Direct observation of adaptive tracking on ecological timescales in *Drosophila*

Key words: adaptive tracking, rapid adaptation, repeated evolution, polygenic adaptation, genomic architecture, balancing selection

One sentence summary: Rapid environmental change drives continuous phenotypic and polygenic adaptation demonstrating the temporal dynamism of adaptation.

Authors: Seth M. Rudman*, Sharon I. Greenblum*, Subhash Rajpurohit*, Nicolas J. Betancourt, Jinjoo Hanna, Susanne Tilk, Tuya Yokoyama, Dmitri A. Petrov, Paul Schmidt

Abstract

Direct observation of evolution in response to natural environmental change can resolve fundamental questions about adaptation including its pace, temporal dynamics, and underlying phenotypic and genomic architecture. We tracked evolution of fitness-associated phenotypes and allele frequencies genome-wide in ten replicate field populations of *Drosophila melanogaster* over ten generations from summer to late fall. Adaptation was evident over each sampling interval (1-4 generations) with exceptionally rapid phenotypic adaptation and large allele frequency shifts at many independent loci. The direction and basis of the adaptive response shifted repeatedly over time, consistent with the action of strong and rapidly fluctuating selection. Overall, we find clear phenotypic and genomic evidence of adaptive tracking occurring synchronously with environmental change, demonstrating the temporally dynamic nature of adaptation.

*** equal contribution**

Continuous adaptation in response to rapidly changing environmental conditions, termed adaptive tracking, could be a crucial mechanism by which populations respond to environmental change. Adaptive tracking has historically received little study due to the impression that adaptive evolutionary change is too slow to track complex and rapidly changing selection pressures in the wild [1]. Moreover, theory suggests that variable and complex selective pressures should in general lead to the evolution of phenotypic plasticity or bet-hedging [2,3]. Yet, evidence of adaptation on ecological timescales from multiple longitudinal field studies and experiments demonstrates that adaptation can indeed occur very rapidly at individual traits or loci in response to strong environmental perturbations [4–10]. Whether this translates into populations undergoing adaptive tracking in response to multifarious ecological changes, when theory predicts that pleiotropy should constrain natural selection and prevent adaptive tracking [11,12], is unknown. If adaptive tracking does indeed occur in such situations it would have broad implications for our understanding of the limits and pace of polygenic adaptation [13], the prevalence of fluctuating selection [14] and its role in the maintenance of genetic variation [15], and the importance of rapid adaptation in ecological outcomes [16].

To identify adaptive tracking it is necessary to directly measure phenotypic and genotypic evolution across replicated field populations in response to ongoing natural environmental change. Ideally an experimental system would provide: 1) the means for highly accurate measurements of even subtle heritable shifts in key independent fitness-related phenotypes and loci under selection, 2) the ability to assay multiple replicate populations exhibiting some degree of ecological and environmental realism to detect parallel genetic and phenotypic changes indicative of adaptation [17], and 3) high resolution temporal sampling to quantify rapid fluctuations in the magnitude and direction of selection as environmental changes occur.

Here, we employ such an experimental system using field mesocosms to measure the extent, pace, repeatability, and genomic basis of adaptive tracking using *Drosophila melanogaster* in the naturally fluctuating, temperate environment of a single growing season in Pennsylvania, USA [10,18,19] (Fig. 1). The design precluded migration and allowed populations to expand to a large adult census size (on the order of 100,000 adults in each replicate at the maximum population size). To initiate the experiment, an outbred baseline population of *D. melanogaster* was derived from a set of 80 inbred strains originally collected in the spring from Pennsylvania. Ten replicate cages were founded on July 15th, 2014 with ~1,000 individuals from the baseline population. All populations were tracked until the first hard frost on November 7th, 2014 and assayed at monthly intervals. Specifically, at five timepoints we measured the evolution of six complex, fitness-associated phenotypes, focusing on a set associated with either reproductive output or stress tolerance (Fig. 1). We used repeated common garden rearing to distinguish evolution from phenotypic plasticity and measured all phenotypes in the F3. We also tracked changes in allele frequencies genome-wide in each replicate using pooled sequencing at five timepoints, employing haplotype-based allele frequency estimation [20] in order to generate highly accurate allele frequency trajectories.

