An examination of the evolve-and-resequence method using *Drosophila simulans*

John K. Kelly[1*] and Kimberly A. Hughes[2]

[1]Department of Ecology and Evolutionary Biology, University of Kansas,
1200 Sunnyside Ave., Lawrence, KS, 66045, USA

[2]Department of Biological Science, Florida State University,
319 Stadium Dr., Tallahassee, FL 32306 USA

[*]Correspondence: jkk@ku.edu

Abstract

We develop a set of analytical and simulation tools for Evolve-and-Resequence (E&R) experiments and apply them to a new study of rapid evolution in *Drosophila simulans*. Likelihood based test statistics combined with permutation suggest parallel evolution of about 120 polymorphisms across the genome. Whole genome simulations calibrated from several Drosophila genetic datasets indicate this response could be generated by as few as 20 loci under strong directional selection, with a corresponding hitch-hiking effect. While this resolves the "excessive significance" paradox typical of E&R experiments, a rank-based correlation analysis of genomic windows that suggests that weaker selection might be acting at a much larger number of sites. Finally, the allele frequency spectrum at sites that experience strong selection in the novel laboratory environment suggests that loci responding to selection may be balanced polymorphisms in nature.

Key words: genomics, Evolve-and-Resequence, *Drosophila simulans*

Introduction

In Evolve-and-Resequence (E&R) experiments, populations evolve within one or more controlled environments and are then surveyed with genomic sequencing (NUZHDIN AND TURNER 2013; LONG *et al.* 2015). The volume of data that emerges from an E&R study is remarkable, typically allele frequency changes at hundreds of thousands of loci within replicated populations. While researchers naturally focus on the small fraction of sites exhibiting the

largest or most consistent changes, a wealth of information resides in the 'background response', i.e. the evolution observed of polymorphisms that are not direct targets of selection (the overwhelming majority of the genome). In this paper, we present an analytical framework for E&R studies, first to provide more detailed predictions regarding whole genome evolution, and second to robustly detect loci under parallel selection across replicate populations. We apply the method to results from a new E&R experiment on *Drosophila simulans* designed to answer two major questions: What is the genomic basis of rapid adaptation to a novel environment? What do the features of the genetic response tell us about the maintenance of polymorphisms in nature?

***The genetic basis of rapid adaptation*—**The traditional view of the tempo of adaptive evolution is that evolutionary change is slow relative to the ecological processes that influence contemporary populations (SLOBODKIN 1980; GILLESPIE 1991). In this paradigm, genetic change does not interact with ecological and demographic processes over the short time scale (few to several generations) encompassed by ecological processes (THOMPSON 1998; HENDRY AND KINNISON 1999; PALUMBI 2001; HAIRSTON *et al.* 2005). However, examples of rapid phenotypic evolution have been known since the mid-20th century (KETTLEWELL 1958. ; FORD 1964; JOHNSTON AND SELANDER 1964) and its prevalence has become increasingly appreciated in recent years. Rapid evolution has profound practical consequences for biological control of pathogens, pests and invasive species, fisheries management, and biodiversity conservation (CONOVER AND MUNCH 2002; DARIMONT *et al.* 2009), especially in the context of accelerating climate change (WARD AND KELLY 2004). Indeed, this growing appreciation for the potential importance of rapid evolution has spawned new subdisciplines such as eco-evolutionary dynamics (ELLNER *et al.* 2011).

Many instances of rapid evolution of ecologically important traits have been documented in invertebrates (ELLNER *et al.* 1999; DABORN *et al.* 2002), vertebrates (REZNICK *et al.* 1997; GRANT 1999), plants (FRANKS AND WEIS 2008), yeast (LANG *et al.* 2013; LEVY *et al.* 2015), and prokaryotes (BARRICK *et al.* 2009). Biochemical (GHALAMBOR *et al.* 2015; HUANG AND AGRAWAL 2016), morphological (LOSOS *et al.* 1997; GRANT 1999), life history (ROSE 1984; HAIRSTON AND WALTON 1986; REZNICK *et al.* 1997), and behavioral (TURNER AND MILLER 2012b; STUART *et al.* 2014) phenotypes can evolve substantially in just a handful of generations when populations experience new selective regimes. However, less is known about the genomic changes that occur during rapid adaptation to novel environments, especially in multicellular eukaryotes (MESSER *et al.* 2016; JAIN AND STEPHAN 2017).

A key question is whether the standing genetic variation within populations is sufficient for adaptation to a novel environment, or if new mutations are required. In sexual eukaryotes, abundant standing variation is indicated by the observation that artificial selection can immediately, and often dramatically, change the mean of almost any variable trait (LEWONTIN 1974). Still, it is possible that natural selection may fail where artificial selection succeeds if the alleles that respond in artificial selection experiments are encumbered with deleterious side effects. E&R studies seem an ideal alternative to artificial selection experiments in this regard. While the researcher largely controls fitness with artificial selection, organisms "select themselves" in an E&R experiment. Pleiotropic effects on general vigor will be a major determinant of selection on alleles with favorable trait effects in an E&R experiment, but much less so in an artificial selection experiment. Admittedly, most previous E&R experiments do not

evaluate the evolutionary potential of natural populations simply because they are initiated from laboratory adapted populations, or small numbers of founders. Here, we describe a replicated E&R experiment using *D. simulans* founder populations initiated with wild-caught individuals to investigate the earliest stages of adaptive evolution.

Genome-wide evolution in E&R studies— E&R experiments using Drosophila have addressed questions about the number and kinds of loci under selection, the relative frequency of hard versus soft selective sweeps, temporal dynamics, and the effect of selection on genome-wide patterns of diversity (BURKE *et al.* 2010; TURNER *et al.* 2011a; OROZCO-TERWENGEL *et al.* 2012a; REMOLINA *et al.* 2012; TURNER AND MILLER 2012b; HUANG *et al.* 2014; TOBLER *et al.* 2014; KANG *et al.* 2016; BARGHI *et al.* 2017; MICHALAK *et al.* 2017; SCHOU *et al.* 2017). In a review of these studies, NUZHDIN AND TURNER (2013) noted a striking "excess of significance" in that thousands of polymorphisms appear to respond to selection. The number of loci that can be simultaneously under direct selection is an important and long-standing controversy in evolutionary genetics (HALDANE 1957; SVED *et al.* 1967; BARTON 1995). Generally however, we expect that the more loci affecting fitness, the smaller the allele frequency change per locus. It is thus paradoxical that so many SNPs exhibit large change. There are numerous potential reasons for excessive significance. Perhaps the simplest is that testing procedures might be anti-conservative if genetic drift and sampling produce greater than predicted changes.

Hitch-hiking (MAYNARD SMITH AND HAIGH 1974) is the most likely driver of excessive significance – many significant tests might be neutral SNPs in Linkage Disequilibria (LD) with selected loci. Hitch-hiking in an E&R study requires an initial association between loci in the

ancestral population(s) and also minimal subsequent recombination over the course of selection. Relevant to both, natural and laboratory adapted *D. melanogaster* populations are polymorphic for large inversions and have dramatically suppressed recombination near centromeres (CORBETT-DETIG AND HARTL 2012; KAPUN *et al.* 2014; TOBLER *et al.* 2014; FRANSSEN *et al.* 2015a), potentially resulting in a large number of false positive candidate SNPs (TOBLER *et al.* 2014; FRANSSEN *et al.* 2015a; BARGHI *et al.* 2017). We chose *D. simulans* for this study to evaluate genomic patterns underlying rapid evolution in a population largely free of inversion polymorphism. LD declines rapidly with the physical distance between sites in *D. simulans* (SIGNOR *et al.* 2018), but as emphasized by NUZHDIN AND TURNER (2013), sampling of haplotypes to form experimental populations can generate higher levels of LD (even at considerable physical distance) than are present in the natural population. The contribution of these sampling-generated associations to parallel evolution of replicate E&R populations can be mitigated by founding experimental replicates from distinct natural population samples.

The genomewide response in an E&R experiment (evolution at both selected and neutral loci) depends on the number and genomic positions of selected loci, how those loci interact to determine fitness, the nature and extent of LD, the recombination map, the reproductive biology, and the experimental design. Given these myriad factors, we do not have a clear picture of how much hitch-hiking is to be expected in a typical E&R experiment and thus a means to infer how many sites are direct targets of selection. To develop predictions, we build a simulation framework to predict the full observed response of an E&R experiment. The design of the experiment (how replicate populations are founded, how many individuals reproduce, and how many generations occur) is directly reiterated in a model that predicts change of every

polymorphism in the genome. This is a parameter rich model, but prior work on *D. simulans* and its close relative *D. melanogaster* underpin essential assumptions (e.g. the recombination map, patterns of LD in nature). Observations from our specific E&R experiment establish other features such as the number and genomic positions of polymorphisms and initial allele frequencies. Finally, we extract essential information not only from the extreme outliers (putative targets of selection), but from observations on the "average SNP." The amount and variability of change at neutral loci dispersed across the genome is an indicator of "genetic draft" (GILLESPIE 2001; NEHER AND SHRAIMAN 2011) and thus of selection. The simulation model provides important insights on the observed experimental results, not only in terms of the number of significant tests but also on the allele frequency spectrum at fitness determining loci.

Methods

I. The experiment.

Founding populations. The ancestral populations of this experiment are from the offspring of wild-collected mated *D. simulans* females. We collected adult *D. simulans* from compost piles at Orchard Pond Organic Farm in Tallahassee, FL (Universal Transverse Mercator Grid coordinates 16N 761030 3386162) between October 28 and November 25, 2014. Adult females were isolated in vials to produce offspring. From each wild-collected female, we collected two male and two female offspring after verifying that male offspring were *D. simulans*. One male and one female offspring were flash frozen immediately to represent the founding generation (see below). These flies were kept at -80° until DNA extraction. The other male and female were used to found replicate lab populations. We initially established six replicate population cages, using one male and one female offspring of 250 wild-caught mated females per replicate

and using the offspring of different wild female progenitors in each replicate. Approximately 3 weeks after founding these populations (Dec 2-3, 2014), the six cages were combined two at a time to form the A, B, and C population replicates. Equal numbers of flies were used from the pair of cages and mixed to create new cages. Thus, each of the A, B, and C populations was founded with approximately 1000 individuals descended from non-overlapping sets of 500 wild-caught, mated female parents.

