

Genomics of Parallel Experimental Evolution in *Drosophila*

J.L. Graves Jr,^{†,1} K.L. Hertweck,^{†,2} M.A. Phillips,^{*,†,3} M.V. Han,^{†,4} L.G. Cabral,³ T.T. Barter,³ L.F. Greer,³ M.K. Burke,⁵ L.D. Mueller,³ and M.R. Rose³

¹Joint School of Nanoscience and Nanoengineering, North Carolina A&T State University and UNC Greensboro, Greensboro, NC

²Department of Biology, The University of Texas at Tyler, Tyler, TX

³Department of Ecology and Evolutionary Biology, University of California, Irvine, CA

⁴School of Life Sciences, University of Nevada, Las Vegas, NV

⁵Department of Integrative Biology, Oregon State University, Corvallis, OR

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: phillim1@uci.edu.

Associate editor: Nadia Singh

Abstract

What are the genomic foundations of adaptation in sexual populations? We address this question using fitness–character and whole-genome sequence data from 30 *Drosophila* laboratory populations. These 30 populations are part of a nearly 40-year laboratory radiation featuring 3 selection regimes, each shared by 10 populations for up to 837 generations, with moderately large effective population sizes. Each of 3 sets of the 10 populations that shared a selection regime consists of 5 populations that have long been maintained under that selection regime, paired with 5 populations that had only recently been subjected to that selection regime. We find a high degree of evolutionary parallelism in fitness phenotypes when most-recent selection regimes are shared, as in previous studies from our laboratory. We also find genomic parallelism with respect to the frequencies of single-nucleotide polymorphisms, transposable elements, insertions, and structural variants, which was expected. Entirely unexpected was a high degree of parallelism for linkage disequilibrium. The evolutionary genetic changes among these sexual populations are rapid and genomically extensive. This pattern may be due to segregating functional genetic variation that is abundantly maintained genome-wide by selection, variation that responds immediately to changes of selection regime.

Key words: experimental evolution, population genomics, adaptation.

Introduction

Genome-wide sequencing of experimentally evolved populations has emerged as a powerful method for parsing the genetic underpinnings of adaptation (Burke et al. 2010; Turner et al. 2011; Tenaillon et al. 2012; Schlötterer et al. 2015). An emerging pattern in this research is a contrast between the genomics of experimental evolution in asexual and sexual populations. Initial adaptation in the most common asexual paradigm, serially cultured *Escherichia coli*, features sequential selective sweeps of new mutations, at least over the first several thousand generations of laboratory selection (e.g., Barrick et al. 2009; Tenaillon et al. 2012; Maddamsetti et al. 2015). In contrast, initial adaptation in the most common sexual paradigm, outbred laboratory *Drosophila melanogaster*, apparently depends on moderate changes in the allele frequencies of standing genetic variation, rather than selective sweeps, at least for as many as 600 generations (Burke et al. 2010; Turner et al. 2011; Burke 2012; Orozco-ter Wengel et al. 2012; Tobler et al. 2014). Here, we present data for fitness characters and whole-genome, pooled, DNA sequences of 30 *Drosophila* laboratory populations taken from a large laboratory radiation of outbred populations (Rose et al. 2004; Mueller et al. 2013).

In this study, our primary goal is to establish whether the emerging contrast between asexual and sexual populations is sustained over a wide range of evolutionary durations, from dozens to nearly 1,000 generations of sustained selection regimes in sexual populations. Our second goal is to determine the degree to which phenotypic and genomic parallelism occurs between populations that share recently imposed versus long-sustained selection regimes. Our third goal is to probe the evolutionary genetic mechanisms underlying parallelism within each set of 10 populations that share a recent selection regime. The terms parallelism and convergence are often confused in the literature (Fong et al. 2005; Arendt and Reznick 2008), and in any case are terms of shifting usage. One such usage is that “convergence” refers to a similar phenotype evolving by different genetic mechanism among distantly related populations and species (Arendt and Reznick 2008), whereas “parallelism” refers to the emergence of similar functional, structural, and genomic features of populations that diverged from a common ancestor (Elias and Tawfik 2012). In this usage, our experimental system is well qualified to examine such “parallelism,” as all 30 study populations share a moderately distant ancestor by the standards of experimental evolution: The Ives population (vid. Ives 1970; Rose et al. 2004; Burke et al. 2016).

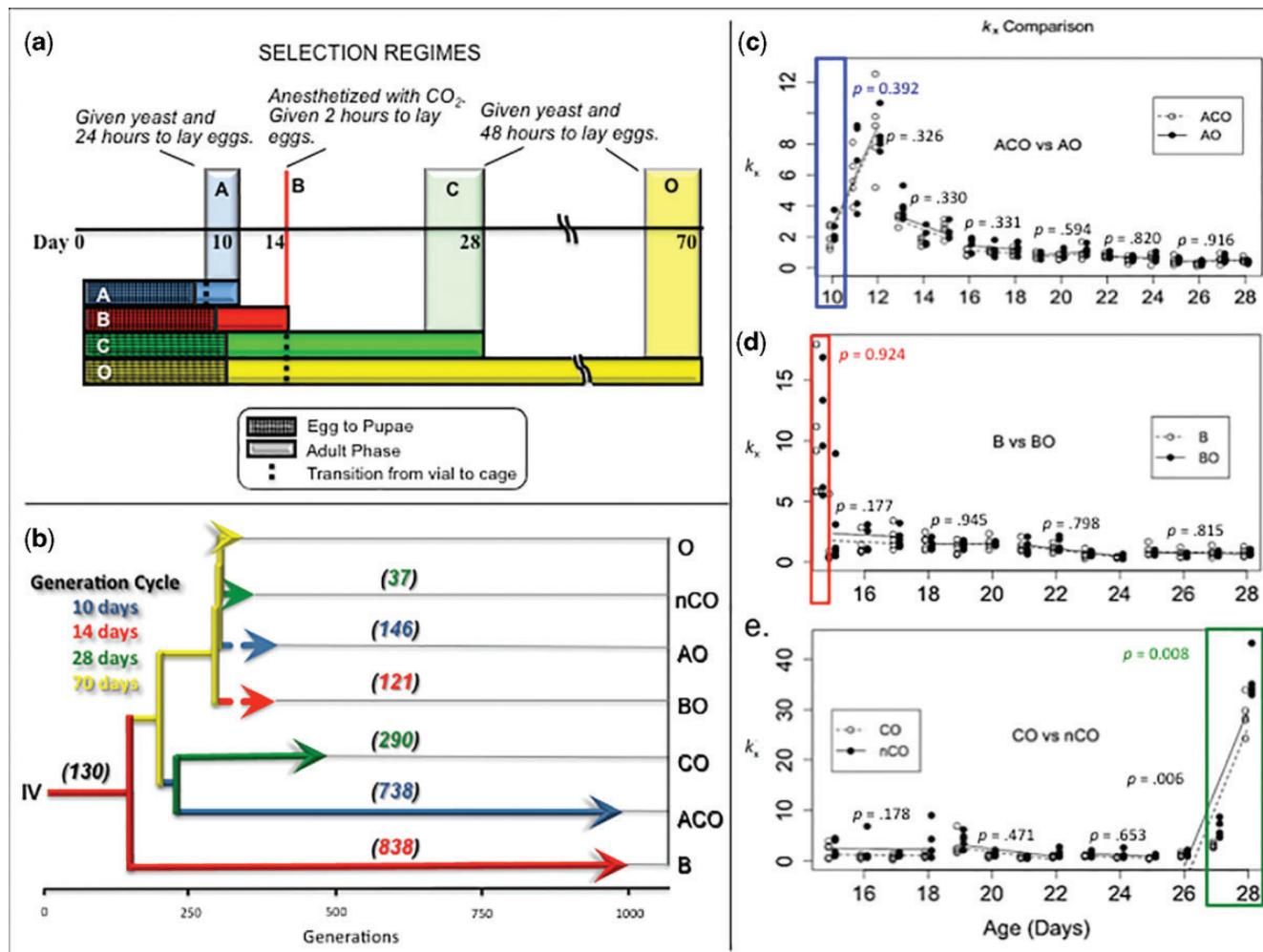


FIG. 1. Experimental design, fecundity, and survival. (A) Protocols for selection regimes, (B) evolutionary history of experimental populations and number of generations under selection regime for six groups of 5-fold replicated populations. (C–E) Comparison of k_x ($k_x = l_x m_x$, where l_x is probability of survival to age x and m_x is fecundity at age x) between long-standing and newly derived populations over a 19-day interval.