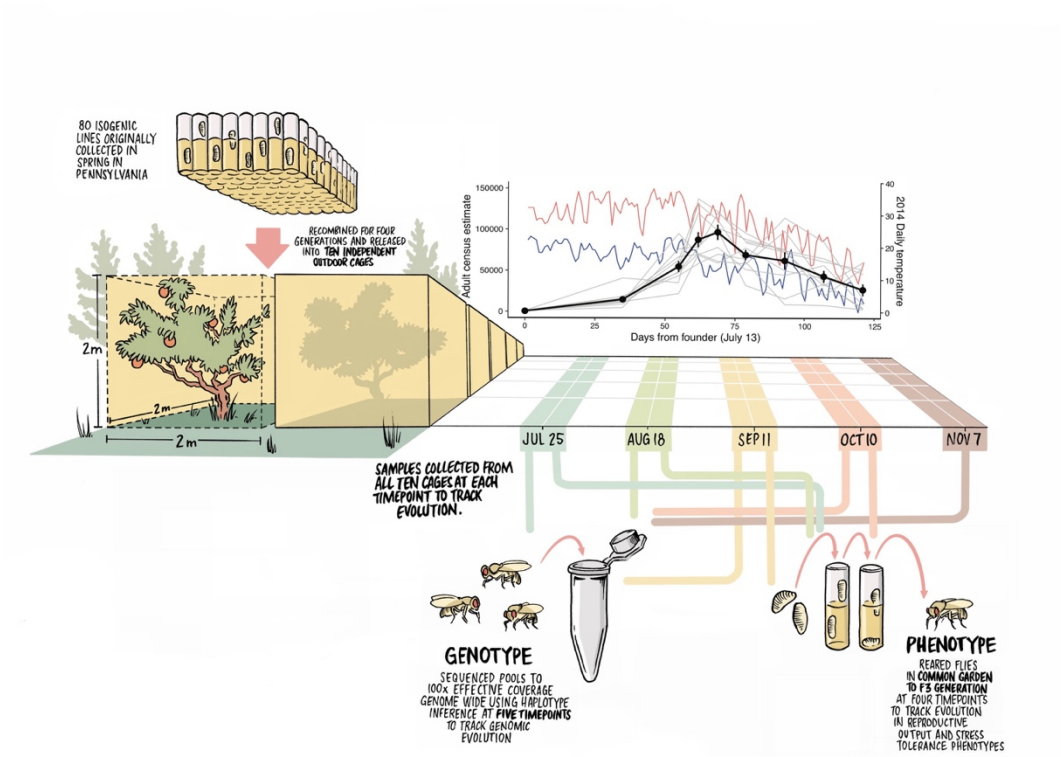

Figure 1: The experiment was designed to reflect ecological and evolutionary realism while testing for adaptation using replicate populations. 80 inbred lines originally collected in spring from an orchard in Pennsylvania were recombined and expanded for four generations into a genetically diverse outbred population in the laboratory. 500 males and 500 females from this outbred population were used to found 10 independent outdoor cages (2m x 2m x 2m). We measured daily minimum and maximum temperatures (blue and red, respectively) and estimated adult population size of each replicate over four months of seasonal change. To study adaptation, we tracked phenotypic evolution by collecting eggs from each replicate, rearing in common garden laboratory conditions for three generations, and then measuring six fitness-associated phenotypes. We conducted this procedure on the founder population and at four subsequent time points to measure phenotypic evolution over time. To study adaptation at the genomic level we sequenced pools of 100 females from each cage to >100x effective coverage at five time points using haplotype inference [20] and assessed changes in allele frequencies.

Population dynamics were largely consistent among the replicates with population size increasing sharply during summer, peaking in early fall, and then declining steadily as minimum daily temperatures declined in late fall (Fig. 1). These population dynamics mimic the patterns observed in *D. melanogaster* populations [21] and many other multivoltine organisms inhabiting temperate natural environments, with summer exponential growth, high densities in late summer to early fall, and late fall population declines. Similarity in the ecological conditions among replicate populations, including abiotic factors and population dynamics, suggests similar selective landscapes may have driven parallel evolution across replicates.

Phenotypic evolution was rapid and parallel, but temporal patterns varied across traits. In order to measure phenotypic evolution we sampled individuals from the founding population and ~2500 eggs from each cage at the first four time points (July 25, August 18, September 11, October 10), reared them in common garden laboratory condition for three generations, and assayed phenotypes in the F3 progeny (Fig. 1). For all six phenotypes, which are known to be polygenic and associated with fitness [22], we observed substantial trait evolution with an average of 23% change in the trait value for each cage across all phenotypes over each time interval.

All six phenotypes also showed evidence of parallel evolution, indicative of adaptation, over time. Four of six phenotypes evolved rapidly, repeatedly, and in a consistent direction across the duration of the experiment (Fecundity: $F_{3,27}$=43.75, p<0.0001; Egg size: $F_{3,27}$=11.5, p<0.0001; Starvation: $F_{4,36}$=129.05, p<0.0001; Chill coma recovery: $F_{4,36}$=197.75, p<0.0001) (Fig. 2). The magnitude of change was often substantial: for example, the average increase in fecundity was 61% over each monthly sampling interval across replicates, representing 1-4 overlapping generations. Desiccation tolerance and development rate also evolved rapidly and in parallel ($F_{4,36}$ =86.66, p<0.0001 Fig. 2C; $F_{4,36}$=98.70, p<0.0001, Fig. 2F), but the direction of evolution varied over time. Fluctuations in the direction of selection on the measured phenotypes, or some unmeasured but correlated phenotypes, had considerable effects on phenotypic trajectories; for desiccation tolerance the amount of evolution measured over the whole experiment (founder to October 10th) was less than what was observed over the first interval (founder to July 25th).

The pace of parallel trait evolution observed over the short timescales examined in this study was unusually fast. As expected, we observed rapid parallel evolution when outbred laboratory populations were introduced into the field enclosures (Founder → T_1). However, we also observed evidence of rapid adaptation between intervals in the enclosures for all six phenotypes, with some showing reversals in the direction of evolution across intervals (Fig. 2 C&F). The rate of phenotypic adaptation, measured as the mean change across replicate populations in the number of standard deviations per generation (Haldanes [23,24]), was calculated for each phenotype over each interval and across the whole experiment (Fig. 2G). The rate of adaptation over the whole experiment ranged from moderate to extremely fast for different traits (0 - 0.8 Haldanes) [25]. However, when calculated over each sampling interval, the rate of adaptation was often comparable or faster than the fastest known pace of phenotypic change measured in any prior field study or experiment (Fig. 2G).

