bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 **Title:** Crossing design shapes patterns of genetic variation in synthetic recombinant populations
2 of *Saccharomyces cerevisiae*
3
4 **Authors:** Mark A. Phillips[1*], Ian C. Kutch[1], Kaitlin M. McHugh[1], Savannah K. Taggard1, and
5 Molly K. Burke[1*]
6
7 **Affiliations:**
8 [1] Department of Integrative Biology,
9 Oregon State University
10 Corvallis, OR 97331
11
12 [*]Corresponding authors: Mark A. Phillips (philmark@oregonstate.edu), Molly K. Burke
13 (molly.burke@oregonstate.edu)
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

32 **Abstract**
33 "Synthetic recombinant" populations have emerged as a useful tool for dissecting the genetics of
34 complex traits. They can be used to derive inbred lines for fine QTL mapping, or the
35 populations themselves can be sampled for experimental evolution. In latter application,
36 investigators generally value maximizing genetic variation in constructed populations. This is
37 because in evolution experiments initiated from such populations, adaptation is primarily fueled
38 by standing genetic variation. Despite this reality, little has been done to systematically evaluate
39 how different methods of constructing synthetic populations shape initial patterns of variation.
40 Here we seek to address this issue by comparing outcomes in synthetic recombinant
41 *Saccharomyces cerevisiae* populations created using one of two strategies: pairwise crossing of
42 isogenic strains or simple mixing of strains in equal proportion. We also explore the impact of
43 the varying the number of parental strains. We find that more genetic variation is initially present
44 and maintained when population construction includes a round of pairwise crossing. As perhaps
45 expected, we also observe that increasing the number of parental strains typically increases
46 genetic diversity. In summary, we suggest that when constructing populations for use in
47 evolution experiments, simply mixing founder strains in equal proportion may limit the adaptive
48 potential.
49
50
51
52
53
54
55
56
57
58
59
60
61
62

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

63 **Introduction**

64 Despite major advances in DNA sequencing technologies, deciphering the genetic basis
65 of complex traits remains a major challenge in modern biology. Efforts to fully characterize the
66 genetic variants underlying traits like human height have had limited success[1], and there is no
67 clear consensus regarding what models best describe the genetic architecture of complex traits as
68 a whole[2]. Alongside traditional QTL mapping and genome-wide association studies, the "evolve-
69 and-resequence" (E&R) approach has emerged as powerful technique that can generate insight
70 into the genetic basis of complex phenotypes, at least in model organisms[3-5]. This approach
71 involves subjecting experimental populations to conditions that target a specific trait, and simply
72 monitoring the genotypic and phenotypic changes that occur over generations. As these
73 experiments typically feature replicate populations evolving in parallel, researchers can use the
74 data they generate to make powerful statistical associations between genomic variants and traits
75 of interest.

76 In order to most appropriately model the evolution of complex traits in humans, E&R
77 studies should employ sexually-reproducing eukaryotic organisms[4,6]. However, despite the
78 promise of the E&R approach for dissecting complex traits, there are still unanswered questions
79 regarding what constitutes optimal experimental design. To that end, simulations have assessed
80 the effects of particular experimental parameters on the outcomes of E&R experiments designed
81 to model the evolution of complex traits[7-9]. These simulations support intuitive practices such as
82 maximizing replication and experimental duration, and they recommend maximizing levels of
83 starting genetic variation in experimental populations (i.e. by including more founder
84 haplotypes). Our work here aims to determine how to best achieve this recommendation in
85 experiments with outcrossing *Saccharomyces cerevisiae*.

86 Most E&R experiments with sexually-reproducing model organisms use *Drosophila*[10-19].
87 However, outcrossing *S. cerevisiae* populations have emerged as an attractive alternative due to
88 shorter generation times, ease of replication, the ability to freeze and "resurrect" samples taken
89 from different timepoints, and comparable genetic resources[20-21]. As such, we have chosen to
90 focus our efforts on determining how to best construct populations of outcrossing *S. cerevisiae*
91 for use in E&R studies with respect to maximizing standing genetic variation and haplotype
92 diversity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

93 As *S. cerevisiae* populations isolated from natural or industrial settings typically lack
94 genetic heterogeneity, generating laboratory populations with standing genetic variation requires
95 crossing distinct isogenic strains. These "synthetic recombinant" or multiparent populations have
96 traditionally been used for QTL mapping[22-23], but they can also serve as the ancestral population
97 in evolution experiments[20-21]. Notably, synthetic recombinant populations have also been used
98 for E&R in other model organisms such as *D. melanogaster* (e.g the *Drosophila* Synthetic
99 Population Resource *cf.* King et al.[24]) and *Caenorhabditis elegans* (e.g. the *C. elegans*
100 Multiparent Experimental Evolution panel, *cf.* Teotónio et al.[25]). In this context, having clearly
101 defined founder genotypes enables the estimation of haplotype frequencies in evolved
102 populations from "Pool-SEQ" data[20,23,26]. And this ability to detect haplotypes responding to
103 selection, as opposed to individual markers, allows for better characterization of linkage and
104 genetic hitchhiking in evolved populations[27] than is possible in studies where founder genotypes
105 are unknown (e.g. Graves et al.[17]). All of this said, to our knowledge little has been done to
106 explicitly assess how to best generate a recombinant population for downstream use in an
107 evolution experiment. While simulations predict that maximizing the number of founding
108 haplotypes and levels of genetic variation will improve experimental outcomes, there are
109 practical and biological limits to the number of haplotypes that can be combined into a single
110 recombinant background and effectively maintained. One of our goals here is to evaluate the
111 costs and benefits of increasing the number of founding haplotypes in a synthetic recombinant
112 population for E&R.
113 Another of our goals is to evaluate how different methods of crossing founder strains
114 impact levels of genetic variation in synthetic recombinant populations. There are two general
115 approaches to this task across model organisms. In the more common approach, individuals
116 sampled from various lines are simply mixed in equal proportions and allowed to mate freely
117 prior to the start of the experiment (e.g. Bhargi et al.[19]). With this strategy, the genetic makeup of
118 the resulting population will depend on mating efficiency between individuals from different
119 founding strains, reproductive output, and chance. However, work with *Drosophila* has shown
120 there is no significant allele frequency differentiation between independently constructed
121 synthetic populations using the same sets of isofemale lines[28]. Such a high level of
122 reproducibility could indicate that mating efficiency and differences in reproductive output
123 among founding lines are negligible, or it could indicate that these differences exist and shape

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

124 the genetic makeup of recombinant populations in a parallel manner. So, while this approach is
125 relatively simple and practical to implement, it may be difficult to maximize the amount of
126 genetic variation and founding haplotype representation in synthetic populations when there are
127 drastic differences in reproductive output and/or mating efficiencies among founding strains. In
128 other words, if there is substantial variation in mating efficiencies among founding strains, this
129 could lead to the over-representation of certain haplotypes belonging to strains that mate most
130 efficiently. By contrast, the second approach for constructing synthetic populations is more
131 complex and involves some level of pairwise crossing between founders[22-25,29]. While this
132 approach is significantly more labor-intensive and time-consuming, it perhaps has advantages in
133 terms of producing populations that have more equal founder haplotype representation and,
134 consequently, higher levels of genetic variation.

135 Here we assess which of these methods produces populations most well-suited for E&R
136 studies in outcrossing *S. cerevisiae*. We have constructed sets of synthetic populations using both
137 approaches and the same founder populations. In addition to crossing approach, we also evaluate
138 how the number of founder populations used impacts the genetic makeup of the resulting
139 population. Specifically, we created sets of populations using 4, 8 and 12 isogenic strains and
140 both construction methods. Our objective is to characterize patterns of genetic variation and
141 haplotype diversity across these populations to provide recommendations on how to best
142 construct synthetic recombinant populations in outcrossing yeast for the specific application of
143 using them in an E&R study.
144

145 Materials and Methods

146 *Population creation*

147 All yeast strains used in this study originated from heterothallic, haploid, barcoded
148 derivatives of the SGRP yeast strain collection[30]. A subset of 12 of these haploid strains,
149 originally isolated from distinct geographic locations worldwide, were used to create the
150 synthetic populations we describe here (See Supplementary Fig. S1 for phylogeny). These 12
151 strains were all genetically modified as described in detail by Linder et al.[23] to enable easy
152 crossing and diploid recovery; these modified strains were kindly provided by Anthony D. Long
153 (UC Irvine) in 2017. Briefly, these strains were modified so that *MAT*a and *MAT*α strains both
154 contain *ho* deletions to prevent mating-type switching, but each contain a different drug-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

155 resistance marker in a pseudogene (YCR043C) tightly linked to the mating type locus (*MAT*a,
156 *ho*Δ, ura3::KANMX-barcode, *ycr043C*::NatMX and *MAT*α, *ho*Δ, ura3::KANMX-barcode,
157 *ycr043C*::HygMX). These genotypes enable haploids of each mating type to be recovered using
158 media supplemented with either hygromycin B or nourseothricin sulfate, and they enable newly
159 mated a/α diploids to be recovered in media supplemented with both drugs.

160 Two different crossing strategies were used to create genetically diverse populations
161 using 4, 8, and 12 strains as founders (Table 1). "K-type" populations (named for what we call
162 the "kitchen sink method", or the practice of pooling isogenic strains together without careful
163 focus on representation) were created by simply pooling equal volumes of saturated overnight
164 cultures of the respective haploid founders and allowing those cells to mate. To accomplish this,
165 single colonies of each haploid founder strain were sampled and grown overnight (at 30ºC / 200
166 rpm) in 1 mL of rich media consisting of 1% yeast extract, 2% peptone, and 2% dextrose (YPD).
167 After ~24 hours, cultures were washed in fresh YPD media, pooled with the relevant other
168 overnight cultures in a 50 mL conical tube, vortexed, and now-mixed cultures were allowed to
169 settle and mate for 90 minutes at room temperature. These cultures were then transferred in
170 200uL aliquots to agar plates containing 100mg/mL nourseothricin sulfate ("NTC"), 300mg/mL
171 hygromycin B ("hyg") as well as 200mg/mL G418; this strategy ensured that that only newly
172 mated diploids would grow. The resulting lawns of mated diploid cells were collected by
173 scraping with a sterile glass slide into a fresh YPD media. This "cell bank" was archived as a
174 frozen stock at -80ºC for each of the K-type populations made with 4, 8, and 12 haploid founders
175 (4K, 8K, and 12K respectively).

