Received: 8 February 2020 | Revised: 29 September 2020 | Accepted: 30 September 2020

DOI: 10.1111/mec.15687

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Increased time sampling in an evolve-and-resequence experiment with outcrossing *Saccharomyces cerevisiae* reveals multiple paths of adaptive change

Mark A. Phillips[1] | Ian C. Kutch[1] | Anthony D. Long[2] | Molly K. Burke[1]

[1]Department of Integrative Biology, Oregon State University, Corvallis, OR, USA

[2]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA

Correspondence
Mark A. Phillips and Molly K. Burke, Department of Integrative Biology, Oregon State University, Corvallis, OR, USA.
Email: mphillips6789@gmail.com; molly.burke@oregonstate.edu

Funding information
National Science Foundation, Grant/Award Number: NSF 1906246; Oregon State University

Abstract

"Evolve and resequence" (E&R) studies combine experimental evolution and whole-genome sequencing to interrogate the genetics underlying adaptation. Due to ease of handling, E&R work with asexual organisms such as bacteria can employ optimized experimental design, with large experiments and many generations of selection. By contrast, E&R experiments with sexually reproducing organisms are more difficult to implement, and design parameters vary dramatically among studies. Thus, efforts have been made to assess how these differences, such as number of independent replicates, or size of experimental populations, impact inference. We add to this work by investigating the role of time sampling—the number of discrete time points sequence data are collected from evolving populations. Using data from an E&R experiment with outcrossing *Saccharomyces cerevisiae* in which populations were sequenced 17 times over ~540 generations, we address the following questions: (a) Do more time points improve the ability to identify candidate regions underlying selection? And (b) does high-resolution sampling provide unique insight into evolutionary processes driving adaptation? We find that while time sampling does not improve the ability to identify candidate regions, high-resolution sampling does provide valuable opportunities to characterize evolutionary dynamics. Increased time sampling reveals three distinct trajectories for adaptive alleles: one consistent with classic population genetic theory (i.e., models assuming constant selection coefficients), and two where trajectories suggest more context-dependent responses (i.e., models involving dynamic selection coefficients). We conclude that while time sampling has limited impact on candidate region identification, sampling eight or more time points has clear benefits for studying complex evolutionary dynamics.

KEYWORDS

adaptation, experimental evolution, genomics, population genomics

1 | INTRODUCTION

Experimental evolution seeks to characterize fundamental evolutionary processes by studying experimental populations across generations as they evolve in response to carefully imposed selective pressures (Rose & Garland 2009). Since its first documented use as a tool to study thermal adaptation (Dallinger, 1887), experimental evolution has furthered our understanding of evolutionary ecology (e.g., Chao et al., 1997; Chao & Levin, 1981; Crill et al., 2000; Kaltz & Bell, 2002; Turner & Chao, 1999), life history evolution (e.g., Luckinbill et al., 1984; Mueller & Ayala, 1981; Rose, 1984; Chippindale et al., 1997), and various aspects of evolutionary

physiology (e.g., Graves et al., 1992; Rose et al., 1992; Gibbs et al., 1997; Swallow et al., 1998; Roff & Fairbairn, 2007). More recently, the combination of experimental evolution with genome sequencing, termed "evolve and resequence" (E&R), has facilitated studying the genetic architecture of complex traits (Bailey & Bataillon, 2016; Long et al., 2015; Schlötterer et al., 2014). The E&R framework has been used to effectively probe the genetics of adaptation and complex traits in both asexual (Barrick & Lenski, 2013; Barrick et al., 2009; Blount et al., 2012; Kishimoto et al., 2010; Maddamsetti et al., 2015; Tenaillon et al., 2012, 2016; Woods et al., 2011) and sexual organisms (Burke et al., 2010, 2014; Franssen et al., 2015; Graves et al., 2017; Huang et al., 2014; Jha et al., 2015; Orozco-ter Wengel et al., 2012; Phillips et al., 2018; Tobler et al., 2014; Turner et al., 2011). One of the most notable findings across this body of work is that the process of adaptation appears to operate fundamentally differently in asexual versus sexual populations. Specifically, fixation of beneficial *de novo* mutations and historical contingency have emerged as defining features of asexual E&R studies, while adaptation from standing genetic variation defines sexual E&R studies (Burke, 2012; Long et al., 2015). Here our focus is on sexual E&R work, where due to practical limitations, it is often not possible to implement ideal experimental design.

At present, sexual E&R studies vary tremendously in experimental design. For instance, many studies feature two to three replicates per selection treatment (Huang et al., 2014; Jha et al., 2015; Orozco-ter Wengel et al., 2012; Turner et al., 2011; Zhou et al., 2011), while others feature five to more than 10 replicates per treatment (Burke et al., 2014; Carnes et al., 2015; Graves et al., 2017; Phillips et al., 2018). In addition, while some experiments feature fewer than 50 generations of selection (Orozco-ter Wengel et al., 2012; Tobler et al., 2014; Franssen et al., 2015; Jha et al., 2015), others feature hundreds (Burke et al., 2010, 2014; Carnes et al., 2015; Graves et al., 2017; Turner et al., 2011; Zhou et al., 2011). Simulation work predicts that the experimental parameters that vary widely across studies should in fact affect their statistical power (Baldwin-Brown et al., 2014; Kessner & Novembre, 2015; Kofler & Schlötterer, 2014). These simulations suggest that E&R efforts aimed at identifying the genetics underlying complex traits should maximize population size, duration and number of replicate populations, and take steps to minimize levels of linkage disequilibrium in starting populations. Replication in particular is thought to be an especially important determinant of statistical power, and strong empirical evidence supports this idea. In their study featuring 12 replicate outcrossing *Saccharomyces cerevisiae* populations sequenced at four time points, Burke et al. (2014) directly show that reducing the number of replicates in their analyses dramatically diminished their ability to dissect genomic regions responding to selection. Here we use an expanded version of the Burke et al. (2014) experiment to test another potentially important determinant of statistical power: the number of time points sampled. With genomic data from 17 time points collected over the course of a 540-generation evolution experiment, we examine how varying the number of time points used in our analyses affects our ability to identify candidate regions using standard regression techniques (e.g., Burke et al., 2014).

Beyond simply identifying candidate genomic regions associated with adaptive phenotypes, the E&R framework also allows the characterization of evolutionary dynamics underlying adaptation. This is especially relevant for experiments that collect data across multiple stages of adaptation, as opportunities to assess long-term evolutionary dynamics from experiments where genomic samples are compared before and after selection, or between selected and control populations, can be quite limited. For instance, the sexual E&R studies featuring the longest amount of time (approaching 1,000 generations of selection) report little evidence of fixation of alleles presumed to be beneficial in the selected populations (Burke et al., 2010; Graves et al., 2017; Phillips et al., 2016). Through population-genetic simulations, Phillips et al. (2016) also show that there is significantly more genetic variation and fewer instances of fixation in selected populations than would be predicted across a wide array of evolutionary scenarios. In these cases, the authors suggest explanations that run contrary to the classic prediction that beneficial alleles and closely linked loci simply increase in frequency until fixation (Maynard Smith & Haigh, 1974). Instead, they argue that perhaps these results reflect the effects of widespread balancing selection, or models of selection in which individual advantageous alleles do not ultimately fix. To give an example of one of the latter, Chevin and Hospital (2008) proposed a model where an initially rare beneficial allele does not reach fixation because its selective advantage is dynamic and inversely proportional to the distance to some new phenotypic optimum, and that new optimum is reached before the allele fixes due to shifts in beneficial alleles at other loci. However, as samples were not taken at multiple time points for sequencing in these studies, the authors of these studies did not have the power to directly assess whether the observed genomic response to selection might support one or more of these alternative hypotheses. For instance, how commonly might allele frequencies show clear initial increases in response to selection, then plateau?

One of the few sexual E&R experiments providing evidence for allelic trajectories that do not conform to classical predictions is a *Drosophila melanogaster* study (Orozco-ter Wengel et al., 2012). Here the authors sample and sequence pools of individuals from their experimental populations at multiple time points over the course of adaptation, and report evidence of two distinct classes of alleles: those that increase continuously over the entire course of selection, and those that exhibit a rapid shift in frequency at the onset of selection but plateau rather than fix. This study was limited to three replicates sequenced at three time points across ~ 40 generations, but it demonstrates that important biological phenomena could be rendered invisible in E&R studies without sufficient time sampling. As such, beyond assessing how time point sampling impacts our ability to identify candidate regions, we also examine how it affects our ability to make inferences about evolutionary processes. We examine the relationship between time point sampling and estimates of selection coefficients at polymorphic sites across the genome using the approach described by Iranmehr et al. (2017). We also examine

how time point sampling affects attempts to bin allele frequency trajectories into categories using an artificial neural network approach, self-organizing maps (SOM) (Kohonen, 1990). Our goal here is to characterize with high resolution the allelic trajectories that present over hundreds of generations of selection.

In summary, here we seek to address the following questions: (a) How might the number of individual time points sampled in an E&R study affect the ability to identify candidate regions underlying selection? (b) What might data from a large number of sampled time points teach us about evolutionary processes that cannot be learned from studies featuring fewer time points? We address these questions by evaluating a uniquely large data set—17 time points over the course of ~ 540 generations—and assessing how systematically excluding data from some of these time points changes our ability to draw inference.