Lab rearing and maintenance. Flies were housed in plexiglass containers measuring 6028 cubic centimeters. Each cage was supplied with six 177-mL plastic bottles containing 50 mL of standard cornmeal-yeast-dextrose media. Every two weeks, we replaced three of the six bottles with bottles containing fresh media; each bottle remained in a cage for four weeks. We replaced cages with clean plexiglass containers every 28 days, in sync with a media change. Censuses of cages were conducted approximately every 5 weeks using digital images. Census values for each population were: A (mean=1277, range=832-1635), B (mean=849, range = 672-1147), and C (mean=1187, range=963-1620).

We maintained populations A, B, and C under constant lighting and temperature conditions (12L:12D, 25°) for approximately 195 days from initial collection (population A: founded from females collected Oct 28 – Nov 1, 2014, descendants preserved May 12, 2015; B: founders collected Nov 4-Nov 11, descendants preserved May 22; C: founders collected Nov 19-Nov 25, descendants preserved Jun 5). That is, descendants of the original founders were sampled approximately 7 months (about 14 generations) after collection of the wild founders. From the last generation of each population, we collected 500 males and 500 females by aspiration. These

flies were snap frozen on dry ice and kept at -80° until DNA extraction. For DNA extraction and sequencing, we pooled the 1000 preserved offspring of the founding females to form "Ancestral" samples A0, B0, and C0. Similarly, we pooled the 1000 flies collected at the end of the experiment to form Descendant samples (A7, B7, C7), with "7" designating months since population founding. Extractions and sequencing libraries were performed simultaneously for all six population samples.

Library prep and level of sequencing. Whole flies (500 males and 500 females from each Ancestral and Descendant population) were homogenized and the DNA was extracted using DNAzol reagent (Thermo Fisher). We fragmented DNA using a Covaris E220 Ultrasonicator and size selected to produce insert lengths of 380-480 bp. One sequencing library was prepared for each population using the NEBNext Ultra DNA Library kit for Illumina (New England Biolabs) following manufacturers recommendations, with each population receiving a unique index. We made a distinct library from flies from Ancestral populations A0, B0, and C0 (NEB indices 13-15, respectively), and Descendant populations A7, B7, and C7 (NEB indices 16, 18, and 19 respectively). Ancestral population samples were multiplexed and sequenced in one lane, and those from the Descendant populations were multiplexed and sequenced in three additional lanes. Because one library (C7) was over-represented in the resulting data, we performed an additional sequencing run using the five remaining libraries (A0, B0, C0, A7, and B7) which were multiplexed and run on a single lane. All sequencing was conducted using an Illumina HiSeq 2500 instrument at the Translational Science Lab at Florida State University, using V3 chemistry. We sequenced 150bp on each of the paired ends. In total, DNA from 6000 flies was sequenced.

Sequence Analysis. We edited read pairs (fastq format) from each population sample using Scythe (https://github.com/vsbuffalo/scythe/) to remove adaptor contamination and then with Sickle (https://github.com/najoshi/sickle/) to trim low quality sequence. We used the mem function of BWA (Li and Durbin 2009) to map read pairs to version r2.02 of the *Drosophila simulans* reference genome, updated from the build published by Hu *et al.* (2013). We used picard-tools-1.102 to eliminate PCR duplicates from the mapping files; an important step given that PCR duplicates represent pseudo-replication in bulked population samples. Prior to variant calling, we applied the RealignerTargetCreator and IndelRealigner to each population bam file. The population bams were input to Varscan v2.3.6 to call SNPs and indels. We piped the output from samtools mpileup (version 1.2) to the varscan functions mpileup2snp (for SNPs) and mpileup2indel (for indels). We obtained the read count (number of alleles) and reference allele frequency at each variant site for each sample. We suppressed indels in downstream analyses as well as all SNPs within 5bp of an indel. We also limited attention to the major chromosomes: X, 2R, 2L, 3R, and 3L.

We scored read depths within each population prior to filtering and found that the median depth at X-linked loci was very close to ¾ the corresponding value for autosomal loci (ratio = 0.77 for Ancestral populations, 0.75 for Descendant populations). For subsequent analysis, we eliminated polymorphisms if the read depth across populations was too low for powerful tests or atypically high across samples. For inclusion of a SNP, we required at least 60 reads per population for X-linked and at least 80 for Autosomal loci. We excluded SNPs if the total read depth within Ancestral and Descendant (considered separately) was greater than the 95th

224 percentile of the corresponding depth distribution. Here, we conducted separate filtering on
225 autosomal sites and X-linked sites as the latter have lower coverage. After filtering, 291272
226 SNPs remained (58647, 49940, 69010, 71289, and 42386 on 2L, 2R, 3L, 3R, and X,
227 respectively).
228
229 **Copy Number Variant Analysis.** We also performed a scan for copy number variants (CNV)
230 using the method of SCHRIDER *et al.* (2013). Specifically, we applied an updated set of python
231 scripts to identify reads pairs with excessively large inserts (https://github.com/kern-
232 lab/poolDiffCNV). An unusually large distance between the first and second read of a read pair
233 suggests an insertion of DNA within the reference genome relative to the sequenced molecule.
234 We determined loci exhibiting this feature and scored reads with and without the putative
235 insertion for each population. After correcting for differences in average read depth among
236 populations, we identified loci with the greatest divergence between ancestral and descendant
237 populations. Finally, we combined results across experimental replicate populations to identify
238 the subset of these loci that exhibited parallel evolution. For putative CNVs exhibiting parallel
239 change in the three replicate populations, we identified those that occurred within known coding
240 or regulatory regions. We extracting the minimum and maximum chromosome position of the
241 CNV across all three populations, and then used the GenomicRanges R package v. 1.30.1
242 (Lawrence et al. 2013) to find genes with coding or regulatory regions that overlapped the CNV
243 position.
244
245 **II. Analysis of evolutionary change at SNPs**.

Null divergence. The raw data for each polymorphism is twelve numbers, the counts of reads for each alternative allele in each of the six populations (A0, B0, C0, A7, B7, C7). The statistical treatment of data is based on transformed allele frequencies (FISHER AND FORD 1947; WALSH AND LYNCH 2017). If p_A and p_D are allele frequencies in Ancestral and Descendant populations, respectively, then $x_A = 2\sin^{-1}(\sqrt{p_A})$ and $x_D = 2\sin^{-1}(\sqrt{p_D})$, with x measured in radians (the A/D subscript denotes Ancestral and Descendant, respectively). This angular transformation is useful because, to a first approximation, the variance in change owing to genetic drift and sampling is independent of the true allele frequency (FISHER AND FORD (1947), Supplemental Figure S1). As a consequence, a common test can be applied across polymorphisms despite differing initial allele frequencies (KELLY *et al.* 2013).

At a neutral SNP, divergence ($x_D - x_A$) is predicted to be normally distributed with mean 0 and variance equal to $\frac{1}{2n_A} + \frac{1}{2n_D} + \frac{1}{m_A} + \frac{1}{m_D} + \frac{t}{2N_e}$, where m = read depth at the locus after sequencing, n = number of diploid individuals sampled for sequencing, N_e is the effective population size, and t is the number of generations. While five of these quantities are known, N_e is not. Also, it may be that draft is as or more important than drift in determining the variance in change of neutral SNPs. For these reasons, we fit a simpler model: $[x_D - x_A] = v + \frac{1}{m_A} + \frac{1}{m_D}$, where ν is the 'null variance.' ν absorbs quantities shared among SNPs within a population, although it could differ between autosomes and X owing to differences in N_e (CHARLESWORTH 2009). With hitch-hiking, it will vary at much smaller genomic scales. We estimate ν on a population and chromosome specific basis using $V^*_{AD} - V_{rd.}$. $V_{rd.}$ is the read-depth variance, the

average of $(\frac{1}{m_A}+\frac{1}{m_D})$ across loci. $V*_{AD}$ is the robust estimator for the variance in divergence between A and D based on the interquartile range of the full distribution (KELLY *et al.* 2013). For estimation of ν, we focus on SNPs with $0.05 < p_A < 0.95$ to eliminate boundary effects.

Testing for parallel divergence. Divergence between ancestral and descendant populations is statistically independent between replicates because each population was founded from distinct flies and there was no gene flow between replicates. Under the null hypothesis, $E[x_A - x_D] = 0$ and the variance is given by ν, m_A, and m_D. The likelihood for the null model at each SNP (LL_0) is the product of three normal densities. We contrast the normal density likelihood of the data under this model to an alternative allowing parallel evolution across replicates: $E[x_D - x_A] = \delta$, where δ is the (shared) change in allele frequency due to selection across replicate populations. For the alternative model, we also require the MLE of δ : $\hat{\delta} = \frac{w_A dx_A + w_B dx_B + w_C dx_C}{w_A + w_B + w_C}$ where the *dx* terms are the observed $(x_A - x_D)$ in each replicate and the *w* are replicate specific weights: $w = \frac{1}{v+\frac{1}{m_A}+\frac{1}{m_D}}$ (see MONNAHAN AND KELLY (2017) for derivation). The Likelihood Ratio Test (LRT) for parallel evolution is $2(LL_1\text{-}LL_0)$. This is not a strict likelihood in the sense that ν are treated as constants and not free parameters, but the error associated with this procedure should be minimal given that ν are estimated from the aggregate of thousands of variant sites (MONNAHAN AND KELLY 2017). Also, we use permutation (and not the parametric chi-square distribution) to assess genome-wide significance of LRT values. To create a permutation replicate, we randomly scrambled observed standardized divergences across SNPs within each replicate population and then applied the LRT test to each SNP. The distribution of