There have been a number of experimental evolution studies of phenotypic convergence or parallelism in *Drosophila* (Service et al. 1988; Teotónio and Rose 2000, 2001; Matos and Avelar 2001; Matos et al. 2002, 2004; Teotónio et al. 2009). Here we employ *Drosophila* stocks that feature 5-fold replicated groups of outbred populations that have been subjected to parallel sequences of selection regimes for decades (Rose et al. 2004; Burke et al. 2016). Three selection regimes were repeatedly imposed on these five-population groups, here called “A-type”, “B-type”, and “C-type” (fig. 1A). These selection treatments differ chiefly with respect to the length of their discrete generations, which are 10, 14, and 28 days, respectively. Fifteen of these populations were created decades ago and have long been subjected to either A, B, or C-type selection (the ACO, B, and CO populations, respectively). Matched to them are 15 populations recently derived from five common ancestral “O” populations and since subjected to one of A, B, or C-type selection (the AO, BO, and nCO populations). In total, this experimental system features 30 populations subsequently assessed using pooled genome-wide sequencing and focal fitness

assays, structured as 3 groups of the 10 populations each sharing the same recent selection history (fig. 1B).

Results

Phenotypes

We performed assays of fecundity and survival during the same generations as those used to collect samples for DNA sequencing, testing for functional parallelism between long-standing and newly derived populations (k_x , fig. 1C–E; l_x , m_x , see supplementary fig. S1 and tables S1 and S2 for P -values, Supplementary Material online). We found statistically significant phenotypic differentiation between groups of populations sharing their most recent selection regime for just one fitness-character: Newly derived nCO populations had superior reproductive output to long-standing CO populations during a single time interval (fig. 1E). Overall, the fitness-character results show a lack of evolutionary differentiation between long-standing and newly derived populations sharing the same distal selection regime. This rapid loss of differentiation is consistent with previous *Drosophila* studies on

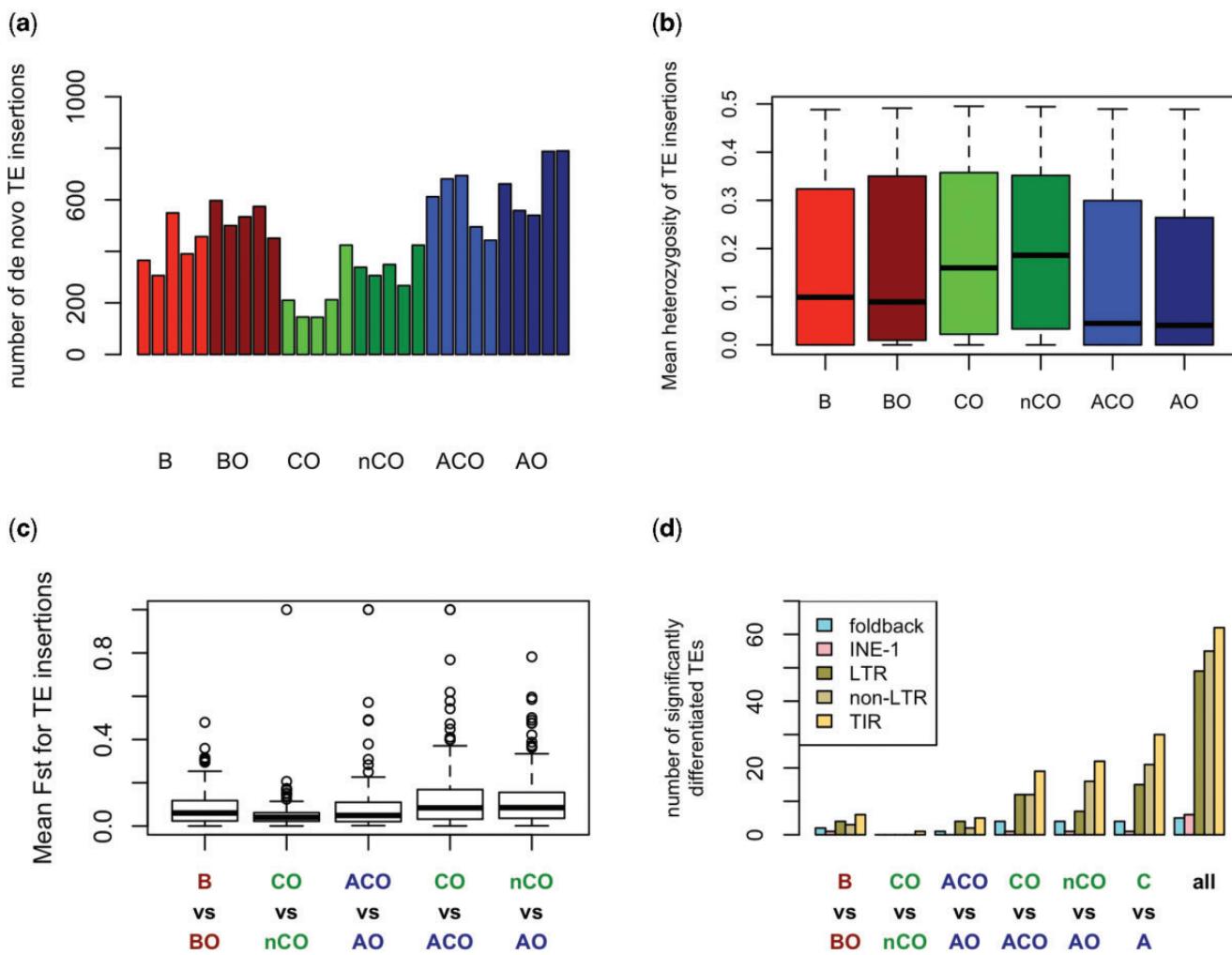


Fig. 2. Transposable element (TE) insertions compared among treatments. (A) Counts of total TE insertions in each population from coverage-normalized PoPopulationTE analysis. (B) Mean genome-wide heterozygosity from T-lex2 insertion analysis across all five replicates by treatment. (C) Mean genome-wide F_{ST} from T-lex2 insertion analysis for within and between treatment comparisons. (D) Significantly differentiated CMH results from T-lex2 insertion analysis for within and between treatment comparison separated by TE type.

the experimental evolution of functional characters (Teótonio and Rose 2000; Rose et al. 2004; Fragata et al. 2013).

Genomic Analysis

Given the lack of phenotypic differentiation with shared recent selection regimes, we aimed to identify potential underlying genomic factors that similarly lack differentiation. Our average sequencing coverage varied between populations, but was $\geq 50X$ for all populations except CO₅ (25X) and B₃ (29X) (supplementary table S3, Supplementary Material online). Genomic analysis focused on three types of variation in the pooled population sequences: 1) single nucleotide polymorphisms (SNPs), 2) structural variants (SVs), and 3) transposable elements (TEs).

We identify ~ 1.01 million SNPs across the major chromosome arms, 1,200 deletions and 490 tandem duplications (SVs), as well as 177 TEs shared with the reference genome which vary in insertion frequency among our populations. Between 144 and 790 TEs total occur in each population (as identified by PoPopulationTE, fig. 2A). Most detected SVs and variable TEs are $>2,000$ bp, with additional frequency

peaks at 5, 7.5 and 9 kb (data not shown), recapitulating the pattern found for TEs among the DGRP populations (Zichner et al. 2013). We find no concrete evidence for the presence of any of the seven inversions for which marker alleles are available (Kapun et al. 2014). For six of these inversions, our populations were fixed for the reference allele and not the inversion-specific allele, suggesting these inversions are not segregating in our population. For the seventh inversion [In(3R)Mo], 10 out of the 150 SNPs that are In(3R)Mo specific are segregating in our population, but at fairly low frequency (0.01–0.29). Given that this is a relatively young insertion, it is not unexpected that some inversion SNPs may still be shared with noninverted chromosomes.