The pace, magnitude, and parallelism of the phenotypic evolution we observed is notable for three reasons. First, the evolutionary rates were calculated based on the phenotypic shifts of the F3 progeny in common garden conditions, thus excluding phenotypic plasticity as the driver of change. Second, because we focus only on the parallel phenotypic shifts across the cages, we estimate the rate of putatively adaptive phenotypic change. Third, these patterns of rapid adaptation were observed for multiple fitness-associated phenotypes, each with a complex and likely distinct genetic architecture [26]. Overall, our results show that strong and temporally variable natural selection can consistently drive rapid and polygenic adaptation of multiple fitness associated phenotypes on the same timescale as the environmental change.

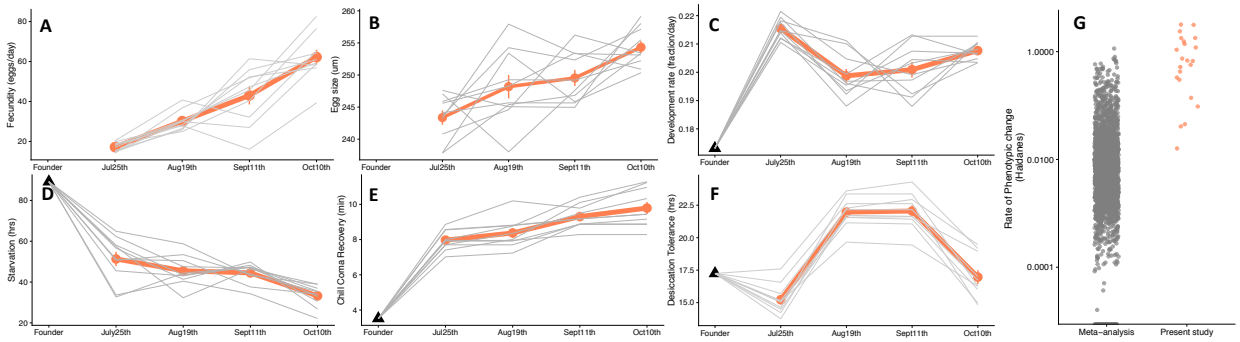

Figure 2: Parallel evolution of stress tolerance traits, reproductive output traits, and comparison of the rate of adaptation. Trajectories of phenotypic evolution for reproductive-associated (A, B, C) and stress resistance traits (D, E, F) as measured after three generations of common garden rearing. Panel A: mean fecundity as number of eggs/female/day, Panel B: mean egg size, Panel C: development rate as the fraction of development to pupation completed in one day (1/(total hours/24)). Panel D: starvation tolerance as time to death by starvation, Panel E: recovery time following chill coma, Panel F: desiccation tolerance as the time to death from desiccation. Black points are the mean phenotypes of the founding population, grey lines represent mean phenotypic trajectories of individual populations, and red lines are the mean of all cage means. Panel G: a comparison of the rates of adaptation from this experiment over individual intervals (red) to rates of phenotypic change from a prior meta-analysis (grey) [25].

To investigate the genomic architecture underlying the observed rapid phenotypic adaptation, we performed whole-genome pooled sequencing of ~100 randomly selected individuals from the baseline population and each replicate population at five timepoints across the experiment (Fig. 1). Allele frequencies at 1.9 M biallelic sites were inferred for each sample via haplotype inference using HAF-pipe [20] (see Materials and Methods) at accuracy levels consistent with an 'effective coverage' of >100x (Supplementary Materials, Fig. S1). This high-resolution dataset yielded strong evidence for rapid genome-wide evolution. Specifically, we observed that the genome-wide estimates of F_{ST} between the founder population and all five monthly timepoints (mean $3.0 \pm 0.2 \times 10^{-3}$ standard error) exceeded expected margins of error based on technical and biological replicates ($2.6 \pm 0.24 \times 10^{-4}$ and $1.8 \pm 0.048 \times 10^{-3}$ respectively, t-test p-values $< 2\times10^{-8}$, Fig. 3A). Furthermore, divergence from the founder population changed significantly over time both genome-wide (Kruskal-Wallace p-value for difference in means across timepoints: $p < 2.3\times10^{-5}$) and for individual chromosomes ($p < 0.006$, Fig. S2B). Given the large population sizes (up to 10^5) it is unlikely that such substantial evolutionary change can be attributed solely to random genetic drift.

Further examination of the magnitude and direction of evolution across the 10 replicate cages showed substantial genomic adaptation, as defined by parallel, and thus deterministic, allele frequency shifts across replicate cages. To test for parallel shifts, we used a leave-one-out cross-validation approach. For each monthly time interval ($T_i \rightarrow T_{i+1}$; i = 1,2,3,4), we used a generalized linear model (GLM) to identify sets of SNPs whose frequency shifted significantly across the 9 training cages, and then tested whether shifts at those SNPs in the 10th left-out cage exceeded shifts at randomly-selected matched control sites. Using this test, we found

widespread parallel genomic adaptation for the first 3 sampling intervals (in 29 out of 30 leave-one-out tests) (Fig.3C). The pattern of parallelism was muted and evolution was more idiosyncratic in $T_4 \rightarrow T_5$. We also repeated the procedure for SNPs that shifted across the whole experiment ($T_1 \rightarrow T_5$), and found a similarly strong signal of parallel adaptation (10 out of 10 times). The magnitude of allele frequency shifts in each interval (2-8%) and over the whole experiment (2-5%) corresponds to very strong effective selection strength at the most parallel sites of ~10-50% per monthly interval (1-3 generations) (Materials and Methods). This pattern was largely repeated when analyzing sites from each chromosome individually (Fig S3). The pronounced parallel shifts in allele frequency across independent populations demonstrate the strong action of natural selection.