176 "S-type" populations (named S due to the manipulation of spores to achieve better
177 representation of founder genotypes) were built to mimic more careful crossing designs in which
178 founding lines are crossed in pairs and/or a round-robin. To accomplish this, each haploid strain
179 was paired with a different strain of the opposite mating type and mated as described above.
180 Successful diploid colonies were isolated, grown overnight in 1 mL of YPD, washed and
181 resuspended in sporulation media (1mL 1% potassium acetate), then cultured for 72 hours at
182 30ºC / 200 rpm. Tetrads from these diploid cells were dissected using the Spore Play dissecting
183 microscope (Singer). The four meiotic products (spores) were then collected, allowed to grow for
184 2 days, and replica plated to plates containing either NTC or hyg to verify the proper segregation
185 of drug resistance markers and thus mating types. Once validated, the meiotic products were

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

186 grown overnight in 1 mL of YPD. Overnight cultures were standardized to the same optical
187 density (OD_{600}) before being pooled in equal volumes in a 50 mL conical tube. Populations were
188 given 90 minutes to mate at room temperature, then were plated on agar plates supplemented
189 with both NTC/hyg/G418 so that only newly mated diploid cells could grow. The resulting lawns
190 of mated diploid cells were collected by scraping with a sterile glass slide into fresh YPD media.
191 This "cell bank" was archived as a frozen stock at -80ºC for each of the S-type populations made
192 with 4, 8 and 12 founders (S4, S8, and S12 respectively; see Supplementary Fig. S2 for crossing
193 schematics).
194
195 *Population maintenance and 12 cycles of outcrossing*
196 After the creation of the 3 "K-type" and 3 "S-type" synthetic recombinant populations
197 described above, all populations were taken through 12 consecutive cycles of intentional
198 outcrossing; in other words, the populations were subjected in parallel to a series of steps that
199 induced regular sporulation, spore isolation, and mating. Detailed methods are described by
200 Burke et al.[31]. Briefly, newly mated diploid cells from the last step of the population creation
201 protocol (i.e. the "cell banks") were grown overnight in 10 mL YPD media. These cultures were
202 washed and resuspended in sporulation media, and incubated with shaking for 72 hours
203 (30°C/200 rpm). Cells then underwent a number of methods to disrupt asci and
204 isolate/randomize spores, including incubation with Y-PER yeast protein extraction reagent
205 (Thermo) to kill vegetative diploids, digestion with 1% zymolyase (Zymo Research) to weaken
206 ascus walls, and well as high-speed shaking with 0.5 mm silica beads (BioSpec) to mechanically
207 agitate the asci. After these steps spores were resuspended in 10 mL YPD and allowed to settle
208 and mate for 90 minutes at room temperature. Diploids were recovered as described above;
209 cultures were transferred to 10 individual YPD agar plates supplemented with NTC/hyg/G418 in
210 200 µL aliquots and incubated at 30°C for 48 hours. The resulting lawns of mated diploid cells
211 were collected by scraping with a sterile glass slide into fresh YPD media. This "cell bank" was
212 again sampled for archiving at –80°C, and used to initiate an overnight culture for the next
213 outcrossing cycle. We estimate that 15–20 asexual generations occurred between every
214 outcrossing cycle of the experiment. Based on counting colonies from dilutions of cultures plated
215 at various benchmarks during the protocol, we expect that 7.5–11 generations elapse during the
216 overnight culture in YPD media, and another 7.5-11 generations elapse during the period of

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

217 diploid recovery on agar plates. Thus, a minimum of 15*12 = 180 cell doublings likely took
218 place over the 12 cycles of outcrossing in each synthetic recombinant population.
219
220 *Genome sequencing and SNP identification*
221 Each of the recombinant K- or S-type population was sequenced at three specific
222 timepoints: initially (we also call this timepoint "cycle 0"), after 6 cycles of outcrossing ("cycle
223 6"), and after 12 cycles of outcrossing ("cycle 12"). We also sequenced each haploid founder
224 strain such that we could estimate the relative contributions of each to the recombinant
225 populations. Each of the founding SGRP strains were plated as haploids on plates containing
226 either NTC or hygromycin to verify the presence of the appropriate drug resistance markers.
227 Individual colonies were isolated from each strain for verification of identifying barcodes at the
228 URA3 locus using Sanger sequencing (Cubillos et al.[30] provide barcode and primer sequences).
229 Once validated, single colonies were again isolated for whole-genome sequencing. One milliliter
230 of YPD media was inoculated with single colonies, grown overnight, and the resulting culture
231 was harvested for gDNA extracted using the Qiagen Puregene Yeast/Bact. Kit. Purified gDNA
232 from each haploid founder was then prepared for sequencing using the Nextera DNA Sample
233 Preparation Kit (Illumina). Some minor modifications to the manufacturer's protocol were
234 implemented to optimize throughput (*cf*. Baym et al.[32]). Genomic DNA libraries were prepared
235 for experimental recombinant populations in the same way and all samples were pooled to
236 generate a single multiplexed library. Because the recombinant (i.e. genetically variable)
237 populations require significantly higher coverage to accurately estimate allele frequencies at
238 variable sites, these populations were added to the library at 10X the molarity of each haploid
239 founder sample. The multiplexed library was run on two SE150 lanes on the HiSeq3000 at the
240 OSU Center for Genomic Research and Biocomputing (CGRB). Data for the 4S populations
241 were previously published in Burke et al.[31] and raw fastq files are available through NCBI SRA
242 (BioProject ID: PRJNA678990). Raw fastqs for all other populations are also available through
243 NCBI SRA (BioPorject ID: PRJNA732717).
244 We have developed a processing pipeline for estimating allele frequencies in each
245 population directly from our pooled sequence data. We used GATK v4.0[33-34] to align raw data
246 to the *S. cerevisiae* S288C reference genome (R64-2-1) and create a single VCF file for all
247 variants identified across all replicate populations, using standard best practices workflows and

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

248 default filtering options. We also downloaded and indexed a reference VCF file with SNP
249 information for a number of distinct natural isolates of S. cerevisiae[35]; this is a recommended
250 best practice for calibrating base quality with GATK v4.0. This VCF file was converted into a
251 SNP frequency table by extracting the AD (allele depth) and DP (unfiltered depth) fields for all
252 SNPs passing quality filters; the former field was used as the minor allele count and the latter
253 was used as the total coverage. The python scripts used to generate and convert VCF files to
254 tables suitable for downstream analyses in R (www.R-project.org) are available through GitHub
255 (see Data Availability statement for details on where to find all major scripts used to process and
256 analyze data).

257 Our general SNP analysis strategy involved portioning the data to create three separate
258 SNP tables with each table corresponding to a set of founders and populations derived from them
259 (e.g. a table containing with the S4 and K4 populations and their founders). In each table, we
260 chose to only include sites with a minimum coverage >20X in the in synthetic populations as a
261 quality control measure. Next, sites were filtered based on data from the founder populations.
262 We excluded all sites that appeared to be polymorphic within a given founder, and sites where a
263 single nucleotide was fixed across all founders. This was done as such occurrences could
264 indicate sequencing error given that our founder strains are haploid and isogenic, and a site is
265 unlikely to be polymorphic in our synthetic populations if it is fixed across all of the founders.
266 After these filters were applied, we retained a collection of high-quality SNPs in each population
267 to subject to further analysis. The total number of SNPs identified in each population is given in
268 Table 1, and the average genome-wide coverage (i.e. depth of sequence coverage) of each
269 population is given in Supplementary Table S1. All populations had mean coverages >50X with
270 all but one population (S4 cycle 0) having greater than 70X mean coverage (Supplementary
271 Table S1).

272

273 *SNP Variation*

274 Our main objective was to evaluate how crossing strategy and the number of founder
275 strains impacts patterns of SNP variation in synthetic recombinant populations. To that end, we
276 assessed SNP-level variation in our recombinant populations using several metrics. First, we
277 simply determined the number of polymorphic sites segregating in each population immediately
278 following their creation (cycle 0), and monitored how that number changed over time ((i.e. after

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

279 6 or 12 outcrossing cycles). This approach of tracking the total number of SNPs should reveal
280 whether particular crossing strategies – i.e. using a certain number of founders, and/or one of the
281 two crossing strategies – consistently produced populations with more SNPs, and whether these
282 SNPs were maintained or lost over 12 outcrossing cycles. We also generated UpSet plots using
283 the UpsetR package[36] in R to visualize patterns of overlap between the total number of SNPs
284 possible for a given combination of founder strains, and the SNPs we observed in our actual
285 populations. We define the total number of possible SNPs as all loci for which at least one of the
286 founding strains used has an allele different from the others; this number will therefore differ
287 among the 4-way, 8-way, and 12-way crosses.

288 In addition to SNP number, we also characterized the distribution of SNP frequencies in
289 each population, which allows more direct comparisons between populations with different
290 numbers of founders but the same crossing strategy, or the same number of founders but
291 different crossing strategy. To do this, we focused on two metrics: the site frequency spectrum
292 (SFS), and genome-wide heterozygosity. Here heterozygosity refers to $2pq$, the product of the
293 reference (i.e. the S288C allele) and alternate allele frequency at a given site multiplied by 2. In
294 addition to looking at differences in mean genome-wide heterozygosity between populations, we
295 also generated sliding window plots showing patterns of variation across each chromosome. To
296 define windows, we used the GenWin package[37] in R with the following parameters:
297 "smoothness = 6000, method = 3." GenWin itself uses a smoothing spline technique to define
298 windows based on breakpoints in the data. While we ultimately used "smoothness = 6000", we
299 did initially try a range of values. Our final selection was made based on what most clearly
300 represented trends in the data. For interested parties, plots with more or less smoothness can be
301 easily generated using data and scripts we have made available through Dryad and Github (See
302 "Data availability" statement for details).

