examples that are strongly selected for in the primary treatment, and weakly selected for in the moderate ethanol stress treatment (Figure 5 and 7, Supplementary Figure 7). In the case of the high ethanol-specific candidates, these perhaps represent instances where more intense selection has in fact produced greater changes in frequencies at target sites. However, interpretation is more complicated for control-specific candidates. While these responses are

presumably adaptative for the selective pressures imposed by experimental protocols common to all three treatments, they appear to be less so with increasing ethanol exposure due to some interaction. All together, we interpret these results as further evidence that different selection intensities can in fact potentiate different trait architectures.

We considered the possibility that in this experiment, adaptation to laboratory protocols may have been stressful enough that the inclusion of ethanol exposure had little additional impact (i.e. all treatments are under very intense selection). This could explain why the general laboratory selection category is accounts for so many of the most significant candidates SNPs we identify. However, growth rate assays make it clear that the ancestral populations and controls are indeed negatively impacted by ethanol exposure (Figure 1). As such, we generally argue that the three experimental treatments experienced distinct and specific selection pressures, and that our results partially support the idea of a positive correlation between selection intensity and allele frequency change.

<u>Stronger selection does not obviously promote selective sweep dynamics</u>

We predicted that stronger selection intensities should produce signatures of adaptation more consistent with selective sweeps (dramatic change observed at fewer loci), while weaker selection intensities should produce signatures of adaptation more consistent with polygenic adaptation (subtle shifts observed at many loci). Patterns of heterozygosity observed across the genome in experimental populations of each treatment do not provide strong support for these predictions. Genomic regions that could be considered most “sweep-like” (i.e. reductions in variation where heterozygosity approaches zero) are rare in all treatments, and regions that best fit this description were generally already low in variation prior to the start of the experiment (Supplementary Figure 8). In the rare regions where we do observe large reductions in heterozygosity associated with one or more of our 5 categories of candidate sites, it is not the case that these regions look more “sweep-like” in one treatment type compared to another; in other words, we find no clear evidence to support the idea that the high ethanol stress is more likely to result in sweep-like patterns than the moderate ethanol or control treatment (the region of reduced heterozygosity mid-way through chromosome 7 provides an illustrative example).

<u>The presence of different selection response categories reinforces that trait architectures vary with selection intensity</u>

As noted previously, we believe the existence of treatment-specific responses to selection across groups indicate that genetic architecture of adaptation varies with selection intensity. For instance, while we do find evidence of “general ethanol selection” responses, there are also a number responses that appear specific to either high or moderate ethanol exposure (red peaks in Figure 2A, orange peaks in Figure 2B). Based on trajectories (Figure 7A and B, Supplementary Figure 7A and B), it is possible that some of these sites are responding to selection in both treatments but how strongly they are favored varies with exposure. However, even with this interpretation, the observed context-dependence supports the idea that trait architecture varies

with selection intensity. This is further supported by the apparent phenotypic trade-offs in the M and H replicates described above.

Surprisingly, the clearest evidence for different trait architectures at different selection intensities comes from our control populations and control-specific responses (Figure 2C). Unlike our moderate or high ethanol specific peaks which are often small or contain SNPs from multiple response categories, control-specific responses include some of our most significant and "pure" peaks. So, while these regions of the genome are strongly associated with the maintenance and laboratory conditions that are shared by all experimental treatments, they appear to be (at most) weakly selected under moderate ethanol stress and not selected at all under high ethanol stress (Figure 7C and Supplementary Figure 7C). Here one possibility is while these alleles do confer advantages under control conditions, they have costs in the presence of ethanol and those costs increase as the level of ethanol exposure increases (i.e. context-dependence due to pleiotropy-associated trade-offs). Alternatively, this pattern could also be explained by negative epistatic interactions that amplify with increasing levels of ethanol stress. While we do not have the ability to distinguish between these two scenarios, both point towards complex relationships between selection intensity and the genetic architecture of adaptative traits.

<u>Selection targets different biological mechanisms across treatments</u>

As shown by our gene enrichment analysis, different biological functions appear to be associated with adaptation across treatments. Focusing on the M and H populations, we find that adaptation to general ethanol stress involves genes related to alcohol catabolic processes, signaling, and responses to abiotic stimuli (Supplementary Table 2). However, we also observe distinct enrichment terms within the M and H categories; while adaptation to high ethanol stress is associated with GO terms relating to the integrity of cell division (e.g. cytokinesis, DNA repair, and budding), these functions may either be unnecessary or come with some cost at lower levels of ethanol stress. Similarly, while adaptation to moderate ethanol stress is associated with GO terms related to metabolic function, these are not implicated at high levels of ethanol exposure. Together, these results provide further evidence that the genetic architecture of adaptation varies with selection intensity due to underlying differences and potential trade-offs in the mechanisms through which adaptation occurs.

With respect to the various mechanisms associated with ethanol resistance in and across the moderate and high ethanol stress treatments, our findings are consistent with past work suggesting ethanol resistance in *S. cerevisiae* is complex and involves many genetic and physiological mechanisms (reviewed in Ding et al. 2009 and Ma and Liu 2010). Signaling, lipid metabolism, carbohydrate homeostasis all stand out as mechanisms implicated by past studies. Work focusing on adaptation to high ethanol exposure in asexual *S. cerevisiae*, in this case peaking at 12% of the culture medium, also specifically implicated mutations in genetic pathways associated with cell cycle and DNA replication (Voordeckers et al. 2015). Here the authors suggest that delayed cell cycle progression allows for greater protection for individual cells. This sort of survival mechanism, which might more accurately be described as stress

tolerance rather than stress resistance, may drive the apparent lower carrying capacity observed in the populations adapted for high ethanol stress in the presence of ethanol. This stands out as an intuitive example of a potential context-dependent adaptative mechanism associated with ethanol stress.

Conclusions

Our results generally support the idea that selection intensity significantly impacts evolutionary dynamics. As such, this should be a major point of consideration when synthesizing results across E&R studies even when they feature the same stress or condition as a focal selective pressure. It should also be considered when attempting to extrapolate findings from E&R studies to natural populations. For instance, candidate sites and dynamics observed in E&R studies with intense selection are perhaps unlikely translate very well to a natural population facing more moderate environmental changes and vice versa. Our observations also indicate a greater need for models of polygenic adaptation to include the possibility of extensive pleiotropy and epistasis than sometimes thought. Furthermore, while here we chose to focus on selection intensity, it is likely that other dynamic circumstances of natural environments, such as temporal heterogeneity in selection, will also affect outcomes in meaningful ways. As such, we believe our findings illuminate a clear need for studies that explicitly demonstrate how experimental and population-genetic factors shape evolutionary dynamics in E&R studies moving forward. Lastly, as with the work of Pfenninger and Foucault (2020), our results demonstrate the importance of appropriate controls even when timeseries data are available. Our controls ultimately prevented us from making a number of spurious associations between large portions of observed genomic responses to selection and ethanol stress resistance.

References

Barghi, N., Hermisson, J., Schlötterer, C. (2020). Polygenic adaptation: a unifying framework to understand positive selection.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.

Baym, M., Kryazhimskiy, S., Liberman, T.D., Chung, H., Desai, M.M., & Kishony, R. (2015). Inexpensive multiplexed library preparation for megabase-sizes genomes. *PLoS One*, 10**,** e0131262.

Beissinger, T.M., Rosa, G.J.M, Kaeppler, S.M., Gianola, D., & de Leon, N. (2015). Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genetics Selection Evolution*, 47, 30 .

Boyle, E.A., Li, Y.I., & Pritchard, J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169, 1177-1186 .

Burke, M.K. (2012), How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proceedings of the Royal Society B*, 279, 5029-5038.

Burke, M.K., Liti, G., & Long, A.D. (2014). Long standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. *Molecular Biology and Evolution, 32*, 3228-3239.

Burke, M.K., McHugh, K.M., & Kutch, I.C. (2020). Heat shock improves random spore analysis in diverse strains of Saccharomyces cerevisiae. *Frontiers in Genetics*, 11, 1636.

Christodoulaki, E., Barghi, N, & Schlötterer, C (2019). Distance to trait optimum is a crucial factor determining the genomic signature of polygenic adaptation. *bioRxiv*, doi.org/10.1101/721340.

Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Ruden, D.M., & Lu, X. (2012). A program for annotating and predictiong the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2, iso-3. *Fly*, 6, 80-92.

Cubillos, F.A., Louis, E.J., & Liti, G. (2009) Generation of a large set of genetically tractable haploid and diploid *Saccharomyces* strains. *FEMS Yeast Research*, 9, 1217-1225.

Ding, J., Huang, X., Zhang, L., Zhao, N, Yang, D., & Zhang, K. (2009). Tolerance and stress response to ethanol in the yeast Saccharomyces cerevisiae. *Applied Microbiology and Biotechnology*, 85, 253-263.

Graves, J.L., Hertweck, K.L., Phillips, M.A., Han, M.V., Cabral, L.G., Barter, T.T., … Rose MR. 2017. Genomics of parallel experimental evolution in Drosophila. *Molecular Biology and Evolution*, 34, 831–842.

Hayward, L.K., & Sella, G. (2021). Polygenic adaptation after a sudden change in environment. bioRxiv, doi.org/10.1101/792952.