Our cross-validation analysis also yielded clear evidence of variation in the magnitude and direction of selection over time, consistent with the observed patterns of phenotypic evolution for some traits (Fig. 2). Specifically, the leave-one-out analysis and the time series genomic data allowed us to examine the full trajectory of alleles detected at any specific time interval (T_{det}). We found that these alleles do often shift significantly more than alleles at control sites (Fig 3, lower panel) at other time intervals; however, the nature of these shifts varied over time. In some left-out cages and at some time intervals, alleles shifted in a direction consistent with their behavior during T_{det} (orange points); however, in other cases the direction flipped, resulting in significant reverse shifts (green points). Reverse shifts were strongest for sites with $T_{det} = T_3 \rightarrow T_4$ (Aug→Sept) during the time when populations expanded most rapidly and reached their maximum. These $T_3 \rightarrow T_4$ parallel sites showed consistent shifts in the *opposite* direction during the preceding interval ($T_2 \rightarrow T_3$, July→Aug) when the populations were still expanding. In many cages, these sites also shifted in the opposite direction during the subsequent ($T_4 \rightarrow T_5$, Oct→Nov) interval when population sizes were declining. These patterns likely reflect the action of rapidly fluctuating selection over the 4 months of the experiment.

A complex and fluctuating selective landscape drives adaptation at multiple timescales, simultaneously, and illustrates that lower-resolution temporal sampling would obscure the pace of adaptive tracking. While sites identified during individual time intervals often showed median shifts of >2% in a single month, the strongest parallel sites detected from lower-resolution sampling (i.e., sampling only at T1 and T5) showed smaller monotonic shifts at each interval (on average, 0.6% per month). Moreover, the magnitude of this discrepancy varied widely over time. The inferred rates of genomic adaptation depended on the temporal resolution, with allele frequency shifts over monthly intervals being greater than those over the whole study. Taken together, these results underscore the dynamic nature of genomic adaptation and the value of high-resolution temporal sampling in revealing the existence of both temporally variable and temporally consistent directional selective forces.