303 It is worth noting that our ability to assess levels of genetic variation across our synthetic
304 populations is limited by the fact we have only collected Pool-SEQ data. Given the complex life-
305 history of the yeast populations in this experiment, which involves periods of 7-15 generations of
306 asexual growth punctuated by discrete outcrossing events, it is not possible for the genotypes of
307 all individuals in the population to be shuffled by recombination every generation. Therefore,
308 asexual lineages will evolve by clonal interference for relatively short periods of time, until the
309 next outcrossing event decouples individual adaptive alleles from a particular genetic

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

310 background. It is possible that during these periods of clonal interference, particular diploid
311 lineages will dominate, and if these lineages are heterozygous at a given locus, that will lead to
312 an artificially elevated heterozygosity value at that SNP. But we do not believe this is a major
313 complication in the understanding of nucleotide diversity in our experiment for several reasons;
314 namely, that our outcrossing protocol includes several measures that maximize outcrossing
315 efficiency (i.e. any asexual diploids that fail to sporulate are killed), the generally high rate of
316 recombination in yeast, and that the periods of asexual growth are short and unlikely to exceed
317 ~20 cell doublings.
318
319 *SNP frequency changes over 12 cycles of outcrossing*
320 Although statistical power in this power is limited due to a lack of replication, we
321 attempted to identify regions of the genome showing obvious responses to selection in each
322 synthetic population. Specifically, we used Pearson's χ^2 test as implemented in the poolSeq[38]
323 package in R to compare SNP frequencies between cycle 0 and cycle 12 in each synthetic
324 population. We chose this particular test based on a benchmarking effort that suggests it is well-
325 suited to detecting selection in E&R experiments lacking replication[9]. After results were
326 generated for each synthetic population, log transformed p-values were plotted for each
327 chromosome across sliding windows. The GenWin package in R (parameters: "smoothness =
328 2000, method = 3") was once again used to define windows based on breakpoints in the data.
329 Plots were then examined to see if there were any genomic regions showing signs of selection
330 based on significance levels relative to the background.
331
332 *Haplotype representation*
333 In addition to describing SNP diversity, we also describe the diversity of founder
334 haplotypes represented in our synthetic populations. We were particularly interested in
335 evaluating whether the S-type strategy might produce populations in which founder haplotypes
336 are more evenly represented (at intermediate frequency) compared to the K-strategy. Given the
337 stochasticity inherent in the K-type strategy, we thought it probable that founder genotypes with
338 especially high sporulation and/or mating efficiencies (i.e. those with the highest reproductive
339 outputs in the outcrossing context) might come to dominate. To this end, we estimated haplotype
340 frequencies in all experimental populations initially, and after 6 and 12 cycles of outcrossing to

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

341 determine how evenly haplotypes were represented, and how this might have changed over time.
342 We used the sliding-window haplotype caller described in Linder et al. (2020) and software the
343 authors have made available as a community resource: https://github.com/tdlong/yeast_SNP-
344 HAP. Our results were generated by using the haplotyper.limSolve.code.R script and estimates
345 were made across 30KB windows with a 1KB stepsize. This particular haplotype caller was
346 developed specifically to estimate haplotype frequencies in multiparent populations when
347 founder haplotypes are known. A full description of the algorithm being used, and results of
348 empirical validation can be found in Linder et al. (2020). To quantify haplotype variation in each
349 population, we calculated haplotype diversity (H) using the following formula: (1) $H = 1 -$
350 $\sum_{i=1}^{n} x_i^2$ where x_i is the frequency of the *i*th haplotype of the *n* founders used to create given
351 population[39]. Though it is worth noting that maximum expected H will vary depending on the
352 number of founders used to create a given population as (2) $H_{max} = 1 - \left(\frac{1}{n}^2 * n\right)$.

353 *Phenotypic characterization of experimental populations*

354 To evaluate the possibility that populations might be phenotypically differentiated, we
355 measured two life-history traits: sporulation efficiency and growth rate._We estimated the 3-day
356 sporulation efficiency for each recombinant population at the beginning and end of the
357 experiment, as this is a life-history trait that might have reasonably responded to the selection
358 imposed by the regular outcrossing protocol. All populations archived at "cycle 0" (i.e. the pool
359 of diploid cells used to initiate each K- or S-type population) and "cycle 12" (i.e. diploid cells
360 recovered from each population after the 12th outcrossing cycle) were revived by plating 1mL of
361 thawed culture onto a YPD agar plate and incubation at 30°C for 48 hours. In order to sample
362 the genetic diversity of each population, a sterile wooden applicator was scraped in a zig-zag
363 pattern across the lawn of cells on each plate to collect a pinhead-sized clump of yeast. Each
364 clump was mixed in 10 mL YPD in a 50 mL conical tube and vortexed. Tubes were then
365 incubated at 30°C/200 rpm for ~24 hours. After confirming that each tube had comparable cell
366 densities – this was done by verifying that the OD_{600} absorbance value of a 1:100 dilution ranged
367 between 0.095-0.2 – cell pellets were collected by spinning for 5 minutes at 5000 rpm. Cell
368 pellets were washed in 1 mL of sterile water, spun down again, and resuspended in 40 mL of
369 minimal sporulation media (1% potassium acetate w/v). Each culture was transferred to sterile
370 250 mL Erlenmeyer flasks and covered loosely with foil, where they were cultured at 30°C/200
371 rpm for ~72 hours to sporulate. After sporulation, aliquots of each culture were loaded onto a

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

372 hemacytometer (Incyto C-Chip, type NI) and visualized under 40x magnification on a Singer
373 SporePlay microscope. For each culture, ~200 cells were counted (specific range: 190-230
374 cells), and sporulation efficiencies were estimated as the proportion of tetrads observed over the
375 total number of cells in the field of view. Sporulation efficiency for each of the 12 recombinant
376 populations (6 "cycle 0" and 6 "cycle 12") was assessed by averaging these proportions over 2-3
377 independent biological replicates.

378 In addition to characterizing sporulation efficiencies for each of the "cycle 0" and "cycle
379 12" recombinant populations, we also measured growth rate with high-throughput absorbance-
380 based assays in liquid YPD. We also included the 12 founder strains in this assay, for
381 comparison with the recombinant populations. S- and K-type recombinant populations were
382 sampled from each freezer recovery plate as described above. Haploid founder strains were
383 revived from freezer stocks by striking for single colonies onto YPD agar plates. Each
384 population or strain was assayed in two biological replicates; recombinant populations were
385 sampled to inoculate two separate overnight cultures in liquid YPD, and strains were sampled by
386 picking two distinct colonies to initiate two separate overnight cultures (one colony per culture).
387 All biological replicates were incubated for ~24 hours at 30°C/200 rpm. The day of the assay,
388 OD_{600} was measured in all cultures and the readings used to standardize them to a target OD_{600}
389 of 0.05 in fresh YPD (observed values ranged 0.042-0.061). 200uL of each culture was aliquoted
390 to separate wells of a 96-well plate, with two technical replicates per biological replicate. The
391 arrangement of technical replicates on the plate was carried out in an attempt to control for
392 possible edge effects. The growth rate assay was carried out in a Tecan Spark Multimode
393 Microplate Reader, set to record the absorbance at 600 nm for each well every 30 minutes for 48
394 hours at 30°C, without plate agitation/aeration. The R-package "Growthcurver" (Sprouffske and
395 Wagner[40]) was used to estimate population growth parameters from the raw data. In order to
396 determine the carrying capacity and doubling time of the culture in each well, the absorbance
397 measurements taken during the assay were fit to the following equation:
398

399 (3): $$N_t = \frac{N_0 K}{N_0 + (K - N_0)e^{-rt}}$$

400 Where N_t is the absorbance reading at time t, N_0 is the initial absorbance, K is the carrying
401 capacity, and r is the growth rate, or doubling time. Here, doubling time refers to the time
402 necessary for the size of a population to double under non-restricted conditions, while carrying

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

403 capacity is the maximum population size under the given conditions. The values for each
404 biological replicate were averaged across technical replicates, and the values for each
405 strain/population were determined by averaging across biological replicates.
406

407 Results

408 *SNP Variation*

409 To assess how crossing strategy and number of founder strains impacts SNP variation, we
410 began by simply counting the number of SNPs present in each of our synthetic populations upon
411 their creation and how that changes over several cycles of recombination (Table 2). As expected,
412 the total number of possible SNPs that can possibly contribute to segregating genetic variation
413 increases with the number of founders used. Looking at our actual populations at cycle 0 and
414 focusing on those created using the same crossing strategy, we also generally find the observed
415 number of SNPs in each population to increase with the number of founders used. The only
416 exception to this pattern is the K12 population where we see dramatic losses in polymorphic sites
417 relative to all other populations. We also typically observe reductions in the number of SNPs in
418 all experimental populations over time. However, we do note higher "stability" (i.e. smaller
419 losses) in the 8-founder populations, and in population S8, we actually observe higher SNP
420 counts in cycle 12 than in cycle 6. This discrepancy is most likely due to a relatively small
421 number of sites at very low frequency in cycle 6 (i.e. too low for our SNP calling to pick up),
422 increasing to detectable levels by cycle 12. Nevertheless, the overall trend still appears to be
423 reductions in the number of polymorphic sites over time. Our data also suggest these reductions
424 are typically more pronounced in populations created using the K-type strategy, and that
425 populations created using the S-type strategy have more polymorphic sites than those created
426 with the K-type strategy.

427 We also examined patterns of overlap between our synthetic populations and what is
428 possible given their respective founders using UpSet plots (Supplementary Fig. S3). UpSet plots
429 are useful because they allow visualization of observations unique to groups; in other words,
430 they highlight observations that are included in a specific group and excluded from all others.
431 As an illustrative example, the fourth vertical bar in Supplementary Fig. S3A-C represents all of
432 the SNPs that could possibly segregate in a population but do not, and this reveals that a greater
433 proportion of possible SNPs are lost in synthetic populations with 12 founders compared to those

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

434 with 4 or 8 (this result is also noted in Table 2). We were interested in using this visualization
435 approach to evaluate whether there might be particular SNP groups that are hallmarks of the S-
436 type or K-type strategy. While many more of the possible SNPs appear in S-type populations
437 relative to the K-type populations for a given set of founders, the UpSet plots indicate that very
438 few SNPS are unique to a particular crossing strategy. We interpret this result as evidence that
439 neither crossing strategy favors specific alleles, and this is true regardless of the number of
440 founder haplotypes used.