Höllinger, I., Pennings, P.S., & Hermisson, J. (2019). Polygenic adaptation: from sweeps to subtle frequency shifts. *PLoS Genetics*, 15, e1008035.

Jha, A.R., Miles, C.M., Lippert, N.R., Brown, C.D., White, KP, & Kretiman, M. (2015). Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in *Drosophila melanogaster*. *Molecular Biology and Evolution*, 32, 2616-2632.

Kawecki, T.J., Erkosar, B., Dupuis, C., Hollis, B., Stillwell, R.C., Kapun, M. (2021). Molecular Biology and Evolution, 28, 2732-2749.

Kessner, D., & Novembre, J. (2015). Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits. *Genetics*, 199, 991-1005.

Linder, R.A., Zabanavar, B., Majumder, A., Shyan-Chiao Hoang, H., Delgado, V., Tran, R., La, V., Leemans, S., & Long, A.D. (2021). bioRxiv: doi.org/10.1101/2021.08.27.457900.

Long AD, Liti G, Luptak A, Tenaillon O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature Reveiews Genetics*, 16, 567-582.

Ma, M., & Liu, Z.L. (2010). Mechanism of ethanol tolerance in *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology*, 87, 829-845.

Pfenninger, M. &, Foucault, Q. (2020). Genomic processes underlying rapid adaptation of a natural *Chironomus riparius* population to unintendedly applied experimental selection pressures. *Molecular Ecology*, 29, 536-548.

Phillips, M.A., & Burke, M.K. (2021). Can laboratory evolution experiments teach us about natural populations? *Molecular Ecology*, 30, 877-879.

Phillips, M.A., Kutch, I.C., Long, A.D., & Burke, M.K. (2020) Increased time sampling in an eolve-and-resequence experiment with outcrossing *Saccharomyces cerevisiae* reveals multiple paths of adaptive change. *Molecular Ecology*, 29, 4898-4912.

Phillips, M.A., Kutch, I.C., McHugh, K.M., Taggard, S.K., & Burke, M.K. (2021). Crossing design shapes patterns of genetic variation in synthetic recombinant populations of *Saccharomyces cerevisiae*. *Scientific Reports*, 11, 19551.

R Core Team. (2021). R: a language and environment for statistical computing. R foundation for Statistical Computing, Vienna, Austria. www.R-project.org.

Schlötterer, C., Kofler, R., Versace, E., Tobler, R., & Franssen, S.U. (2015). Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*, 114, 431-440.

Sprouffske, K. & Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable meterics from microbial growth curves. *BMC Bioinformatics*, 17, 172 (2016).

Stetter, M.G., Thornton, K., & Ross-Ibarra, J. (2018). Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. *PLoS Genetics*, 14, e1007794.

Taus, T., Futschik, A., & Schlötterer, C. (2017). Quantifying selection with pool-seq time series data. *Molecular Biology and Evolution*, 34, 3023-3034.

Van der Auwera, GA., & O'Connor, B.D. (2020). Genomics in the cloud: using Docker, GATK, WDL, and Terra. O'reilly Media.

Visscher, P.M., & Yang, J. (2016) A plethora of pleiotropy across complex traits. *Nature Genetics*, 48, 707-708.

Vlachos, C., & Kofler, R. (2019). Optimizng the power to identify the genetic basis of complex traits with evolve and resequence studies. Molecular Biology and Evolution, 36, 2890-2905.

Vlachos, C., Burny, C., Pelizzola, M., Borges, R., Futschik, A., Kofler, R., & Schlötterer, C. (2019). Benchmarking software tools for detecting and quantifying selection in evolve and resequecing studies. *Genome Biology*, 20, 169.

Voordeckers, K., Komine, J., Das, A., Espinosa-Cantu, A., De Maeyer, D., Arslan, A., Van Pee, M., van der Zande, E, Meert, W., Yang, Y., Zhu, B., Marchal, K., DeLuna, A., Verstrepen, K.J. (2015). Adaptation to high ethanol reveals complex evolutionary pathways. *PloS Genetics*, 11, e1005635.

Wiberg, R.A., Gaggiotti, O.E., Morrisssey, M.B., Ritchie M.G. (2017). Identifying consistent allele frequency differences in studies of stratified populations. Methods in Ecology and Evolution, 8, 1899-1909.

Data Availability

The raw sequence files generated over the course of this project are available through NCBI SRA (BioProject ID: PRJNA839395) and scripts used to process raw data and perform SNP calling are available through Github (https://github.com/mollyburke/Burke-Lab-SNP-calling-pipeline). Core scripts necessary to reproduce our results are available through Github (https://github.com/mphillips67/Genomics-Ethanol-Stress-Outcrossing-Yeast) and major input and results files are available through Dryad (https://doi.org/10.5061/dryad.tqjq2bw27).

Author Contributions

M.A.P. and M.K.B. conceived of this project. M.A.P., R.K.B., M.S., and S.Z. did the lab work necessary to generate the data sets used in this study. M.A.P. carried out the data analysis, and M.A.P. and M.K.B. wrote the manuscript.

Acknowledgements

We thank Oregon State University's Center for Quantitative Life Sciences for use of their computational and sequencing resources. This work was supported by startup funds provided to M.K.B. by the College of Science at Oregon State University, and M.A.P. was supported by a National Science Foundation Postdoctoral Fellowship (NSF 1906246).

Figures

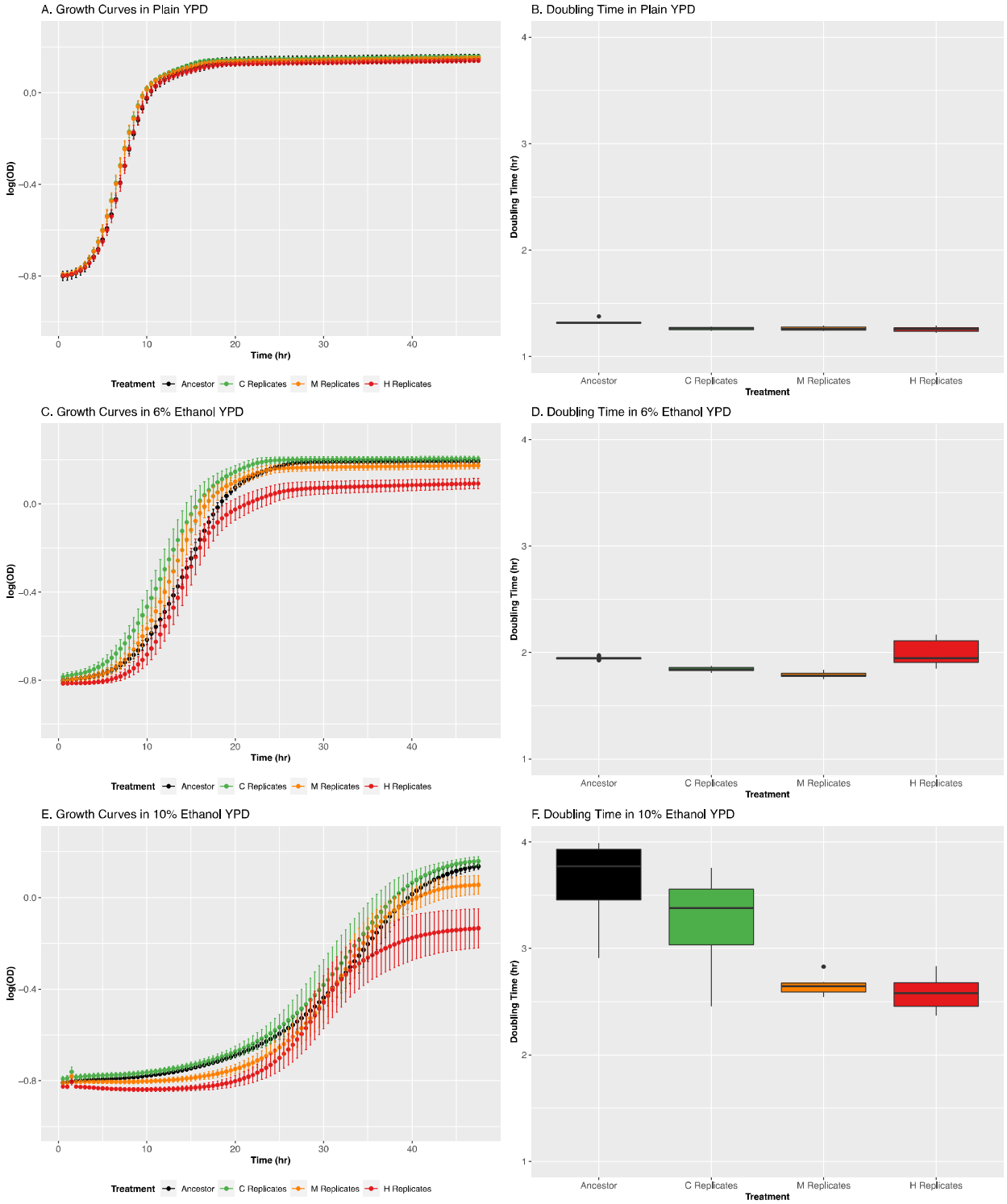


Figure 1. Growth rates and doubling times for ancestral population and experimental populations after 15 cycles of adaptation in plain YPD (A and B), 6% ethanol YPD (C and D), and 10% ethanol YPD (E and F). Nine technical replicates for ancestor, and 10 randomly chosen replicates from each treatment were used in these assays.