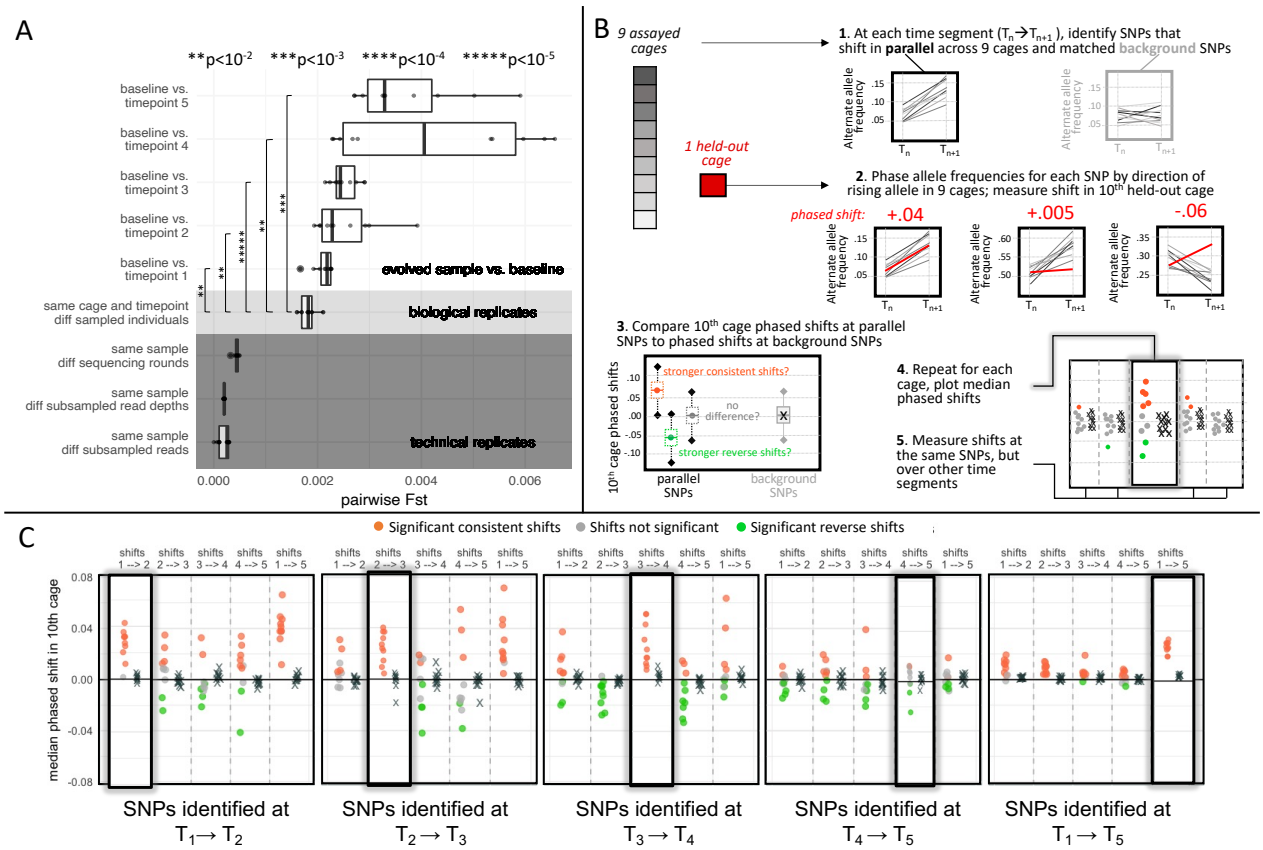

Figure 3: A) Distributions of genome-wide mean pairwise Fst values between technical replicates (light gray; same flies, different reads), biological replicates (dark gray; different flies, same timepoint), and experimental samples from different timepoints compared to baseline (yellow). Note that negligible Fst values between pairs of technical replicates are consistent with extreme precision of HAFs, suggesting that the variance in allele frequency estimates for biological replicates is primarily driven by sampling of different individuals. Asterisks represent the significance of divergence over time compared to biological replicates (t-test). B) Graphical description of the leave-one out 10-fold cross-validation process for significant sites. In each round, significantly parallel sites (FDR <0.05, effect size>2%) at each time segment were identified using 9 of the 10 cages, then the shift at those sites in the 10th left-out cage was measured, after phasing such that positive values represent shifts in the same direction as the 9 assayed cages and negative values represent shifts in the reverse direction. The set of phased shifts at parallel sites was compared to phased shifts at background sites matched for chromosome and initial frequency and assigned to one of three significance bins: consistent (orange) or reverse (green), or no significant difference from background (gray). Shifts at these same sites over other time segments were also measured, phased, and assigned to significance bins. C) The median shift for each set of parallel sites (circles) and background sites (x marks) is plotted for each left-out cage. Each block of 5 panels represents shifts at the same sets of sites, those identified as parallel at the time segment labeled below the block. Shifts measured at that same time segment are highlighted in the panel with a dark shadowed outline.

The number and genomic location of causal loci involved in adaptation is central to understanding the mechanics of the adaptive process [27]. To quantify the genomic architecture

of adaptation we examined the distribution of parallel sites across the genome and developed an algorithm to differentiate putatively independent targets of selection from the sites whose shifts could largely be ascribed to linkage disequilibrium and genetic draft. We first fit allele frequencies from all 10 cages to a GLM and identified significantly parallel sites (Fig. S4) at each time segment (n=4,274) and across the whole experiment (n=5,036), yielding 9,310 significant shifts overall (Fig. 4A, Table 1; Materials and Methods). As expected from the leave-one-out analysis, the sets were largely non-overlapping: the 9,310 detected parallel shifts occurred at 9,000 unique SNPs. Moreover, at each time interval and across the whole experiment, parallel sites were both strongly clustered (empirical $p<0.01$; Fig. S5) and showed significantly higher average linkage values than the matched control sites (paired t-test p-value $< 10^{-16}$; Fig. S6) (Material and Methods), suggesting that the majority of parallel sites were only linked to causal loci rather than being causal themselves.

We next identified the minimum number of independent genetic loci under selection using an algorithm that aggregated the parallel sites into clusters of linked sites (Materials and Methods, Fig. S6). This algorithm clustered 8,214 parallel SNPs detected across all the time segments (~90% of all SNPs at FDR <0.05) into 165 unlinked independent clusters (Fig. 4A, Table S1). These clusters were found on every chromosome and at every time segment. Although inversions can drive patterns of adaptation in Drosophila [28,29], no inversion markers were found among the parallel sites, and only 3 of the 165 clusters were strongly linked to inversions with average $R^2 > 0.1$ (Fig. S7). Combining clusters from all time segments, 61% of all assayed SNPs and 62% of the genome was contained in at least one cluster, highlighting the pervasive impact of short-term adaptive evolution at tens to hundreds of independent selected sites on allele frequencies genome-wide.

The genomic distribution and frequency shifts of these clusters suggested rapid changes in the targets and direction of selection over time. Specifically, 36 of the 90 clusters (40%) identified at a specific monthly time interval did not overlap any clusters identified at other monthly intervals, suggesting that selection at these loci was limited to one month. Among the remaining 54 clusters that did overlap with at least one other cluster from a different time segment, only 27 (50%) contained SNPs that were significantly linked to SNPs in an overlapping cluster. By merging groups of overlapping linked clusters, these 27 clusters formed 9 distinct 'superclusters' (Fig. 4; Fig. S8), representing genomic regions in which allele frequencies shifted significantly in multiple monthly intervals. Strikingly, in 5 of the 6 superclusters involving a cluster identified at $T_3 \rightarrow T_4$ linked to a cluster identified at $T_4 \rightarrow T_5$, 90% of SNPs flipped direction between the two time intervals, and in the 6th cluster >80% flipped direction, together totaling 10,464 SNPs that flipped direction in these six regions between $T_3 \rightarrow T_4$ and $T_4 \rightarrow T_5$. A smaller majority of SNPs (67%) flipped in the supercluster formed by a cluster identified at $T_2 \rightarrow T_3$ linked to a cluster identified at $T_3 \rightarrow T_4$. Finally, in the two superclusters involving sets of linked clusters from 3 different time segments ($T_2 \rightarrow T_3$, $T_3 \rightarrow T_4$, $T_4 \rightarrow T_5$), together covering over 5Mb of chromosome arm 3L, most SNPs (72% and 85%, respectively) flipped direction *twice*. We further confirmed that similar dynamics characterized the full set of putatively causal SNPs by choosing the SNP with the strongest parallelism p-value in each cluster and examining its trajectory (Fig. 4B). While the initial frequencies of these marker SNPs (Fig. S9) and exact shape of their trajectories varied widely, we observed a consistent trend: markers for the clusters identified at an individual monthly time interval often changed little during other months or even moved in the opposite

direction (especially clusters identified at $T_3 \rightarrow T_4$), whereas markers for clusters identified across the whole experiment tended to shift evenly and monotonically over time. The analysis of overlapping clusters and marker SNPs reveals similar patterns to individual SNP based analyses, together supporting an oligogenic and rapid adaptive response to momentary selection pressures that often results in strong and rapidly fluctuating selection.