2 | METHODS

2.1 | Yeast experimental evolution

The E&R experiment from which current data were collected is described in Burke et al. (2014). Briefly, a highly recombinant diploid population called the SGRP-4X (Cubillos et al., 2013) was used as the ancestral population for the selection experiment. This base population was created by crossing four geographically and genetically distinct "founder" strains from the *Saccharomyces* Resequencing Project: a European wine strain (DBVPG6765), a West African palm wine strain (DBVPG6044), a North American strain isolated from an oak tree (YPS128) and a Japanese sake strain (Y12). The SGRP-4X base population is highly efficient at sporulation and mating and contains auxotrophic markers that allow for mated diploids to be easily recovered in selective media. This base population was split into 12 replicate populations that were handled in parallel over the course of the selection experiment. These populations experienced growth in rich yeast–peptone–dextrose (YPD) media for 3 days per week, followed by a regime that forced populations to sporulate and mate in restrictive media. The experiment was carried out for 18 weeks, during which time there was one "outcrossing event" and an estimated ~ 30 cell doublings per week, for a total of 540 asexual generations. Every week prior to sporulation, a portion of each replicate population was archived at − 80°C at a final concentration of 15% glycerol. Additional details of the selection regime are described in Burke et al. (2014). For the current study, frozen stocks were revived for each of the 12 replicate populations at 17 distinct time points (the ancestor plus 16 of the 18 weeks).

2.2 | Genome sequencing, SNP identification and coverage assessment

To collect DNA from archived populations, 50 μl of each frozen stock (an estimated 5–50 million cells) was revived on a YPD agar plate and incubated for 48 hr at 37°C. The resulting lawns were sampled with wooden applicators into YPD liquid media and incubated at 37°C for 24 hr. For DNA extraction, 1 ml of this saturated liquid culture was processed with a Qiagen Gentra Puregene Yeast/Bacteria kit. Next-generation sequencing libraries were constructed using the Nextera Library Preparation Kit (Illumina). Libraries were combined in equal molarities and run on HiSeq3000 instruments at the OSU Center for Genomic Research and Biocomputing (10 lanes), as well as at the UCI High-Throughput Sequencing Facility (one lane). We have developed a processing pipeline for estimating allele frequencies in each population directly from our pooled sequence data. We used GATK version 4.0 (McKenna et al., 2010; Poplin et al., 2018) to align raw data (Phillips et al. 2020) to the *Saccharomyces cerevisiae* S288C reference genome (R64-2-1) using default filters and created a single VCF file for all variants identified across all replicate populations. We also downloaded and indexed a reference VCF file with single nucleotide polymorphism (SNP) information for 42 distinct natural isolates of *S. cerevisiae* (Bergström et al., 2014); including a VCF of known SNPs is a recommended best practice for calibrating base quality with GATK version 4.0. This VCF file was converted into an SNP frequency table suitable for downstream analysis with R (www.R-project.org). To further curate our SNP list, we chose to only include sites with a minimum coverage > 10× in each population and with minor allele frequencies between 0.02 and 0.98 across all populations. This additional filtering reduced our SNP table from 118,142 to 96,610 sites.

To assess coverage variation, we first calculated the coefficient of variation (standard deviation in coverage divided by mean coverage) for each sample. In next generation sequencing, this is a useful metric for assessing the uniformity of coverage within a sample (Feng et al., 2015). Values < 1 are ideal and indicate a low variance and uniform coverage across the genome (i.e., no indication of sequencing bias), while values > 1 indicate high variance and skewed coverage distributions. In addition to assessing coverage uniformity, we examined coverage patterns across the genome as an indirect means of identifying potential structural variation (duplications, deletions, etc.). Specifically, we looked at normalized coverage in each sample (coverage at a given site divided by the mean coverage of the sample) averaged across 10-kb windows and checked for regions where normalized coverage deviated from one.

2.3 | Data quality control details

Two recent studies (Iranmehr et al., 2017; Vlachos et al., 2019) that made use of the Burke et al. (2014) data set found that particular time points from five replicate populations displayed unusual patterns in their genome-wide site frequency spectra relative to the rest of the data set. As these patterns could not be easily explained from the available data, Iranmehr et al. (2017) and Vlachos et al. (2019) excluded these populations from their analysis. To address concerns regarding data quality, we evaluated genome-wide site frequency spectra in all new data. Based on our findings, we also opted to exclude these populations from our analysis (see

Supporting Information 1 for results). So, while we have made all data publicly available in the interest of transparency, the analyses described below were limited to data from seven replicate populations (replicates 3, 7, 8, 9, 10, 11 and 12).

2.4 | Regression analyses of allele frequency change

Following the approach of Burke et al. (2014), we used linear regression to model the square root arcsine-transformed allele frequencies as a means of identifying SNPs that changed consistently over time across replicate populations. The idea behind this approach is that allele frequencies that change in parallel across multiple replicate populations are probably targets of selection, and unlikely to have changed due solely to drift. While adaptive replicate-specific genomic responses have been observed in some E&R studies (e.g., Barghi et al., 2019), we chose to focus our efforts on identifying sites that show consistent responses across replicates. While replicate-specific candidate sites could potentially be validated as adaptive using functional genomic techniques, such techniques are not especially straightforward in outbred populations such as ours. Therefore, our best strategy for generating a list of strong candidate sites involves the increased confidence that comes with observing a repeatable response to selection. Many previous E&R studies have found evidence for widespread parallelism in allele frequency changes presumed to be adaptive (e.g., Burke et al., 2010, 2014; Graves et al., 2017; Huang et al., 2014; Jha et al., 2015; Turner et al., 2011), so focusing on the sites where we observe evidence of parallelism has a clear precedent in the literature.

We evaluated a number of models ranging from simple models where cycle ("cyc") was treated as a continuous regressor, to more complex models that included quadratic terms and population replicate ("rep") as a factor. As in Burke et al. (2014), we did not find any benefit to using more complex models (data not shown, but available upon request) and ultimately used a simple model of transformed allele frequency (Y) as a function of time:

$$Y = \beta_0 + \beta_1 \mathrm{cyc} + \in \quad (1)$$

To evaluate the impact of number of time points sampled on our ability to identify targets of selection, we compared results from the regression analysis using the entire data set (17 time points in total) to results from the analysis on four, eight and 13 time points (Table S1 lists the time points used in each iteration of the analysis). To establish significance thresholds for a given iteration of the analysis, we made Q-Q plots comparing observed p-values against expected null values and identified the point at which values began to diverge from the line of best fit (Figure S1). The best fit lines were generated using the "lm" function in R. This approach is similar to those used in genome-wide association studies (e.g., Wellcome Trust Case Control Consortium 2007).

After identifying candidate regions, we also estimated frequencies of founder haplotypes associated with our most significant SNP. This is possible because the genotypes for the four isogenic founder strains are unique and completely known. For any small region of the genome we can estimate the most likely set of founder haplotype frequencies that explain observed SNP frequency estimates within that region. This was accomplished following the methods described in Burke et al. (2014); briefly, we used sets of SNPs contained within 10-kb windows to determine the most likely set of founder haplotype frequencies that would produce the observed set of SNP frequencies, for each time point. Here our goal was to illustrate the degree to which data from additional time points allows better characterization of haplotype frequency changes driving adaptation.

2.5 | Self-organizing maps

Self-organizing maps (Kohonen, 1990) as implemented in the SOM package in R (Yan, 2016) were used to identify different classes of allele frequency trajectories present in our data set. We opted to use SOM instead of a conventional principal component analysis due its advantages from a visualization standpoint (Chattopadhyay et al., 2011) and in extracting patterns from noisy data (Liu et al., 2006). To evaluate how the number of time points used in the analysis impacted our ability to accurately define SNP trajectories, we compared results from applying the approach using all 17 time points and results from using four, eight and 13 time points (Table S1). As it was not feasible to apply this method across every polymorphic site in our data set, we ran tests under the following conditions: (a) candidate SNPs based on relevant regression analysis (e.g., for the four-time point test we used the candidate sites from the regression analysis using four time points); and (b) randomly selected SNPs matching the number of candidate sites from appropriate regression analysis (e.g., number of randomly selected SNPs in the four-time point test matched the number of candidate SNPs from the four-time point regression analysis). We chose to apply SOM to random SNPs to provide relevant comparisons with results from tests on candidate SNPs; the goal of this approach is to confirm that patterns enriched among our candidate sites were not simply patterns prevalent across the data set as a whole. For all of our analyses using SOM, we phased SNP frequencies so we were referencing the frequency of the less common nucleotide across the replicates at a given position. This was done to orient our trajectories in the same direction.

For SOM analyses, data matrices were created for each iteration of the data set where each row was an individual observation (i.e., allele frequency trajectory for a given SNP in a given population). As is standard practice with this particular approach, the data were normalized so that each row had a mean of 0 and a variance of 1 before applying the SOM algorithm (Yan, 2016). The "som" function was then applied to each data matrix using the following arguments: xdim = 3, and ydim = 1 (default settings were used for all other arguments). These parameters specify creating of a map with a total of three nodes in a 1×3 configuration.

We chose this map size (M) and configuration through a process of trial and error aimed at finding a configuration that

balanced minimizing map size and redundancy (i.e., nodes showing very similar trajectory patterns) without sacrificing definition in the trajectories represented. Tian et al. (2014) suggest that map size should be a function of the square root of the number of observations (in our case, SNP trajectories). Given the number of observations in our analyses, this formula typically resulted in $M > 500$ which returns uninformative maps with high levels of redundancy. Figure S2 shows an example of a large map of the candidate trajectories from the full 17-time point regression analysis that illustrates this issue (note: the 10×10 map shown here is still considerably smaller than what the formula suggests). In reality, there is no definitive way of selecting an optimal map size when performing an SOM analysis and the process usually involves a great deal of trial and error to find a configuration that minimizes redundant nodes (Kohonen, 2014). Testing smaller and smaller map configurations, we ultimately found that a 1×3 configuration captures the primary patterns that describe our candidate trajectories. While it was not practical to include figures for all of the configurations we examined, interested readers should be able to easily generate these results using the scripts and files we have made available through GitHub and Dryad, respectively.