441 Next, we looked at how the SFS varied across different populations and how they
442 changed over time. As shown in Fig. 1, the S-type populations tend to exhibit less skewed
443 frequency distributions compared to the K-type population, and they are also more stable over
444 time. This is particularly evident in the 12-founder populations. In these we see that by the final
445 cycle of recombination in the K12 populations, there is an extreme skew in the SFS with most
446 sites exhibiting very high or low SNP frequencies. This is also consistent with the results shown
447 in Table 2 where we see much higher levels of fixation over time in K12 than in any other
448 population. This contrast is present but far less extreme in populations created with 4 or 8
449 founder strains. In general, at cycle 0 all populations deviate substantially from the SFS that we
450 would expect if the respective founders combined in perfectly equal proportions (Supplementary
451 Fig. S4). These deviations suggest that drift and/or selection are impacting the genetic makeup of
452 synthetic populations from the moment they are established.

453 Finally, we assessed the effects of crossing strategy and number of founders on genome-
454 wide heterozygosity. The clearest pattern we observed is that after 12 cycles of recombination, S-
455 type populations exhibited greater overall levels of heterozygosity across the genome compared
456 to their K-type counterparts (Table 3). After 12 cycles of outcrossing, we also see large stretches
457 along chromosomes where heterozygosity is near zero in K4 and K12 which is not the case in
458 their S-type counterparts (Fig. 2). Other patterns in the data are less clear, however; for instance,
459 we do not find stretches of the genome where variation has been expunged in the K8 population,
460 compared to the S8 counterpart (Fig 2C-D). Looking at the S-type populations alone, we find
461 that by cycle 12 the S4 populations have experienced a greater loss of heterozygosity than the S8
462 and S12 populations (Table 3). However, differences between S8 and S12 populations are far
463 less severe with the former having slightly higher mean heterozygosity. As such, there is no clear
464 positive relationship between heterozygosity and number of founders. This pattern largely breaks

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

465 down in the K populations. The K4 population experiences a greater loss of heterozygosity than
466 K8 by cycle 12, but then K12 experiences the most severe declines in heterozygosity by cycle 12
467 as expected given the other measures of SNP variation we have looked at thus far.
468
469 *SNP frequency changes over 12 cycles of outcrossing*
470 We used Pearson's χ^2 test to compare SNP frequencies between cycle 0 and cycle 12 of
471 each individual population to see if there were any regions of the genome showing obvious
472 responses to selection imposed by our outcrossing maintenance protocol. Looking across results
473 for each population, we do not find any genomic regions that show consistent responses to
474 selection (Fig. 3). However, we do find instances in individual populations where there are clear
475 peaks in significance relative to the rest of the genome (e.g. Fig. 3A and C-E). Widespread
476 fixation in K4 and K12 make it difficult to identify such peaks (Fig. 3B and F).
477 We also compared our results to prior work which identified regions of the genome and
478 genes associated with selection for frequent outcrossing, Cubillos et al.[22] and Burke et al.[20]. Our
479 rationale for doing this is that as our current study features no within-treatment replication, it is
480 weak to implicate signatures of adaptation by itself; but, obvious overlap with other studies
481 would provide indirect evidence for adaptation. Similarly, if a particular population (or
482 population type) implicated more candidate regions from the literature than the others, this could
483 provide evidence that a particular crossing strategy might lead to the best outcomes in an E&R
484 study, with respect to identifying regions of interest. We do find instances where peaks in
485 significance align with previously identified candidate regions (e.g. Fig. 3A, C-D), but there is
486 no clear pattern where peaks consistently line up with a given candidate region across multiple
487 populations. There are also many more instances where we do not find peaks that align with the
488 previously identified candidate regions. So generally speaking, while this comparative approach
489 provides opportunities for describing particular regions, perhaps those of *a priori* interest, we
490 cannot conclude that observed changes in SNP frequencies in any population are likely
491 signatures of selection for forced outcrossing. On the other hand, we also observe that the 8-
492 founder populations generally implicate more candidate genes from the literature, compared to
493 any other population (Fig. 3C-D). This suggests that perhaps using 8 founding haplotypes in a
494 population for an E&R project leads to better outcomes, in terms of identifying candidate
495 regions, than using 4 or 12. Notably, the K8 population and S8 population implicated similar

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

496 numbers of peaks, so it is not clear that one crossing strategy is better than the other in this
497 respect.
498
499 *Haplotype Representation*
500 Using sequence data from all experimental populations and our founder strains, we
501 estimated haplotype frequencies across the genome to assess how crossing strategy and number
502 of founders impact haplotype representation initially, as well as after 6 or 12 cycles of
503 outcrossing. Notably, estimates for K12 are made using far fewer SNPs due to the extreme
504 levels of fixation seen in this population and are almost certainly less reliable than those from
505 other populations. We first assessed the haplotype frequencies observed after 12 cycles of
506 outcrossing, as this provides a birds-eye view of the amount of haplotype diversity present at the
507 end of the experiment (Supplementary Figs. S5-S10). While haplotype frequencies fluctuate
508 across the genome in all populations, mean genome-wide haplotype frequency estimates point to
509 clearer patterns. In the 4- and 12-founder populations, we find that the S-type populations have
510 haplotype frequencies that are more evenly-distributed compared to the K-type populations
511 (Table 4). This is also reflected in levels of haplotype diversity for these populations, where we
512 find that the S-type populations typically have greater levels of diversity across the genome
513 compared to the K-type populations (Fig. 4). Mean haplotype diversity is typically greater in the
514 S-type populations, closer to maximum expected values, and has smaller variance across the
515 genome (Supplementary Table S2). These differences can largely be attributed to the almost
516 complete loss of particular haplotypes in the K-type populations versus their S-type counterparts.
517 For instance, in 4S we observe nearly equal representation of the founding haplotypes (Table 4;
518 Supplementary Fig. S5), but in K4 we observe that two of the founding haplotypes dominate and
519 the other two are almost entirely lost (Table 4; Supplementary Fig. S6). The S8 and K8
520 populations appear more similar in measures of haplotype diversity (Fig. 4; Supplementary
521 Table S2, but as with K4, we once again that two haplotypes, DBVPG6044 and YPS128, are
522 almost entirely lost (Supplementary Fig. S8; Table 4) in a way not seen in S8 (Supplementary
523 Fig. S7; Table 4). The S12 (Supplementary Fig. S9) and K12 (Supplementary Fig. S10) results
524 show similar patterns, but interpretation is complicated by the fact K12 haplotype estimates were
525 generated using a very limited number of SNPs. However, across our study as a whole,
526 haplotype representation does appear to benefit from an S-type crossing strategy. Using fewer

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

527 founders as well as an S-type crossing strategy yields distributions that most closely match what
528 we would expect under equal blending.
529 While patterns of haplotype representation in the "cycle 12" populations speak to how
530 well haplotype diversity is maintained over time, evaluating haplotype representation at earlier
531 parts of the experiment also reveals insight into how the K- versus S- strategies impact haplotype
532 diversity. The near-complete absence of the YPS128 and DBVPG6044 haplotypes in K4 after
533 12 outcrossing cycles begs the question of exactly when these haplotypes were lost. In fact,
534 these two haplotypes were not observed at high frequencies even immediately after the
535 population's founding, which suggests that these two strains simply failed to mate with the other
536 strains in the pool (Supplementary Figure S11). The general pattern of haplotype frequencies in
537 this population over time suggests that these two haplotypes gradually diverge from starting
538 frequencies near ~0.5 at cycle 0 (Supplementary Figure S11), such that at cycle 6 frequencies are
539 more variable (Supplementary Figure S12) and by cycle 12 the frequencies are all nearly fixed at
540 any given position along the genome. This is in stark contrast to what we see in S4 where all
541 founders are well-represented at each timepoint (Supplementary Figure S7, S13-14). However,
542 patterns become more complicated in populations created using more founders. For instance, the
543 haplotype frequencies in the initial K8 population (Supplementary Figure S15) suggest that
544 YPS128 was lost immediately, as we also do not observe it after 6 cycles (Supplementary Figure
545 S16) or 12 cycles (Supplementary Figure S8) of outcrossing. But Y12, which is also appears lost
546 in the initial population, increases after 6 and 12 cycles to become one of the most prevalent
547 haplotypes in the population. We also see the opposite pattern for DBVPG6044. The observation
548 that a single founder is lost for good in one K-type population but recovers in another is difficult
549 to explain, but it is consistent with the idea that the K-type strategy is generally more
550 unpredictable, and perhaps more prone to haplotype loss than the S-type strategy.
551 Next, we evaluated whether any of our observed patterns of haplotype representation
552 could be easily explained by selection. To do this we compared our findings to the results of
553 Burke et al.[20]. In this study, twelve replicate populations were created using the same four
554 founding strains as the S4 and K4 populations, and were subjected to a similar outcrossing
555 regime for 18 cycles (Note: the synthetic population the replicates were derived from was created
556 using a S-type strategy and underwent 12 cycles of recombination before replicates were
557 generated). Based on their analysis of genomic data taken from each population across three

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

558 timepoints, Burke et al. (2014) identified five major candidate regions associated with adaptation
559 to forced outcrossing. Four of these regions (regions A, B, C, and E in the paper) were also found
560 to be clearly defined by a given haplotype at the end of the experiment (See Supplementary
561 Figure S4 in Burke et al.[20]). As the founders used to create the synthetic populations of Burke et
562 al. (2014) are present in all of the populations featured in this study, we can reasonably interpret
563 changes in haplotype frequency in these four candidate regions in our populations as evidence of
564 selection. In other words, as our populations here are unreplicated, and we cannot use strong-
565 inference approaches with them directly, we sought parallels between haplotype frequency
566 differences observed here and those described as adaptive in a very similar experiment.