288 divergences (across SNPs within each population) is preserved by this procedure and so the test
289 is based entirely on consistency of response of the same SNP across replicates. We extracted the
290 most extreme (largest) LRT from each permuted dataset and repeated the procedure 1000 times.
291
292 We applied a second scan for parallel adaptation at the scale of genomic windows. The LRT
293 described above requires the same SNP to show a strong parallel response. This may miss
294 selected sites because allele frequency estimates are encumbered with substantial error owing to
295 finite sequencing depth. Closely linked sites might all respond to selection (owing to hitch-
296 hiking) but show differing signals across replicate populations owing to differing estimation
297 error. To capture signals of parallel evolution in genomic windows, we compared the rank order
298 of divergence in windows, using the rank-rank hypergeometric (RRHO) approach of PLAISIER *et*
299 *al.* (2010). This method has been used to compare ranked lists from gene expression (TELES *et*
300 *al.* 2013; PENA *et al.* 2017) and genome-wide association studies (MEDLAND *et al.* 2014; FRANKE
301 *et al.* 2016). Here we used this rank-based approach to identify genomic regions characterized by
302 high divergence in all replicate populations, but where the actual sites of highest divergence
303 might differ among populations. We first identified, within each replicate, the site in each
304 window with the maximum absolute value of divergence between ancestor and descendant. We
305 assigned that maximum value to the window. We excluded regions near centromeres and
306 telomeres to remove regions likely to have low recombination (regions retained: 2L;2,400,000-
307 19,600,000; 2R:24,000,000-200,000,000; 3L:2,400,000-21,000,000 ; 3R:24,000,000-25,000000 ,
308 X:2,000,000-19,600,000). We then ranked these maximum divergence values, with windows
309 exhibiting the greatest decrease in the reference allele frequency during laboratory adaptation at
310 the top of the ranked list and those with the greatest increase in reference allele frequency at the

bottom. The rank order of windows in populations A, B, and C were then compared by evaluating overlap in ranks at many different thresholds. For example, the overlap of the top 100 windows in two populations was compared, then the overlap in the top 200 windows, and so on, until every possible combination of thresholds was evaluated. The hypergeometric probability of the number of overlapping items was calculated for every threshold, applying the BENJAMINI AND YEKUTIELI (2001) correction for multiple tests. We conducted this analysis at three different window sizes (100-kb, 10-kb, and 1-kb) to assess our expectation that effects of hitch-hiking would be minimal at the largest window size and more pronounced as window size decreased.

III. The Simulator.

The program tracks each chromosome of each individual in the population as a series of binary values (the allele present at each locus). The number and position of each locus is given by the number/position of each polymorphism observed in the experiment. To optimize use of computer memory, the program compresses sites following SUKUMARAN AND HOLDER (2011). The population is defined as N_m male and N_f female adults that are formed each generation (only one X in males). Generation zero is founded by selecting haplotypes from the "natural population" using observed Ancestral allele frequencies combined with hash tables to accommodate LD (see Supplemental Table S1). The hash tables were estimated using the large collection of *D. simulans* genomes recently published by SIGNOR *et al.* (2018) and account for LD differences between low recombination regions (near centromeres and telomeres) and between the X and autosomes.

Each subsequent generation is formed by randomly selecting parents and synthesizing gametes from those parents to create a new set of N_m males and N_f females. We assume no recombination in males. In females, recombination probabilities per chromosome arm are specified as 2L: 0.570, 2R: 0.655, 3L: 0.555, 3R: 0.830, and X: 0.590 following TRUE *et al.* (1996). The location of crossovers, when they occur, is probabilistic. We use the location distribution from *D. melanogaster* (COMERON *et al.* (2012); Supplemental Table S3) which exhibits reduced rates near telomeres and centromeres.

In simulations with selection, individuals are chosen with probabilities proportional to their fitness (which is a function of the genotype at relevant sites). We first consider a truncation selection model – individuals have fitness 1 if their genotypic score (a simple sum of effects across loci) exceeds the threshold and 0 otherwise. We second consider a multiplicative model where each site affecting fitness has a selection coefficient (s) with 1, 1+s, 1+2s, as the locus specific fitness values. Individual fitness (the relative probability of selection for reproduction) is a product across loci. For both selection models, we assume that hemizygous male genotypes have the "homozygous" effect, e.g. 1 and 1+2s.

A simulation replicate is three distinct populations propagated for 14 generations. In each population, we simulate data collection (sequencing) to mimic the actual experiment. Read depths at each polymorphic site are equivalent to the observed values. We then subject the simulated data to the same analysis pipeline applied to the real data (estimation of null variance parameters followed by LRT tests on all sites). For each parameter set (specified values for N_m

and N_f, the number and location of selected sites, selection coefficients at each site), we simulate the entire experiment 1000 times to determine the range of outcomes.

Results

Divergence of transformed allele frequency is highly normal and changes in transformed allele frequencies are essentially homoscedastic (Supplemental Figure S1), which justifies use of the angular transform (FISHER AND FORD 1947) . However, the null variance in divergence (ν) varies substantially among chromosomes and populations (Table 1). If the average ν = 0.0183 is equated to its expected value under neutral genetic drift ($\frac{1}{2n_A} + \frac{1}{2n_D} + \frac{t}{2N_e}$), we obtain N_e = 405, which is much lower than adult census sizes (ca. 1000 flies). Also, the variance in ν among chromosomes and populations is much greater than predicted by neutral drift dynamics (see simulation results below).

Against the background distribution, 126 polymorphisms provide strong evidence for parallel adaption in captivity (Supplemental Table S2). The LRT value for these SNPs is larger than the most extreme single test in 95% of the permuted datasets (LRT cutoff = 46.64 yields permutation $p < 0.05$; Supplemental Table S4). Importantly, these sites do not all represent distinct "evolutionary events" given that significant variants are often closely linked. Figure 1 illustrates the largest LRT value per 100kb window along each chromosome. There is clearly an aggregation of strong signal in regions of low recombination. The clumping of LRT in these regions nicely parallels the pattern of high LD among distant sites (Supplemental Figure S3).

If we "bin" significant tests that are relatively closely linked, then perhaps 20-30 distinct "loci" are evident as foci of selection across the genome. The striking feature of selected variants is that nearly all are loci where the two alleles segregate at intermediate frequencies in the natural population (Figure 2). The overall allele frequency spectrum (AFS) in the ancestral population (blue bars) is typical of natural population samples of *D. simulans* (SIGNOR *et al.* 2018). Less than half of SNPs have a minor allele frequency (MAF) greater than 0.1. However, among SNPs testing positive for parallel selection (orange bars), 98% have MAF > 0.1 in the ancestral population and 88% have MAF > 0.2.

Rank-based analysis of divergence of genomic windows shows high similarity between populations, even after excluding regions of low recombination. Using 100-kb windows, 903 windows remained after removing centromeric and telomeric regions. At this window size, we expect effects of hitch-hiking to be limited (see Figure 1 where, outside of centromeric and telomeric regions, high LRT regions tend not to be clumped). Nevertheless, the ranks of windows with respect to their maximum divergence value was more similar than expected by chance. In the heat maps in Figure 3, a pattern of warm colors (small probability values) in cells in the upper-right and lower-left quadrats indicates that more windows were ranked similarly in both populations than expected by chance (Plaisier et al. 2010). Especially for windows showing most extreme divergence (top right and bottom left corners of heat maps), the overlap in rankings is associated with probability values below 10^{-10}, after correction for multiple hypergeometric tests. At this window size, similarity in ranks for windows in which the reference allele increased in frequency during laboratory adaptation is especially apparent (upper right quadrat). One quarter to one third of all genomic windows exhibited parallel divergence in this direction in all

pairwise comparisons between populations (p values < 10^{-15}). Using smaller windows (1 kb and 10 kb), replicate populations exhibited even more similar ranks, indicated by very low probabilities at most overlap thresholds (Supplemental Figure S2).

We identified 303 putative CNV that evolved in parallel in the three replicate populations. These regions overlapped known coding and/or regulatory regions of 396 *D. simulans* genes (Supplemental Table S5). Few *D simulans* genes have been annotated with functional information, so we obtained *D. melanogaster* orthologs for the 311 genes for which the orthologs are known (downloaded from Flybase Release 2018_02, on May 9, 2018). Among these were orthologs of genes involved in metamorphosis (*neverland*), circadian rhythm and sleep (*Ac3*, *Nipped-B*), stress response (*mth*), immunity (*PGRP-SC1a*, *Trx-2*), sensory perception (odorant receptors *Or42a*, *Or13a*), and chromatin modification (*gpp*, *brm*, *Nipped-A*, and *CG12659*) (see Supplemental Table S5 for complete list of orthologs and functional annotation).

Simulation results. We evaluated a series of parameter sets against four distinct aspects of the observed results. First, how well do simulation outcomes match the observed mean and variability in the null variance (v) across chromosomes and replicate populations (comparison to Table 1)? Second, do simulations yield a comparable number of significant LRT tests for parallel evolution (comparison to Supplemental Table S2)? Third, are significant tests distributed across the genome in a way similar to the real data (comparison to Figure 1)? Fourth, is the allele frequency distribution of significant loci similar to Figure 2?

Strictly neutral evolution using the census population sizes for adults (N_m = 500, N_f = 500) fails to match the observed evolutionary response in all regards. The mean *v* for simulated data is less than half the observed mean (Supplemental Table S6.A) and exhibits higher values for the X than autosomes (not observed in Table 1). In the simulations, the standard deviation across chromosomes and population replicates in simulations is less than 1/5 the observed value. Thirty-seven of 1000 simulations of the experiment produced an LRT value that exceeded our threshold for the real data (46.64) which suggests the test is slightly conservative (null expectation is 50 of 1000). Also, the distribution for individual LRT tests (within and across simulations) closely matches the parametric expectation. With neutral evolution, the LRT should follow a chi-square distribution with 1 degree of freedom.

If we simulate neutral evolution but reduce adult population sizes (N_m = N_f = 210), the mean *v* of autosomes is close to observed, but the simulated mean for the X is too high (0.23) and the variance among chromosomes too low (Supplemental Tables S6.B). The X versus autosome discrepancy for mean *v* can be addressed introducing a large sex ratio bias. If we simulate neutral evolution with N_m = 120 (much lower than census) and N_f = 800 (much higher than census), the simulation mean *v* is 0.17 for autosomes and 0.18 for X (Supplemental Tables S6.C). These are reasonable given Table 1. However, the variance in *v* from these simulations remains too low relative to observed. Most importantly, the neutral simulations never yield many high LRT tests for parallel selection. There are 895 tests with LRT > 20 in real data versus an average of 3-5 in neutral simulations.