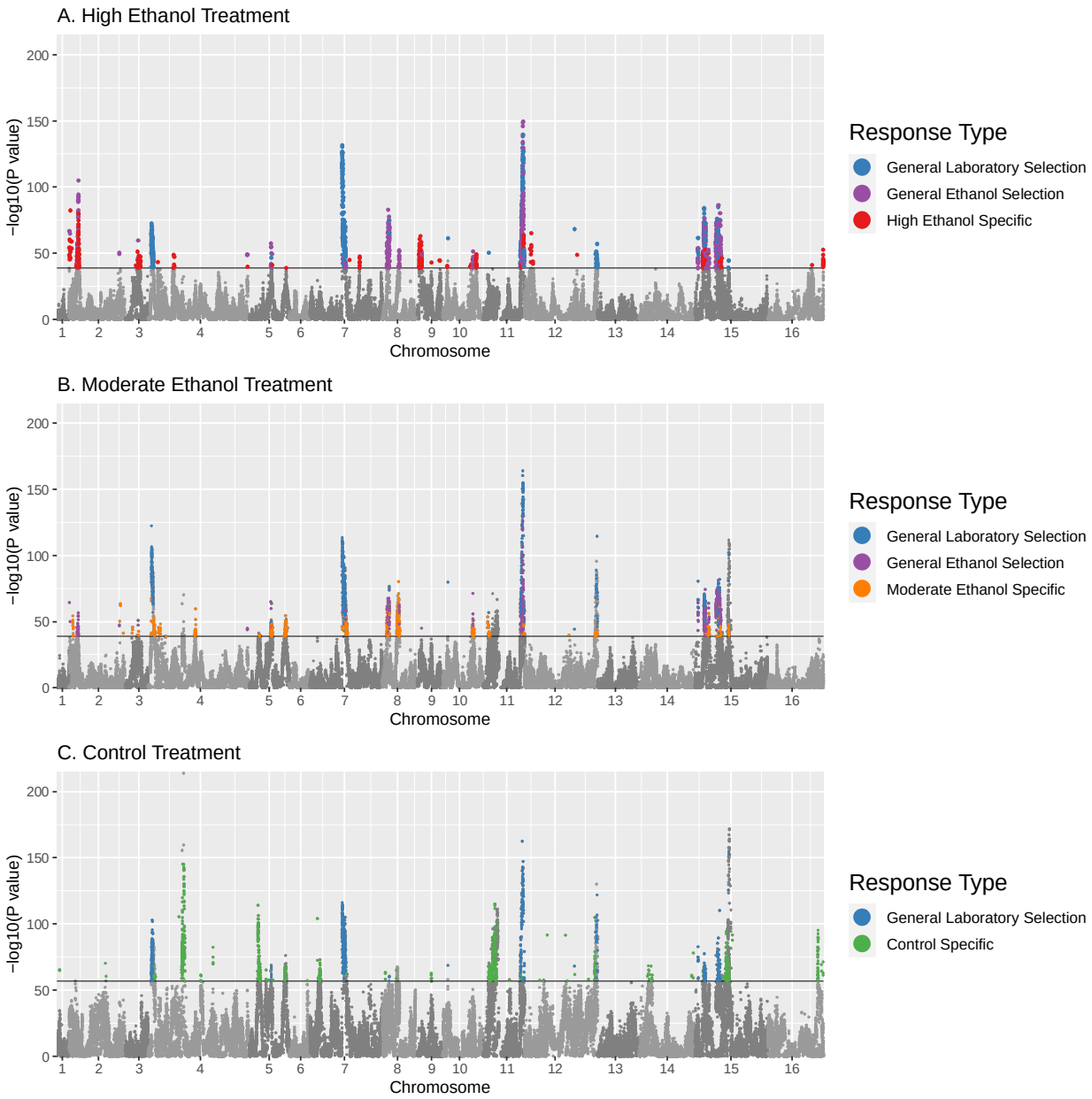


Figure 2. CMH results comparing SNP frequencies between cycles 1 and 15 in (A) High ethanol stress, (B) Moderate ethanol stress, and (C) Control treatments. Black lines represent permutation derived significance thresholds, and color coding of significantly differentiated sites corresponds response categories described in Table 1.

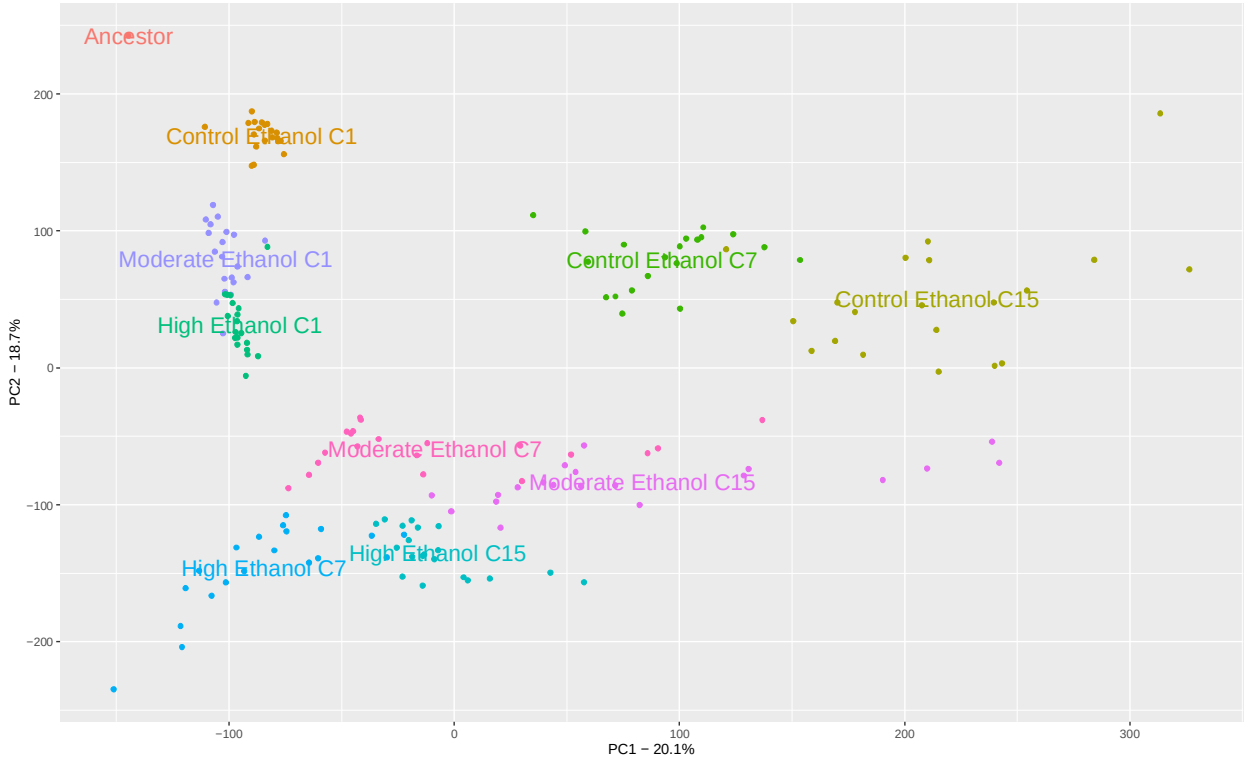


Figure 3. PCA plot showing how different time points for replacement population in each treatment cluster based on SNP frequencies.