To address concerns that the patterns identified by our SOM analysis are perhaps an artefact of using linear regression to identify candidate sites (i.e., biased towards linear trajectories), we also applied SOM to candidate sites identified by the Cochran–Mantel–Haenszel (CMH) test. The CMH test is commonly used in E&R studies, and simulations suggest it accurately identifies genomic targets of selection (Vlachos et al., 2019). Unlike our regression approach, the CMH test only requires data from the first and last time point of the experiment. As such, it has the power to reveal candidate sites that exhibit nonlinear responses to selection, should they exist. We performed a CMH analysis on our data based on its implementation in the "poolSeq" package in R (Taus et al., 2017), using the "cmh.test" function and data from cycle 1 and cycle 18. Using cycle 1 rather than "cycle 0"—the single ancestral population that all replicates were derived from—seemed more appropriate given the structure and assumptions of the CMH test. We then established a genome-wide significance threshold using the same approach as we did with our regression analysis to generate a list of candidate sites. With this list of CMH-test-generated candidate SNPs, we used SOM as described above to categorize the full allelic trajectories of those SNPs.

2.6 | Estimating selection coefficients

Selection coefficients (S) were estimated for all polymorphic sites in our data set using the method described in Iranmehr et al. (2017). This method, Composition of Likelihoods for Evolve-And-Resequence, or "CLEAR," was designed specifically to estimate population-genetic parameters such as effective population size (N_e) and selection coefficients from time-series genomic data generated by E&R experiments. Simulation work also suggests CLEAR is robust to experimental design and coverage variation (Iranmehr et al., 2017) and provides more accurate results than other comparable approaches (Vlachos et al., 2019). All analyses using this approach were conducted using scripts made publicly available by Iranmehr et al. (2017): https://github.com/airanmehr/CLEAR.

Using CLEAR, we generated N_e and S estimates using all 17 time points and compared them to results using four, eight and 13 time points (cf. Table S1). To run the analysis, a "sync" file containing data from relevant time points from our replicate populations for each iteration of the analysis was created based on the results of our SNP calling. This was done because the "sync" format, created by Kofler et al. (2011), serves as one of the primary input formats when running CLEAR. Lastly, CLEAR was run using default settings (e.g., python CLEAR.py --sync file.sync --out file.output). While in principle we could have analysed each replicate individually in each iteration of our CLEAR analysis, here, as before, we take advantage of our experiment's replicate structure and focus on sites where parallelism increases confidence in our results. After obtaining S estimates for each SNP in the data set, we compared expected SNP trajectories given those estimates to observed trajectories across our replicate populations. To generate expected SNP trajectories, we used a simple model for estimating allele frequencies over time assuming additivity (no dominance):

$$p_{t+1} = \frac{wAAp^2 + wAApq}{wAAp^2 + wAa2pq + waaq^2} \quad (2)$$

where $wAA = 1 + s$, $wAa = 1 + (0.5 * S)$, and $waa = 1$ (Hartl & Clark, 2007).

3 | RESULTS

3.1 | SNP identification and coverage assessment

Mean SNP coverage across our raw data set (i.e., prior to any sort of filtering based on coverage) was ~66× and mean coverage per sample ranged from 25× to 143×, with all but two samples having over 30× and all but 11 of our 193 samples having over 40× (Figure S3, Table S1). The coefficient of variation for each sample was less than 1, indicating there is low variance in coverage across the genome within each sample (Table S2). Next, as a means of detecting potential structural variation in the system (duplications, deletions, etc.) we examined normalized coverage across the genome averaged over 10-kb windows in each sample looking for regions where values diverged greatly from 1. We found one such region located on chromosome 12. Normalized coverage in this region was consistently higher than 1 in all samples (ranging from ~3 to ~7), while the next highest regions in other areas of the genome all had values < 2. The high-coverage SNPs that produced this result were all found in NTS1–2, a nontranscribed region of the rDNA repeat located between the 3′ ETS and RDN5. These inflated coverages are therefore probably an artefact due to the highly repetitive nature of this region.

3.2 | Regression analysis

Here we compared results from running our regression analysis on the entire data set (17 time points in total) to results from running the analysis on four, eight and 13 time points to see if there were any major differences in the candidate regions identified. As seen in Figure 1, we generally identify the same major candidate regions regardless of how many time points are included. While there appears to be a positive relationship between the number of candidate SNPs detected and the number of time points used in the analysis, the differences in total number of SNPs identified are minor when analyses feature eight or more time points (Table 1). By comparison, significantly fewer sites are identified when only four time points are included in the analysis. While our analyses featuring different numbers of time points implicate the same major genomic regions, individual candidate sites identified from down-sampled analyses do not always match those returned by the analysis of the full data set (Table 1). There is again a positive relationship between the level of overlap (i.e., number of SNPs identified in a given down-sampled analysis that are also identified in the analysis using all time points). For instance, when four time points are used, only 74% of the identified SNPs match the list generated by the analysis including data from all time points. However, the level of overlap increased to 90% when 13 time points are used. So, while the number of time points used in this style of analysis does not appear to dramatically impact our ability to detect candidate regions, it does appear to impact our ability to accurately identify specific candidate SNPs.

In addition to the results described above, we estimated haplotype frequencies associated with the most significant marker from our regression analysis using all 17 time points (chr16:456156). We qualitatively assessed how our interpretation of haplotype trajectories changes when the original four time points published in Burke et al. (2014) are used (Figure S4) versus when all 17 time points are used (Figure S5). We find that at the level of haplotypes, increasing the number of time points used does not necessarily impact our understanding of which founder haplotype is primarily driving the genomic response to selection. Consistent with Burke et al. (2014), we find that the haplotype frequencies associated with the Japanese sake founder strain (Y12) drive observed SNP frequency changes (cf. Figure 2b of Burke et al., 2014). While sequences from additional time points do not complicate this result, they do reveal with much higher resolution the path of haplotype frequency change—which in this case is an initial increase followed by a high-frequency plateau (Figures S4 and S5). However, the overall shapes of the trajectories are not qualitatively impacted by the addition of more time points. While not shown here, these general observations also extend to

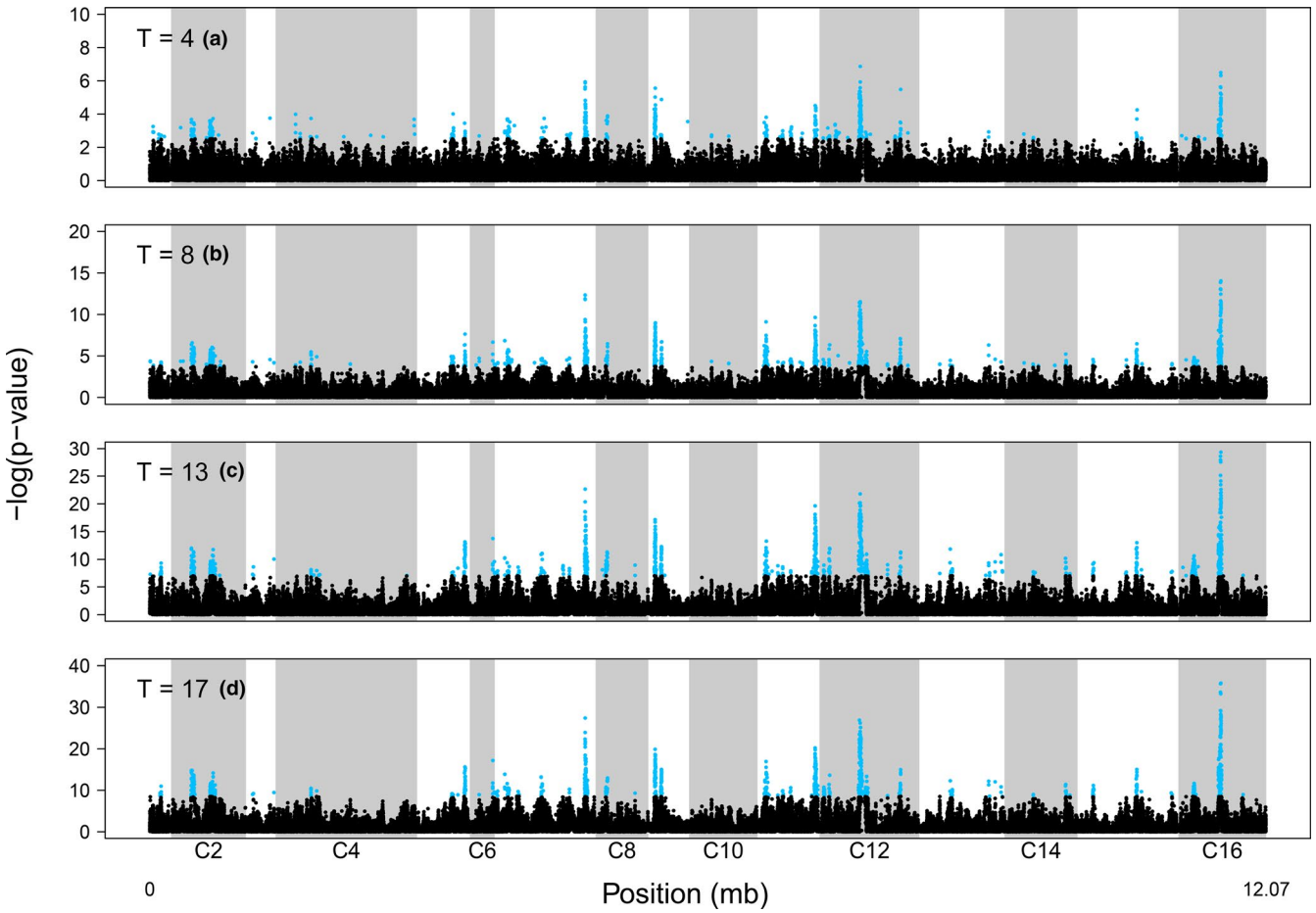

FIGURE 1 Results from regression analysis using four (a), eight (b), 13 (c) or all 17 (d) time points. SNPs that show significant and consistent changes in frequency across all replicates are shown in blue

TABLE 1 Number of candidate SNPs identified by regression analysis including data from decreasing numbers of time points

Time points sampled	Number of candidate SNPs	Overlap with full-data set analysis
All (17)	1,164	
13	1,096	887 (90%)
8	937	740 (86%)
4	555	396 (74%)

"Overlap" refers to the number of candidate SNPs identified in down-sampled analyses that are also identified by analysis including all time points. Sampling schemes always include the initial and final time points and are evenly spaced apart; Table S1 provides the details of all time points used.

what we see when we look at high- and low-resolution haplotype trajectories at other prominent peaks from our regression analysis. As such, we conclude that high-resolution time sampling does not do a great deal to enhance our understanding of haplotype trajectories.