There is an enormous range of possibilities for simulations with selection. We have considered variations on the number of loci affecting fitness, the magnitude and direction of their respective effects, the chromosomal locations of these loci, and how different loci combine to determine fitness. Here, we will limit attention to a few specific cases that predict the general features of the observed evolutionary response in the experiment. Figure 4 illustrates four simulation replicates for evolution on chromosome 2 (comparable in content to the top panel of Figure 1). This is a case where 20 loci influence fitness, 4 positioned uniformly over each chromosome arm (8 sites across 2L and 2R are shown in Figure 4). In each replicate, the initial allele frequencies at fitness determining sites were sampled from the observed distribution for significant sites (orange bars in Figure 2). Each locus contributes equally and additively to the genotypic score subject to truncation selection. With the top 75% selected each generation (375 males and 375 females), the mean and standard deviation of v are close to their observed values (0.018 and 0.0022, respectively; Supplemental Table S6.D). The number of large LRT tests for parallel selection is less than observed. Genomewide, an average of 320 sites yield an LRT greater than 20 (895 in the real data). However, some simulation replicates yield LRT distributions that closely match the observed. Moreover, plots of maximum LRT test per 100kb (Figure 4) reiterate key features of the observed data. In particular, most replicates demonstrate the aggregation of strong signal in regions of low recombination. The hitch-hiking effect is variable among replicates (since initial allele frequencies at selected sites and natural LD patterns with those sites are resampled with each replicate), but can be as pronounced as in the real data (Figure 1). Since the fitness-determining loci do not always yield high LRT (and if they did, would ***directly*** produce only 20 high LRTs), the simulations clearly show how hitch-hiking can reiterate the abundance of significant tests noted by NUZHDIN AND TURNER (2013).

467

468 We manipulated the initial distribution of allele frequencies at selected sites to address the
469 question that emerges from Figure 2: Does the intermediate frequency of significant variants
470 from the experiment (orange bars in Figure 2) imply that intermediate frequency variants are the
471 primary targets of selection in the laboratory environment? We conducted otherwise equivalent
472 simulations but with initial allele frequencies of selected sites sampled from the genome-wide
473 distribution (blue bars in Figure 2; Supplemental Table S6.E). As previously, the minor allele
474 was equally likely to increase or decrease genotypic score. The broken lines (green) in Figure 2
475 illustrate the AFS of sites in simulation experiments with LRT exceeding the genome-wide
476 threshold from the real experiment. There is a "pull to the middle" in these simulations; most
477 selected sites have a MAF < 0.1, but those with higher initial frequencies are more likely to yield
478 high LRT (Figure 2). However, despite this ascertainment effect (among selected sites, those
479 with more intermediate frequency are more likely to be detected), the real data contain an excess
480 of SNPs with MAF > 0.3 that yield significant tests (orange versus green in Figure 3). Also
481 relevant is that the pull to the middle in simulations with selection from the background AFS
482 reflects a biologically important feature – the loci with higher MAF are generating greater
483 variance in fitness when the population experiences the novel laboratory environment.

484

485 The 20-locus model is only one of a great many that might predict the general features of the
486 observed response, but it is instructive as a basis for comparison. For example, if we increase the
487 strength of selection (50% of adults instead of 75% reproducing), but maintain all else the same,
488 the draft effect becomes excessive relative to the data (Supplemental Table S6.F). In these
489 simulations, the mean v is 4 times greater than observed and the standard deviation of v is

inflated 13 fold. If we maintain 75% selected but allow 30 loci to affect fitness (6 per chromosome arm), the mean and standard deviation of *v* are very similar to the 20-locus model, and thus close to the real data. However, the number of high LRT tests is reduced because selection ***per locus*** is weaker and thus so is the magnitude of allele frequency change. If we replace truncation selection with multiplicative selection with s = 0.1 at each of the 20 loci, the simulations yield approximately the observed value for mean *v*. However, the standard deviation for *v* is too large by 60% and the number of high LRT scores is reduced by about a third relative to the base case (Supplemental Table S6.G). In general, we see that truncation selection allows a more pronounced parallel response than multiplicative selection for a given level of draft (for similar values of the mean and standard deviation of *v*).

Discussion

This E&R experiment, while short in duration, yielded a pronounced evolutionary response. Over a 100 SNPs, clustered into perhaps 20-30 distinct loci across the genome, exhibit parallel change. The allele frequency spectrum (AFS) of selected sites is strongly biased toward intermediate allele frequencies and it is clear that the source natural population harbors abundant standing variation, allowing rapid adaptation in a novel environment. Accumulating examples of rapid phenotypic and genomic evolution and observations that allele frequencies can vary cyclically in natural Drosophila populations (Bergland et al. 2014) have led to the suggestion that many of the foundational principles of molecular population genetics, based on neutral and nearly-neutral theory, might require revision (Messer et al. 2016; Hermisson et al. 2017). Our study provides evidence that with genome wide selection in a novel environment, genetic draft is

more important than genetic drift. We could not find a neutral model that accurately predicts the evolution of the "average SNP" (the determinants of *v*); the many polymorphisms that were neutral or nearly so in this experiment. Below, we discuss these results in relation to the maintenance of polymorphisms in nature and the paradox of excessive significance in E&R studies.

***Hitch-hiking and excessive significance*—**NUZHDIN AND TURNER (2013) argue that the large number of significant tests in recent E&R experiments, e.g. (TURNER *et al.* 2011b; OROZCO-TERWENGEL *et al.* 2012b; TURNER AND MILLER 2012a), must be due to over-estimating the number of loci under selection. Because a limited number of haplotypes are sampled to initiate E&R experiments, non-random associations between loci can be generated even at loci far apart in the genome. This sampling effect, combined with traditional hitch-hiking, could produce large changes in allele frequency at many sites that are not the direct targets of selection. For the present experiment, we established each replicate population with a distinct sampling of genotypes from nature, which should reduce the scope for haplotype-sampled LD to generate false positives. The finite number of haplotypes that survive in each experimental population will yield idiosyncratic associations between distant SNPs, but these associations should be population specific. Hitch-hiking may produce strong correlated responses at neutral polymorphisms, but it is less likely to generate consistent *parallel* changes across replicates. Perhaps for these reasons, the number of significantly diverged SNPs in this experiment is reduced by an order of magnitude from previous studies (from thousands to 126).

Another potential reason we observed a low number of significant SNPs relative to other E&R experiments is the nature of the *D. simulans* genome. Inversions are rare in *D. simulans* (Lemeunier and Aulard 1992), recombination suppression near centromeres is less pronounced than in *D. melanogaster* and the genome-wide recombination rate is 30% higher than in *D. melanogaster* (True et al. 1996). In a recent investigation comparing genomic change in *D. melanogaster* and *D. simulans* after ~60 generations of adaptation to hot conditions in a laboratory environment, patterns were substantially different in the two species (Barghi et al. 2017). *D. simulans* had fewer candidate SNPs, and the regions of the genome implicated in response to selection were narrower and more distinct. Strikingly, almost all of chromosome arm 3R in *D. melanogaster* (which contains several overlapping segregating inversions) exhibited a pattern consistent with selection. In *D. simulans*, which lacks similar inversions on 3R, several narrow, distinct regions on this chromosome arm exhibited such a pattern. The authors attributed many of these differences in the frequency of segregating inversions and in centromeric recombination suppression.

Figure 1 suggests there are 20-30 major loci responding to parallel selection with several linked SNPs per locus generating significant LRT tests. However, there might be a much larger number of loci that responded in parallel, but that failed to reach the conservative threshold we used in our SNP-specific LRT tests. Some support for this possibility is provided by the rank-based analyses of genomic windows. Even at the largest window size (100 kb), where hitch-hiking effects should be minimal, several hundred windows exhibited parallel patterns of divergence, suggesting that more than 20 loci responded in parallel in this experiment. If sampling variation led to variation among populations in the individual SNPs that produced strong signals of

558 divergence at loci undergoing parallel evolution, those loci would not be detected in our SNP-
559 based analyses, but would be detected in the window-based analyses. Alternately, haplotype
560 sampling could produce an inflated signal of parallel evolution in these window-based tests, but
561 as argued above, using independent founders in different replicate populations should mitigate
562 that sampling effect.
563
564 Among our genome-wide significant SNPs (Supplemental Table S2), we cannot distinguish the
565 actual targets of selection from hitch-hikers. While this is a clear limitation of the experiment,
566 the intermediate frequency result (Figure 2) is not undermined by hitch-hiking. Change at a
567 hitch-hiking locus will only match change at the selected locus if allelic association is maximal
568 (and remains unbroken by recombination) and allele frequencies are similar at the two
569 loci. Imagine a rare allele that is exclusively associated with one of the two alleles at a neutral,
570 intermediate frequency polymorphism. If the rare allele at locus A (A') becomes favorable, the
571 change of the hitch-hiking allele (B') will likely be much smaller than for A' because most
572 copies of B' are initially linked to the deleterious A allele. While the loci of this example are
573 perfectly associated in one sense (maximal value for Lewontin's D' (LEWONTIN 1964)), r^2 is
574 quite low, and r^2 better predicts the magnitude of correlated change. Of course, favorable rare
575 alleles can cause extensive hitch-hiking by other rare alleles when captured in the same founding
576 haplotype (see FRANSSEN *et al.* (2015b) for an illustration of this effect).
577
578 ***Environmental heterogeneity and the maintenance of polymorphism*—**We have emphasized
579 that the lab environment is novel and thus a selective challenge to wild *D. simulans*. However,
580 the lab environment is also relatively constant and homogeneous. The experiment thus tests

whether maintaining a genetically diverse population in a constant environment leads to predictable changes in allele frequency. Polymorphisms maintained by environmental heterogeneity should not be stable if heterogeneity is eliminated. If there is heterogeneity in the natural environment (e.g., seasonal variation in temperature or spatial structure in resource availability) and if genotypes vary in their optima for these variables, then variation is maintained as a 'multi-niche polymorphism' (LEVENE 1953). At such loci, we cannot predict which allele would increase in captivity without detailed information about genotype-specific tolerances. However, it is likely that one genotype will accidently match the lab environment better than alternatives, and as a consequence, a multi-niche polymorphism will break down if the population is maintained in a single niche. Relevant to this expectation, we observed that expected heterozygosity, 2p(1-p), declined at over 75% of the sites significant for parallel evolution (Supplemental Table S2). While dramatic evolution is routinely observed in captive and/or domesticated populations (DRISCOLL *et al.* 2009; TRUT *et al.* 2009; MEYER AND PURUGGANAN 2013), it is usually impossible to distinguish adaptive changes (to the specific laboratory environment) from those owing to genetic drift and inbreeding or directed (artificial) selection.