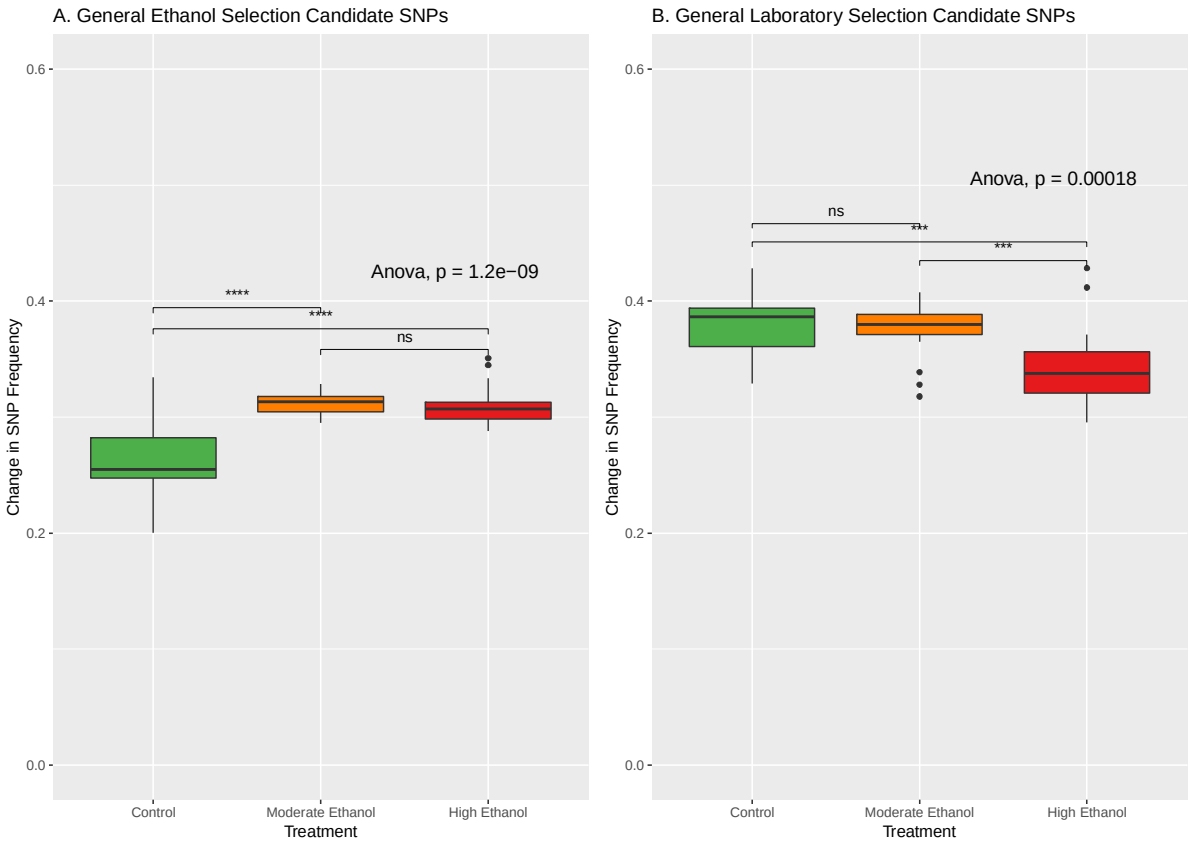


Figure 4. Boxplots showing the absolute mean change in SNP frequency for each replicate in each treatment for (A) General ethanol response candidate SNPs and (B) General Laboratory Selection Candidate SNPs.

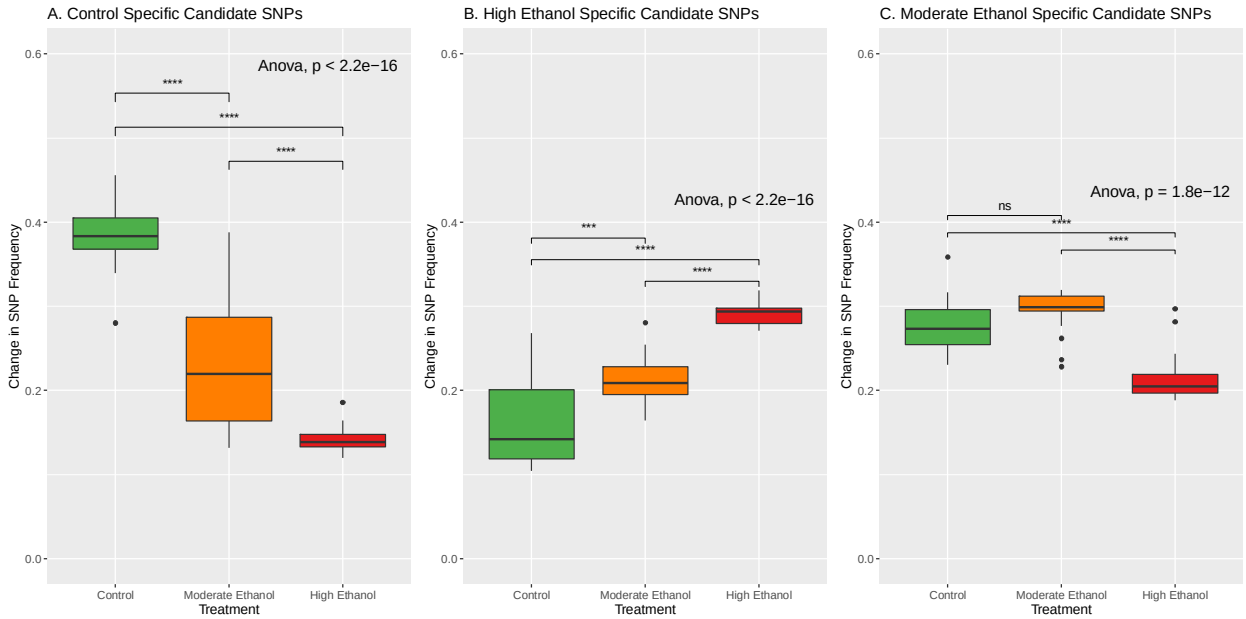


Figure 5. Boxplots showing the absolute mean change in SNP frequency for each replicate in each treatment for (A) Control Specific Candidate SNPs, (B) High Ethanol Stress Candidate SNPs, and (C) Moderate Ethanol Stress Candidate SNPs.

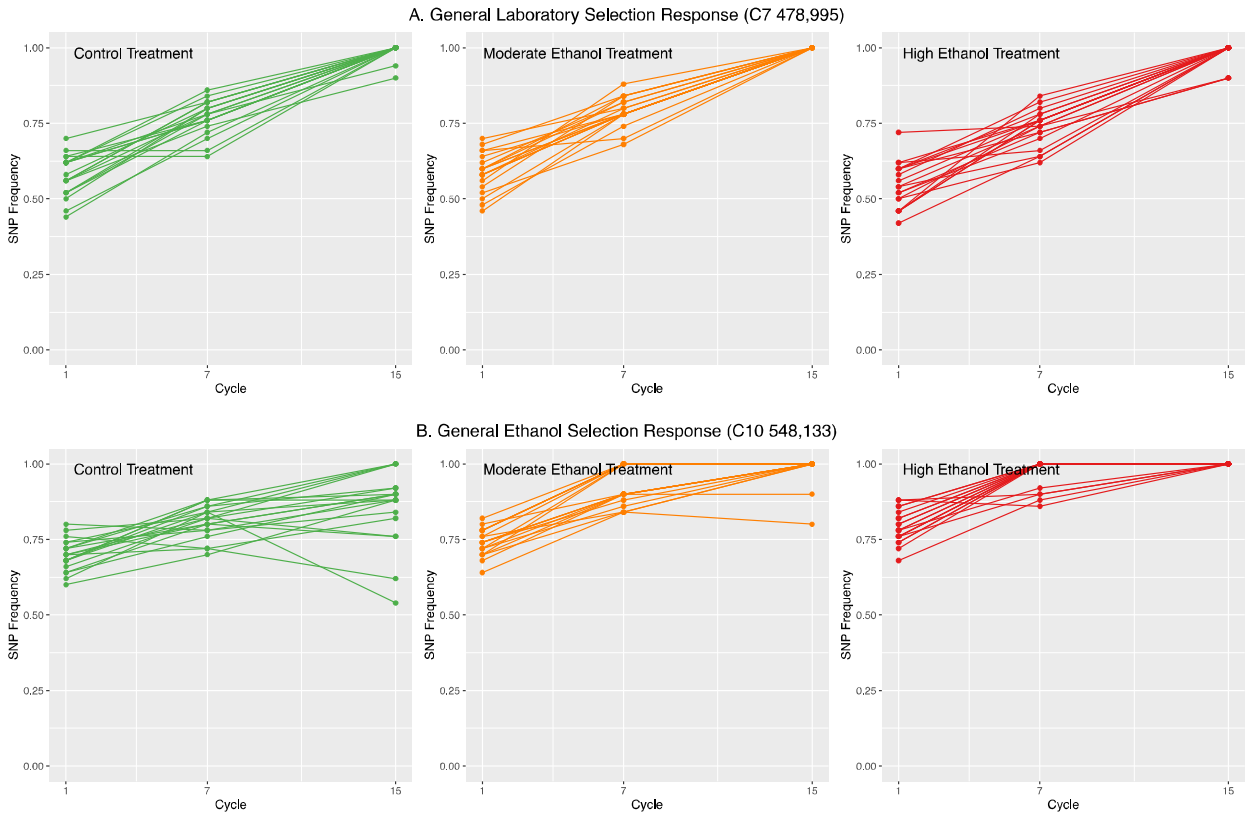


Figure 6. SNP frequency trajectories for representative (A) General Laboratory Selection response site, and (B) General Ethanol Selection response sites. Trajectories for all 20 replicates within each treatment are plotted.