3.3 | Categorizing SNP trajectories

Here we compared results from applying the SOM algorithm to SNP trajectories including all 17 time points to those featuring only four, eight and 13 time points to evaluate how different sampling schemes impacted our ability to accurately define common SNP trajectories. As practical limitations prevented us from applying this method using every SNP in our data set, we ran tests under the following conditions: (a) candidate SNPs based on relevant regression analysis; and (b) randomly selected SNPs matching the number of candidate sites from appropriate regression analysis.

We find that SOM can clearly identify and categorize SNP trajectories present in the data, when given collections of both random SNPs and candidate SNPs from our regression analyses (Figure 2). Broadly, we find distinct differences in the patterns identified from candidate SNPs versus random SNPs regardless of how many time points are used in a given analysis. However, increasing the number of time points used in a given analysis does produce more detailed trajectories across the board. When we apply the SOM algorithm to candidate SNPs and use eight or more time points, we find three common patterns: Category 1, little to no change during early stages of adaptation followed by an increase; Category 2, a gradual increase over the course of the experiment; and Category 3, an increase followed by a plateau (Figure 2b–e). For Category 3 trajectories, we find that while plateaus at later stages of adaptation do typically occur at high frequencies, they do not always involve fixation (see Figure 3 for details of SNP frequency distribution across this category, and Figure S6 for frequencies across replicates for a single representative site).

We also used SOM to categorize the trajectories of candidate sites returned from our CMH analysis. This was done to assess whether our main SOM results might be biased toward linear patterns, due to our use of linear regression to identify candidate sites (see Materials and Methods for more details). We find that despite some minor differences, the linear regression and CMH test approaches generally reveal the same significant candidate regions (Figure S7). We also find that SOM returns the same categories of allele frequency trajectories, whether we use candidate sites derived from our CMH or linear regression analyses (Figure S8). We therefore conclude that the SOM results presented in Figure 2 are robust to multiple strategies for genome-wide screening of candidate sites, and not merely an artefact of using linear regression.

One question that arises from our SOM results using candidate sites is: do SNPs within peaks identified in Figure 1 exhibit the same evolutionary dynamics? To address this question, we replotted the results shown in Figure 1(e), colour-coding each candidate SNP based on the most common trajectory category from the SOM analysis (Figure 4). For a candidate site to be placed in given category, SNP trajectories from four or more populations must fall into that category. Given the results displayed in Figure 2(e), one natural prediction is for our three categories to be proportionally represented in Figure 4. And we find this to be largely true, with Category 3 SNPs being the most represented followed by 1 and 2.

As seen in Figure 4, we find that the largest peaks appear to be dominated by a single trajectory type. To characterize this result more deeply, we examined individual SNP trajectories for each replicate across the five largest peaks present in Figure 1(e) and identified the most significant site within each peak. We then plotted the smoothed allele frequency trajectory at that site along with sites occurring within ± 4 kb of that position (Figures S9–S13). These adjacent sites were plotted as a means of giving some sense of how trajectory patterns shift with distance from our most significant markers. These figures lead us to three general observations. First, SNP trajectories appear largely consistent with expectations from SOM categories. Second, candidate sites do not always reach fixation. Third, there is often significant heterogeneity in SNP trajectories among replicate populations.

3.4 | Selection coefficient and effective population size estimates

S and N_e were estimated for our populations using CLEAR (Iranmehr et al., 2017). To evaluate how varying the number of time points used in our analysis impacted these estimates, we compared results from running the analysis with all 17 time points to results from running the analysis on four, eight and 13. Starting with our N_e estimates, we find that regardless of how may time points were used in a given iteration of our analysis, estimates always returned 2,000 haplotypes. This is consistent with Iranmehr et al.'s (2017) analysis of the four time points from this experiment previously published. Thus, the number of time points used in a given analysis does not appear to have any meaningful impact on the accuracy of N_e estimates.

We also find that varying the number of time points used in our analysis does not appear to dramatically impact estimates of S

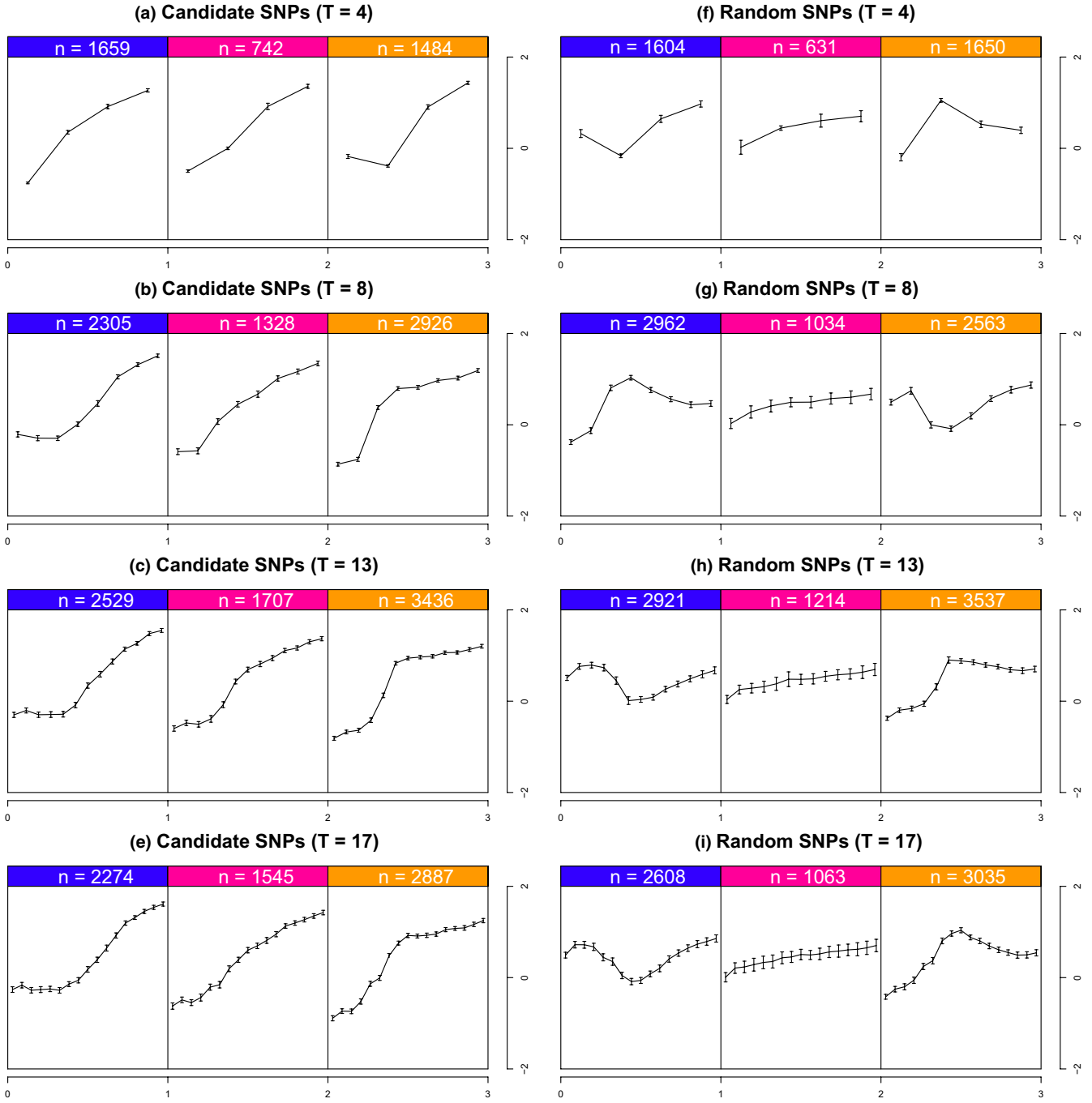

FIGURE 2 SOM results from analyses using four, eight, 13 and 17 time points. Results are from analyses performed on candidate sites from respective regression analysis (a–e), or an equivalent number of randomly selected sites. The number of sites that fall into each category is displayed above each individual plot, and error bars represent standard deviation. Trajectory patterns associated with candidate SNPs were binned into three broad categories (shown in e) for subsequent analysis: little initial change followed by an increase shown in blue (Category 1), a gradual and continuous increase over the course of the experiment shown in pink (Category 2), and an increase followed by a plateau shown in orange (Category 3)

(Figure 5). Looking across the genome, we find that estimates consistently peak in the same regions across all iterations of our analysis. Additionally, we find that estimates from down-sampled analyses and the full analysis using all time points are highly correlated (Figure S14, R^2 ranging from 0.95 to 0.99). We do observe that estimates are more highly correlated when more time points are used, but these differences appear to be fairly marginal.