We hypothesize that the polymorphisms significant for parallel selection are likely maintained by balancing selection of some form (including but not limited to environmental heterogeneity) in nature (BERGLAND *et al.* 2014; CHARLESWORTH 2015). The AFS at significant loci is strikingly different from the genomic background. The overall AFS (blue bars in Figure 2) exhibits the expected preponderance of rare alleles (MORIYAMA AND POWELL 1996; PRZEWORSKI *et al.* 2001), consistent with the idea that most polymorphisms are neutral or nearly neutral (WRIGHT

1931; OHTA 1976). However, the polymorphisms that responded to selection are overwhelmingly intermediate frequency (orange bars in Figure 2). If selection were consistently favoring one of the two alleles at these loci in the natural population, it is extremely unlikely that the minor allele would persist at frequencies greater than 20%. Of course, neutral alleles can sometimes drift to intermediate frequencies. However, the fact that these polymorphisms have clear phenotypic effects (implied by rapid evolution in the lab) undermines the argument that they are neutral in nature.

Ascertainment and cause are two important considerations for interpreting the AFS of significant sites. Regarding ascertainment, the question is whether the excess of intermediate frequency polymorphisms among selected loci is because it is easier to detect change at these loci. For the great majority of polymorphisms, the answer is clearly no. In terms of transformed allele frequency (on which testing is based), the average magnitude of change is as large for rare alleles as for common (Supplemental Figure S1). The angular transformation effectively normalizes the effects of genetic drift and experimental sampling on allele frequency change (FISHER AND FORD 1947). However, it does not eliminate dependency of selection-driven change on allele frequency. For a given selection coefficient, the variance in fitness generated by a locus is maximal at intermediate allele frequency. The difference between blue and green in Figure 2 illustrates that we are more likely to detect intermediate frequency alleles because they experience greater change. This reflects the simple fact that loci with more intermediate allele frequencies generate greater variance in fitness, all else equal. However, despite this ascertainment effect, there is still an excess of SNPs with significant tests with MAF > 0.3 in the real data (orange versus green in Figure 3).

627

628 Rapid evolution in a uniform environment may suggest a role for environmental heterogeneity,
629 but the experiment does not identify the most important aspects of heterogeneity. Many
630 selective agents are consistent with the results. For example, consider a polymorphism that is
631 maintained by spatial or temporal variation in natural enemies (pathogens or predators). One
632 allele confers resistance but at metabolic expense. This allele is favored when
633 predators/pathogens are abundant but disfavored in their absence. In the captive environment,
634 the natural enemy is absent and the low-defense allele is now under positive directional
635 selection. Now consider a different sort of antagonistic pleiotropy; a polymorphism maintained
636 by a life-history tradeoff where one allele has high viability at the expense of lower fecundity. If
637 we eliminate the agents of viability selection in the lab, such that nearly everyone survives, the
638 high-fecundity allele will now be under positive directional selection. Frequency-dependent
639 selection arising from competitive or social interactions (ANTONOVICS AND KAREIVA 1988) can
640 also maintain polymorphism that the lab environment could remove or reduce. While these
641 examples are varied in terms of the proximate causes of fitness variation, all involve trade-offs
642 present in nature that are eliminated in the laboratory environment.

643

644 ***Summary***—Predicting evolutionary response to a changing environment, and its genetic basis, is
645 one of the major goals of modern evolutionary genetics. However, recent genome-wide analyses
646 of evolution occurring in real time have failed to identify the targets of selection. Here we find
647 that both hitch-hiking and haplotype sampling effects are likely to lead to large over-estimation
648 of the number of sites that respond to selection in evolve-and-resequence studies. We also
649 propose new analytical and simulation approaches that mitigate this over-estimation. These

650 methods allow us to infer that loci that responded in parallel to a novel environment were
651 predominantly those maintained at intermediate frequency in the ancestral wild population. Even
652 so, these techniques do not allow us to distinguish the actual targets of selection from hitch-
653 hikers.
654
655 **Acknowledgments:** We thank R. Unckless, S. Macdonald, D. Houle, and J. Sztepanacz for input
656 on the project and/or manuscript. We thank Christopher Souders and Lauren Reynolds for lab
657 work and the KU ACF for computing support. **Funding:** This work was supported by grants
658 from the National Institutes of Health to J.K. (R01 GM073990) and NSF (IOS 1257735 and
659 DEB 1740466) to K.H.

660 **Literature Cited**

661

662 Antonovics, J., and P. Kareiva, 1988 Frequency-dependent selection and competition - empirical
663 approaches. Phil. Trans. Roy. Soc. Lond. B 319**:** 601-613.
664 Barghi, N., R. Tobler, V. Nolte and C. Schlotterer, 2017 Drosophila simulans: A Species with
665 Improved Resolution in Evolve and Resequence Studies. G3 (Bethesda) 7**:** 2337-2343.
666 Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh *et al.*, 2009 Genome evolution and
667 adaptation in a long-term experiment with *Escherichia coli*. Nature 461**:** 1243-U1274.
668 Barton, N. H., 1995 Linkage and the Limits to Natural-Selection. Genetics 140**:** 821-841.
669 Benjamini, Y., and D. Yekutieli, 2001 The control of the false discovery rate in multiple testing
670 under dependency. Annals of Statistics 29**:** 1165-1188.
671 Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt and D. A. Petrov, 2014 Genomic
672 Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in
673 Drosophila. PLOS Genetics 10**:** e1004775.
674 Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose *et al.*, 2010 Genome-
675 wide analysis of a long-term evolution experiment with Drosophila. Nature 467**:** 587-
676 U111.
677 Charlesworth, B., 2009 Effective population size and patterns of molecular evolution and
678 variation. Nat Rev Genet 10**:** 195-205.
679 Charlesworth, B., 2015 Causes of natural variation in fitness: Evidence from studies of
680 Drosophila populations. Proceedings of the National Academy of Sciences 112**:** 1662-
681 1669.
682 Comeron, J. M., R. Ratnappan and S. Bailin, 2012 The Many Landscapes of Recombination in
683 Drosophila melanogaster. PLOS Genetics 8**:** e1002905.
684 Conover, D. O., and S. B. Munch, 2002 Sustaining fisheries yields over evolutionary time scales.
685 Science 297**:** 94-96.
686 Corbett-Detig, R. B., and D. L. Hartl, 2012 Population genomics of inversion polymorphisms in
687 Drosophila melanogaster. PLoS Genet 8**:** e1003056.
688 Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil *et al.*, 2002 A single p450 allele
689 associated with insecticide resistance in Drosophila. Science 297**:** 2253-2256.
690 Darimont, C. T., S. M. Carlson, M. T. Kinnison, P. C. Paquet, T. E. Reimchen *et al.*, 2009
691 Human predators outpace other agents of trait change in the wild. Proceedings of the
692 National Academy of Sciences of the United States of America 106**:** 952-954.
693 Driscoll, C. A., D. W. Macdonald and S. J. O'Brien, 2009 From wild animals to domestic pets,
694 an evolutionary view of domestication. Proceedings of the National Academy of Sciences
695 106**:** 9971-9978.
696 Ellner, S. P., M. A. Geber and N. G. Hairston, Jr., 2011 Does rapid evolution matter? Measuring
697 the rate of contemporary evolution and its impacts on ecological dynamics. Ecol Lett 14**:**
698 603-614.
699 Ellner, S. P., N. G. Hairston, Jr., C. M. Kearns and D. Babai, 1999 The roles of fluctuating
700 selection and long-term diapause in microevolution of diapause timing in a freshwater
701 copepod. Evolution 53**:** 111-122.
702 Fisher, R. A., and E. B. Ford, 1947 The spread of a gene in natural conditions in a colony of the
703 moth *Panaxia dominula*. Heredity 1**:** 143-174.

Ford, E. B., 1964 *Ecological Genetics*. Methuen and Company, London, UK.

Franke, B., J. L. Stein, S. Ripke, V. Anttila, D. P. Hibar *et al.*, 2016 Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. Nat Neurosci 19**:** 420-431.

Franks, S. J., and A. E. Weis, 2008 A change in climate causes rapid evolution of multiple life-history traits and their interactions in an annual plant. Journal of Evolutionary Biology 21**:** 1321-1334.

Franssen, S. U., V. Nolte, R. Tobler and C. Schlotterer, 2015a Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental *Drosophila melanogaster* populations. Mol Biol Evol 32**:** 495-509.

Franssen, S. U., V. Nolte, R. Tobler and C. Schlötterer, 2015b Patterns of Linkage Disequilibrium and Long Range Hitchhiking in Evolving Experimental Drosophila melanogaster Populations. Molecular Biology and Evolution 32**:** 495-509.

Ghalambor, C. K., K. Hoke, E. Ruell, E. K. Fisher, D. N. reznick *et al.*, 2015 Non-adaptive plasticity and rapid adaptive evolution of the guppy transcriptome. Nature 525**:** 372–375.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, Oxford.

Gillespie, J. H., 2001 IS THE POPULATION SIZE OF A SPECIES RELEVANT TO ITS EVOLUTION? Evolution 55**:** 2161-2169.

Grant, P. R., 1999 *Ecology and Evoution of Darwin's Finches*. Princeton University Press, Princeton, N.J.

Hairston, N. G., S. P. Ellner, M. A. Geber, T. Yoshida and J. A. Fox, 2005 Rapid evolution and the convergence of ecological and evolutionary time. Ecology Letters 8**:** 1114-1127.

Hairston, N. G., and W. E. Walton, 1986 Rapid evolution of a life history trait. Proc Natl Acad Sci U S A 83**:** 4831-4833.

Haldane, J. B. S., 1957 The cost of natural selection. J. Genet. 55**:** 511-524.

Hendry, A. P., and M. T. Kinnison, 1999 Perspective: The pace of modern life: Measuring rates of contemporary microevolution. Evolution 53**:** 1637-1653.