567 As shown in Table 5, we find some evidence that the haplotypes driving adaptation in the
568 Burke et al.[20] study also show substantial increases in frequency in our populations. In candidate
569 regions A (C9: 950,000-975,000) and E (C13:445,000-460,000), we find clear evidence of the
570 Y12 haplotype frequency increasing across all of our populations as it did in the Burke et al.
571 (2014) study. Given how dramatic some of these changes are from cycle 0 to cycle 12, we think
572 it reasonable to speculate that these haplotypes might also be driving adaptation here. However,
573 this pattern in not corroborated for candidate regions B (C9: 65,000-80,000) and C
574 (C11:615,000-620,000). Here the haplotypes driving adaptation in the Burke et al.[20] study
575 actually decrease in frequency in most instances.
576

577 *Phenotypic characterization of experimental populations*

578 Sporulation efficiencies were estimated in all recombinant populations initially, and after
579 12 cycles of outcrossing (Supplementary Table S3). We find that population K12 at the end of
580 the experiment (after 12 outcrossing cycles) has the lowest sporulation efficiency at ~10%, while
581 all other populations, including the initial K12 population, had sporulation efficiencies exceeding
582 30%. Excluding K12 cycle 12, estimates range from 30% to 55% with a mean of 38%. Looking
583 across these estimates, we find no clear relationships of sporulation efficiency to either: i)
584 number founders used; ii) K- versus S-type build strategy; or iii) cycle 0 versus cycle 12. But,
585 variation among assays of biological replicates of the same population was high enough (see
586 standard deviations provided in Supplementary Table S3), our ability to make definitive
587 comparisons between these estimates is limited. As such, we can only say that sporulation
588 efficiency in K12 cycle 12 is greatly reduced when compared to the other recombinant

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

589 populations. We also assessed growth rates in rich media for all 12 founder strains, and the
590 recombinant populations initially and after 12 cycles of outcrossing. We observe variation in the
591 founder strains with doubling times ranging from ~1 hour to ~1.3 hours, but we see no obvious
592 relationships between founders used to create a specific recombinant population and its initial
593 growth rate (Supplementary Table S4). For instance, two of our slowest growing founder strains,
594 YJM975 and 273614N, are only used when creating K12 and S12. However, the 12-founder
595 populations do not grow more slowly than the 4- or 8- founder populations. We do see evidence
596 of a trend where the recombinant populations grow more slowly at cycle 12 than they did
597 initially, but changes are not particularly large (ranging from -0.01 to -0.22 hours with a mean of
598 -0.10).
599

600 **Discussion**

601 *Careful crossing of founder strains increases SNP-level variation*

602 Here we primarily sought to assess how crossing strategy - pairwise crossing of founder
603 strains versus mixing populations in equal proportion - impacted levels of SNP and haplotype
604 variation. We considered both the total number of polymorphic sites and levels of heterozygosity
605 at those sites in synthetic populations created using these two approaches. We consistently found
606 that populations generated by imposing a round of careful crossing have more polymorphic sites
607 (Table 2), and less variation is lost at those sites over time, compared to populations created by
608 simply mixing founders in equal proportion (Fig. 2 and Table 3). We also found that increasing
609 the number of founder strains used to create a given synthetic population also generally resulted
610 in higher levels of SNP variation. However, the total percentage of potential polymorphic sites
611 lost was typically higher when more populations were used (Table 2). And combining the K-type
612 crossing strategy with a large number of founders resulted in dramatic skews in the site
613 frequency spectrum and losses in SNP variation (Fig. 1). Lastly, we find no evidence that
614 specific sets of variants are favored in one crossing strategy versus another when the same
615 founders are used (Supplementary Fig S3). Instead, we suggest that differences in outcomes
616 between crossing strategies are most likely due to strain-specific differences in mating
617 efficiencies and reproductive output (i.e. when populations are simply mixed together, the most
618 compatible strains dominate).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

619 Overall, these findings have led us to two general recommendations. First, if one
620 endeavors to produce populations with the highest possible levels of total SNP variation, many
621 founders and a crossing strategy that involves at least one round of pairwise crossing should be
622 considered. However, if one's goal is to preserve as much of the variation in founding genotypes
623 as possible, a crossing strategy with fewer founders might be more desirable.
624
625 *Careful crossing of founder strains results in more balanced haplotype representation*
626 Haplotype frequency estimates for our synthetic populations suggest that a strategy
627 involving pairwise crosses results in better representation and maintenance of founder genotypes
628 (Fig. 4, Supplementary Figs. S5-S18, and Table 4). It is clear than when haploid strains are
629 simply pooled, there is potential for subsets of the founder strains to dominate and skew
630 haplotype representation in the resulting populations. As such, we recommend a careful (i.e.
631 pairwise) crossing strategy when seeking to maximize founder haplotype representation. In
632 addition, using fewer founder haplotypes also generally results in more even representation
633 among them, though we would not necessarily recommend against using more founders unless it
634 is crucial to achieve an even haplotype representation. In fact, we would argue that there are
635 empirical benefits to a more varied distribution of haplotype frequencies segregating within a
636 synthetic population. Specifically, in an E&R context, such a population creates opportunities to
637 observe selection on both rare and common haplotypes, and the relative contributions each to the
638 evolution of quantitative traits remains an unresolved question (e.g. Bloom et al.[41]).
639 What might be driving the extreme pattern of haplotype loss we observe in the K-type
640 populations compared to their S-type counterparts? Here we outline two possibilities, using the
641 K4 and S4 populations to illustrate. In the former, we find that the YPS128 and DBVPG6044
642 haplotypes are almost entirely missing while in the latter all founder haplotypes are evenly
643 represented. We think that the most obvious mechanism underlying this pattern is the
644 reproductive capacity of the founding strains; in other words, if particular strains inefficiently
645 sporulate and/or mate, or are not compatible with other strains, haplotype loss should result. The
646 two missing haplotypes in K4 appear to have been lost immediately after the population's
647 creation (Supplementary Figure S11), which would support the idea that these two genotypes are
648 simply disadvantageous under environmental circumstances which require cells to sexually
649 reproduce to survive. A second, non-mutually exclusive possibility that could drive the haplotype

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

650 loss we observed is the emergence of an asexual diploid lineage that can evade our forced
651 outcrossing protocols and become the majority genotype through clonal interference. Others
652 working with recombinant *S. cerevisiae* have reported evidence of such "cheater" lineages
653 (Linder et al.[42]; note: their crossing protocol, described in Linder et al.[23], represents a middle
654 ground between our K and S strategy). Since we observe intermediate levels of haplotype
655 representation in the K4 population mid-way through the experiment (Supplementary Figure
656 S12), this suggests that if a diploid cheater lineage emerged, this happened over a long
657 evolutionary timescale; in other words, this cannot explain the early haplotype losses we
658 observed in the population, but it might explain long-term loss of haplotype diversity. At the
659 same time, one would expect a expect a cheater genotype to take over very rapidly in the
660 population, which we did not observe. In summary, we think our experiments provide
661 compelling evidence that differences in sexual reproduction between strains will lead to
662 haplotype loss in the K-type populations, and that this loss can be prevented by using an S-type
663 strategy. While the evolution of cheating could reasonably occur in any of our populations, and
664 also result in haplotype loss, we report no strong evidence that this is more likely to happen in K-
665 type versus S-type populations.
666
667 <u>*Evidence for adaptation to outcrossing*</u>
668 While we interpret the differences between the different crossing strategies (i.e. between
669 all S-type and K-type populations) as being primarily driven by initial differences in sporulating
670 and mating efficiencies between strains and subsequent genetic drift, selection may also play a
671 role. For instance, we do find evidence that two specific regions of the genome exhibit similar
672 dynamics in haplotype frequencies, compared to similar prior work; Burke et al.[20] previously
673 implicated these regions as potentially driving adaptation for frequent outcrossing. We also
674 compared our results to those of Linder et al.[23] which features synthetic populations constructed
675 using the same founder strains present in this study. However, here we did not find it was the
676 case that haplotypes most common in their populations on average are also the most common in
677 our populations. For instance, Y12 is a common haplotype across all of our populations (Table 4)
678 with mean genome-wide frequencies ranging from 0.2 to ~0.5, but it does not exceed a frequency
679 of 0.06 in either of their synthetic populations (*cf.* Table 3 of Linder et al.[20]). Similarly, YPS128
680 which has a mean frequency of 0.41 in one of the their populations appears at frequencies below

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

681 0.11 in all of our populations except S4. That being said, it is worth noting that while there is
682 overlap in the strains used between these studies, maintenance protocols are different, and their
683 populations include several strains absent in our study. So, we can only say that no common
684 founders appear to universally favored when selection for frequent outcrossing is imposed.
685 Our analysis comparing initial SNP frequencies to those observed after 12 rounds of
686 outcrossing in each population also yields possible evidence for adaptation, though our
687 interpretation here is cautious. We do observe regions that produce peaks in significance in some
688 populations that overlap regions described by other studies to underlie traits related to
689 outcrossing (e.g. Fig. 3D; peak in C16 in K8). However, peaks are rarely recapitulated across
690 two or more populations (Fig. 3) and in some populations we observe no clearly defined peaks at
691 all (e.g. Fig. 3B; population K4). As such, our results are at best suggestive. But, we find it
692 notable that a crossing strategy using 8 founding haplotypes leads to the most overlap with other
693 candidate regions identified in the literature. While this experiment has limited ability to
694 associate particular genomic regions and phenotypes related to outcrossing, this observation
695 leads us to speculate that using an intermediate number of founding haplotypes (e.g. 8) may lead
696 to an increased ability to localize candidate regions in an E&R experiment.
697 Given that any adaptation we did observe can only be ascribed to the outcrossing process
698 and associated handling protocols, we conclude that there is likely a great deal of value in
699 allowing newly-established synthetic populations to experience several cycles of outcrossing –
700 this can also be thought of as laboratory domestication – before any sort of new selective
701 pressure is imposed. To be explicit, if some other selection pressure was intentionally imposed
702 on our populations immediately following cycle 0, it would be very difficult to dissect the
703 specific genetic changes that might occur due to that pressure, other laboratory handling steps, or
704 general selection for reproduction via outcrossing.
705
706 *Crossing strategy and number of founder strains does not obviously impact sporulation or*
707 *growth rate*
708 We assayed sporulation efficiencies and growth rates in recombinant populations as these
709 are both important fitness related characters that may be impacted by crossing strategy, or that
710 might respond to the selection imposed by many cycles of forced outcrossing. Looking at
711 sporulation efficiencies (Supplementary Table S3), we do not find any obvious associations

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

712 between these estimates and a particular crossing strategy. We also find no clear evidence that
713 sporulation efficiency increases over the course of the experiment, which is somewhat surprising
714 given the forced outcrossing that defines our maintenance protocol. The only major finding that
715 emerges is that sporulation efficiency is much lower in K12 cycle 12 than what we observe in
716 any other population, including the initial K12 population. We speculate that this is perhaps
717 related to the wholesale loss of genetic variation in K12, or this may indicate that asexual diploid
718 lineages representing only a fraction of total variation among founders have risen to prominence.
719 However, our ability to make any definitive statements about how crossing strategy or number of
720 founders shapes sporulation efficiency as a life-history trait is limited.