Given that SOM results suggest that there are three major patterns among our candidate sites, we also compared expected SNP trajectories given our S estimates for the most significant sites displaying Category 1, 2 and 3 patterns from Figure 4(e). (Note: these sites also have some of our highest S values when estimates are made using all time points at 0.007, 0.014 and 0.012 respectively.) In the most significant Category 1 site, the expected trajectory given

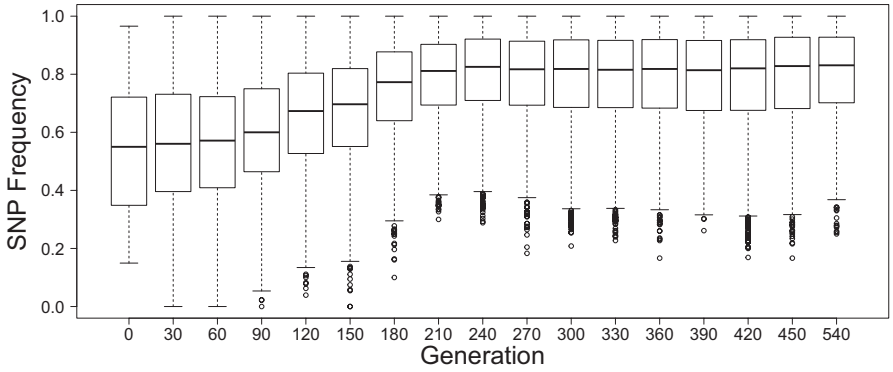

FIGURE 3 Boxplot of SNP frequencies at each time point for Category 3 sites. This category is defined by an initial increase followed by a plateau. As shown in the figure, plateaus are not always the result of fixation

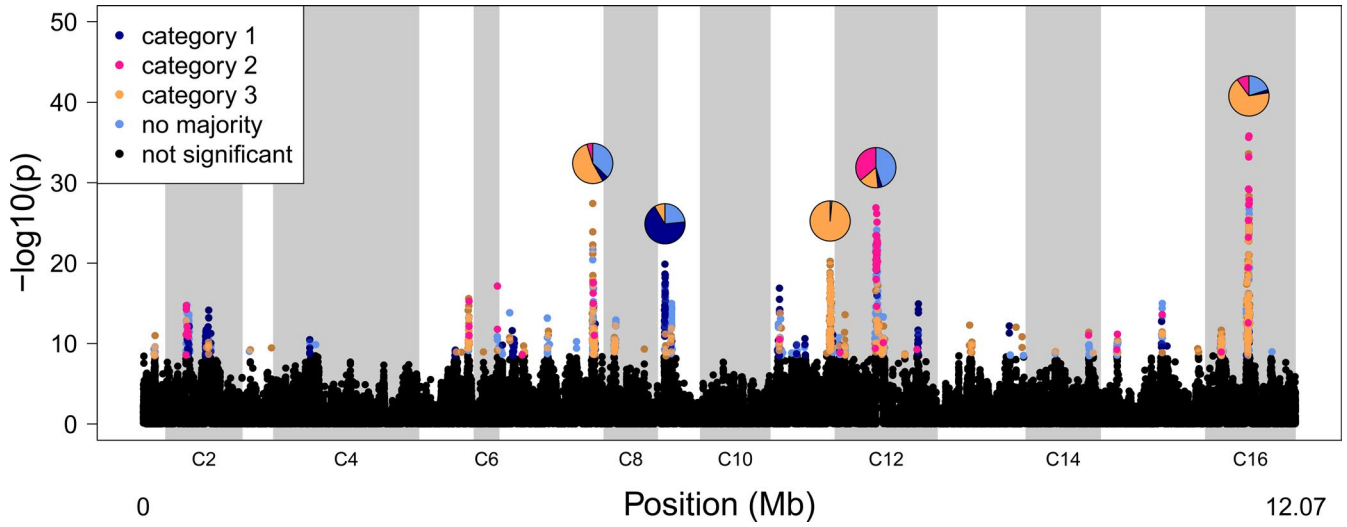

FIGURE 4 Results from regression analysis using all 17 time points. Candidate sites are colour-coded based on the most common trajectory pattern across all populations at a given site from SOM analysis. Dark blue indicates four or more populations have a Category 1 pattern (little to no change followed by an increase), pink indicates a Category 2 pattern (gradual increase) and orange indicates a Category 3 pattern (gradual increase followed by plateau). Points in light blue represent sites for which there is no majority pattern observed across replicates, and points in black sites with *p*-values below our significance threshold. Pie charts above the five most significant peaks show the proportion of significant sites in each region exhibiting each pattern

our *S* estimates do not qualitatively match observed trajectories, particularly during early portions of the experiment (Figure 6a). This is also the case for our most significant Category 3 SNP, except here the mismatch occurs primarily during middle portions of the experiment (Figure 6c). However, expected and observed trajectories are more consistent for the most significant Category 2 SNP (Figure 6b).

4 | DISCUSSION

4.1 | Increasing time-sampling alone does not improve candidate region discovery using regression analysis

Here we empirically assess how the number of time points sampled in an E&R study impacts our ability to identify regions of the genome that respond to selection. To this end, we compare results from a complete, time point-rich *Saccharomyces cerevisiae* E&R data set to results calculated using subsets of the data set. Our conclusions are premised on the assumption that candidate sites identified from the most time point-rich data set best represent the true targets of selection in this system.

From regression analyses, we find that increasing the number of time points used in a given analysis does not dramatically improve our ability to detect broad candidate regions (Figure 1). A such, our results suggest little need for high-resolution time sampling (e.g., 10 or more time points) in E&R studies aiming to identify broad candidate regions. This conclusion is further supported by the results of a separate CMH analysis (Figure S7). Despite using only the first and last time points of the experiment in this analysis, we find the same major peaks identified by the regression results. Taken together, these findings indicate a limited need for extensive time point sampling when the major goal of an E&R study is to identify candidate regions. Given that our ability to identify candidate regions is largely determined by the absolute allele frequency differences over the course of the experiment and consistency across replicates, we conclude that optimizing the length of an experiment should be prioritized over devoting resources to high-resolution

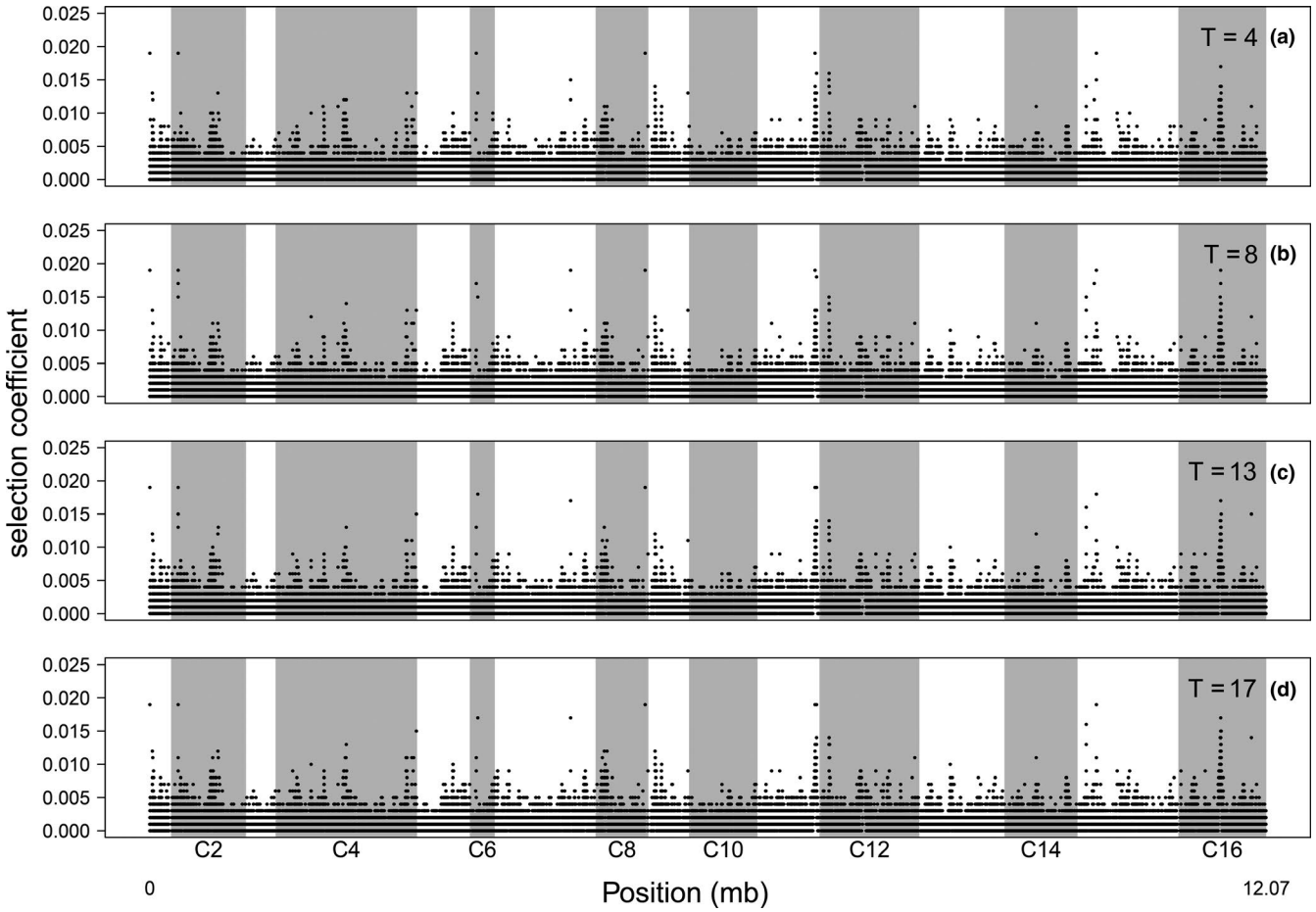

FIGURE 5 Selection coefficient estimates for all polymorphic sites in our data set from analyses using four (a), eight (b), 13 (c) and all 17 (d) time points