Hu, T. T., M. B. Eisen, K. R. Thornton and P. Andolfatto, 2013 A second-generation assembly of the Drosophila simulans genome provides new insights into patterns of lineage-specific divergence. Genome Research 23**:** 89-98.

Huang, Y., and A. F. Agrawal, 2016 Experimental evolution of gene expression and plasticity in alternative selective regimes. PLoS Genet 12**:** e1006336.

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

Jain, K., and W. Stephan, 2017 Rapid adaptation of a polygenic trait after a sudden environmental shift. Genetics 206**:** 389-406.

Johnston, R. F., and R. K. Selander, 1964 House sparrows: rapid evolution of races in North America. . Science 144**:** .548-550.

Kang, L., D. D. Aggarwal, E. Rashkovetsky, A. B. Korol and P. Michalak, 2016 Rapid genomic changes in Drosophila melanogaster adapting to desiccation stress in an experimental evolution system. BMC Genomics 17**:** 233.

Kapun, M., H. van Schalkwyk, B. McAllister, T. Flatt and C. Schlotterer, 2014 Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of Drosophila melanogaster. Mol Ecol 23**:** 1813-1827.

Kelly, J. K., B. Koseva and J. P. Mojica, 2013 The Genomic Signal of Partial Sweeps in Mimulus guttatus. Genome Biology and Evolution 5**:** 1457-1469.

Kettlewell, H. D., 1958. A survey of the frequencies of *Biston betularia* (L.)(Lep.) and its melanic forms in Great Britain. Heredity 12**:** 51-72.

Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock *et al.*, 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500**:** 571-574.

Levene, H., 1953 Genetic equilibrium when more than one ecological niche is available. Amer. Natur. 87**:** 331-333.

Levy, S. F., J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher *et al.*, 2015 Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature 519**:** 181-186.

Lewontin, R. C., 1964 THE INTERACTION OF SELECTION AND LINKAGE. I. GENERAL CONSIDERATIONS; HETEROTIC MODELS. Genetics 49**:** 49-67.

Lewontin, R. C., 1974 *The genetic basis of evolutionary change*. Columbia University Press, New York, NY.

Long, A., G. Liti, A. Luptak and O. Tenaillon, 2015 Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nature Reviews Genetics 16**:** 567.

Losos, J. B., K. I. Warheit and T. W. Shoener, 1997 Adaptive differentiation following experimental island colonization in *Anolis* lizards. Nature 387**:** 70-73.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genetic research 23**:** 23-35.

Medland, S. E., N. Jahanshad, B. M. Neale and P. M. Thompson, 2014 Whole-genome analyses of whole-brain data: working within an expanded search space. Nature neuroscience 17**:** 791-800.

Messer, P. W., S. P. Ellner and N. G. Hairston, Jr., 2016 Can population genetics adapt to rapid evolution? Trends Genet 32**:** 408-418.

Meyer, R. S., and M. D. Purugganan, 2013 Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet 14**:** 840-852.

Michalak, P., L. Kang, P. M. Sarup, M. F. Schou and V. Loeschcke, 2017 Nucleotide diversity inflation as a genome-wide response to experimental lifespan extension in Drosophila melanogaster. BMC Genomics 18**:** 84.

Monnahan, P. J., and J. K. Kelly, 2017 The Genomic Architecture of Flowering Time Varies Across Space and Time in Mimulus guttatus. Genetics.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific Nuclear-Dna Variation In Drosophila. Molecular Biology and Evolution 13**:** 261-277.

Neher, R. A., and B. I. Shraiman, 2011 Genetic Draft and Quasi-Neutrality in Large Facultatively Sexual Populations. Genetics 188**:** 975-996.

Nuzhdin, S. V., and T. L. Turner, 2013 Promises and limitations of hitchhiking mapping. Current Opinion in Genetics & Development 23**:** 694-699.

Ohta, T., 1976 Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theor. Pop. Biol. 10**:** 254-275.

Orozco-terWengel, P., M. Kapun, V. Nolte, R. Kofler, T. Flatt *et al.*, 2012a Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Mol Ecol 21**:** 4931-4941.

Orozco-terWengel, P., M. Kapun, V. Nolte, R. Kofler, T. Flatt *et al.*, 2012b Adaptation of Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. Molecular Ecology 21**:** 4931-4941.

795 Palumbi, S. R., 2001 *The Evolution Explosion: How Humans Cause Rapid Evolutionary Change*.
796 W.W. Norton, New York, NY, USA.
797 Pena, C. J., H. G. Kronman, D. M. Walker, H. M. Cates, R. C. Bagot *et al.*, 2017 Early life stress
798 confers lifelong stress susceptibility Science 356**:** 1185-1188.
799 Plaisier, S. B., R. Taschereau, J. A. Wong and T. G. Graeber, 2010 Rank-rank hypergeometric
800 overlap: identification of statistically significant overlap between gene-expression
801 signatures. Nucleic Acids Res 38**:** e169.
802 Przeworski, M., J. D. Wall and P. Andolfatto, 2001 Recombination and the Frequency Spectrum
803 in Drosophila melanogaster and Drosophila simulans. Molecular Biology and Evolution
804 18**:** 291-298.
805 Remolina, S. C., P. L. Chang, J. Leips, S. V. Nuzhdin and K. A. Hughes, 2012 Genomic basis of
806 aging and life-history evolution in Drosophila melanogaster. Evolution 66**:** 3390-3403.
807 Reznick, D. N., F. H. Shaw, F. H. Rodd and R. G. Shaw, 1997 Evaluation of the rate of evolution
808 in natural populations of guppies (*Poecilia reticulata*). Science 275**:** 1934-1937.
809 Rose, M. R., 1984 Laboratory evolution of postponed senescence in *Drosophila melanogaster*.
810 Evolution 38**:** 1004-1010.
811 Schou, M. F., V. Loeschcke, J. Bechsgaard, C. Schlotterer and T. N. Kristensen, 2017
812 Unexpected high genetic diversity in small populations suggests maintenance by
813 associative overdominance. Mol Ecol.
814 Schrider, D. R., D. J. Begun and M. W. Hahn, 2013 Detecting Highly Differentiated Copy-
815 Number Variants from Pooled Population Sequencing, pp. 344-355 in *Pacific Symposium*
816 *on Biocomputing*.
817 Signor, S. A., F. N. New and S. Nuzhdin, 2018 A Large Panel of Drosophila simulans Reveals
818 an Abundance of Common Variants. Genome Biology and Evolution 10**:** 189-206.
819 Slobodkin, L. B., 1980 *Growth and Regulation in Animal Populations*. Dover, New York, NY.
820 Stuart, Y. E., T. S. Campbell, P. A. Hohenlohe, R. G. Reynolds, L. J. Revell *et al.*, 2014 Rapid
821 evolution of a native species following invasion by a congener. Science 346**:** 463-466.
822 Sukumaran, J., and M. T. Holder, 2011 Ginkgo: spatially-explicit simulator of complex
823 phylogeographic histories. Molecular Ecology Resources 11**:** 364-369.
824 Sved, J. A., T. E. Reed and W. F. Bodmer, 1967 The number of balanced populations that can be
825 maintained in a natural population. Genetics 55**:** 469-481.
826 Teles, R. M. B., T. G. Graeber, S. R. Krutzik, D. Montoya, M. Schenk *et al.*, 2013 Type I
827 Interferon Suppresses Type II Interferon–Triggered Human Anti-Mycobacterial
828 Responses. Science 339**:** 1448-1453.
829 Thompson, J. N., 1998 Rapid evolution as an ecological process. Trends in Ecology & Evolution
830 13**:** 329-332.
831 Tobler, R., S. U. Franssen, R. Kofler, P. Orozco-Terwengel, V. Nolte *et al.*, 2014 Massive
832 habitat-specific genomic response in D. melanogaster populations during experimental
833 evolution in hot and cold environments. Mol Biol Evol 31**:** 364-375.
834 True, J. R., J. M. Mercer and C. C. Laurie, 1996 Differences in crossover frequency and
835 distribution among three sibling species of Drosophila. Genetics 142**:** 507-523.
836 Trut, L., I. Oskina and A. Kharlamova, 2009 Animal evolution during domestication: the
837 domesticated fox as a model. BioEssays : news and reviews in molecular, cellular and
838 developmental biology 31**:** 349-360.
839 Turner, T. L., and P. M. Miller, 2012a Investigating Natural Variation in Drosophila
840 Courtship Song by the Evolve and Resequence Approach. Genetics 191**:** 633-642.

841 Turner, T. L., and P. M. Miller, 2012b Investigating natural variation in Drosophila courtship
842 song by the evolve and resequence approach. Genetics 191**:** 633-642.
843 Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice and A. M. Tarone, 2011a Population-based
844 resequencing of experimentally evolved populations reveals the genetic basis of body
845 size variation in *Drosophila melanogaster*. Plos Genetics 7.
846 Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice and A. M. Tarone, 2011b Population-
847 Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis
848 of Body Size Variation in Drosophila melanogaster. PLOS Genetics 7**:** e1001336.
849 Walsh, B., and M. Lynch, 2017 *Evolution and Selection of Quantitative Traits*.
850 http://nitro.biosci.arizona.edu/zbook/NewVolume_2/pdf/Chapter09.pdf.
851 Ward, J. K., and J. K. Kelly, 2004 Scaling up evolutionary responses to elevated CO_2: lessons
852 from Arapidopsis. Ecology Letters 7**:** 427-440.
853 Wright, S., 1931 Evolution in mendelian populations. Genetics 16**:** 97-159.
854

855

856 **Figure 1. The maximum LRT test value for parallel selection per 100kb window along**
857 **each chromosome. Red lines indicate centromeres.**
858
859 **Figure 2. The allele frequency spectrum is given for all SNPs in the ancestral populations**
860 **(blue bars) and for those that tested significant for parallel selection (orange bars). The**
861 **green broken lines denote proportions observed in a simulation of parallel selection on 20**
862 **loci with initial allele frequencies sampled from genome wide distribution (see text).**
863
864 **Figure 3. Rank-rank hypergeometric tests of overlap in divergence between populations. The rank**
865 **order of maximum divergence values in 100kb windows is compared. The color in each cell**
866 **indicates the hypergeometric probability that the rank ordering of windows is independent in the**
867 **two populations, after applying Benjamini-Yekutieli correction for multiple tests (Plaisier et al.**
868 **2010). Each cell in the map represents 10 100kb windows, with windows ranked from those**
869 **showing the greatest decrease in the reference allele frequency during laboratory adaptation**
870 **(left/bottom) to those showing the greatest increase in reference allele frequency (right/top).**
871 **Pattern of warm colors (small probability values) in cells in lower-left (upper-right) quadrats**
872 **indicates that populations being compared tend to have the same windows ranking among the top**
873 **(bottom) k windows, for many values of k.**
874
875 **Figure 4. The LRT test is applied to simulated data with 20 loci under selection**
876 **(parameters reported in text). The maximum LRT test value for parallel selection per**
877 **100kb window along chromosome 2. Each panel is a different simulation replicate.**
878
879
880