Extensive heterozygosity is maintained in our populations, despite hundreds of generations in the laboratory, as previously found by Burke et al. (2010) for the five ACO populations and the five CO populations. In terms of heterozygosity calculated from our SNP data, we find few large regions where genetic variation has been depleted when heterozygosity is calculated over 100 kb windows (fig. 3). Regions where heterozygosity has been reduced to low levels are predominately

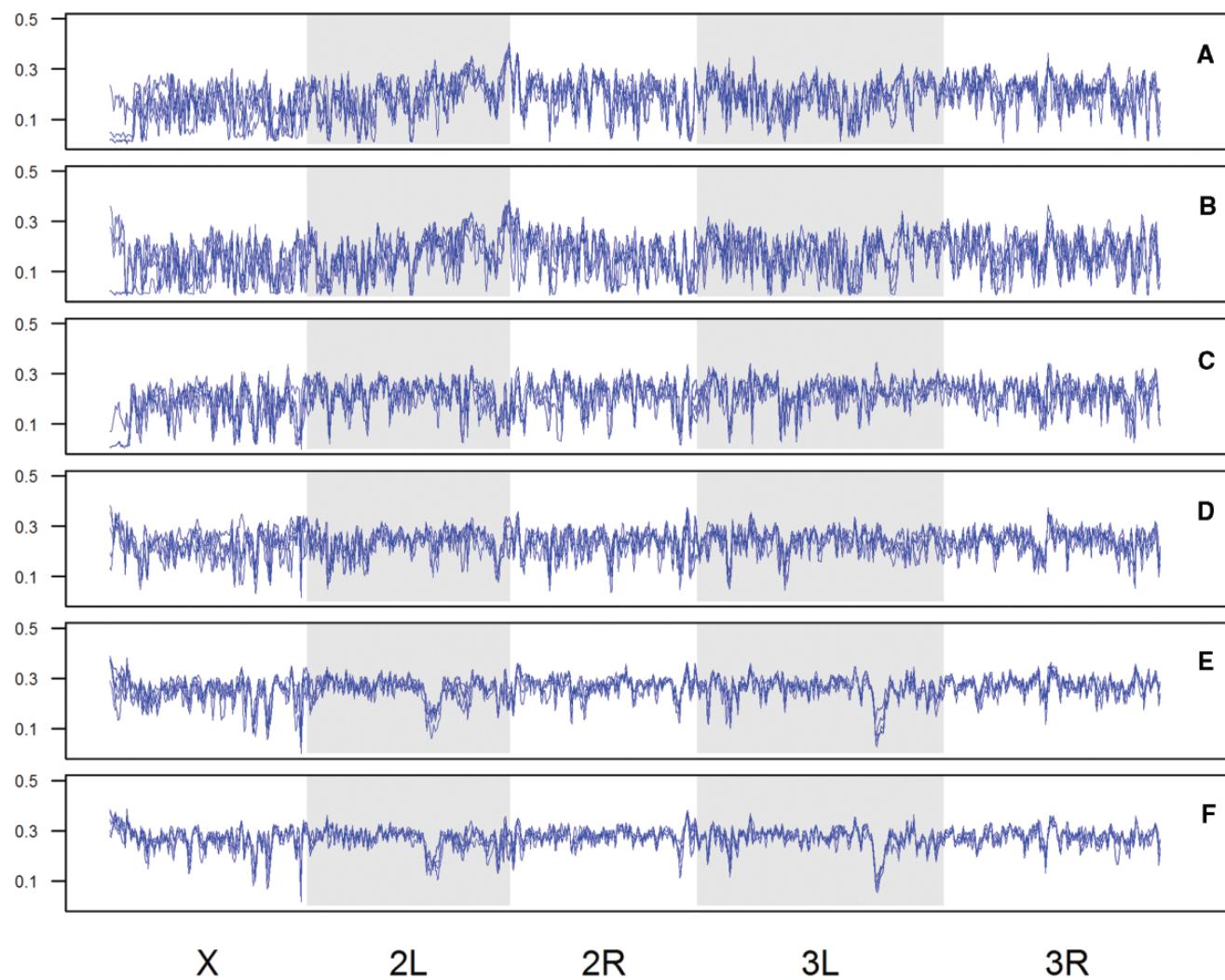


Fig. 3. Heterozygosity calculated from SNP data plotted across 100 kb nonoverlapping windows across all major chromosome arms for ACO_{1–5} (A), AO_{1–5} (B), B_{1–5} (C), BO_{1–5} (D), CO_{1–5} (E), and nCO_{1–5} (F). Each panel shows results from all five replicate populations plotted individually.

found in the A-type populations. This result is robust to reductions in window size (see [supplementary figs. S2 and S3, Supplementary Material](#) online for 50 and 30 kb results, respectively). Mean heterozygosity is also lower in populations subjected to A-type selection (SNPs: [supplementary table S4, Supplementary Material](#) online; TEs: [fig. 2B](#); SVs: [supplementary table S5, Supplementary Material](#) online), which could be due to more intense A-type selection. For heterozygosity calculated from our SNP data, these differences were found to be statistically significant based on *t*-tests comparing mean genome wide heterozygosity between groups of populations (see [supplementary table S6, Supplementary Material](#) online for *P*-values). We find that more variation is maintained in C-type populations than both A-type and B-type populations, with more in B-type populations than A-type.

As for patterns of similarity in variation across replicates, we find a high degree of similarity between replicate populations with parallel evolutionary histories, as indicated by mean genome wide F_{ST} estimates that are all less than 0.10 from SNP data ([supplementary table S7, Supplementary Material](#) online). Given the range of these values

(0.041–0.087), as well as correspondingly low F_{ST} estimates from TEs ([fig. 2C](#)), we suggest that [Wright \(1978\)](#) would have described the variation between replicates within each selective treatment as “small”. This pattern is recapitulated by visual examination of heterozygosity for replicate populations ([supplementary fig. S4, Supplementary Material](#) online). We also display this finding by plotting the variance in allele frequencies, per SNP, for each of the selection treatments ([supplementary fig. S5, Supplementary Material](#) online). Observed variances in raw SNP frequencies are very low across all treatments, implicating parallel evolution of standing variants among the five replicates of each treatment. Although we do not find any evidence for widespread reductions in SNP heterozygosity across our 100 kb windows, there are a number of local depressions in heterozygosity in all of our populations. We also find many localized depressions in SNP heterozygosity, defined as 100 kb windows with heterozygosity less than 0.2, which are consistent across given groups of replicate populations ([supplementary tables S8–S13, Supplementary Material](#) online). More regions like this are found in our populations subjected to A-type selection

(ACO: 105 regions, AO: 161 regions), than populations subjected to B-type (B: 59, BO: 11), and C-type selection (CO: 5, nCO: 8). This is suggestive of selection acting at parallel at sites within these regions (vid. Oleksyk et al. 2010). This result is consistent with other *Drosophila* work showing that adaptation in experimentally evolved populations is driven by selection on standing genetic variation (Burke et al. 2010; Turner et al. 2011; Orozco-ter Wengel et al. 2012; Tobler et al. 2014). Heterozygosity levels are similar between replicate populations of the same treatments for both TEs and SVs (fig. 2B; supplementary table S5, Supplementary Material online). Moreover, total number of TEs in each population identified by PoPopulationTE (fig. 2A) varied significantly in between-treatment comparisons (CO vs. ACO: $P = 0.0108$; nCO vs. AO: $P = 0.0021$). Comparisons of new (AO) versus longstanding (ACO) A-type populations showed no such differentiation ($P = 0.4688$). C-type populations did show evidence of differentiation in TE load ($P = 0.0406$), although this we attribute to many fewer generations of C-type selection (fig. 1B).

Previous work with these populations sequenced only ACO and CO populations, and pooled DNA across replicates such that direct observations of parallelism within evolutionary histories at the nucleotide level were not possible (Burke et al. 2010). Burke et al. (2010) did sequence a single replicate population (ACO1) and allele frequencies in this single replicate were highly similar to allele frequencies in the entire pool of ACO flies. The present work corroborates this observation and expands upon it considerably. Although widespread migration between replicates of each selection treatment would also produce the patterns we observe here, we contend that evolutionary parallelism at the SNP level is the more likely underlying scenario. For further discussion of how our results compare with those from Burke et al. 2010; see supplementary fig. S6, Supplementary Material online.

To formally and systematically assess levels of differentiation between the six groups of 5-fold replicated populations, we performed Cochran–Mantel–Haenszel (CMH) tests comparing SNP, SV or TE frequencies between and among groups that shared distal selection regimes. We find lower levels of differentiation among newly derived and long-standing populations recently subjected to the *same* type of selection, versus comparisons involving populations recently subjected to *different* types of selection as indicated by more significantly differentiated variants in the latter (SNPs: figs. 4 and 5; supplementary fig. S7, Supplementary Material online; TEs: fig. 2D; SVs: supplementary table S14, Supplementary Material online).