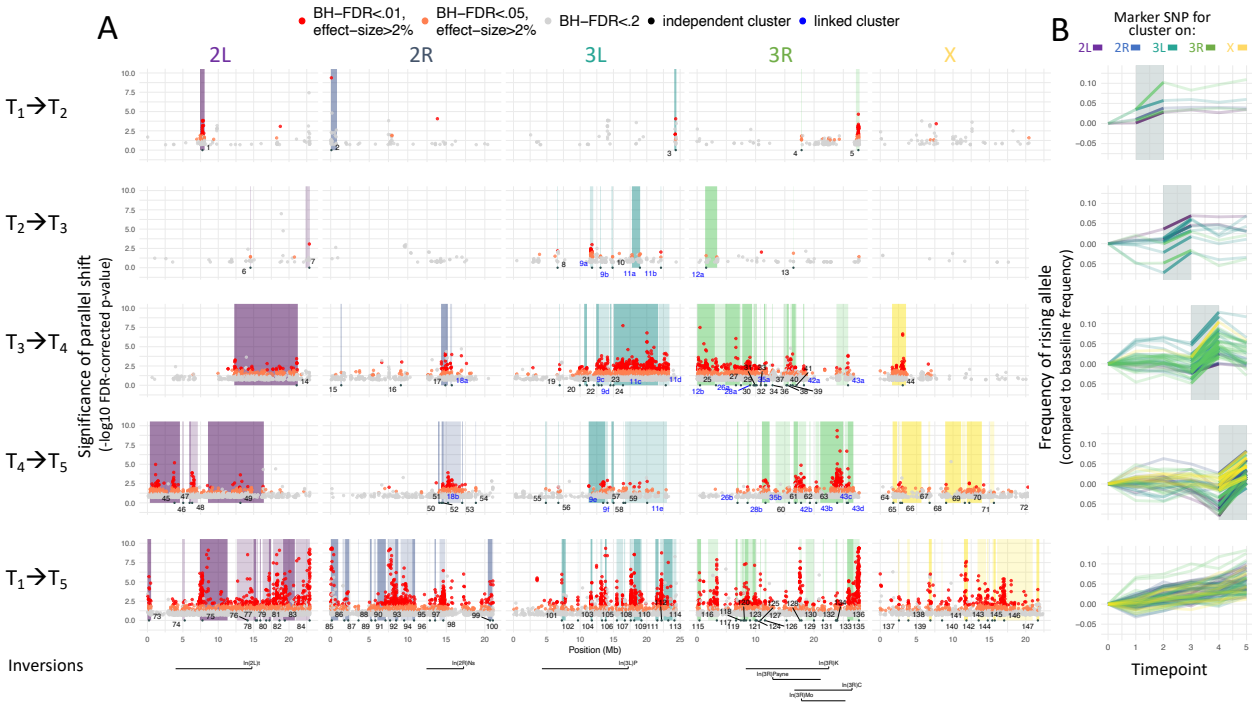

Figure 4: A) Manhattan plot of significant parallel allele frequency shifts over time in 10 replicate cages. Each dot shows the -log10 of the FDR-corrected p-value (y-axis) for the significance of the allele frequency shift at a given SNP position (x-axis) over a given time segment of the experiment (rows). Only SNPs with an FDR <0.2 are shown, and dots are colored according to 3 significance bins (legend). Shaded areas indicate regions of the genome that are likely driven by the same causal site, as defined by a clustering algorithm accounting for SNP linkage. Each clustered genome block is identified by a number marking the position of the top parallel SNP. Clusters from different time segments ('superclusters') that are significantly linked are given the same number, labeled in blue. The position of seven common chromosomal inversions are indicated below. B) Allele frequency trajectories are shown for the top marker SNP from each cluster. Each trajectory is translated to show allele frequency change relative to initial frequency in the baseline population, and phased to show the frequency of the rising allele at the time segment in which the cluster was identified.

The phenotypic and genomic patterns observed in this study are consistent with a form of adaptive tracking in which (i) populations adapt in response to continuous environmental shifts, (ii) parallel evolution is driven by strong selection on multiple phenotypes and on a substantial number (tens to hundreds) of strongly selected genetic variants, (iii) the identity of the phenotypes and variants under selection changes considerably over short timescales, and (iv)

selection operates at multiple timescales, acting in a consistent direction across the whole experiment on some variants and phenotypes, and rapidly fluctuating in direction and magnitude at others (*30*). This fluctuating selection leads to inferred rates of adaptation being slower when measured from the beginning to the end of the experiment when compared to single monthly intervals. Thus, our data suggest that it is fluctuating selection that leads to the general pattern of evolution being faster when measured over shorter periods of time (*13, 31*).

The rapid adaptation in our field cages is consistent with prior observations of seasonal evolution in natural temperate populations of *D. melanogaster* (*21*, *32–34*). However, with additional temporal resolution and replication we detect rapidly fluctuating patterns of adaptation that suggests populations of *D. melanogaster* are continuously and adaptively tracking the environment. These patterns also imply that segregating functional variation is abundant and that much of the segregating variation in fitness is likely due to balancing selection (*35*), including temporally fluctuating selection that maintains genetic variation (*14*, *36*, *37*).