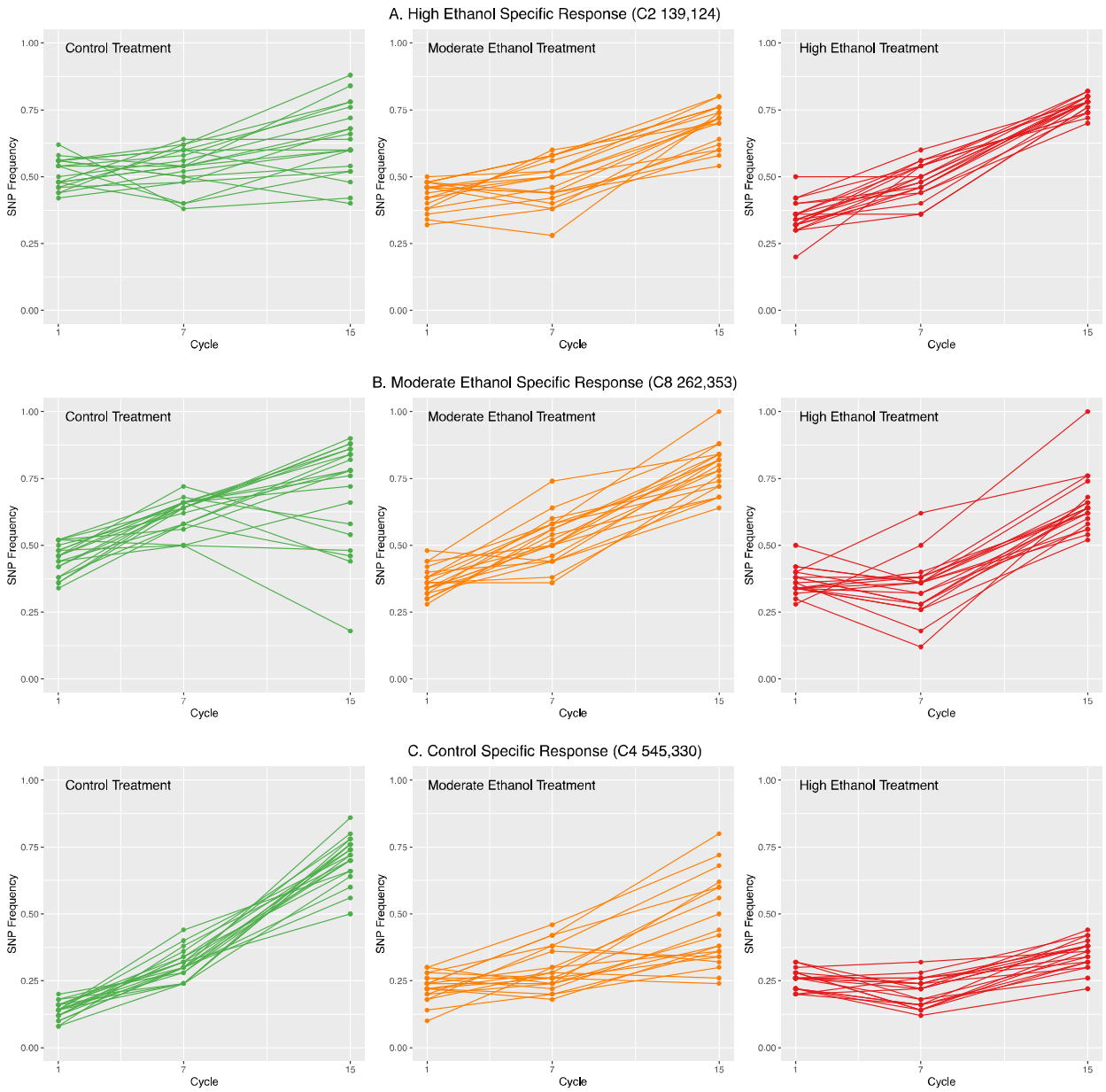


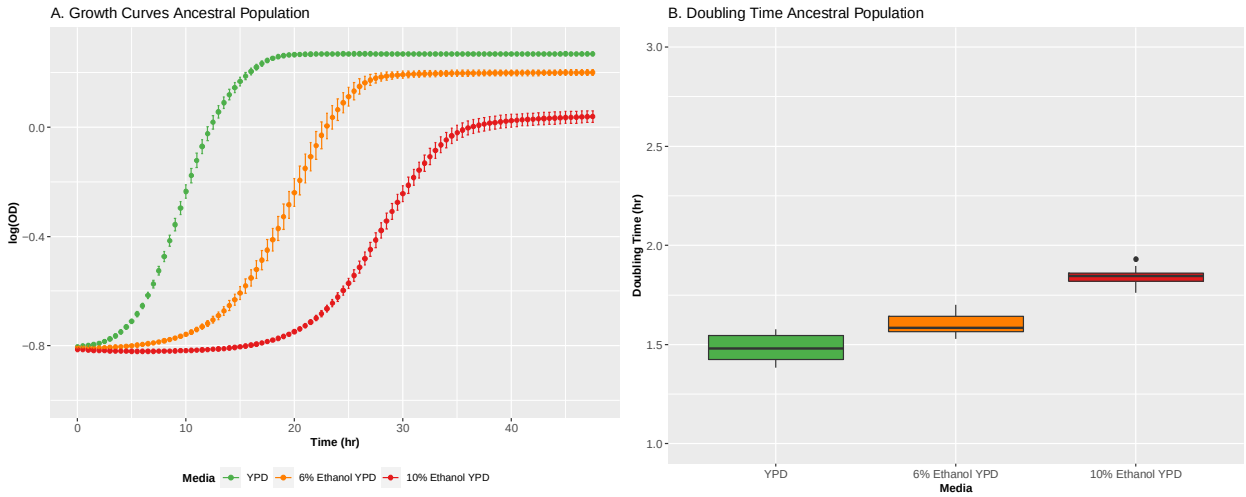
Figure 7. SNP frequency trajectories for representative (A) High Ethanol Stress Specific response site, (B) Moderate Ethanol Stress Specific response, and (C) Control Specific response sites. Trajectories for all 20 replicates within each treatment are plotted.

Tables

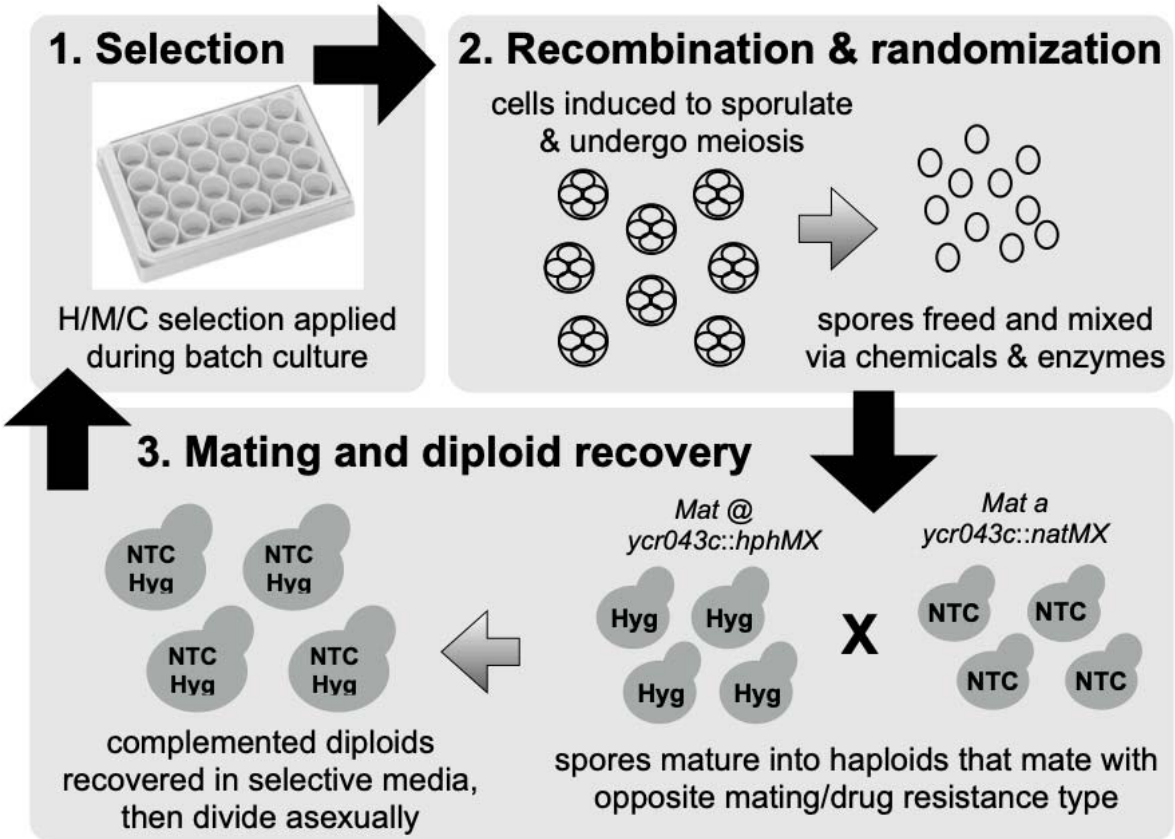
Table 1. The five major response types used to categorize our candidate sites and how they are defined based on how results overlap across within and between treatment statistical comparisons.

RESPONSE TYPE	DEFINITION
GENERAL LABORATORY SELECTION	Significantly differentiated sites that overlap across all three treatments based on comparisons between cycles 1 and 15 for each treatment. Sites that are differentiated between Controls versus High and Moderate Ethanol Stress treatments are also removed to account for directionality.
GENERAL ETHANOL SELECTION	Significantly differentiated sites that overlap across High and Moderate Ethanol Stress treatments based on comparisons between cycles 1 and 15 for each. Any sites shared with Control cycle 1 versus 15 are removed, along with sites that are differentiated between High versus Moderate Ethanol stress during cycle 15.
HIGH ETHANOL STRESS SPECIFIC RESPONSE	Significantly differentiated sites that are unique to High Ethanol Stress treatment based on cycle 1 versus cycle 15 comparisons (i.e. removing overlap with Control and Moderate Ethanol Stress cycle 1 versus 15 comparisons).
MODERATE ETHANOL STRESS SPECIFIC RESPONSE	Significantly differentiated sites that are unique to Moderate Ethanol Stress treatment based on cycle 1 versus cycle 15 comparisons (i.e. removing overlap with Control and High Ethanol Stress cycle 1 versus 15 comparisons).
CONTROL SPECIFIC RESPONSE	Significantly differentiated sites that are unique to Control treatment based on cycle 1 versus cycle 15 comparisons (i.e. removing overlap with High and Moderate Ethanol Stress cycle 1 versus 15 comparisons).