We performed nine statistically conservative comparisons of SNP differentiation between populations that did not share recent selection regimes. In six of these tests, dozens to hundreds of significant differentiated SNPs are found (fig. 4; supplementary fig. S7, Supplementary Material online; table 1). The two comparisons between groups of five B-type and five C-type groups did not yield significantly differentiated SNPs. We do, however, find significantly differentiated SNPs when we compare all 10 B-type populations with all 10 C-type populations (fig. 4). In contrast, across the three comparisons of populations that shared a recent selection regime (ACO vs.

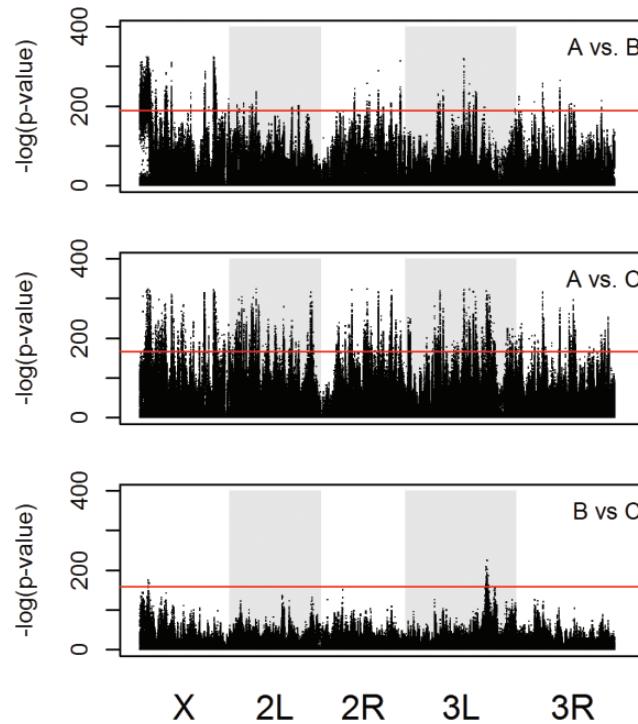


FIG. 4. Results from CMH tests comparing SNP frequencies between populations that are from different selection treatments. For these comparisons, we grouped populations based purely on their most recent selection regime. For example, we compared all 10 A-types (ACO and AO populations) with all 10 B-types (B and BO populations). Results are plotted as $-\log(P\text{-values})$ across all major chromosome arms. Significance thresholds for each comparison are derived from permutation tests and are indicated by a red line.

Table 1. Counts of Significantly Differentiated SNPs from CMH Tests Comparing Frequencies between and among Selective Regimes.

Comparison	Significance Threshold ^a	Number of Significant SNPs
ACO versus CO	5.67×10^{-145}	412
AO versus nCO	1.64×10^{-166}	143
A versus C	8.21×10^{-167}	10,109
ACO versus B	1.25×10^{-158}	1,146
AO versus BO	2.77×10^{-149}	42
A versus B	8.56×10^{-189}	4,152
CO versus B	1.11×10^{-183}	0
nCO versus BO	1.78×10^{-120}	0
B versus C	6.40×10^{-159}	64
ACO versus AO	2.91×10^{-230}	0
B versus BO	2.18×10^{-201}	0
CO versus nCO	1.70×10^{-126}	0

^aSignificance thresholds for each test were determined based on our permutation approach to correcting for multiple comparisons.

AO; B vs. BO; CO vs. nCO), but had different evolutionary histories, we detect no significantly differentiated SNPs (fig. 5).

In our between-treatment comparisons, we find more significantly differentiated SNPs in our comparisons between long-standing populations (e.g. ACO vs. CO) versus comparisons between newly derived populations (e.g. AO vs. nCO) (supplementary fig. S7, Supplementary Material online; table 1). This finding is consistent with previous simulations of

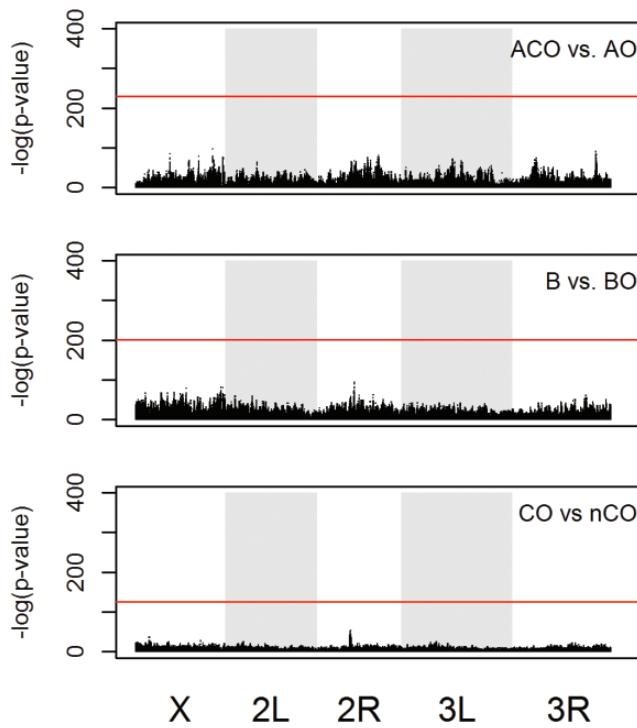


FIG. 5. Results from CMH tests comparing SNP frequencies between populations that share a recent selection regime. Results are plotted as $-\log(P\text{-values})$ across all major chromosome arms. Significance thresholds for each comparison are derived from permutation tests and are indicated by a red line.

experimental evolution in sexual populations, which suggest that increasing number of generations under selection increases the power to detect SNPs underlying responses to directional selection (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014).

When we increase replication by treating our longstanding and newly derived populations that share a recent selection regime as equivalent replicates (e.g., all 10 A-types, merging ACO with AO populations), we detect more significantly differentiated SNPs. This finding is again consistent with previous simulations of experimental evolution, which demonstrate a strong impact of varying the number of evolving replicates on the experimental detection of SNPs underlying responses to directional selection (Baldwin-Brown et al. 2014; Kofler and Schlötterer 2014). It is also consistent with the results of a similar analysis of yeast experimental evolution (Burke et al. 2014).

We also statistically tested the effect of selection treatments on our replicates using ANOVA on SNP allele frequency, considering selection regimes and age of establishment as factors (supplementary fig. S8, Supplementary Material online). The variance in allele frequency could be explained by the effect of selection regime in 0.4% of the variant loci. In contrast, differences in allele frequency change could be explained by the age of the branching for only 0.005% of loci (supplementary fig. S9, Supplementary Material online). Although the number of causal SNPs is overestimated for both of these factors because of linkage, selection regime appears to affect two orders of magnitude more loci than age of establishment. This result

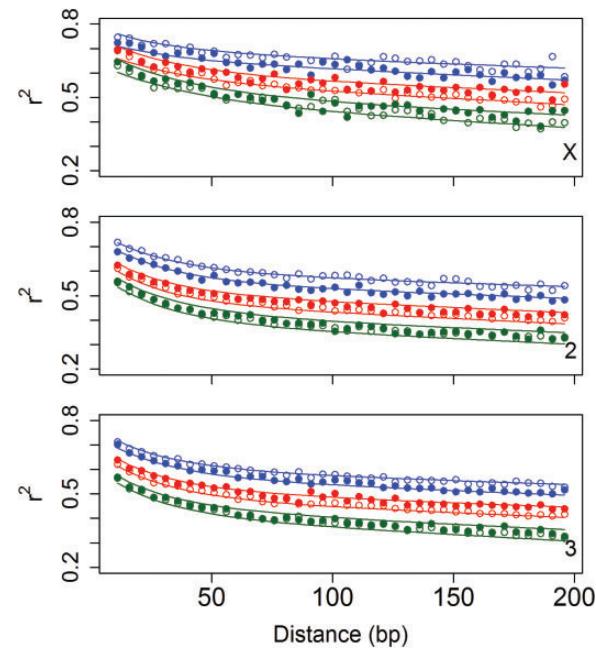


FIG. 6. Mean LD (r^2) estimates from LDx for chromosomes X, 2, and 3, plotted against distance from a focal SNP for six groups of populations. Closed circles are estimates for long-standing populations (ACO, B, and CO); open circles are estimates for newly derived populations (AO, BO, and nCO). A, B, and C-type populations are blue, red, and green, respectively. Colored lines represent simultaneous confidence intervals based on predictions from the biexponential model to which the data were fitted.

suggests that selection regime preponderantly influenced the genetic differentiation of these populations.