721 Comparing growth rates of founder strains to recombinant populations (Supplementary
722 Table S4) similarly does not reveal clear evidence linking growth rates in founder strains to
723 either K or S-type populations (i.e. recombinant populations have similar doubling times
724 regardless of which strains were used or how they were combined). It is also not the case that our
725 fastest growing founders are better represented when looking haplotype estimates or vice versa.
726 There is a consistent trend of slower growth rates in cycle 12 versus cycle 0 for the recombinant
727 populations. Differences are small, but this is still perhaps suggestive of some sort of trade-off
728 between growth and other fitness characters as populations adapt to our maintenance protocols.
729 In other words, it is conceivable that the demands for high levels of outcrossing might result in
730 populations that invest more in sexual reproduction, and less in budding. However, we find no
731 consistent patterns when comparing difference between S and K strategies or number of
732 founders. As such, we find no evidence that these factors are shaping growth rates in the
733 recombinant populations.
734

735 *Conclusions*

736 The results of simulated E&R studies in which populations are sexually reproducing and
737 adaptation is driven by standing genetic variation have led to general experimental design
738 recommendations that maximize genetic variation in the ancestral population, specifically by
739 increasing the number of starting haplotypes[7-8]. Here, we provide empirical results that also these
740 recommendations. Across the metrics we examined, we consistently find that a crossing strategy
741 involving careful pairwise crosses leads to populations with more standing genetic variation than
742 those produced by simply mixing founder genotypes in equal proportion. As such, using this sort

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

743 of strategy would be our primary recommendation for researchers aiming to establish
744 recombinant populations from clonal strains or isogenic lines for use in E&R studies. Using
745 more founders also results in greater total levels of genetic variation but comes at the cost of
746 maximizing representation of all possible alleles from a given set of founders. So, the number of
747 founders to used should be chosen based upon the specific goals of an E&R study, and the
748 questions it hopes to test. Finally, our results should be placed in the context of yeast biology.
749 Meaning, given the high degree of genetic differentiation between our yeast strains and the
750 variation among them for mating and sporulation efficiencies, the patterns we observe are likely
751 more extreme than what might be expected when creating a *Drosophila* synthetic population
752 using either the K-type or S-type approach. However, it is certainly conceivable that mating
753 preferences and genetic incompatibilities among isogenic *Drosophila* lines could impact levels
754 of genetic variation and haplotype representation when they are crossed. Our findings therefore
755 perhaps warrant consideration even when creating synthetic populations in non-yeast systems.
756
757 **Data Availability**
758 The raw sequence files generated over the course of this project are available through NCBI
759 SRA (BioProject ID: PRJNA732717). Core data files (tables with SNP and haplotype
760 frequencies, results of statistical analysis, etc.) are available through Dryad
761 (doi:10.5061/dryad.g79cnp5qg). Script used to process raw data and perform SNP calling are
762 available through Github (https://github.com/mollyburke/Burke-Lab-SNP-calling-pipeline), as
763 are the core scripts necessary to reproduce our results (https://github.com/mphillips67/Build-
764 Paper).
765
766 **Acknowledgements**
767 We thank undergraduate researcher Julia Reinsch for her help with sporulation assays, and Dr.
768 Anthony D. Long for his feedback on the haplotype estimation analysis. We also thank Oregon
769 State University's Center for Genome Research and Biocomputing for use of their computational
770 and sequencing resources. This work was supported by startup funds provided to M.K.B. by
771 College of Science at Oregon State University, and M.A.P. was supported by a National Science
772 Foundation Postdoctoral Fellowship (NSF 1906246).
773

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

774 **Author Contributions**

775 M.K.B. and I.C.K. conceived of the project. I.C.K. performed the lab work necessary to generate
776 the genomic data sets featured in this study, and K.M.M. and S.K.T. generated all phenotypic
777 data and results. M.A.P. and M.K.B. formulated the analytic strategy for the genomic data, and
778 M.A.P. performed all major genomic analyses. M.A.P. and M.K.B. wrote the manuscript.

779

780 **References**

781 1. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**,
782 747-753 (2009).

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

785 3. Schlötterer, C., Tolber, R., Kofler, R., & Nolte, V. Sequencing pools of inidviduals – mining
786 genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* **15**, 749-763 (2014).

787 4. Long A.D, Liti, G., Luptak, A. & Tenaillon O. Elucidating the molecular architecture of
788 adaptation via evolve and resequence experiments. *Nat. Rev. Genet.,* **16**, 567-582 (2015).

789 5. Bailey, S.F. & Bataillon, T. Can the experimental evolution programme help us elucidate the
790 genetic basis of adaptation in nature? *Mol. Ecol.* **25**, 203-216 (2016).

791 6. Burke, M.K. How does adaptation sweep through the genome? Insights from long-term
792 selection experiments. *Proc. Roy. Soc. B.,* **279**, 5029-5038 (2012).

793 7. Baldwin-Brown, J.G., Long, A.D. & Thornton, K.R. The power to detect quantitative trait
794 loci using resequenced, experimentally evolved populations of diploid, sexual organisms.
795 *Mol. Biol. Evol.* **31**, 1040-1055 (2014).

796 8. Kofler, R. & Schlötterer, C. A guide for the design of evolve and resquencing studies. *Mol.*
797 *Biol. Evol.* **31**,473-482 (2014).

798 9. Vlachos, C. & Kofler, R. Optimizing the power to identify the genetic basis of complex traits
799 with evolve and resequence studies. *Mol. Biol. Evol.* **35**, 2890-2905 (2019).

800 10. Burke, M.K. *et al.* Genome-wide analysis of a long-term evolution experiment
801 with *Drosophila*. *Nature* **467**, 587-590 (2010).

802 11. Turner, T.L., Stewart, A., Fields, A.T., Rice, W.R. & Tarone, A.M. Population-based
803 resequencing of experimentally evolved populations reveals the genetic basis of body size
804 variation in *Drosophila melanogaster*. *PLoS Genet.,* **7**, e10001336 (2011).

805 12. Orozco-ter Wengel, P. *et al.* Adaptation of *Drosophila* to a novel laboratory environment
806 reveals temporally heterogeneous trajectories of selected traits. *Mol. Ecol.* **21**, 4931–4941
807 (2012).

808 13. Tobler, R. *et al.* Massive habitat-specific genomic response in D. *melanogaster* populations
809 during experimental evolution in hot and cold environments. *Mol. Biol. Evol.* **31**, 364-375
810 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

811 14. Huang, Y., Wright, S.I. & Agrawal, A.F. Genome-wide patterns of genetic variation within
812 and among alternative selective regimes. *PLoS Genet.* **10**, e1004527 (2014).

813 15. Jha, A.R. *et al.* Whole-genome resequencing of experimental populations reveals polygenic
814 basis of egg-size variation in *Drosophila melanogaster*. *Mol. Biol. Evol.* **32**, 2616-2632
815 (2015**).**

816 16. Franssen, S.U., Nolte, V., Tobler, R. & Schlötterer C. Patterns of linkage disequilibrium and
817 long range hitchhiking in evolving experimental *Drosophila* melanogaster populations. *Mol.*
818 *Biol. Evol.* **32**, 495-509 (2015).

819 17. Graves, J.L. *et al.* Genomics of parallel experimental evolution in Drosophila. *Mol. Biol.*
820 *Evol.* **34**, 831–842 (2017).

821 18. Phillips, M.A. *et al.* Effects of evolutionary history on genome wide and phenotypic
822 convergence in Drosophila populations. *BMC Genomics* **19**, 743 (2018).

823 19. Barghi, N. *et al.* Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biol.*
824 **17**, e3000128 (2019).

825 20. Burke, M.K., Liti, G. & Long, A.D. Long standing genetic variation drives repeatable
826 experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. *Mol. Biol.*
827 *Evol.* **32**, 3228-3239 (2014).

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

831 22. Cubillos, F.A *et al.* High-resolution mapping of complex traits with a four-parent advanced
832 intercross yeast population. *Genetics* **195**, 1141-1155 (2013).

833 23. Linder, R.A., Majumbder, A., Chakraborty, M. & Long, A.D. Two synthetic 18-
834 wayoutcrossed populations of diploid budding yeast with utility for complex trait dissection.
835 *Genetics* **215**, 323-342 (2020).

836 24. King, E.G. *et al.* (2012). Genetic dissection of a model complex trait using the Drosophila
837 synthetic population resource. *Genome Res.*, **22**, 1558-1566.

838 25. Teotónio, H., Carvalho, S., Manoel, D., Roque, M. & Chelo, I.M. Evolution of outcrossing in
839 experimental populations of *Caenorhabditis elegans*. *PLoS One*, **7**, e35811 (2012)

840 26. Michalak, P., Kang, L., Schou, M.F., Garner, H. & Loeschke, V. Genomic signatures of
841 experimental adaptive radiation in Drosophila. *Mol. Ecol.* **28**, 600– 614 (2018).

842 27. Barghi. N. & Schlötterer, C. Shifting the paradigm in evolve and resequence studies, from
843 analysis of single nucleotide polymorphism to selected haplotype blocks. *Mol. Ecol.* **28**, 521-
844 524 (2019).

845 28. Nouhaud, P., Tobler, R. Nolte, V. & Schlötterer, C. Ancestral population reconstitution from
846 isofemale lines as a tool for experimental evolution. *Eco. Evol.* **6**, 7169-7175 (2016).

847 29. Harbison, S.T., Negron Serrano, Y.L., Hansen, N.F. & Lobell, A.S. Selection for long and
848 short sleep duration in *Drosophila melanogaster* reveals the complex genetic network
849 underlying natural variation in sleep. *PLoS Genet.* 13, e1007098 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

852 31. Burke, M.K., McHugh, K.M. & Kutch, I.C. Heat shock improves random spore isolation in
853 diverse strains of *Saccharomyces cerevisiae. Front. Genet.* **11**: 597482 (2020).