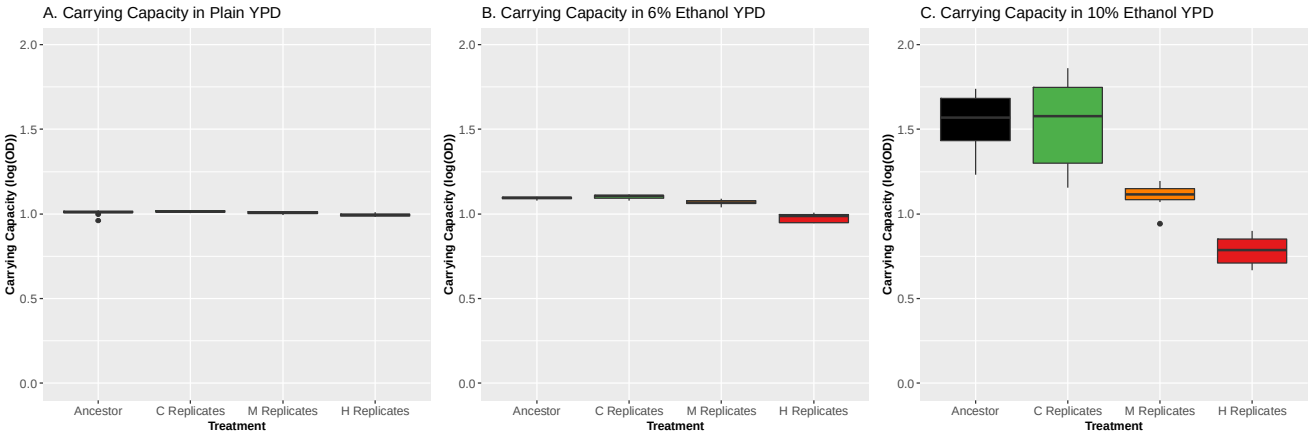
Supplementary Figures



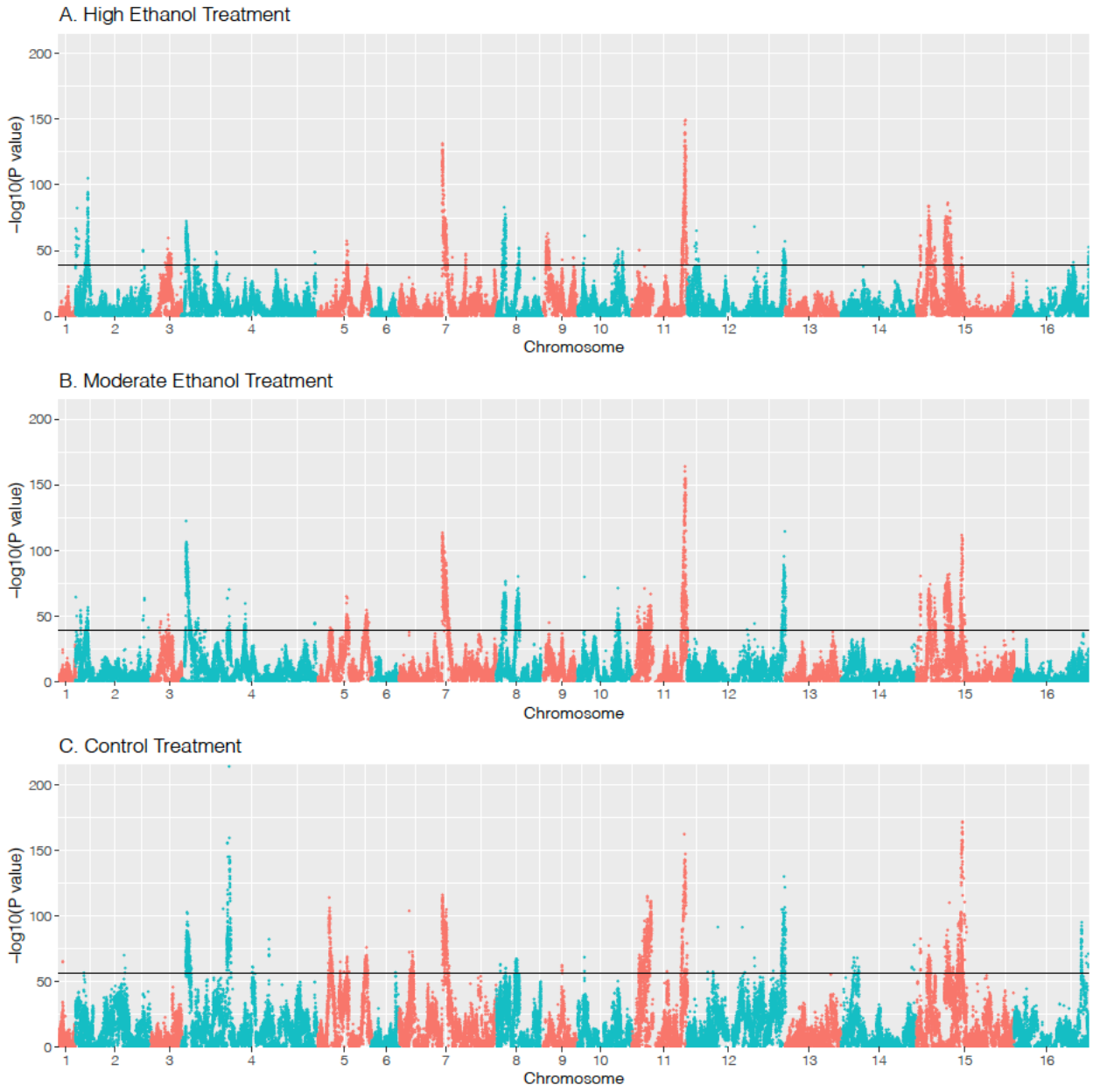
Supplementary Figure 1. Growth rates (A) and doubling times (B) for ancestral population in plain YPD, 6% ethanol YPD, and 10% ethanol YPD.



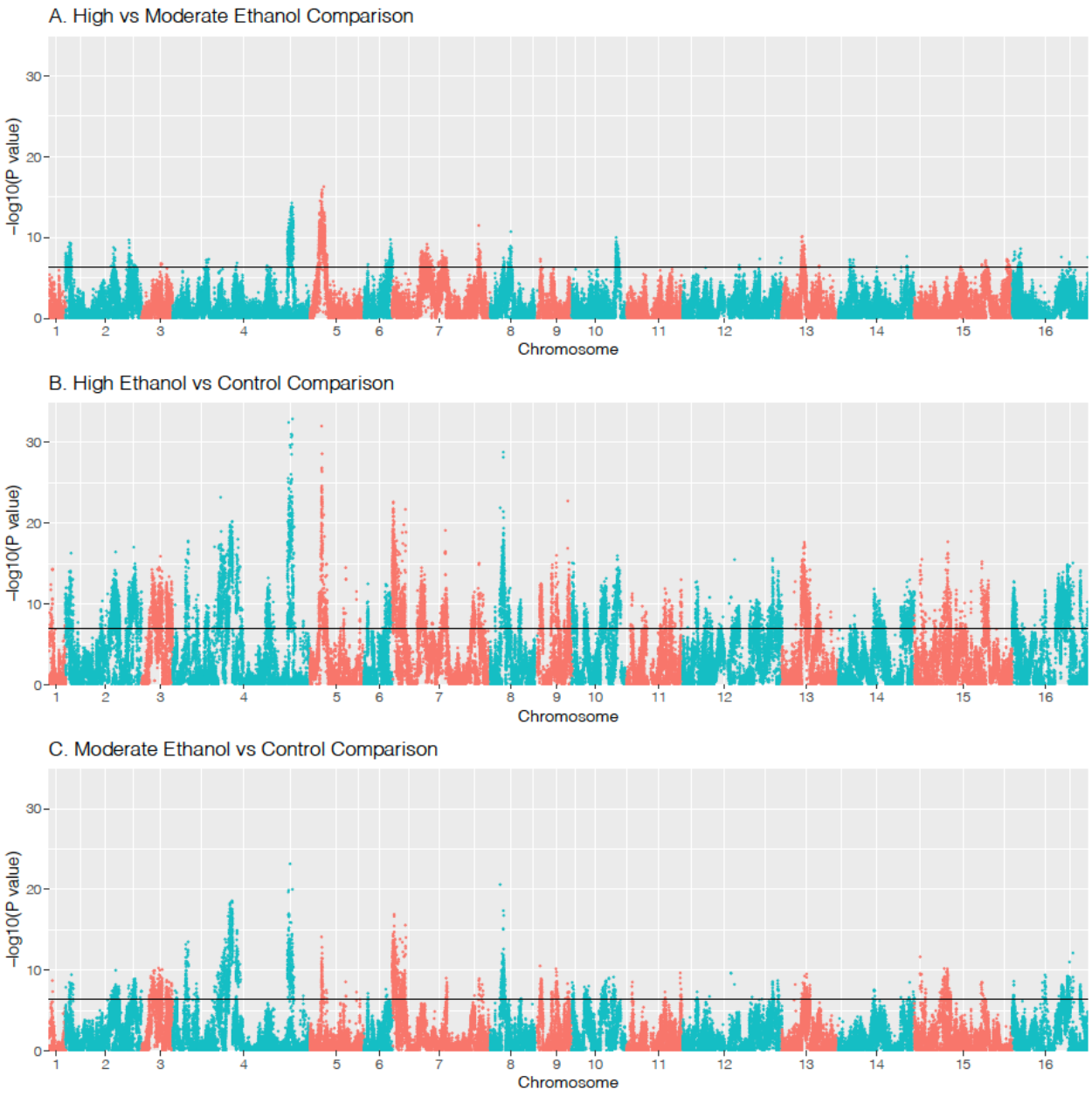
Supplementary Figure 2. General schematic showing how the H, M, and C populations were maintained over the course of this study.