A key question in studies of sexual laboratory populations with low to moderate effective population sizes is level of linkage disequilibrium (LD). If there are high levels of LD, the number of genomic sites targeted by selection may be far smaller than the number of SNPs exhibiting statistically significant differentiation. We characterized linkage in our populations using LDx (Feder et al. 2012) and fit these estimates to a biexponential model. Populations subjected to different selection regimes had statistically distinct patterns of LD decay, but newly derived and long-standing populations subjected to the same selection did not (fig. 6). Furthermore, LD was higher in A-type populations compared with B and C-type populations. This pattern cannot be explained by number of generations, because long-standing ACO populations had many more generations for recombination to break down LD compared with C-type and BO populations. In fact, LD was lowest in the C-type populations, which have the fewest generations under laboratory cultivation. It appears that LD (like SNP, TE, and SV frequency differentiation) is highly parallel, depending primarily on recent selection history. LD patterns may be determined by relative intensities of selection at loci across the entire genome. In particular, consistently higher levels of LD in the A-type populations may be due to more intense selection, as suggested by the low number of individuals surviving to reproduction during this type of selection's initial generations. However,

formal simulation would be required to support this conjecture, as we lack the haplotype sequencing data required to characterize long range LD (cf. [Franssen et al. 2015](#)).

Discussion

The results of our functional and genomic analyses show that experimental evolution of moderately outbred *Drosophila* involves reproducible and extensive changes across the fruit fly genome. Marked functional and genomic differentiation between newly derived populations that do *not* share recent selection regimes is comparable to that found between long-standing populations that do not share selection regimes, despite large disparities in number of generations under divergent selection. This finding supports the contention that convergent, or “parallel” depending on favored usage, evolution can result from polymorphisms that were shared in an ancestral population (cf. [Stern 2013](#)), as all selected lines in this study were derived from a common ancestral population ([Rose et al. 2004](#); [Burke et al. 2016](#)).

This consistent pattern across six evolutionary histories supports the hypothesis that outbred sexual populations rapidly respond to selection in a reproducible manner, because they maintain functional genetic variation at many sites across their genomes. We observe clear evolutionary differentiation in response to selection that is not associated with complete local elimination of genetic variation near sites of genomic change. This provides evidence that evolutionary change in sexual populations is not driven by hard selective sweeps on newly arisen beneficial mutations of large effect ([Maynard Smith and Haigh 1974](#); [Burke 2012](#)), even over periods of more than 800 generations. These dynamics differ from those observed in long-term evolution experiments with *E. coli*, which are dominated by hard selective sweeps, clonal interference, and clonal replacement (e.g., [Maddamsetti et al. 2015](#)), and which also find different alleles driving change in different replicate populations ([Woods et al. 2006](#); [Tenaillon et al. 2012](#)). Similar results to ours have surfaced repeatedly in genomic studies of *Drosophila* lab evolution ([Turner et al. 2011](#); [Orozco-ter Wengel et al. 2012](#); [Tobler et al. 2014](#)), though those studies featured notably less replication than our study, fewer generations of selection, and notably less-extensive genomic responses to selection. We suggest that the genomic foundations for the experimental evolution of outbred sexual populations are different in kind from those of strictly clonal paradigms of experimental evolution (cf. [Barrick et al. 2009](#); [Tenaillon et al. 2012](#); [Maddamsetti et al. 2015](#)), in which selective sweeps by newly arisen mutants are the chief determinants of adaptation. The scale of the our study underscores the following conclusions about the evolution of outbred sexual populations: 1) their evolution can be fast; 2) their evolution can be repeatable from nucleotide to fitness, thanks to standing genetic variation; 3) their adaptation does not wait for new functional mutations and subsequent hard selective sweeps.

It might be proposed that a single haplotype is the target of selection for each of the A, B, and C selection regimes. But if that were the case, the rapid convergence on A, B, and C

phenotypes and genotypes by the recently derived A, B, and C lines (AO, BO, and nCO, respectively) suggests that the selection coefficients associated with these three selection regimes are large in magnitude. That in turn requires that the 15 populations long-subjected to these three selection regimes (ACO, B, and CO) should be approaching fixation at many sites across the genome, a pattern that is not apparent in the heterozygosity data for the B and CO populations, at least. Therefore, we suggest that our results indicate that most of the genomic sites under selection in our study are undergoing a shift between stable balanced polymorphisms, not selective sweeps toward fixation of a single genome-wide haplotype.

Any laboratory evolution study faces the challenging question of its applicability to the evolution of real-world populations in nature. Our point of view on this question is that we do *not* think that it is likely that sexual populations in nature generally undergo stable selection regimes for up to 1,000 generations. Thus our study is not intended to directly model what occurs in natural populations. Instead, our study is intended to use extreme and unusual evolutionary histories to test competing evolutionary genetic hypotheses. Most such hypotheses are hard to test using the shifting, short-term, and ambiguous data available from most studies of evolution in nature. Our strategy has long been based on strong-inference experimentation (cf. [Platt 1964](#)), which requires laboratory control of culture regimes over long periods, as in the classic LTEE study of [Lenski et al. \(1991\)](#). Thus, for example, if hard selective sweeps were the predominant mode of adaptation to the selection regimes we used, then the 15 populations that featured up to 1,000 generations of sustained selection should have produced numerous genomic regions with negligible heterozygosity. As we did not observe such patterns, we consider this a refutation of the hypothesis that adaptation predominantly proceeds by hard selective sweeps in sexual populations.

Materials and Methods

Experimental Evolution System

The laboratory phylogeny used for this experiment consisted of 30 *D. melanogaster* populations that underwent selection for more than three decades, with five replicate populations maintained for each of six different evolutionary histories. The estimated effective population size for each population is near 1,000 for populations maintained with reproduction in either vials or cages ([Mueller et al. 2013](#)). For the 30 populations studied here, we had three selection treatments which differed chiefly with respect to the length of their discrete generations: The duration from egg-collection that began each generation to the egg-laying that started the following generation. A-type selection requires rapid larval development, with flies newly emerged from rearing vials transferred to population cages for egg-laying, in order to start the next discrete generation at about 10 days of age. B-type selection involves population maintenance exclusively in rearing vials, with laying adults harvested at 14 days for a few hours of egg-laying to start the next discrete generation. C-type selection is like A-type selection, except that adults are collected at 14 days from egg, and then these adults are kept in cages for

12 days with unyeasted plates of medium, followed by 2 days of yeasted plates on which egg-laying occurs.

Fifteen populations of A, B, and C-type populations were created long ago with respect to generation number: Five “ACO” populations that had undergone 737 generations of A-type selection at the time of the experiments reported here; 5 “B” populations that had undergone 837 generations of B-type selection; and five “CO” populations that had 290 generations of C-type selection. Corresponding to them are 15 populations all derived from 5 “O-type” populations (vid. Rose 1984) that were cultured from eggs laid by females 9–10 weeks old: 5 “AO” populations given 146 generations of A selection; 5 “BO” given 121 generations of B selection; and 5 “nCO” given 37 generations of C selection. Their derivation is described in detail in Burke et al. 2016. These generation numbers for the three selection regimes are all calibrated to the generation at which flies were sampled for pooled DNA sequencing.

Phenotypic Assay

The flies assayed phenotypically were taken from the same generation as those we used for pooled DNA sequencing. Assays of fecundity and age-specific survival were conducted in cages holding adults during the 19-day interval between day 9 (from egg) and day 28 (from egg). This time period includes the longest duration that any adult fly is allowed to live in our present stock system, which no longer includes the O populations. Fitness components were calculated from the data collected from the assay cages during this 19-day interval. As is our normal practice, all phenotypic assays were performed after two generations of common-garden rearing using 14-day B-type culture (vid. Rose et al. 2004).