881 **Table 1. The estimated value of ν for each chromosome in each replicate population.**

Population	chromosome	Estimated null variance of divergence (ν)
A	2L	0.0174
A	2R	0.0171
A	3L	0.0228
A	3R	0.0190
A	X	0.0152
B	2L	0.0170
B	2R	0.0189
B	3L	0.0227
B	3R	0.0173
B	X	0.0179
C	2L	0.0159
C	2R	0.0163
C	3L	0.0200
C	3R	0.0176
C	X	0.0191

882

883

884

885

886

887 **Supplemental Materials.**

888

889 **Table S1. The hash tables for two locus haplotype (gamete) frequencies are reported as a**
890 **function of allele frequencies (minor allele at both loci) and the distance between loci.**
891 **Separate tables are given for autosomes and the X chromosome. The calculation**
892 **procedures are described on front sheet.**

893

894 **Table S2. The 126 polymorphisms with an LRT test for parallel selection that exceed**
895 **46.64. This threshold was established by permutation: The most extreme single test was**
896 **less than 46.46 in 95% of the permuted datasets.**

897

898 **Table S3. The crossovers location distribution used for simulation based on data from D.**
899 **melanogaster from COMERON et al. (2012). Rates are reduced rates near telomeres and**
900 **centromeres.**

901

902 **Table S4. The maximum LRT per chromosome (and genomewide) is reported for 1000**
903 **permutations of the data.**

904

905 **Table S5. Putative CNV that evolved in parallel in the three replicate populations that**
906 **overlap known coding and/or regulatory regions of *D. simulans* genes.**

907

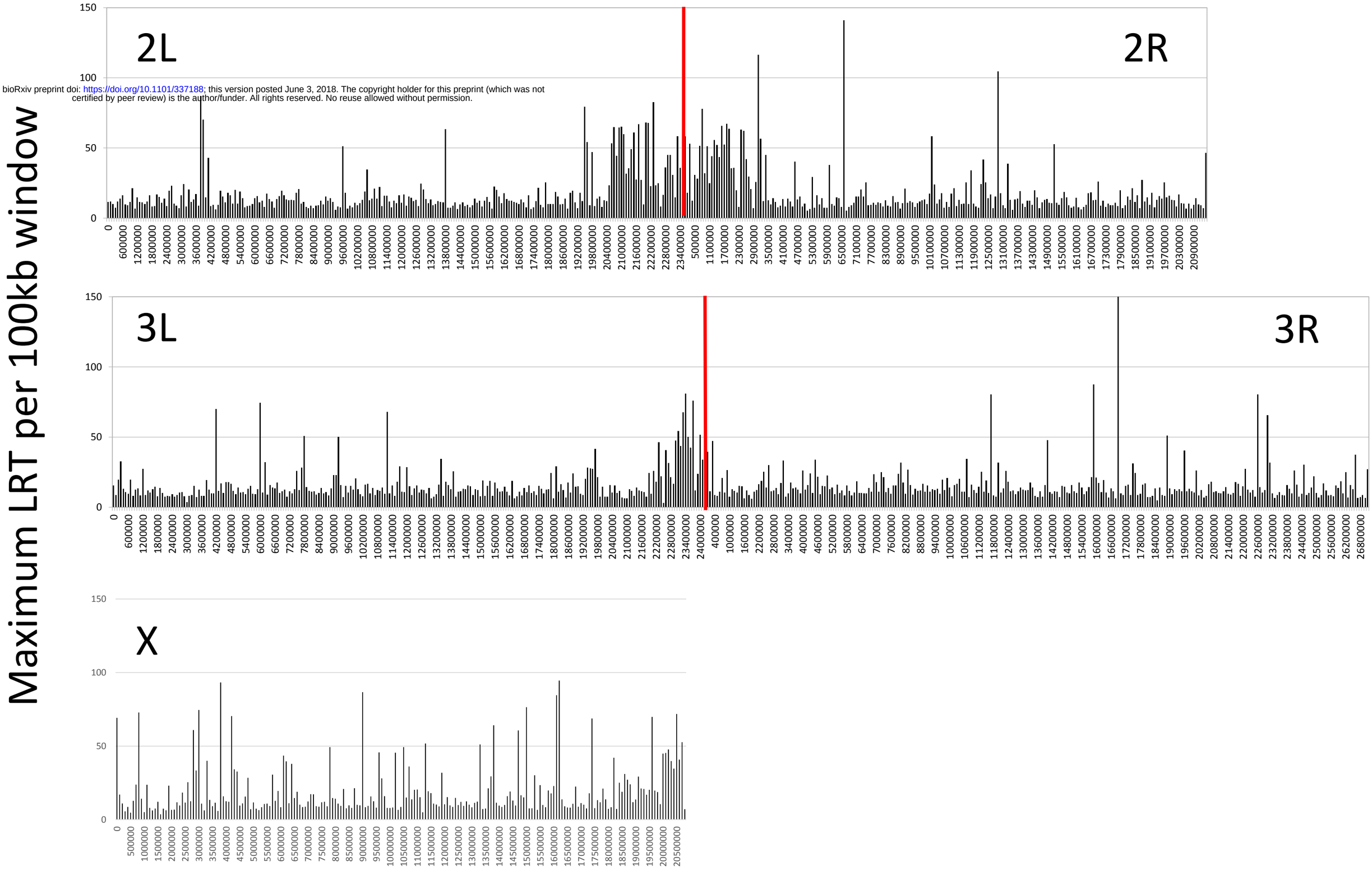
908 **Table S6. A summary of results from simulation – each parameter set is a distinct letter**
909 **(A-G) with statistics based on 1000 replicates of the entire experiment (three replicate**

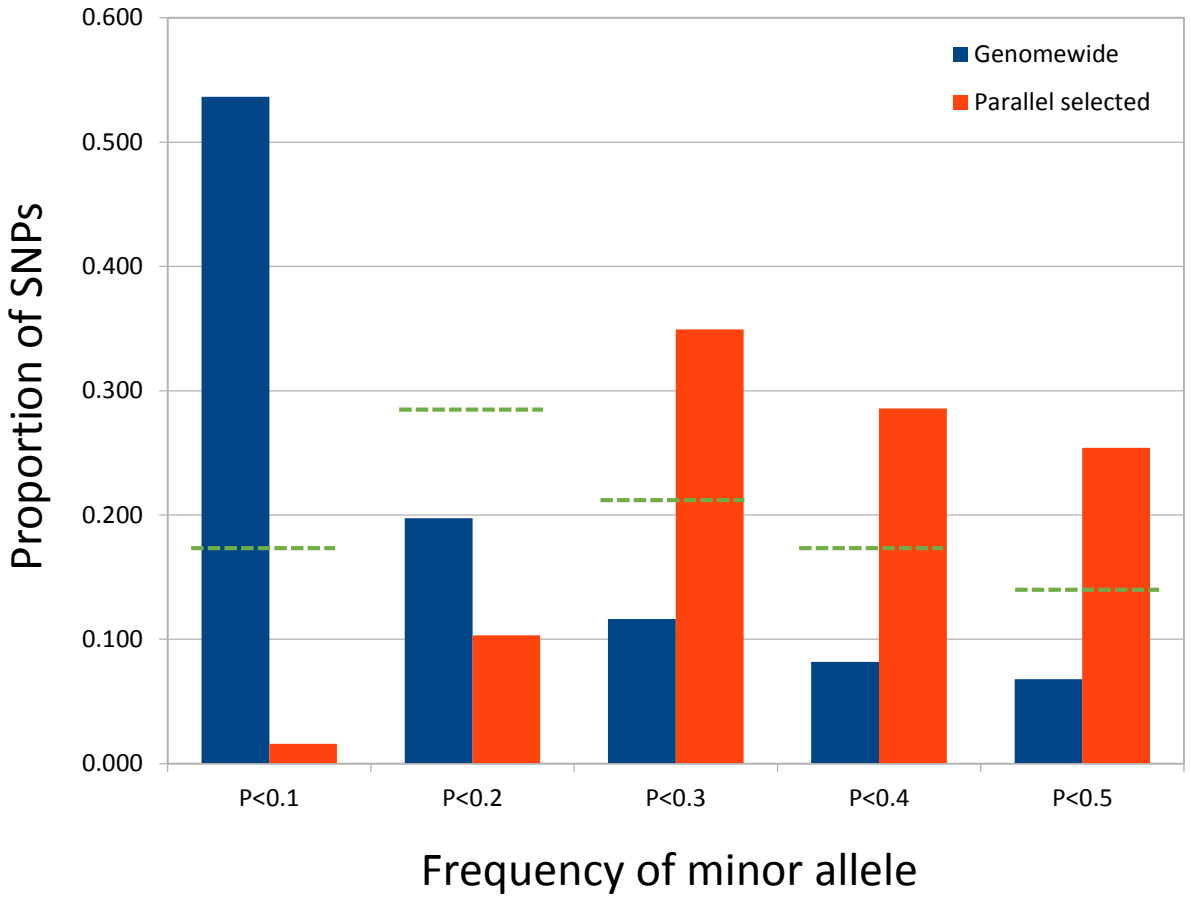
populations per experiment). The first part of each panel reports the mean and standard deviation (SD) of the null variance (v) from each simulation on both the autosomes and X chromosome. The mean is across all chromosomes and replicates. The SD was calculated both within each experiment (across chromosomes) and across all chromosomes and replicates, SD (within) and SD (overall), respectively. The second part of each panel reports the distribution of LRT for parallel evolution at both autosomes and on the X-chromosome. Here, the numbers are sums across all simulation replicates with overall calculations for the fraction of tests with LRT > 20 and LRT > 30, respectively.