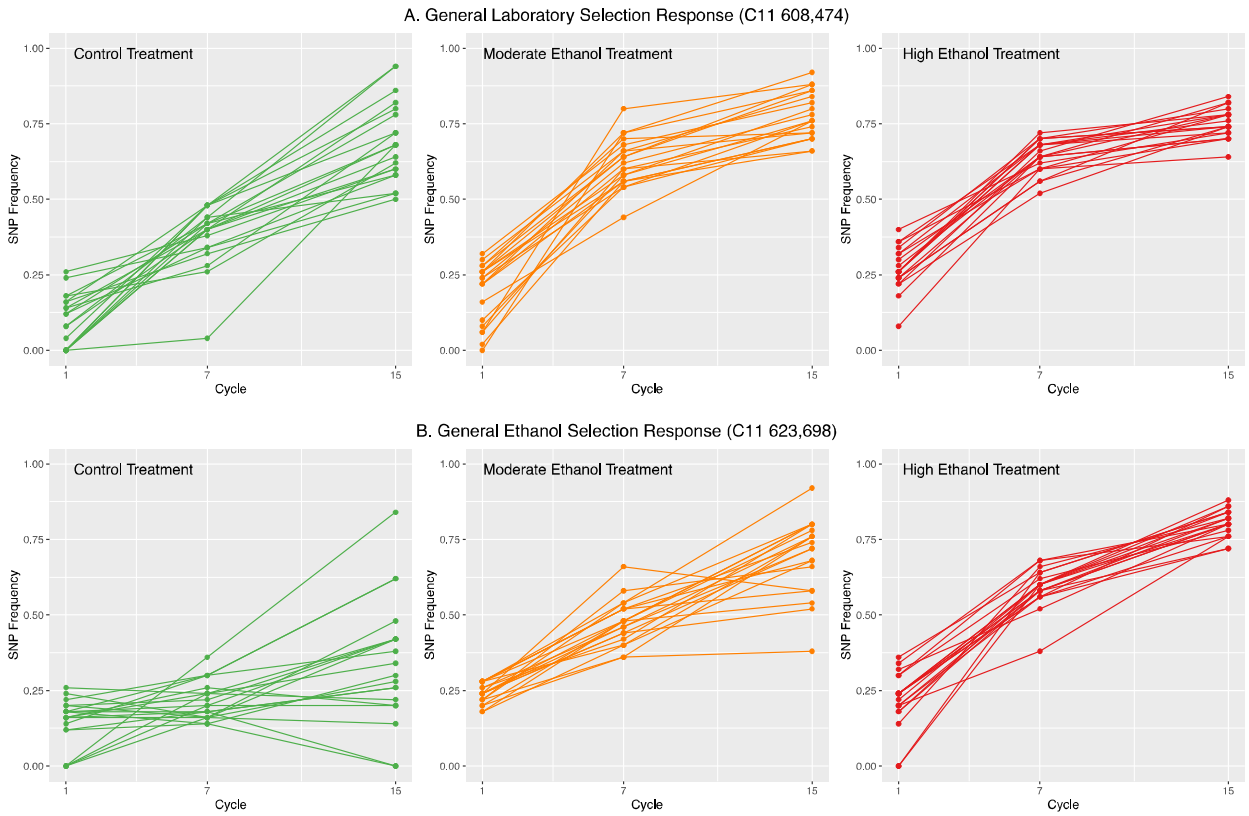
Supplementary Figure 3. Carrying capacity estimates based on growth rate assays for ancestral population and experimental populations after 15 cycles of adaptation in plain YPD (A), 6% ethanol YPD (B), and 10% ethanol YPD (C). Nine technical replicates for ancestor, and 10 randomly chosen replicates from each treatment were used in these assays.



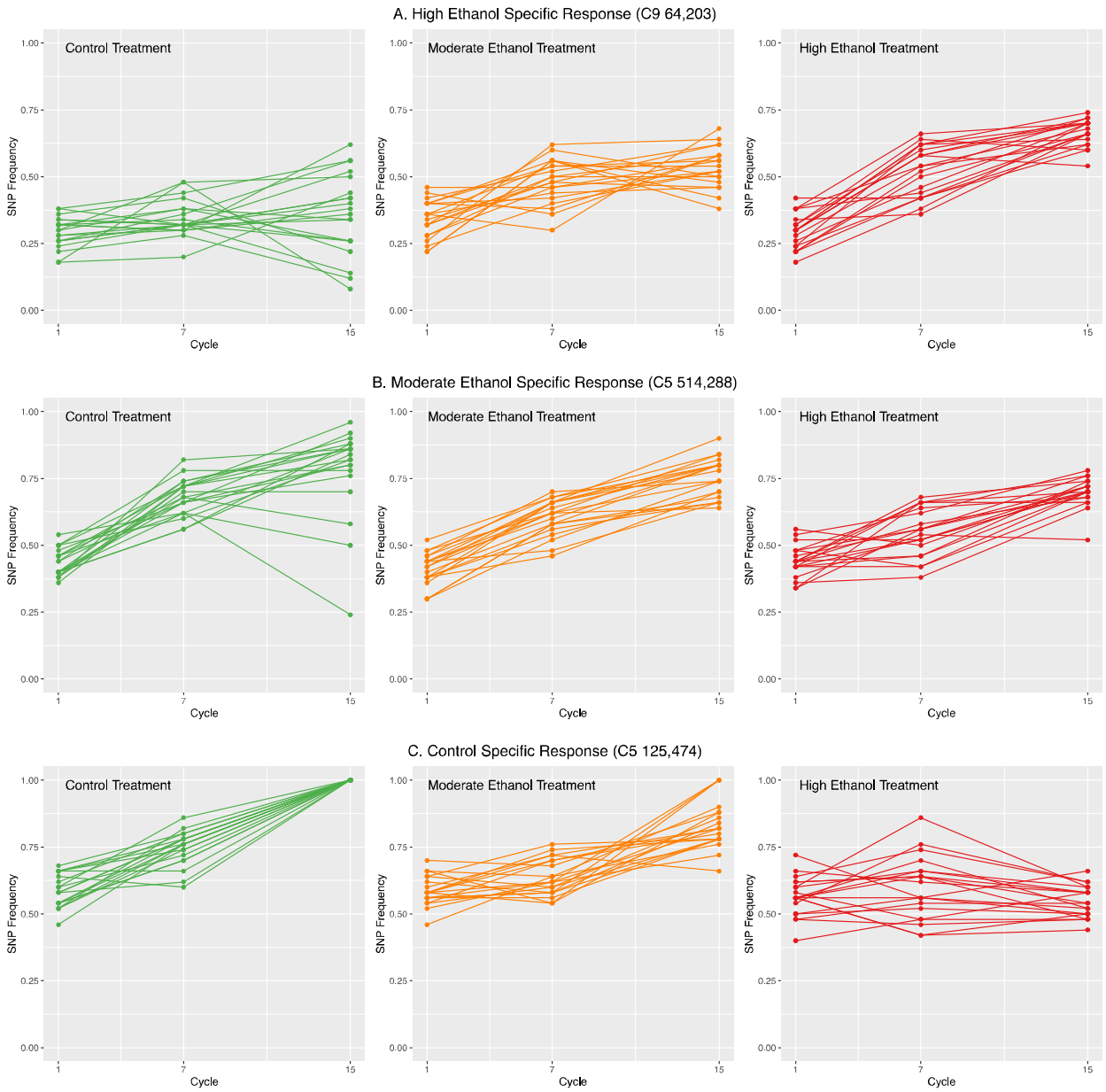
Supplementary Figure 4. CMH results comparing SNP frequencies between cycles 1 and 15 in (A) High ethanol stress, (B) Moderate ethanol stress, and (C) Control treatments. Black lines represent permutation derived significance thresholds.