To test for phenotypic differentiation between newly derived and long standing replicates of the same treatment, we tested effects of selection on fecundity over three to four consecutive ages. The observations consisted of fecundity at a particular age (t) but within a small age interval ($k = 1, 2, \dots, m$). Within each interval, mortality or fecundity rates were modeled by a straight line and allowing selection regime [$j = 1$ (ACO or B or CO), 2 (AO or BO or nCO)] to affect the intercept of that line but not the slope. However, slopes were allowed to vary between intervals. Populations ($i = 1, \dots, 10$) were assumed to contribute random variation to these measures. With this notation the fecundity at age- t , interval- k , selection regime- j , and population- i , is y_{ijk} and is described by

$$y_{ijk} = \alpha + \beta_k + \delta_j \gamma_j + (\omega + \pi_k \delta_k) t + \delta_k \delta_j \mu_{jk} + c_i + \varepsilon_{ijk} \quad (1)$$

where $\delta_s = 0$ if $s = 1$ and 1 otherwise and c_i and ε_{ijk} are independent standard normal random variables with variance σ_c^2 and σ_ε^2 , respectively. The effects of selection on the intercept are assessed by considering the magnitude and variance of both γ_j and μ_{jk} .

DNA Extraction and Sequencing

Genomic DNA was extracted from samples of 120 female flies collected from each of the 30 individual populations

(ACO_{1–5}, AO_{1–5}, CO_{1–5}, nCO_{1–5}, B_{1–5}, BO_{1–5}) using the Qiagen/Gentra Puregene kit, following the manufacturer's protocol for bulk DNA purification. The 30 gDNA pools were prepared as standard 200–300 bp fragment libraries for Illumina sequencing and constructed such that each five replicate populations of a treatment (e.g., ACO_{1–5}) were given unique barcodes, normalized, and pooled together. Each 5-plex library was run on individual PE100 lanes of an Illumina HiSeq 2000 at the UNC High Throughput Sequencing Facility. Resulting data were 100 bp paired-end reads. Each population was sequenced two times; data from both runs were combined for some analyses as described below. Combining reads from two independent sequencing runs likely alleviate the effects of possible bias introduced from running all replicates for each population in the same lane. Moreover, preliminary comparisons between analyses resulting from only the first run and analyses from the combined data did not yield substantive differences in our results (data not shown).

TE Analysis

We used two methods to characterize the response of TEs to selective pressure in our replicate populations. Scripts used to run both sets of software can be found at <https://github.com/k8hertweck/flyPopGenomics>. First, we assessed total number of TE insertions in each of our pooled populations using PopoolationTE (Kofler et al. 2012). The goal of this approach was to characterize the overall TE load (e.g., total number of insertions) in each population. Given that this approach is especially sensitive to varying depth of coverage, we subsampled each population to 20 million reads prior to analysis. We combined annotations from the *Drosophila* 5.51 (FB2013_03) release with canonical transposon consensus sequence set 9.44 (<https://github.com/cbergman/transposons>) to screen our data for TE insertions. We used PopoolationTE to identify insertions with a minimum read count of 5 and minimum map quality of 20. TE insertions for which there was evidence on both the forward and reverse flanking regions were tallied for each population independently. Paired *t*-tests were used to test whether populations had significantly different numbers of TE insertions from this subsampled (e.g., coverage-normalized) analysis.

Second, we used T-lex2 (Fiston-Lavier et al. 2010) to estimate frequency of TE insertions in our experimental stocks relative to a known set of TE reference sequences. The goal of this approach was to obtain high-confidence assessments of changes in TE insertion frequency for a known subset of TEs while avoiding overlap with SNP and SV analyses (which masked these same TEs). Our reference TE sequences included >5,000 TEs from the *Drosophila* r5.53 genome which were manually curated and reannotated using T-lex2 (annotations available from <http://petrov.stanford.edu/cgi-bin/Tlex.html>). T-lex2 filters all annotated TEs with repetitive sequences in the flanking regions, so our analysis included 2,947 TEs for which we estimated insertion frequencies. T-lex2 analysis was conducted using the Duke Compute Cluster at Duke University using only the first replicate of Illumina sequencing results, which provided sufficient depth of coverage to assess these known insertions and avoided

computational problems with the mapping parameters T-lex2 implements.

T-lex2 returns two types of read counts from which to estimate insertion frequency. First, the absence of insertion is estimated from the number of reads spanning the junction of flanking regions of an annotated TE (read_number_absence in the Tresults file). Second, the presence of an insertion is supported by reads which span each flanking region and the respective neighboring region of the TE (left_read_number_presence and right_read_number_presence). We assume read counts represent individual samples of the population. Given that presence detection had two sites to which reads may map, the counts from both sites were averaged for downstream analyses. Two TE insertions (Fbti0060253 and Fbti0062820) possessed very high numbers of reads mapped relative to other insertions, suggesting repetitive sequences in the flanking regions (Fiston-Lavier et al. 2010), so these insertions were discarded as false positives. Heterozygosity, F_{ST} , and CMH were calculated for the 177 variable insertions using the same procedures as for SNPs (see below).

SNP Analysis

Read Mapping and Preprocessing. We first trimmed the reads to remove low-quality bases using a script provided in the PoPopulation software package (Kofler, Orozco-terWengel, et al. 2011). We then mapped reads with BWA (version 0.7.6) (Li and Durbin 2009) against the *D. melanogaster* reference genome (release 5.51) with the following mapping parameters: -n 0.01 (error rate), -o 2 (gap opening), -d 12 and -e 12 (gap length), and -l 150 to effectively disable the seed option. We then converted the resulting alignment files to SAM format using the BWA sampe command. We filtered the SAM files for reads mapped in proper pairs with a minimum mapping quality of 20 and converted them to the BAM format using SAMtools (Li et al. 2009). The rmdup command in SAMtools was then used to remove potential PCR duplicates. The two BAM files from each population's two sequencing runs were merged using BAMtools to maximize coverage (Barnett et al. 2011). These merged BAM files were then all combined in the mpileup format once again using SAMtools. Using PoPopulation2 (Kofler, Pandey, et al. 2011), the resulting mpileup was converted to a "synchronized" files, which is a format that allele counts for all bases in the reference genome and for all populations being analyzed. We used RepeatMasker 4.0.3 (<http://www.repeatmasker.org>) to create a gff file to mask low complexity regions of the *D. melanogaster* genome version 5.51. All SVs (identified with Delly, described below) and TEs (annotations used in T-lex2, described above) were masked prior to SNP calling.

Heterozygosity

We calculated and plotted heterozygosity across the five major chromosome arms to see if we could find any evidence of selective sweeps and to determine if there was convergence in overall patterns of variation. To do this, SNPs were first called across all 30 populations used in this study from our synchronized file. SNPs where discarded if coverage in any of the

populations was less than 20X or greater than 500X. We also required a minimum minor allele frequency of 2% across all eight populations. Based on these parameters, ~1.01 million SNPs were identified across the major chromosome arms. The average SNP coverage at each across our 30 populations ranged from 28X to 108X, with all but two of our populations (CO₅ and B₃ at 28X and 31X, respectively) and having coverage greater than 30X (Table S3 for more detailed coverage information). A SNP table with major and minor allele counts for each SNP in each population was then generated. Using these counts, heterozygosities were calculated and plotted over 100 kb nonoverlapping windows. We also performed t-tests comparing mean genome-wide heterozygosities between different groups of populations.

We also took steps to identify potential bias in our SNP calling procedure given that average coverage varied across the 30 populations. In particular, the CO₅ and B₃ sequence data have much lower average coverage than the other populations. To test for possible bias, we repeated our SNP calling procedure with the added exception that a minimum coverage of 20 was relaxed to 10 for the CO₅ and B₃ populations; it was unchanged for the remaining 28 populations. This ultimately resulted in an additional ~300,000 SNPs being identified, suggesting that the inclusion of these low-coverage populations combined with our minimum coverage requirement of 20 is indeed a source of bias. However, as allele frequency estimates suffer at low coverages, we chose to be more conservative and maintain this requirement for SNPs included in our analyses.

F_{ST} Estimates. F_{ST} estimates for replicate populations were obtained using the formula: $F_{ST} = (H_T - H_S)/H_T$ where H_T is heterozygosity based on total population allele frequencies, and H_S is the average subpopulation heterozygosity in each of the replicate populations (Hedrick, 2009). F_{ST} estimates were made at every polymorphic site in the data set for a given set of replicate populations. This was done to quantify the level of similarity between replicates of our six sets of populations. The mean genome wide F_{ST} for the six sets of populations are reported in [supplementary table S7, Supplementary Material](#) online.