854 32. Baym, M. *et al.* Inexpensive multiplexed library preparation for megabase-sizes genomes.
855 *PLoS One* **10,** e0131262 (2015).

856 33. Mkenna *et al.* The genome analysis toolkit: a mapreduce framework for analyzing next-
857 generation DNA sequencing data. *Genome Res.* **20**,1297-1303 (2010)

858 34. Poplin *et al.* A universal SNP and small-indel variant caller using deep neural networks. *Nat.*
859 *Biotechnol.* **36**, 983-987 (2018).

860 35. Bergström, A. *et al.* A high-definition view of functional genetic variation from natural yeast
861 genomes. *Mol. Biol. Evol.* **31**, 872-888 (2014).

862 36. Conway, J.R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of
863 intersecting sets and their properties. *Bioinformatics*, **33**,2938-2940 (2017).

864 37. Beissinger, T.M., Rosa, G.J.M, Kaeppler, S.M., Gianola, D. & de Leon, N. Defining
865 window-boundaries for genomic analyses using smoothing spline techniques. *Genet Sel.*
866 *Evol.* **47**, 30 (2015).

867 38. Taus, T., Futschik, A. & Schlötterer, C. Quantifying selection with pool-seq time series data.
868 *Mol. Biol. Evol.,* **34**, 3023-3034 (2017).

869 39. Nei, M. & Tajima, F. DNA polymorphism detectable by restriction endonucleases. *Genetics*
870 **97**, 145-163 (1981).

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

873 41. Bloom, J.S. *et al.* Rare variants contribute disproportionately to quantitative trait variation in
874 yeast. *eLife* **8**,e49212 (2019).

875 42. Linder, R.A. *et al.* Adaptation in outbred sexual yeast is repeatable, polygenic, and favors
876 rare haplotypes. *bioXiv*, doi.org/10.1101/2021.08.27.457900 (2021).

877

878

879

880

881

882

883

884

885

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

886 **Figures**

887

888 Figure 1. Site frequency spectra (SFS) for all populations at each timepoint samples were taken
889 for DNA sequencing; immediately after construction, or "cycle 0" (left panels), after 6 cycles of
890 outcrossing (middle panels), and after 12 cycles of outcrossing (right panels).

891

892 Figure 2. Sliding window heterozygosity for all populations featured in this study after 12 cycles
893 of outcrossing.

894

895 Figure 3. Results from Pearson's χ^2 test comparing SNP frequencies initially (cycle 0), and after
896 12 cycles of outcrossing for all populations featured in this study. In each panel, red and green
897 boxes indicate regions of genome and genes associated with selection for frequent outcrossing
898 from Cubillos et al. (2013) and Burke et al. (2014), respectively

899

900 Figure 4. Haplotype diversity for all populations featured in this study after 12 cycles of
901 outcrossing. The dotted line in each panel indicates the maximum expected haplotype diversity
902 for each population.

903

904

905

906

907

908

909

910

911

912

913

914

915

916

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

917 Tables

918 Table 1. Strains used to create the synthetic populations featured in this study. Strains are
919 arranged in the rows in the table to indicate which specific pairs were crossed in the S-type
920 populations.

921

Populations	**Founder Strains**	
	Mat a	**Mat *alpha***
S4 & K4	DBVPG6765	YPS128
	DBVPG6044	Y12
S8 & K8	DBVPG6765	YJM981
	DBVPG6044	L_1528
	BC187	Y12
	L_1374	YPS128
S12 & K12	DBVPG6765	YJM981
	DBVPG6044	Y12
	BC187	L_1528
	SK1	273614N
	L_1374	YPS128
	YJM975	UWOPS05_217_3

922
923
924
925
926
927
928
929
930
931
932
933

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

934

935 Table 2. Number of possible SNPs and those actually observed in all synthetic populations at
936 each cycle where samples were taken for sequencing. Percentages are relative to total possible
937 number of SNPs for a given set of founders (n=4, 8, or 12). We estimate that approximately 15-
938 20 asexual generations occur between each cycle of outcrossing.
939

	S4	**K4**	**S8**	**K8**	**S12**	**K12**
Possible SNPs	112,738		122,663		155,516	
Cycle 0	91,658 **(81%)**	71,771 **(64%)**	98,864 **(81%)**	94,965 **(77%)**	101,134 **(65%)**	50,352 **(32%)**
Cycle 6	75,712 **(67%)**	69,954 **(62%)**	78,078 **(64%)**	62,537 **(51%)**	79,070 **(51%)**	38,853 **(25%)**
Cycle 12	57,405 **(51%)**	39,085 **(35%)**	80,941 **(66%)**	61,636 **(50%)**	73,894 **(48%)**	5,171 **(3%)**

940
941
942
943
944
945
946
947
948
949
950
951
952
953

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

954
955
956
957
958 Table 3. Mean genome-wide heterozygosity for all synthetic populations at each cycle where
959 samples were taken for sequencing. Mean heterozygosity is often, but not always, higher in the
960 S-type populations (shaded gray) relative to their K-type counterparts at a given timepoint.
961

	S4	**K4**	**S8**	**K8**	**S12**	**K12**
Cycle 0	0.38	0.47	0.36	0.28	0.36	0.31
Cycle 6	0.32	0.41	0.33	0.25	0.32	0.25
Cycle 12	0.20	0.17	0.31	0.25	0.28	0.04

962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

981
982
983
984
985
986 Table 4. Mean genome-wide haplotype frequencies for all founder populations in each synthetic
987 population based on SNP frequencies observed after 12 cycles of outcrossing (*note*: shaded cells
988 indicate that a founder not used for that particular population).

Founder strain	K4	S4	K8	S8	K12	S12
DBVPG6765	0.486	0.237	0.019	0.027	0.001	0.127
DBVPG6044	0.004	0.291	0.004	0.144	0.169	0.186
BC187			0.178	0.016	0.001	0.01
SK1					0.353	0.12
L_1374			0.23	0.23	0.001	0.011
YJM975					0.001	0.105
YPS128	0.012	0.249	0.009	0.111	0.002	0.005
Y12	0.498	0.222	0.331	0.325	0.468	0.21
273614N					0.001	0.065
L_1528			0.203	0.023	0.001	0.012
UWOPS05_217_3					0.002	0.018
YJM981			0.027	0.197	0.001	0.131

989
990
991
992
993
994
995

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

996
997
998
999
1000 Table 5. Comparison of putatively adaptive haplotypes from Burke et al. (2014) with observed
1001 haplotype frequency changes in synthetic populations used in this study. While both studies
1002 employed similar protocols Burke et al. (2014) implemented 18 outcrossing cycles while we only
1003 implemented 12. The 2014 study is used as a point of reference here because it features 12-fold
1004 replication, while in the present study features no within-treatment replication. We specifically
1005 targeted the window containing the most significant marker in each region based on the Burke et
1006 al. (2014) results. Notable (> 10%) increases in focal haplotype frequency are shown in green.
1007

Candidate region	Haplotype change in Burke et al. (2014)		Haplotype change in populations of the present study					
	Driving haplotype	Mean increase	**S4**	**K4**	**S8**	**K8**	**S12**	**K12**
A- C07:950000-975000	Y12	+0.23	+0.49	+0.29	+0.36	+0.09	+0.16	+0.66
B- C09:65000-80000	DBVPG6044	+0.24	+0.70	-0.01	-0.20	-0.16	+0.15	+0.49
C- C11:615000-620000	DBVPG6044	+0.29	-0.16	+0.01	-0.06	-0.15	-0.11	+0.49
E- C16:445000-460000	Y12	+0.59	+0.29	+0.51	+0.11	+0.93	+0.50	+0.98

1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445861; this version posted October 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