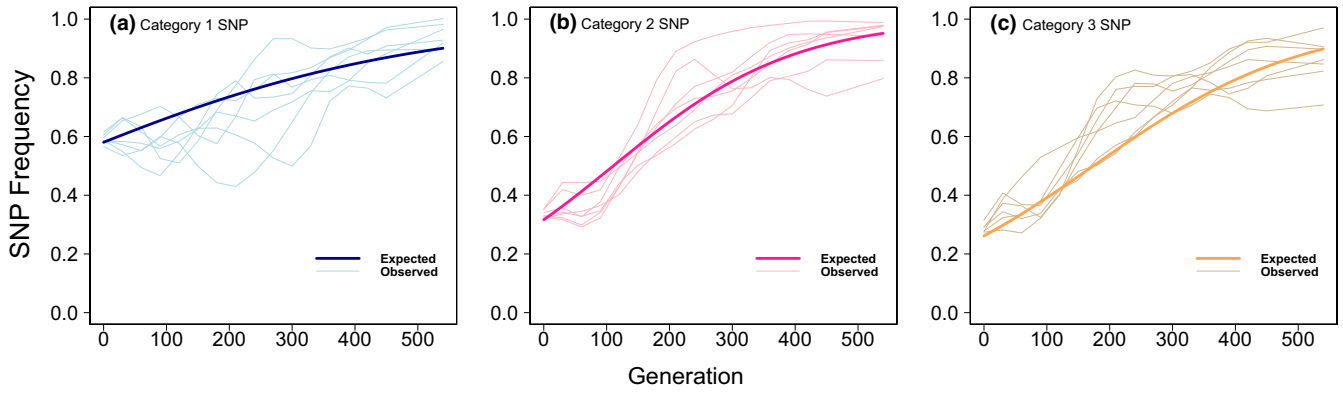

FIGURE 6 Expected and observed SNP trajectories for chr12:435756 (a), chr16:456156 (b) and chr16:452113 (c). These are the most significant Category 1, 2 and 3 SNPs in our data set based on regression and SOM analysis using all time points. Expected SNP trajectories are based on *S* estimates at these sites, 0.007, 0.014 and 0.012 respectively. Here we see that while the observed trajectories qualitatively match that expected for the Category 2 SNP, this is not the case for the Category 1 and 3 SNPs

time point sampling—provided that the experiment is highly replicated. This conclusion is consistent with recent work in *Drosophila* that empirically demonstrates the importance of experimental length in accurately identifying targets of selection (Langmüller and Schlötterer, 2020). In general, we expect these recommendations should also apply to E&R work in which the goal is to identify haplotype blocks underlying selection, rather than candidate SNPs (Barghi & Schlötterer, 2019; Michalak et al., 2018). This recommendation is further supported when considering that using more time points has virtually no impact on our interpretation of the primary haplotype underlying the genomic response to selection within our most significant candidate region (Figures S20 and S21).

In terms of what the above means for regression results previously presented in Burke et al. (2014), it is unlikely those findings are plagued by false positives despite being derived from an analysis of only four time points. While Burke et al. (2014) and the present study use the same regression approach, the former uses a permutation-based approach to establish significance thresholds versus the Q-Q plot-based method used here. We ultimately made this decision because the permutation approach of Burke et al. (2014) was not appropriate for analyses using large numbers of observations, and a method that could be uniformly applied to all versions of the analysis was preferable. When we compare the four time point regression analysis presented here to those of Burke et al. (2014), we find that the permutation approach produces a significantly higher significance threshold and, consequently, only a subset of the candidate regions found in this study are identified. As such, the analysis featured in Burke et al. (2014) is at worst too conservative.

While we tend to identify the same candidate regions regardless of the number of time points used in a given analysis, individual SNPs identified across different analyses varied notably. Moreover, using the complete analysis involving all time points as a reference point, we find that discrepancies at the SNP level increase as fewer time points are used in a given analysis (Table 1). As our regression approach is structured to identify SNPs that change in frequency consistently across replicate populations, one possible explanation for this discrepancy is that analyses with few time points might reflect inconsistencies across replicate populations due to chance and/or sampling error. As such, sampling many time points may be of value to studies interested in tracking specific, perhaps of a priori interest, candidate sites versus revealing larger candidate regions.

4.2 | Self-organizing maps reveal the shapes of allele frequency change

Self-organizing maps allowed us to visualize SNP frequency changes over the course of adaptation and categorize these trajectories according to how commonly they were observed. As such, we find that SOM can be a useful tool in E&R studies. In using SOM, we find that increasing the number of time points has clear benefits. For example, the plateau trajectory that emerges as common among candidate sites in the complete data set does not appear as clearly when only four time points are used. However, the benefit of additional time points tapers off when using more than eight time points (Figure 2).

When applying SOM to random collections of SNPs, we unsurprisingly get maps that appear more consistent with stochastic processes rather than selection. However, applying SOM to candidate sites from regression analyses led to a number of insights regarding the evolutionary processes driving adaptation. SOM results were consistent whether specific lists of candidate sites were generated by regression or CMH analysis, which suggests that these insights are not simply an artefact of our chosen test statistic (Figure S8). As best seen in Figure 2(e), the majority of SNP trajectories at candidate sites fall into one of three categories: little to no change in frequency during early stages of adaptation followed by an increase at later stages (Category 1), a gradual continuous increase (Category 2), or gradual increase in frequency during early stages of adaptation followed by a plateau around generation 270 (Category 3). Figure 4 also shows that as one might expect given linkage, sites with a given trajectory pattern tend to cluster with one another.

We find clear examples of Category 1 SNPs throughout the genome—notably in the most significant peak from our regression analysis on chromosome 9 (Figure S10)—that show little change or even decrease in frequency during early stages of adaptation but then increase at later stages. The prevalence of this pattern suggests some level of context-dependence and supports the idea that models featuring constant selection coefficients may not sufficiently account for the complexities underlying adaptation in sexual populations (Chevin & Hospital, 2008; Orozco-ter Wengel et al., 2012). This notion is further reinforced by comparing observed SNP trajectories at the most significant Category 1 site to textbook expectations given our estimate of S at this site (Figure 6a). Here the discrepancy between observed and expected, particularly during early cycles of the experiment, suggests a model assuming a constant S cannot fully account for the observed patterns. In contrast, Category 2 SNPs that exhibit gradual and continuous increases appear much more consistent with classic models (Figure 6b) and perhaps more representative of what we would expect in a typical soft sweep scenario (Burke, 2012).

The Category 3 trajectory pattern wherein SNPs show an initial increase in frequency but little to no change at later stages of adaptation also appears at odds with expectations from conventional population-genetic theory. For instance, we find that this pattern cannot always be attributed to candidate sites fixing before the end of the experiment (Figure 3). As such, we would argue this directly supports the hypothesis that alleles under selection will not necessarily fix in sexual populations (Burke et al., 2010; Chevin & Hospital, 2008; Phillips et al., 2016). A comparison of the expected versus observed SNP trajectory for our most significant SNP in this category given our S estimates (Figure 6c) again suggests that this sort of trajectory diverges from what we would expect from conventional models assuming a constant S.

It is worth noting we do not find any evidence here of the scenario described in Orozco-ter Wengel et al. (2012) where initially rare variants rapidly increase but stabilize at intermediate frequencies at later stages of adaptation. However, as our ancestral population was constructed by crossing only four strains of *S. cerevisiae*, most alleles occur at intermediate frequencies. In fact, only ~ 1% of polymorphic sites in the ancestor have starting frequencies less than 0.10 or greater than 0.90. Consequently, there is perhaps limited opportunity to observe the specific type of trajectory reported by Orozco-ter Wengel et al. (2012).

To generate insights into the patterns revealed by combining SOM and regression analysis (Figure 4), it was useful to simply examine the actual SNP trajectories within our most significant candidate peaks (Figures S9–S13). These figures provide clear evidence of heterogeneity among trajectory types within a given

peak, a pattern suggested by Figure 4. While the underlying source of this heterogeneity is unknown, possible explanations for it could include selected alleles being carried on more than one haplotype, or multiple targets of selection in close proximity to one another. In addition, these figures reveal that SNP trajectories in a given region exhibit variation within and among replicate populations. This apparent heterogeneity at the replicate level suggests that the high degree of parallelism often found in sexual E&R studies, in which sequence data are typically only collected at the end of the experiment (e.g., Graves et al., 2017), does not necessarily extend to all stages of adaptation. This is particularly interesting in light of findings from a recent E&R study in *Drosophila simulans* suggesting polygenic adaptation can result in reduced parallelism in the response to selection across replicate populations (Barghi et al., 2019). In their study, the authors report the existence of alleles that appear to respond to selection in some replicates but not others. As our own analysis focuses on sites that show parallel responses across replicates, we cannot comment on this particular issue. Instead, we observe a different form of heterogeneity in which a candidate site may show similar starting and ending frequencies in multiple replicate populations, but different replicate-specific evolutionary trajectories. This phenomenon may represent another consequence of polygenic adaptation.

All of this said, it is worth considering that our experiment does not feature a simple selective pressure (low nutrient environment, thermal stress, etc.). As such, our efforts to use this data set and SOM to answer questions regarding the underlying evolutionary processes driving adaptation may not utilize the full power of the E&R approach. However, we maintain that our work provides a valuable proof of concept that in-depth characterization of allele frequency trajectories from sequence data sampled at different stages of adaptation can indeed generate insights into the processes driving adaptation. Applying this approach in experiments featuring selection regimes where one might expect nonlinear or replicate-specific allelic trajectories a priori could be an especially fruitful future direction, such as in studies imposing selection for life history characters where extensive balancing selection owing to antagonistic pleiotropy is expected (Michalak et al., 2017).