SNP Differentiation. PoPopulation2 was used to obtain measures of SNP differentiation between newly derived and long-standing populations within and between treatments (Kofler, Pandey, et al. 2011). More specifically, we used this software package to perform CMH tests of differentiation to compare SNP frequencies between our various groups of replicate populations. The specific comparisons performed were as follows: ACO versus AO, ACO versus B, AO versus BO, ACO versus CO, AO versus nCO, B versus BO, B versus CO, BO versus nCO, CO versus nCO, A versus C (all A-types versus all C-types), A versus B, and B versus C. For each comparison, relevant populations were paired by their replicate number (e.g., ACO₁ with AO₁, ACO₂ with AO₂, etc.), which reflects patterns ancestry in the case of all but the five B populations.

The CMH test was performed at each site polymorphic across our 30 populations. Results for all positions not found in our SNP table were then discarded; thus all positions failing

to meet our SNP calling criteria as stated above were removed. To generate null distributions for *P*-values generated by each comparisons (i.e., distributions of these *P*-values associated with a null expectation of genetic drift rather than selection), we used a permutation approach. For a given comparison, the relevant populations were randomly assigned to one of the two groups and the CMH test was then performed at each polymorphic position in the shuffled data set. This was done a 1,000 times and the smallest *P*-value was recorded. The quantile function in *R* was then used to define thresholds that define the genome-wide false-positive rate, per site, at 5%. This was done for each the 12 comparisons we performed.

To give a concrete example of the permutation procedure, we will focus on the ACO_{1–5} versus CO_{1–5} comparison. For a given permutation, 5 of the 10 populations were randomly assigned to group X and the remaining to group Y. A permuted data set could look something like group X = CO₁, ACO₂, ACO₃, CO₃, CO₄ and group Y = CO₅, ACO₁, ACO₄, ACO₅, CO₂. When the CMH test is run, the populations would be paired based on their order in groups X and Y (eg. CO₁-CO₅, ACO₂-ACO₁, ACO₃-ACO₄, etc.). Given that the pairings matter and we have 10 populations, this gives 3,628,800 possible permutations. For each of the shuffled data sets, the CMH test was performed between groups X and Y at each polymorphic position and the smallest *P*-value generated across the entire genome was recorded. Once we had a list of the smallest *P*-values generated across 1,000 permutations, we used the quantile function in *R* to define thresholds that give a genome-wide false-positive rate, per site, of 5%. That is, our test gives us a Type I error rate of 0.05 for each and every site considered significantly differentiated. Note that this test is not designed to minimize Type II errors across the genome, making it undoubtedly statistically conservative.

Analysis of Variance. Our experimental system has the following distal-selection treatment comparisons: A-B, A-C, and B-C, each of these involving 10 populations versus 10 populations. The A-B and B-C comparisons actually share a very long evolutionary history. The whole B branch and the branch from root to common ancestor of A and C are commonly included in both of the A-B and B-C comparisons. In order to remove the nonindependence from shared branches, we need to define phylogenetically independent contrasts (PICs) A-C and AC-B, so that there are no shared branches between the contrasts (supplementary fig. S8, Supplementary Material online). PICs are calculated by the difference in allele frequency for the leaf values, and the difference in the branch-length-weighted allele frequency for the inner nodes. Branch lengths are equal to the number of generations spanning each branch. Then the PICs are standardized by dividing each PIC by the square roots of sums of branch lengths. Sums of branch lengths used are shown in supplementary fig. S8, Supplementary Material online.

Provided we standardize the contrasts by the branch lengths, we can assume that all contrasts are independent samples coming from the same distribution. The null model assumes that the per-generation difference has the same

mean (mean of 0) and variance across the branches. This means that the allele frequency will move in the Brownian motion along the branches, but there will be no systematic direction detectable in the allele frequency change among the branches. We used two-way ANOVA to test against this null model. Two-way ANOVA was done with standardized PIC as the dependent variable and selection treatment and age as the two factors (PIC ~ treatment + age). $\alpha = 0.05$ was Bonferroni corrected with the number of SNP loci tested (0.05/1,142,812 = 4.37e–08).

LD Analysis. We calculated LD using LDx, a method which uses an approximate maximum likelihood approach to estimate LD (r^2) from pooled resequencing data (Feder et al. 2012), using HPC resources provided by the Texas Advanced Computing Center (TACC) at the University of Texas at Austin. To view the complete scripts used for this analysis; see <https://github.com/k8hertweck/flyPopGenomics>. To prepare data for analysis in LDx, we called SNPs on each merged bam file using mpileup and filtered the output using default options (samtools mpileup -ulf dmel.RELEASE5 *.bam | bcftools view -v snps | vcfutils.pl varFilter > *.flt.vcf). We then ran LDx using default options, except for adjusting the insert size to 200 and minimum read depth to 20 (perl LDx.pl -I 20 -h 100 -s 200 -q 20 -a 0.1 -i 11 *.sam *.flt.vcf > *.flt.out). Because LDx compares SNPs within read pairs, we only report LD calculations for distances of 200 bp or less. Distances less than 11 bp were also discarded because of small sample sizes.

To evaluate differences in patterns of LD decay between our populations, we fit the data to a biexponential model in *R* using the “SSbiexp” function and the “nlme” package (Pinheiro et al. 2015). Data from each chromosome (2, 3, and X) was handled independently. We chose to use a biexponential model after evaluating quadratic, cubic, and quartic models. To do this, we split the data into a training set (80% of the data) and a test set (remaining 20%). We then calculated the mean squared error (MSE) for each model and found that the biexponential model had the lowest MSE.

We initially partitioned variation in LD into the fixed effects of selection regime and whether or not the populations were long standing (ACO, B, or CO) or newly derived (AO, BO, nCO) and the random effects of population. However, we found that whether or not the populations were long standing or newly derived did not have an effect on parameter estimates. So this was dropped from the model. The random effects over populations are due to both sampling and genetically based differences that arise due to genetic drift. We tested models with population variation in subsets of parameters and with a constant within-population variation. The model chosen had the lowest Akaike and Bayesian information criterion (Pinheiro and Bates 2000).

For each chromosome, we constructed confidence intervals based on model predictions with a coverage level that applied to all observed points. For each population we had maximum likelihood parameters estimates and their covariance matrix estimates, which were assumed to have normal distributions. From these distributions, we drew samples of

the parameter vectors, $\tilde{\mu}_i$, ($i = 1, \dots, m$). For each sampled parameter vector, we made predictions $f(\tilde{\mu}_i, t)$, for all values of t . From these m predictions, we generated order statistics, $f^j(\tilde{\mu}_i, t)$, where $f^1(\tilde{\mu}_i, t)$ is the smallest predicted value at t and $f^m(\tilde{\mu}_i, t)$ is the largest. If there are k -points that we want to include in a simultaneous interval, then from the Bonferroni inequality (Miller 1966) the confidence level is elevated to $1 - 0.05/k$. From the order statistics, we then used $f^l(\tilde{\mu}_i, t)$ as the lower confidence limit and $f^u(\tilde{\mu}_i, t)$ as the upper confidence limit where, $l = \text{round}(m0.05/(2k))$ and $u = (m + 1 - l)$. For our purposes, $m = 10,000$ and $k = 38$. We have LD estimates for distances of 10–200 bp, but here we only used distances that were multiples of 5.

SV Analysis of Deletions, Tandem Duplications, and Inversions

Only the data from the first round of sequencing was used for our SV analysis. We called deletions and tandem duplications using a modified version of DELLY (Rausch et al. 2012) that is able to collect more than 1,000 split reads. Discordant read pairs were identified, and then the breakpoints were refined using split reads. We removed false split reads by making sure there is no microhomology across the breakpoint. We required that there be at least five-split reads supporting the SV across all populations. We filtered the SVs to have a size range of $150 \text{ bp} < \text{SV} < 10,000 \text{ bp}$. To infer the allele frequency we wrote a script that counted split reads versus correctly mapped reads. We counted all split reads with $\text{MAPQ} \geq 20$ that match across the two breakpoints for ≥ 6 bps on both sides as supportive of the SV allele. We counted all properly aligned reads with $\text{MAPQ} \geq 20$ that also had ≥ 6 bps matches on both sides of the single breakpoint it overlapped as reference type allele. Then counts were used as proxy for samples of alleles. Due to the uncertainty in frequency estimation when the total counts were low, we required that all allele frequency inference was done on variants that had at least 25 total reads within each treatment \times replicate population. Heterozygosity, F_{ST} , and CMH followed the same procedures described above for SNPs.