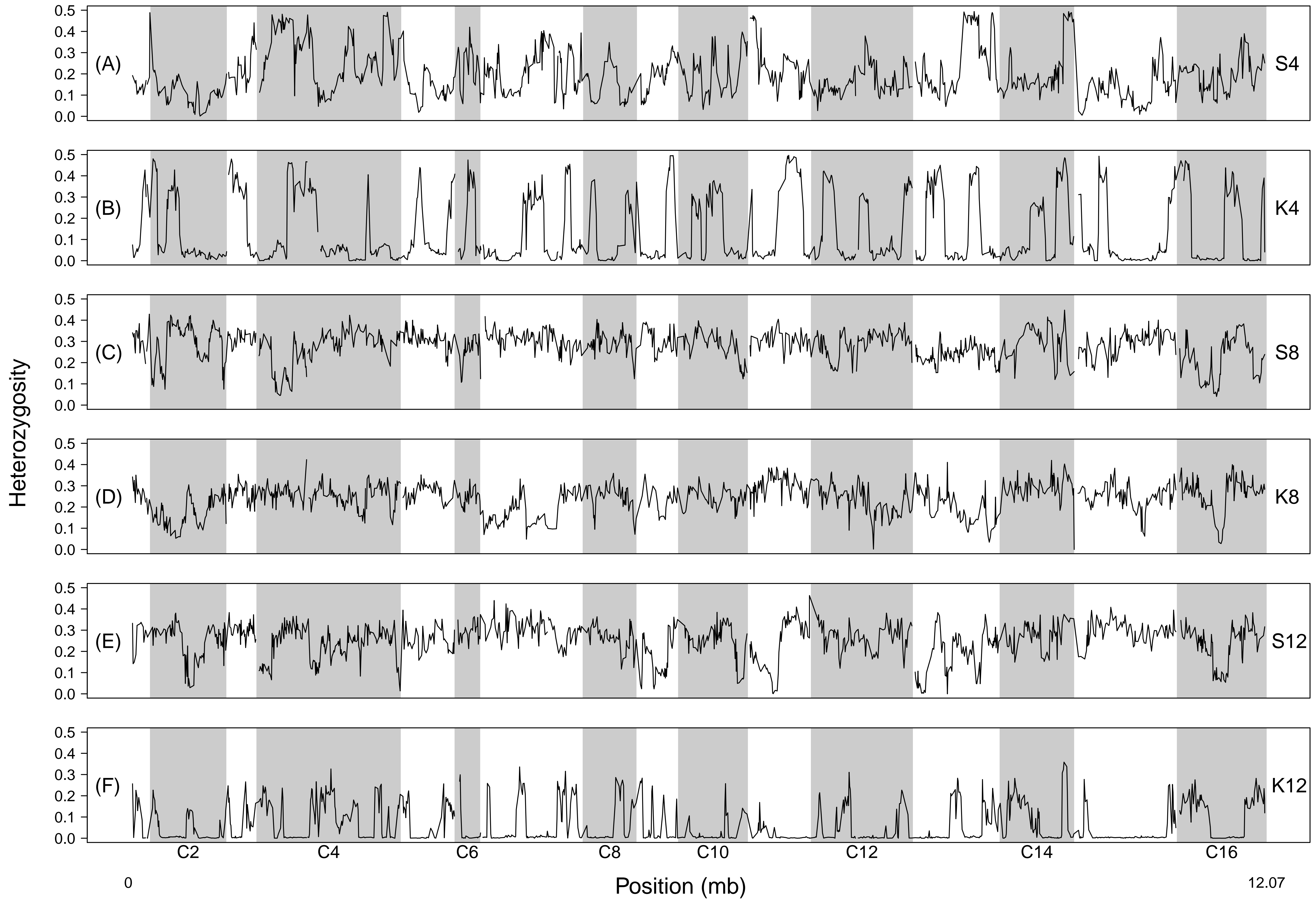
1019

1020

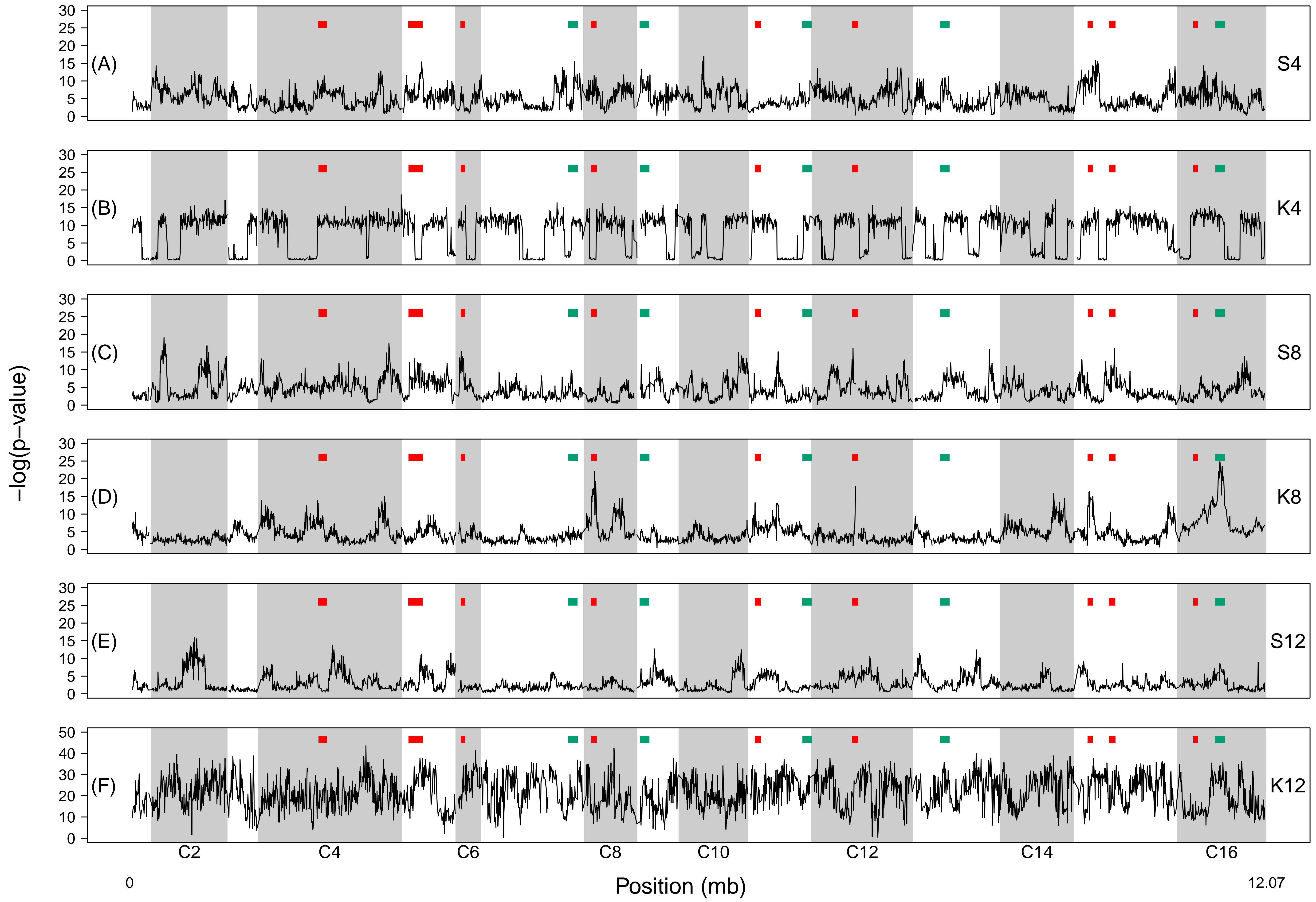
1021

1022

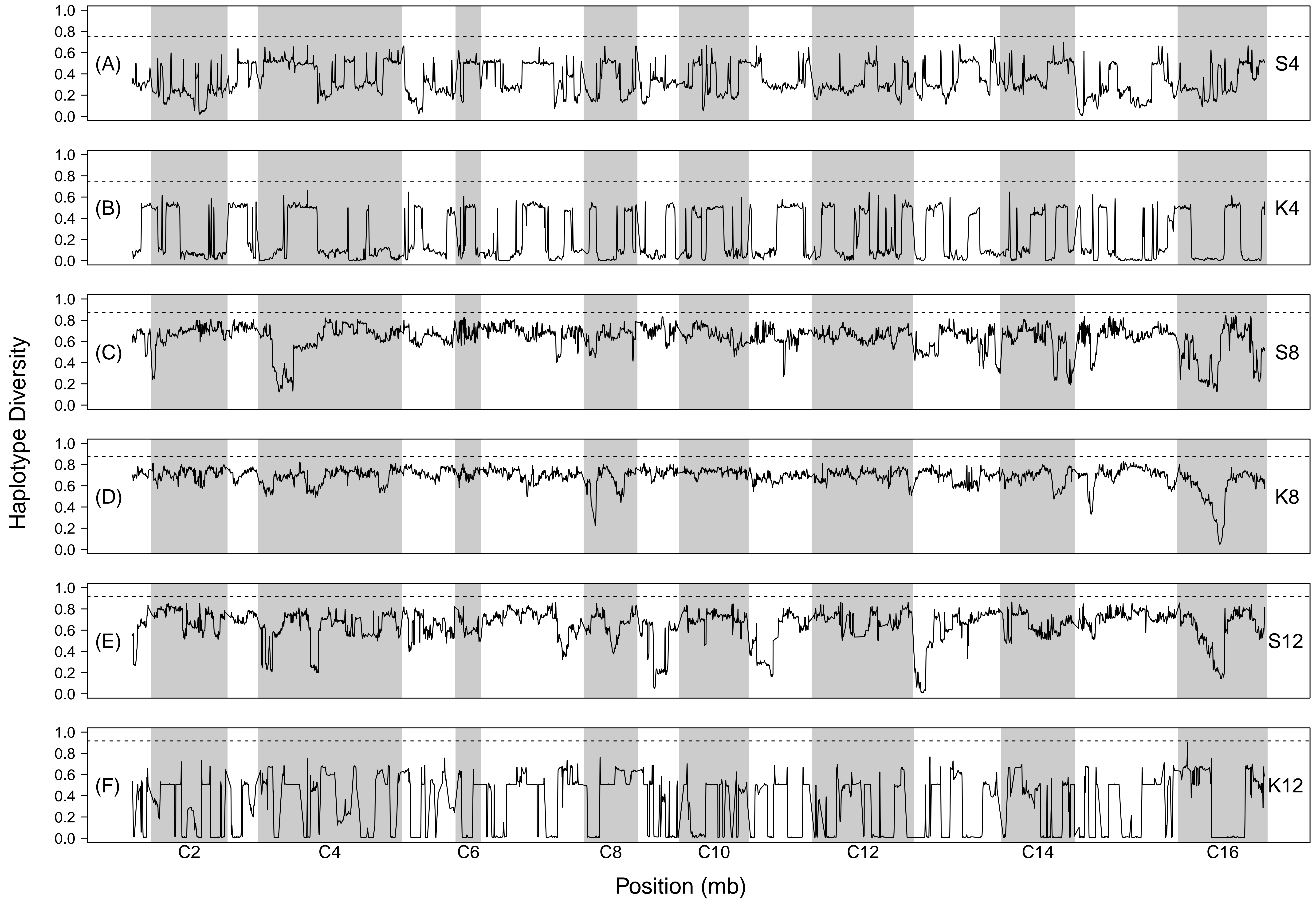
K4_0
K4_6
K4_12
S4_0
S4_6
S4_12
K8_0
K8_6
K8_12
S8_0
S8_6
S8_12
K12_0
K12_6
K12_12
S12_0
S12_6
S12_12
60000
40000
20000
0
Counts
0.00
0.25
0.50
0.75
1.00
SNP Frequency



Heterozygosity
(A)
S4
(B)
K4
(C)
S8
(D)
K8
(E)
S12
(F)
K12
0.0
0.1
0.2
0.3
0.4
0.5
C2
C4
C6
C8
C10
C12
C14
C16
0
12.07
Position (mb)



-log(p−value)
(A)
S4
(B)
K4
(C)
S8
(D)
K8
(E)
S12
(F)
K12
C2
C4
C6
C8
C10
C12
C14
C16
0
12.07
Position (mb)


Haplotype Diversity
1.0
0.8
0.6
0.4
0.2
0.0
(A)
S4
(B)
K4
(C)
S8
(D)
K8
(E)
S12
(F)
K12
C2
C4
C6
C8
C10
C12
C14
C16
Position (mb)