4.3 | Increasing time-sampling alone does not impact selection coefficient estimation

In general, we find largely consistent *S* estimates, regardless of how many time points are used to derive those estimates; however, we do find some evidence for increased consistency when the number of time points used in a given analysis is increased. In other words, correlations between estimates from the complete data set and those calculated from subsets of the data set increase when more time points are used (Figure S14). However, given that results from the four-time point analysis are still highly correlated with those using all 17 ($R^2 = 0.95$), we conclude that increasing time point sampling does not meaningfully change estimates of selection coefficients. This conclusion corroborates the results of Taus et al. (2017), who analysed a range of simulated data sets and found that while increasing experimental parameters such as population size and replication improves the accuracy of selection parameter estimates from time series data, the number of time points has relatively minor effects. Notably, while our data suggest that increased time sampling does not change estimates of selection coefficients (as generated by existing software tools), they also cast some doubt on the accuracy of these estimates.

As discussed above, we find that Category 1 and Category 3 trajectories cannot fully be explained by theoretical predictions using *S* estimates from representative sites in those categories. So, while *S* estimates derived from CLEAR do not change regardless of how many time points are used in the analysis, they may not be fully accurate as the underlying methodology does not allow for the possibility of dynamic selection coefficients, nor for the related idea that selection coefficients at individual sites may be determined by the frequency of other selected sites (Burke & Long, 2012; Chevin & Hospital, 2008; Orozco-ter Wengel et al., 2012). However, should methods to accommodate these possibilities be developed, it seems reasonable to expect that the number of time points sampled could be more consequential than we report here. Low-resolution sampling may fail to fully characterize when and how allele frequencies change across different stages of adaptation. Thus, estimating *S* from a time point-rich data set such as this one cannot provide direct evidence in favour of a particular model of allele frequency change, at least using existing methods.

Lastly, when we compare genome-wide *S* estimates to our regression results (Figures 1 and 5), we find that there is a fair amount of overlap between peaks (i.e., our highest *S* estimates are often located in candidate regions from our regression analysis). However, this overlap is not perfect. Resolving this particular issue would require identifying the true targets of selection across our complete list of candidate regions, which in turn requires extensive experimental validation of those candidates (e.g., a large series of knockout experiments evaluating the impact of candidate genes on fitness across common genetic backgrounds). As such an undertaking falls outside the scope of this particular study, we cannot directly comment on why this relationship sometimes breaks down and can only speculate that perhaps it involves the assumption of a constant *S* when using CLEAR.

5 | Conclusion

Our results regarding the role of time point sampling in E&R studies can be generalized in ways that lead to experimental design recommendations. We believe our conclusions and recommendations should apply to studies featuring many replicates, large populations and moderate selection regimes. However, we remain agnostic about the utility of these recommendations for studies that deviate strongly from this sort of design (those with small populations, more intense selection regimes, etc.).

Within the context of E&R studies that do feature many replicates, large population sizes and moderate selection, we find that the number of time points sampled does not impact the ability to detect candidate regions. We identify the same broad candidate regions regardless of how many time points are analysed. Similarly, the number of time points does not meaningfully impact *S* estimates when using a method like CLEAR. However, our categorization of high-resolution allele frequency trajectories suggests that these *S* estimates may be fundamentally inaccurate, due to the assumptions underlying CLEAR, notably the assumption of static, rather than dynamic, selection coefficients. Conclusions such as this would not be possible without high-resolution time series data; therefore, we report that increasing the number of time points leads to an improved ability to characterize evolutionary processes underlying adaptation.

As for the question of when in an E&R study should populations be sampled, that will largely depend on the model organism used. *Drosophila* E&R studies tend to run for dozens of generations (Franssen et al., 2015; Huang et al., 2014; Jha et al., 2015; Orozco-ter Wengel et al., 2012; Tobler et al., 2014) compared to hundreds in this study; thus none of the sampling schemes in Table S1 would be appropriate. Instead, sampling schemes should be tailored to known rates of adaptation and the duration of a given experiment. To give a concrete example, in *Drosophila* E&R studies most of the phenotypic and genomic response to selection we can expect to observe occurs within ~40 generations (Graves et al., 2017). As such, our results here suggest that an appropriate sampling scheme for a *Drosophila* experiment might involve ~50 generations of selection with samples being taken for sequencing every seven to 10 generations.

Finally, this study of E&R data from 17 time points collected from 12 replicate populations over the course of hundreds of generations represents the largest data set of this kind in a sexual eukaryote. Due to the massive size of this data set, we observed many patterns inconsistent with simple interpretations of the effects of our selection protocol. With so many individual data points, we have a unique opportunity to evaluate these patterns in the context of the rest of the data set. Ultimately, we chose to exclude many of these data points in an effort to remain conservative. However, we argue that, in general, an added benefit of sequencing many time points in E&R work is the ability to develop a deeper understanding of individual experiments. While simulations of evolutionary experiments are very useful in terms of shaping our expectations of the consequences of adaptation, empirical data teach us more about what happens in practice. However, laboratory experiments are never perfect, and populations in nature do not obey simple rules. Only from rich data sets can we hope to untangle the signal from the noise.

ACKNOWLEDGEMENTS

We thank Gianni Liti for sharing the SGRP-4X population, from which all of the replicate populations featured in this study were derived. We also thank Shawn O'Neil and Dana Gibbon for providing valuable discussion and help with our bioinformatic pipeline. OSU's Center for Genome Research and Biocomputing and UC Irvine's Genomics and High Throughput Facility sequenced the many samples taken over the course of this study. This work was supported by startup funds to M.K.B. (OSU's College of Science), and M.A.P. is supported by an NSF Postdoctoral Fellowship (NSF 1906246).

AUTHOR CONTRIBUTIONS

M.K.B. conceived of and supervised the project. M.K.B. and I.C.K. did the laboratory work necessary to generate the data set used this study. M.A.P., M.K.B. and A.D.L. formulated the analytical strategy featured in this study. M.A.P. carried out all primary data analyses. M.A.P. and M.K.B. wrote the manuscript.

DATA AVAILABILITY STATEMENT

The raw sequence files generated over the course of this project are available through NCBI SRA (BioProject ID: PRJNA554138). Core data and result files (tables with SNP and haplotype frequencies, results from regression analysis, and *S* estimates) are available through Dryad (https://doi.org/10.5061/dryad.8gtht76mz). Core scripts necessary to reproduce our results are available through Github (https://github.com/mphillips67/DeepSeq_TimeSeries_Project).

ORCID

Mark A. Phillips https://orcid.org/0000-0002-2319-7913
Molly K. Burke https://orcid.org/0000-0002-4280-0247

REFERENCES

Bailey, S. F., & Bataillon, T. (2016). Can the experimental evolution programme help us elucidate the genetic basis of adaptation in nature? *Molecular Ecology*, *25*, 203–216.

Baldwin-Brown, J. G., Long, A. D., & Thornton, K. R. (2014). The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. *Molecular Biology and Evolution*, *31*, 1040–1055.

Barghi, N., & Schlötterer, C. (2019). Shifting the paradigm in evolve and resequence studies, from analysis of single nucleotide polymorphism to selected haplotype blocks. *Molecular Ecology*, *28*, 521–524.

Barghi, N., Tobler, R., Nolte, V., Jaksic, A. M., Mallard, F., Otte, K. A., Dolezal, M., Taus, T., & Kofler, R. & Schlötterer, C. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biology*, *17*, e3000128.

Barrick, J. E., & Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nature Reviews Genetics*, *14*, 827–839.

Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schnedier, D., Lenski, R. E. & Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, *461*, 1243–1274.

Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, *31*, 872–888.

Blount, Z. D., Barrick, J. E., Davidson, C. J., & Lenski, R. E. (2012). Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature*, *489*, 513–518.

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

Burke, M. K., Dunham, J. P., Sharestani, P., Thornton, K. R., Rose, M. R., & Long, A. D. (2010). Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature*, *467*, 587–590.

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

Burke, M. K., & Long, A. D. (2012). What paths do advantageous alleles take during short-term evolutionary change? *Molecular Ecology*, *21*, 4913–4916. https://doi.org/10.1111/j.1365-294X.2012.05745.x

Carnes, M. U., Campbell, T., Huang, W., Butler, D. G., Carbone, M. A., Duncan, L. H., Harbajan, S. V., King, E. M., Peterson, K. R., Weitzel, A., Zhou, S., & Mackay, T. F. C. (2015). The genomic basis of postponed senescence in *Drosophila melanogaster*. *PLoS One*, *10*, e0138569. https://doi.org/10.1371/journal.pone.0138569

Chao, L., & Levin, B. R. (1981). Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proceedings of the National Academy of Sciences USA*, *78*, 6324–6328. https://doi.org/10.1073/pnas.78.10.6324

Chao, L., Levin, B. R., & Stewart, F. M. (1997). A complex community in a simple habitat, an experimental study with bacteria and phage. *Ecology*, *58*, 369–378. https://doi.org/10.2307/1935611

Chattopadhyay, M., Dan, P. K., & Majumdar, S. (2011). Principal component analysis and self organizing map visual clustering of machine-part cell formation in cellular manufacturing system. *Systems Research Forum*, *5*, 25–51.

Chevin, L. M., & Hospital, F. (2008). Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics*, *180*, 1645–1660.

Chippindale, A. K., Alipaz, J. A., Chen, H.-W., & Rose, M. R. (1997). Experimental evolution of accelerated development in Drosophila. 1. *Development Speed and Larval Survival*. *Evolution*, *51*, 1536–1551.

Crill, W. D., Wichman, H. A., & Bull, J. J. (2000). Evolutionary reversals during viral adaptation to alternating hosts. *Genetics*, *154*, 27–37.