We show that it is possible to observe adaptive tracking in real time, providing a new lens to study the synchronous ecological and evolutionary dynamics of natural populations. While we focus here on *D. melanogaster*, many other taxa exhibit substantial intraspecific genetic variation and experience shifting environments over generational timescales. Determining whether adaptive tracking is a general feature of natural populations and identifying the genetic, demographic, and environmental parameters that most impact adaptive tracking are key areas of future work. Additional experimental work will be needed to define the exact causal loci and phenotypes under selection and re-construct the complex web of genotypic and phenotypic interactions that underlie adaptive tracking in nature. Ultimately, we need to understand the complex interplay among environmental change, population dynamics, standing genetic variation, and trait architecture that allows populations to respond so rapidly across many phenotypic dimensions to ever-shifting and complex environmental challenges.

Methods

Establishment of experimental populations

To examine the pace, magnitude, parallelism, and genomic architecture of adaptation in response to a temporally variable environment we created genetically diverse founder populations that were seeded into each outdoor replicate. The founding population was constructed by combining 10 males and 10 females from each of 80 fully sequenced *Drosophila melanogaster* inbred lines into large cages in May 2014. These inbred lines were derived from wild-caught individuals collected June 1, 2012 from Linvilla Orchards, Media PA USA (*21*). Each line was subsequently inbred for 20 generations by full-sib mating, then 30-50 individuals from each line were pooled for whole genome sequencing. Sequencing and variant calling were performed as described in (*20*), with the addition that genomic DNA from certain lines was resequenced with Illumina HiSeq X to increase coverage to a minimum of 10x for all lines. Mapped and de-duplicated bam files from all original and resequencing runs can be found on SRA under BioProject PRJNA722305. To initiate the baseline population in this experiment, we allowed 4 generations of unmanipulated recombination and population expansion to facilitate

recombination between lines before using 500 males and 500 female flies to found each of 10 field cages.

Field cages are 2m x 2m x 2m mesh enclosures located outdoors (Philadelphia, PA) and they feature a natural insect and microbial community and contain one dwarf peach tree. The only food source and egg-laying substrate was 250ml of Drosophila media ('Spradling cornmeal recipe') contained in 1.5lb aluminum loaf pans that were added every second day for the duration of the experiment (July 13th - November 7th, 2014). Loaf pans of media within experimental cages were protected from rain and direct sun.

Measurement of population size and evolution of fitness associated phenotypes

Census size of adults was estimated in each replicate over the course of the experiment by photographing an equal amount of the surface area (approximately 2.5%) in each cage (12 total census estimates). The number of adult *D. melanogaster* in each photograph for each cage was counted and multiplied by 40 to correct for total surface area and obtain census estimates.

To assess the rate and direction of phenotypic evolution over the course of the experiment we collected ~2500 eggs from each cage, brought them to the laboratory, reared them for an additional 2 generations in a common garden (25˚C, 12L:12D) while maintaining population sizes at ~2500 individuals. Fitness associated phenotypes were measured on density and age-controlled replicates in the F3 generation. Fecundity was measured as the number of eggs laid by a group of five females each day over a three day period with twenty replicate vials for each cage at each time. Egg length was measured using a microscope and image processing software (Schneider et al. 2012) on at least 15 eggs (average of 27) from each cage at each time point. Larval development rate was tracked as the time from when eggs were laid until pupation in three replicate vials from each cage at each time point with 30 eggs in each vial. Starvation tolerance was measured as time to starvation in three replicate vials containing moist cotton (1.5 ml water) and 10 female flies with three replicates for each cage at each time point. Desiccation tolerance was measured as time to death in desiccation chambers containing 10 female flies with three replicates for each cage at each time point. Chill coma recovery was measured as the time it took for flies buried in ice and placed in a 4C incubator for 2h to resume an upright stance at 25C and was measured using groups of 10 female flies for each cage at each time point. We assayed evolution in each of these phenotypes in the founding population (founder assays failed for fecundity and egg size) and at four times during the experiment: day 11 (7/25/14), day 38 (8/19/14), day 61 (9/11/14), and day 90 (10/10/14). Census and phenotypic evolution data have been uploaded to Dryad XXXX.

Calculation of evolutionary rates and statistical analysis of phenotypic data to test for evolutionary parallelism

We calculated evolutionary rates in haldanes by dividing the trait change over each interval by the pooled standard deviation and then by the number of generations elapsed (*23, 24*). We calculated the rate of adaptation as the parallel change across replicates. To do so we took the average trait change across all 10 replicates and calculated a single rate in haldanes. Haldanes were calculated for all six phenotypes for each experimental interval and over the whole experiment. We compared the rates of evolution measured in our experiment to those from a

meta-analysis of evolutionary rates from field populations that focused on contemporary evolution (less than 200 generations) (*25*). The meta-analysis was focused on phenotypic change, which includes both genetic and environmental (plastic) effects, as few prior studies used common garden experiments to measure the rate of evolution.

To test for parallel phenotypic evolution in each of the six phenotypes we carried out separate linear mixed effect models (e.g. lme(phenotype measured ~ time, random=~1|cage/time)) and tested for significance using ANOVA with the nlme package in R.