Large chromosomal inversions have been assessed in numerous natural *Drosophila* populations and potentially have adaptive significance (e.g., Prevosti et al. 1988). There are currently no reliable methods to assess inversion polymorphisms from next generation sequence data alone for inversions with sizes that are large enough to be of interest. There have been attempts to identify potential inversion breakpoints based on discordant reads and F_{ST} calculated from SNPs (Corbett-Detig et al. 2012), but the method is inappropriate for pooled sequence data. Another attempt to study inversion polymorphisms was to identify SNPs that could be used to impute the inversion type (Kapun et al. 2014), based on sequencing and SNP calling for populations with known inversions and karyotyping of new populations. From this effort, there have been inversion specific marker alleles identified for seven inversions: In(2L)t, In(2R)Ns, In(3L)P, In(3R)Mo, In(3R)C, In(3R)Payne, and In(3R)K. We used these marker alleles to check for these seven inversions in our dataset.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The raw fastq files used in this project are available through NCBI SRA (BioProject: PRJNA285429). All other files (SNP table, annotation files, P-value tables, etc.) are available through Dryad (doi:10.5061/dryad.4nc28). This work was supported by Joint School of Nanoscience & Nanoengineering North Carolina A&T State University and UNC Greensboro, which supplied funds to generate the sequence data used in this project. This work was also supported by National Science Foundation through the National Evolutionary Synthesis Center (NESCent), the Department of Education, and the University of California Multicampus Research Programs and Initiatives (MRPI) under grant numbers NSF #EF-0905606, P200A20144, MRPI 142389, respectively.

References

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol.* 23(1):26–32.

Baldwin-Brown JG, Long AD, Thornton KR. 2014. The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. *Mol Biol Evol.* 31(4):1040–1055.

Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27:1691–1692.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1274.

Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. 2010. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467:587–590.

Burke MK. 2012. How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proc Roy Soc B* 279:5029–5038.

Burke MK, Liti G, Long AD. 2014. Long standing genetic variation drives repeatable experimental evolution in outcrossing populations of *Saccharomyces cerevisiae*. *Mol Biol Evol.* 32:3228–3239.

Burke MK, Barter TT, Cabral LG, Kezos JN, Phillips MA, Rutledge GA, Phung KH, Chen RH, Nguyen HD, Mueller LD, Rose MR. 2016. Rapid convergence and divergence of life-history in experimentally evolved *Drosophila Melanogaster*. *Evolution* 70:2085–2098.

Corbett-Detig RB, Cardeno C, Langley CH. 2012. Sequence-based detection and breakpoint assembly of polymorphic inversions. *Genetics* 192:131–137.

Elias M, Tawfik DS. 2012. Divergence and convergence in enzyme evolution: parallel evolution of paraoxonases from quorum-quenching lactonases. *J Biol Chem.* 287(1):11–20.

Feder AF, Petrov DA, Bergland AO. 2012. LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. *PLoS ONE* 7. doi:10.1371/journal.pone.0048588.

Fiston-Lavier AS, Carrigan M, Petrov DA, Gonzalez J. 2010. T-lex: a program for fast and accurate assessment of transposable element presence using next-generation sequencing data. *Nucleic Acids Res.* 39:e36–e36.

Fong SS, Joyce AR, Palsson BØ. 2005. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* 15(10):1365–1372.

Fragata I, Simoes P, Lopez-Cunha M, Lima M, Kellen B, Barbaro M, Santos J, Rose MR, Santos M, Matos M. 2013. Laboratory selection quickly

erases historical differences. *PLoS ONE* 9:e96227. 10.1371/journal.pone.0096227

Franssen SU, Volte N, Tobler R, Schlötterer C. 2015. Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental *Drosophila melanogaster* populations. *Mol Biol Evol*. 32:495–509.

Hedrick PW. 2009. Genetics of populations. Massachusetts: Jones & Bartlett Learning Press.

Ives PT. 1970. Further studies of the South Amherst population of *Drosophila melanogaster*. *Evolution* 24:507–518.

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

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE* 6:e15925.

Kofler R, Pandey RV, Schlötterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27:3435–3436.

Kofler R, Betancourt AJ, Schlötterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet* 8(1):e1002487.

Kofler R, Schlötterer C. 2014. A guide for the design of evolve and resequencing studies. *Mol Biol Evol*. 31:474–483.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am Nat*. 138:1315–1341.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.

Maddamsetti R, Lenski RE, Barrick JE. 2015. Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* 200:619–631.

Matos M, Simoes P, Duarte A, Rego C, Avelar T, Rose MR. 2004. Convergence to novel environment: comparative method versus experimental evolution. *Evolution* 58(7):1503–1510.

Matos M, Avelar T. 2001. Adaptation to the laboratory: comments on Srgo and Partride. *Am Nat*. 158:655–656.

Matos M, Avelar T, Rose MR. 2002. Variation in the rate of convergent evolution: adaptation to the laboratory environment in *Drosophila subobscura*. *J Evol Biol*. 15:673–683.

Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favorable gene. *Genet Res*. 23:23–35.

Miller RG. 1966. Simultaneous statistical inference. New York: McGraw-Hill.

Mueller LD, Joshi A, Santos M, Rose MR. 2013. Effective population size and evolutionary dynamics in outbred laboratory populations of *Drosophila*. *J Genet*. 92:349–361.

Oleksyk TK, Smith MW, O'Brien SJ. 2010. Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond*. 365:185–205.

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. *Mol Ecol*. 21:4931–4941.

Pinheiro JC, Bates DM. 2000. Mixed-effects models in S and S-PLUS. New York: Springer.

Pinheiro JC, Bates DM, DebRoy S, Sarkar D, R Core Team. 2015. nlme: linear and nonlinear mixed effects models. R package version 3.1-122. Available from: <http://CRAN.R-project.org/package=nlme>.

Platt JR. 1964. Strong inference. *Science* 146:347–353.

Prevosti A, Ribo G, Serra L, Aguade M, Balaña J, Monclús M, Mestres F. 1988. Colonization of America by *Drosophila subobscura*: experiment in natural populations that supports the adaptive role of chromosomal-inversion polymorphism. *Proc Natl Acad Sci*. 85:5597–5600.

Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28:i333.i9.

Rose MR. 1984. Laboratory evolution of postponed senescence in *Drosophila melanogaster*. *Evolution* 38:1004–1010.

Rose MR, Passananti HB, Matos M. 2004. Methuselah flies: a case study in the evolution of aging. Singapore: World Scientific Publishing.

Service PM, Hutchinson EW, Rose MR. 1988. Multiple genetic mechanisms for the evolution of senescence in *Drosophila melanogaster*. *Evolution* 42:708–716.

Schlötterer C, Kofler R, Versace E, Tobler R, Franseen SU. 2015. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity* 114:331–440.

Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet*. 14:751–764.

Teotónio H, Rose MR. 2000. Variation in the reversibility of evolution. *Nature* 408:463–466.

Teotónio H, Rose MR. 2001. Perspective: reverse evolution. *Evolution* 55:653–660.

Teotónio H, Chelo IM, Bradiá M, Rose MR, Long AD. 2009. Experimental evolution reveals natural selection on standing genetic variation. *Nat Genet*. 41:251–257.

Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* 335:457–461.

Tobler R, Franssen SU, Kofler R, Orozco-ter Wengel P, Nolte V, Hermissen J, Schlötterer C. 2014. Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Mol Biol Evol*. 31:364–375.

Turner TL, Steward AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genet*. 7:e10001336.

Wright S. 1978. Evolution and genetics of populations. Volume 4: Variability within and among natural populations. Chicago, London: The University of Chicago Press.

Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE. 2006. Tests of parallel molecular evolution with *Escherichia coli*. *Proc Natl Acad Sci*. 103:9107–9112.

Zichner T, Garfield DA, Rausch T, Stutz AM, Cannavo E, Braun M, Furlong EE, Korbel JO. 2013. Impact of genomic structural variation in *Drosophila melanogaster* based on population scale sequencing. *Genome Res*. 23:568–579.