Cubillos, F. A., Parts, L., Salinas, F., Bergström, A., Scovacricchi, E., Zia, A., Illingworth, C. J. R., Mustonen, V., Ibstedt, S., Warringer, J., Louis, E. J., Durbin, R., & Liti, G. (2013). High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. *Genetics*, *195*, 1141–1155. https://doi.org/10.1534/genetics.113.155515

Dallinger, W. (1887). The president's address. *Journal of the Royal Microscopical Society*, *10*, 184–199. https://doi.org/10.1111/j.1365-2818.1887.tb01566.x

Feng, Y., Chen, D., Wang, G. L., Zhang, V. W., & Wong, L. J. (2015). Improved molecular diagnosis by the detection of exonic deletions with target gene capture and deep sequencing. *Genetics in Medicine*, *17*, 99–107. https://doi.org/10.1038/gim.2014.80

Franssen, S. U., Nolte, V., Tobler, R., & Schlötterer, C. (2015). Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental *Drosophila* melanogaster populations. *Molecular Biology and Evolution*, *32*, 495–509.

Garland, T., & Rose, M. R. (2009). *Experimental evolution*. University of California Press.

Gibbs, A. G., Chippindale, A. K., & Rose, M. R. (1997). Physiological mechanisms of evolved desiccation resistance in Drosophila melanogaster. *Journal of Experimental Biology*, *200*, 1821–1832.

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

Graves, J. L., Toolson, E. C., Jeong, C., Vu, L. N., & Rose, M. R. (1992). Desiccation, flight, glycogen, and posted senescence in *Drosophila melanogaster*. *Physiological Zoology*, *65*, 268–286.

Hartl, D. L., & Clark, A. G. (2007). *Principles of population genetics*. :Sinauer Associates Incorporated.

Huang, Y., Wright, S. I., & Agrawal, A. F. (2014). Genome-wide patterns of genetic variation within and among alternative selective regimes. *PLoS Genetics*, *10*, e1004527. https://doi.org/10.1371/journal.pgen.1004527

Iranmehr, A., Akbari, A., Schlötterer, C., & Bafna, V. (2017). Clear: Composition of likelihoods for evolve and resequence experiments. *Genetics*, *206*, 1011–1023. https://doi.org/10.1534/genetics.116.197566

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

Kaltz, O., & Bell, G. (2002). The ecology and genetics of fitness in Chlamydomonas. XII. Repeated sexual episodes increase rates of adaptation to novel environments. *Evolution*, *56*, 1743–1753. https://doi.org/10.1111/j.0014-3820.2002.tb00188.x

Kessner, D., & Novembre, J. (2015). Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits. *Genetics*, *199*, 991–1005. https://doi.org/10.1534/genetics.115.175075

Kishimoto, T., Iijima, L., Tatsumi, M., Ono, N., Oyake, A., Hashimoto, T., Matsuo, M., Okubo, M., Suzuki, S., Mori, K., Kashiwagi, A., Furusawa, C., Ying, B.-W., & Yomo, T. (2010). Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution. *Plos Genetics*, *6*, e1001164. https://doi.org/10.1371/journal.pgen.1001164

Kofler, R., Pandey, V., & Schlötterer, C. (2011). Popoolation 2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-seq). *Bioinformatics*, *27*, 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Kofler, R., & Schlötterer, C. (2014). A guide for the design of evolve and resquencing studies. *Molecular Biology and Evolution*, *31*, 473–482.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*, 1454–1480. https://doi.org/10.1109/5.58325

Kohonen, T. (2014). *Matlab implementation and applications of the self-organizing map*. Unigrafia Oy.

Langmüller A. M., & Schlötterer C. (2020). Low concordance of short-term and long-term selection responses in experimental Drosophila populations. *Molecular Ecology*, *29*,(18), 3466–3475. http://dx.doi.org/10.1111/mec.15579

Liu, Y., Weisberg, R. H., & Mooers, C. N. (2006). Performance evaluation of the self-organzing map for feature extraction. *Journal of Geophysical Research: Oceans*, *111*, C05018.

Long, A. D., Liti, G., Luptak, A., & Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature Reveiews Genetics*, *16*, 567–582. https://doi.org/10.1038/nrg3937

Luckinbill, L. S., Arking, R., Clare, M. J., Cirocco, W. C., & Buck, S. A. (1984). Selection for delayed senescence in *Drosophila melanogaster*. *Evolution*, *38*, 996–1003.

Maddamsetti, R., Lenski, R. E., & Barrick, J. E. (2015). Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics*, *200*, 619–663.

Maynard-Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*, 23–35. https://doi.org/10.1017/S0016672300014634

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

Michalak, P., Kang, L., Pernille, P. M., Schou, M. F., & Loeschke, V. (2017). Nucleotide diversity inflation as a genome-wide response to experimental lifespan extension in *Drosophila melanogaster*. *BMC Genomics*, *18*, 84. https://doi.org/10.1186/s12864-017-3485-0

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

Mueller, L. D., & Ayala, F. J. (1981). Trade-off between r-selection and k-selection in Drosophila populations. *Proceedings of the National Academy of Sciences USA*, *78*, 1303–1305. https://doi.org/10.1073/pnas.78.2.1303

Orozco-ter Wengel, P., Kapun, M., Nolte, V., Kofler, R., Flatt, T., & Schlötterer, C. (2012). Adaptation of *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories of selected traits. *Molecular Ecology*, *21*, 4931–4941.

Phillips, M. A., Kutch, I. C., Long, A. D., & Burke, M. B. (2020). Increased time sampling in an evolve-and-resequence experiment with outcrossing Saccharomyces cerevisiae reveals multiple paths of adaptive change. NCBI SRA, BioProject ID: PRJNA554138.

Phillips, M. A., Long, A. D., Greenspan, Z. S., Greer, L. F., Burke, M. K., Villeponteau, B., Matsagas, K. C., Rizza, C. L., Mueller, L. D., & Rose, M. R. (2016). Genome-wide analysis of long-term evolutionary domestication in *Drosophila melanogaster*. *Scientific Reports*, *6*, 39281. https://doi.org/10.1038/srep39281

Phillips, M. A., Rutledge, G. A., Kezos, J. N., Greenspan, Z. S., Talbott, A., Matty, S., Arain, H., Mueller, L. D., Rose, M. R., & Shahrestani, P. (2018). Effects of evolutionary history on genome wide and phenotypic convergence in Drosophila populations. *BMC Genomics*, *19*, 743. https://doi.org/10.1186/s12864-018-5118-7

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., Roazen, D. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. https://doi.org/10.1101/201178

Roff, D. A., & Fairbairn, D. J. (2007). Laboratory evolution of the migratory polymorphism in the sand cricket: Combining physiology with quantitative genetics. *Physiological and Biochemical Zoology*, *80*, 358–369. https://doi.org/10.1086/518012

Rose, M. R. (1984). Laboratory evolution of postponed senescence in *Drosophila melanogaster*. *Evolution*, *38*, 1004–1010.

Rose, M. R., Vu, L. N., Park, S. U., & Graves, J. L. (1992). Selection for stress resistance increases longevity in *Drosophila melanogaster*. *Experimental Gerontology*, *27*, 241–250.

Schlötterer, C., Tolber, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of inidviduals—mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, *15*, 749–763.

Swallow, J. G., Garland, T. Jr, Carter, P. A., Zhan, W., & Sieck, G. C. (1998). Effects of voluntary activity and genetic selection on aerobic capacity in house mice (*Mus domesticus*). *Journal of Applied Physiology*, *84*, 69–76.

Taus, T., Futschik, A., & Schlötterer, C. (2017). Quantifying selection with pool-seq time series data. *Molecular Biology and Evolution*, *34*, 3023–3034. https://doi.org/10.1093/molbev/msx225

Tenaillon, O., Barrick, J. E., Ribeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., Wu, G. C., Wielgoss, S., Cruveiller, S., Médigue, C., Schneider, D., & Lenski, R. E. (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*, *11*, 165–170. https://doi.org/10.1038/nature18959

Tenaillon, O., Rodriguez-Verdugo, A., Gaut, R. L., McDonald, P., Bennett, A. F., Long, A. D., & Gaut, B. S. (2012). The molecular diversity of adaptive convergence. *Science*, *335*, 457–461. https://doi.org/10.1126/science.1212986

The Wellcome Trust Case Control Consortium (2006). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*, 661–678.

Tian, J., Azarian, M. H., & Michael, P. (2014). Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm. Second European Conference of the Prognostics and Health management Society 2014.

Tobler, R., Franssen, S. U., Kofler, R., Orozco-terWengel, P., Nolte, V., Hermisson, J., & Schlötterer, C. (2014). Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Molecular Biology and Evolution*, *31*, 364–375. https://doi.org/10.1093/molbev/mst205

Turner, P. E., & Chao, L. (1999). Prisoner's Dilemma in an RNA virus. *Nature*, *398*, 441–443. https://doi.org/10.1038/18913

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

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

Woods, R. J., Barrick, J. E., Cooper, T. F., Shrestha, U., Kauth, M. R., & Lenski, R. E. (2011). Second-order selection for evolvability in a large *Escherichia coli* population. *Science*, *331*, 1433–1436. https://doi.org/10.1126/science.1198914

Yan, J. (2016) SOM. R-project CRAN.

Zhou, D., Udpa, N., Gersten, M., Visk, D. W., Bashir, A., Xue, J., Frazer, K. A., Posakony, J. W., Subramaniam, S., Bafna, V., & Haddad, G. G. (2011). Experimental selection of hypoxia-tolerant *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences USA*, *108*, 2349–2354. https://doi.org/10.1073/pnas.1010643108

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Phillips MA, Kutch IC, Long AD, Burke MK. Increased time sampling in an evolve-and-resequence experiment with outcrossing *Saccharomyces cerevisiae* reveals multiple paths of adaptive change. *Mol Ecol*. 2020;29:4898–4912. https://doi.org/10.1111/mec.15687