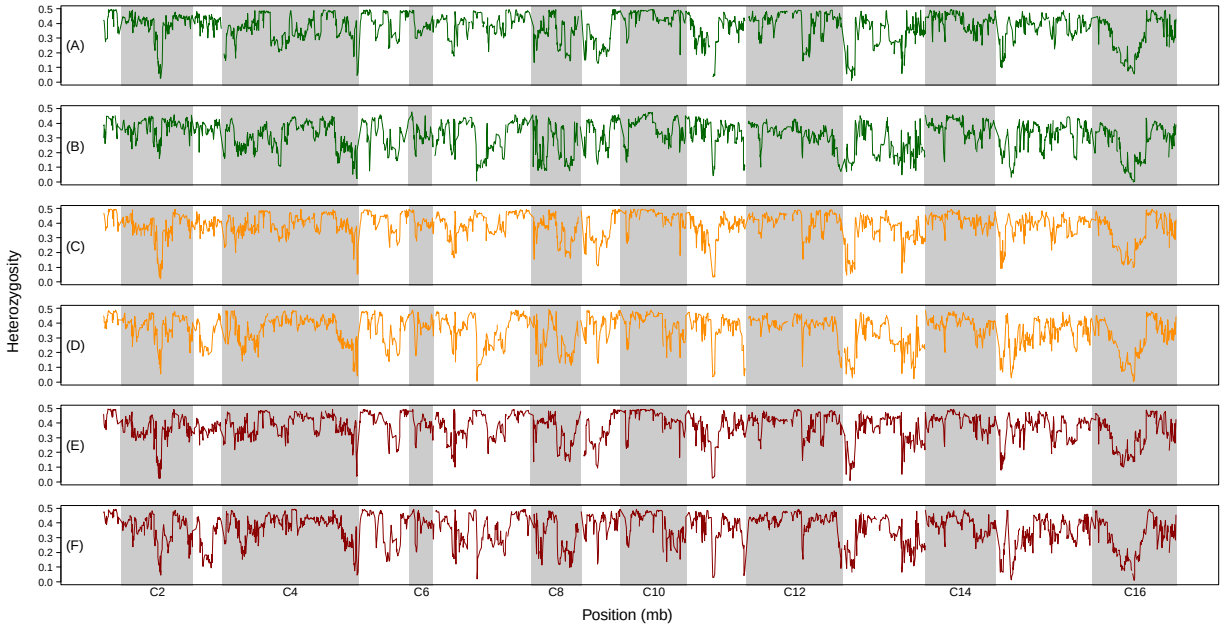
Supplementary Figure 5. GLMM results from comparing SNP frequencies cycle 15 SNP frequencies between (A) High and Moderate Ethanol Stress treatments, (B) High Ethanol Stress and Control treatments, and (C) Moderate Ethanol Stress and Control treatments. Black lines represent permutation derived significance threshold.



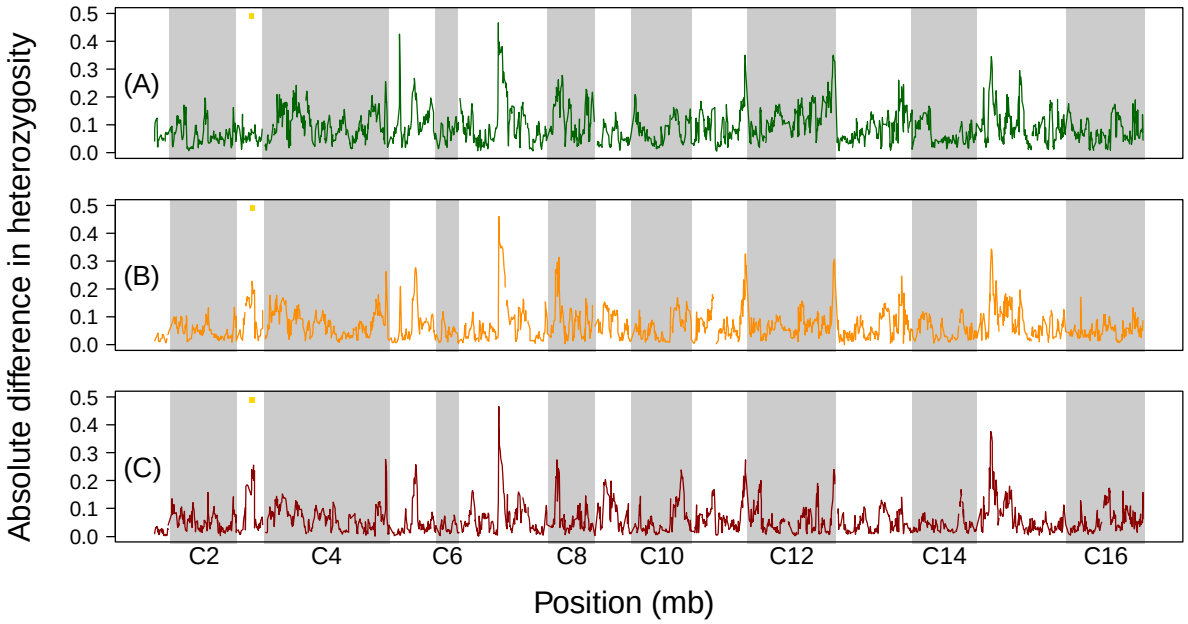
Supplementary Figure 6. SNP frequency trajectories for representative (A) General Laboratory Selection response site, and (B) General Ethanol Selection response sites. Trajectories for all 20 replicates within each treatment are plotted.



Supplementary Figure 7. SNP frequency trajectories for representative (A) High Ethanol Stress Specific response site, (B) Moderate Ethanol Stress Specific response, and (C) Control Specific response sites. Trajectories for all 20 replicates within each treatment are plotted.



Supplementary Figure 8. Mean heterozygosity for $C_{1\text{-}20}$ replicates (A and B), $M_{1\text{-}20}$ replicates (C and D), and $H_{1\text{-}20}$ replicates (E and F) plotted across the genome during cycles 1 and 15. Panels A, C, and E were generated cycle 1 data, while B, D, and F are from cycle 15.



Supplementary Figure 9. Absolute difference in mean heterozygosity between cycles 1 and 15 for $C_{1\text{-}20}$ replicates (A), $M_{1\text{-}20}$ replicates (B), and $H_{1\text{-}20}$ replicates (C) plotted across the genome. The gold square over chromosome 3 in each panel represents a region where there are candidate SNP's in both ethanol treatments, but not in the controls.

Supplementary Tables

Supplementary Table 1. Average sequence coverage for each timepoint of each replicate sequenced.

Replicate	Cycle	Mean Coverage
Ancestor	-	88.38
C1	1	53.00
	7	78.78
	15	144.82
C2	1	57.18
	7	89.32
	15	53.21
C3	1	56.26
	7	76.21
	15	118.98
C4	1	137.77
	7	95.03
	15	87.49
C5	1	104.24
	7	138.28
	15	91.45
C6	1	62.48
	7	150.01
	15	115.99
C7	1	59.82
	7	108.10
	15	96.40
C8	1	86.30
	7	163.51
	15	130.89
C9	1	151.69
	7	108.04
	15	51.33
C10	1	121.45
	7	193.92

	15	54.42
C11	1	72.61
	7	202.69
	15	54.11
C12	1	76.27
	7	50.01
	15	88.65
C13	1	131.62
	7	93.09
	15	117.77
C14	1	97.99
	7	51.87
	15	68.71
C15	1	80.67
	7	91.39
	15	74.77
C16	1	59.72
	7	92.11
	15	65.09
C17	1	108.88
	7	69.87
	15	73.76
C18	1	58.25
	7	75.39
	15	69.35
C19	1	74.23
	7	82.11
	15	60.47
C20	1	65.85
	7	92.91
	15	73.25
M1	1	148.45
	7	75.50
	15	63.77
M2	1	151.08
	7	68.35

	15	126.24
M3	1	202.20
	7	70.62
	15	151.19
M4	1	88.88
	7	180.83
	15	48.66
M5	1	93.68
	7	143.10
	15	146.67
M6	1	97.64
	7	62.12
	15	138.83
M7	1	148.94
	7	90.99
	15	293.71
M8	1	153.92
	7	75.30
	15	118.00
M9	1	65.18
	7	94.02
	15	70.62
M10	1	161.58
	7	89.92
	15	83.01
M11	1	95.15
	7	69.57
	15	126.96
M12	1	77.64
	7	133.60
	15	106.45
M13	1	107.26
	7	72.31
	15	134.66
M14	1	118.18
	7	107.43

	15	143.51
M15	1	87.80
	7	57.45
	15	81.69
M16	1	127.69
	7	100.79
	15	170.28
M17	1	89.98
	7	108.87
	15	100.55
M18	1	71.52
	7	76.33
	15	59.61
M19	1	85.46
	7	110.75
	15	74.14
M20	1	101.33
	7	94.21
	15	58.27
H1	1	99.97
	7	137.61
	15	92.27
H2	1	102.67
	7	133.16
	15	78.03
H3	1	86.47
	7	149.63
	15	182.16
H4	1	104.77
	7	96.76
	15	197.48
H5	1	80.32
	7	78.75
	15	108.65
H6	1	82.64
	7	77.99

	15	173.67
H7	1	50.45
	7	62.33
	15	48.20
H8	1	54.53
	7	68.14
	15	55.64
H9	1	71.14
	7	71.95
	15	85.80
H10	1	88.05
	7	80.73
	15	78.34
H11	1	62.50
	7	93.30
	15	119.56
H12	1	59.88
	7	69.80
	15	96.67
H13	1	78.81
	7	92.58
	15	65.30
H14	1	98.72
	7	134.69
	15	117.62
H15	1	61.00
	7	69.39
	15	64.24
H16	1	47.84
	7	123.63
	15	81.27
H17	1	61.66
	7	161.66
	15	103.26
H18	1	102.09
	7	134.52

	15	125.07
H19	1	74.00
	7	145.69
	15	174.77
H20	1	73.81
	7	97.08
	15	175.79