Figure S1. (A) The distribution of changes in z between Ancestral and Descendant populations of replicate C (all SNPs). Here, z is the difference in transformed allele frequency ($x_A - x_D$) divided by the standard deviation. (B) The average absolute change in z (blue bars) and of untransformed allele frequency (orange bars) as a function of initial allele frequency. The latter is multiplied by 20 to be on same scale as the former.

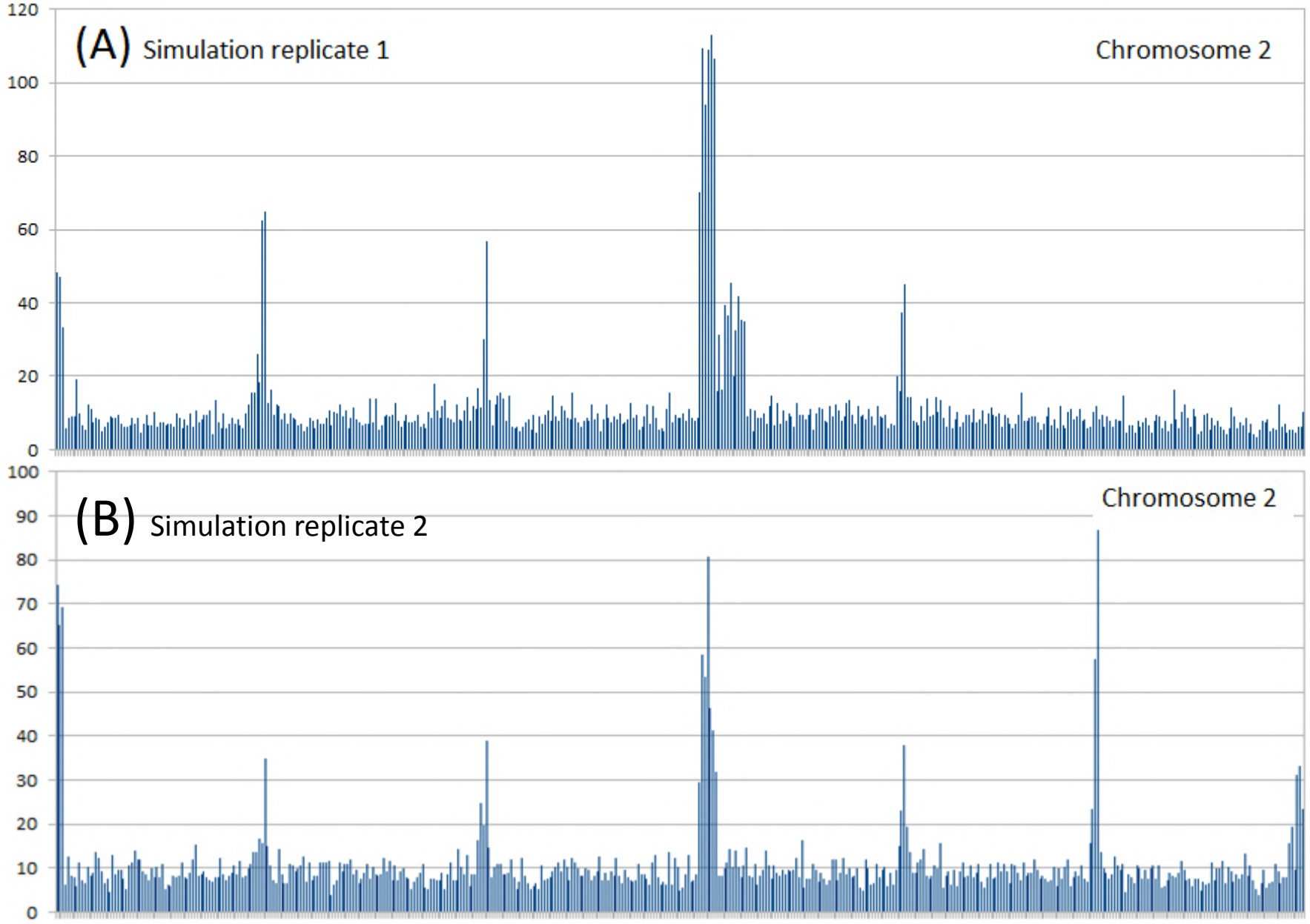
Figure S2. Rank-rank hypergeometric tests of overlap in divergence between populations at different window sizes. The rank order of windows with respect to the maximum divergence within the window is compared, using 10kb and 1kb windows. The color in each cell indicates the hypergeometric probability that the rank ordering of windows is independent in the two populations, after applying Benjamini-Yekutieli correction for multiple tests (Plaisier et al. 2010). Each cell represents 75 10kb windows or 400 1kb windows, respectively, with windows ranked from those showing the greatest decrease in the reference allele frequency during laboratory adaptation (left/bottom) to those showing

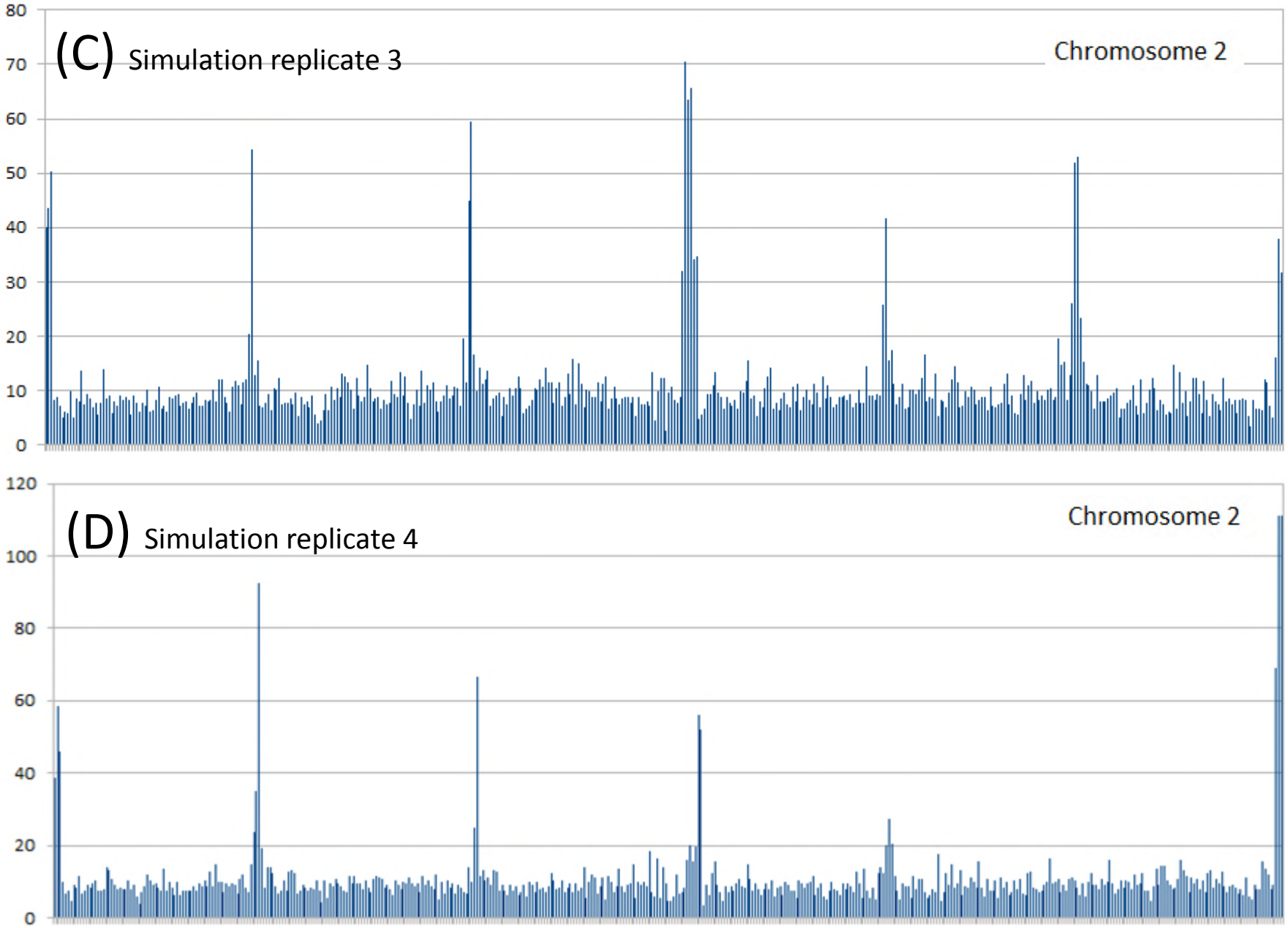
933 **the greatest increase in reference allele frequency (right/top). Pattern of warm colors**
934 **(small probability values) in cells along the lower-left (upper-right) quadrats indicates that**
935 **the rank order of windows in the two populations is more similar than expected by chance.**
936 **Note difference in color scale in the two different analyses shown here and in the analysis of**
937 **100kb windows shown in Figure 3 in the main text.**
938
939 **Figure S3. The average value for LD measured as r^2 is reported as a function of**
940 **chromosomal location. The mean refers to pairs of loci separated by 100-300kb and where**
941 **the minor allele at each locus is greater than 10%. The red line is the centromere for**
942 **chromosomes 2 and 3.**
943
944 **Supplemental.File.1.tar.gz :: The python programs LD.analysis1.py, LD.analysis2.py,**
945 **LD.analysis3.py used for Table S1.**
946
947




0.600
0.500
0.400
0.300
0.200
0.100
0.000
Proportion of SNPs
Genomewide
Parallel selected
P<0.1
P<0.2
P<0.3
P<0.4
P<0.5
Frequency of minor allele

Population B
Population A
Population C
Population B
Population C
Population A
38
30
23
15
8
0
-log(P-value)


(A) Simulation replicate 1
Chromosome 2
120
100
80
60
40
20
0
(B) Simulation replicate 2
Chromosome 2
100
90
80
70
60
50
40
30
20
10
0


(C) Simulation replicate 3
Chromosome 2
80
70
60
50
40
30
20
10
0
(D) Simulation replicate 4
Chromosome 2
120
100
80
60
40
20
0