Genomic sequencing, SNP calls, and bioinformatic analysis

100 female flies from each of the 10 field cages were sampled at 5 monthly time points. Individuals from each sample were pooled and libraries were prepped using a Covaris protocol, then size-selected using an e-gel. Two e-gel bands from each sample were sequenced separately (1 from the 450-500 band and 1 from the 500-550 band) on a HiSeq3000 with 150-bp paired-end reads. Truseq adapter sequences and bases with quality <20 were trimmed with skewer and overlapping forward and reverse reads were merged using PEAR. Resulting reads were mapped to the *Drosophila melanogaster* reference genome v 5.39 with BWA (default parameters). Reads were deduplicated using Picardtools and realigned around indels using GATK v4. Pairs of bam files from the same sample were merged with samtools. Final average per-sample read depth was 7.3x. Haplotype-derived allele frequencies (HAFs) were then calculated via local inference using the 80 genotyped founder strains, as described in (*20*). Haplotype inference was conducted in sliding windows across the genome, using an adaptive window size to reflect the expected length of un-recombined haplotype blocks given the estimated number of generations since population founding. Heterozygous calls in the founder lines were included in the inference calculation, and missing calls were imputed. HAFs from samples were compiled and filtered to contain only sites at which at least one baseline sample and at least one evolved cage sample had a minor allele frequency >1%.

High coverage sequencing

4 biological replicate samples from the baseline population, each a random sample of 100 flies from the same baseline population, were sequenced at high coverage. Baseline library preps were created using a modified Nextera protocol (*38*) and sequenced on a HiSeq4000 with target 100x coverage. Additionally, timepoint-5 evolved samples from 8 of the 10 cages were re-sequenced at high coverage (in addition to separate sequencing at low coverage with the rest of the evolved samples) using a KAPA hyperprep and a HighSeq4000. Processing for both the baseline and high-coverage timepoint-5 samples followed the same workflow. All adapter sequences were trimmed with skewer with default parameters andminimum quality Q=20. Overlapping forward and reverse reads were merged using PEAR. Resulting reads were mapped to the *Drosophila melanogaster* reference genome v 5.39 with BWA (default parameters). Reads were deduplicated using Picardtools and realigned around indels using GATK v4. Raw allele frequencies at each SNP site were then calculated using Popoolation (*39*) and custom bash scripts.

Identifying significant parallel SNPs

A generalized linear model with a quasibinomial error model was fit to allele frequencies at each SNP to assess the parallelism of shifts in allele frequency over each time interval using the R function 'glm'. To consider any outlier as significant we required sites to show >2% average change in allele frequency across all replicate cages over the time interval, and Benjamini-Hochberg false discovery rate corrected p-value <.05. We also created an empirical false discovery rate correction by shuffling the sample time point labels and re-running GLMs, however this rate proved to be less stringent and therefore was not used in the analysis.

Leave-one out cross validation analysis

In each round, a GLM was fit using allele frequencies from 9 training cages, and parallel sites were identified at each time segment as described above. For each parallel site, a matched control site was identified on the same chromosome that had an initial frequency in the baseline population within 5% of the parallel site. At each parallel and control site, the allele frequency shift over each time segment in the 10th left-out cage was calculated and phased such that a shift in the same direction as the training cages was given a positive sign and a shift in the opposite direction was given a negative sign. A t-test was conducted for each time segment to determine if the set of phased shifts at parallel sites was significantly different than shifts across all control sites. In Figure 3, we plotted the median phased shift for each set of sites at each time segment, and colored the point for parallel sites if the t-test p-value was < 0.05 after false discovery rate correction.

Defining SNP clusters

A GLM model was fit to allele frequencies from all 10 cages at each site as described above, to assess the parallelism of the shift over each time interval. Each site was assigned a score for each time interval according to the following criteria: 0 = [FDR >0.2], 1 = [FDR<.2 or effect size <2%], 2 = [FDR<.05, effect size >2%], 3 = [FDR<.01, effect size >2%]. While only sites receiving a score of 2 or 3 were defined as 'significant' in the analysis, lower scoring sites were helpful in identifying large regions of elevated parallelism. Average SNP scores were calculated for sliding windows of 500 SNPs (offset=100 SNPs), and significantly enriched windows were defined as those with an empirical FDR <.05 compared to the distribution of window scores obtained by randomly shuffling sites across the genome. Overlapping enriched windows were then merged. Next, linkage was calculated between all pairs of significant SNPs less than 3 Mb apart from the same time interval. Linkage was defined as the squared correlation coefficient from a Pearson correlation of founder genotypes at the two sites, with genotypes coded as 0, 0.5, 1, or NA for missing data. Neighboring windows with average SNP-pair linkage >0.03 were merged into clusters, and the process was repeated iteratively until no neighboring clusters within 3Mb exceeded an average linkage of 0.03.

Defining superclusters

A list was generated of all pairs of clusters identified at different time segments that overlapped by at least one SNP. Clusters identified across the whole experiment ($T_1 \rightarrow T_5$) were excluded from this list, resulting in 44 pairs of overlapping clusters. For each pair of clusters, linkage (R^2) values between all inter-cluster pairs of significant SNPs within 3Mb of each other were calculated and compared to linkage values for a set of randomly selected control SNP pairs matched for chromosome, initial frequencies, and inter-SNP distance. If linkage values for the cluster SNPs were significantly higher than linkage values for the matched control SNPs

(Benjamini-Hochsburg FDR <.05), the clusters were considered significantly linked. Any individual pairs of linked clusters that shared a cluster in common were merged into linked cluster sets to form the final list of superclusters.

REFERENCES

1. L. B. Slobodkin, *Growth and regulation of animal populations* (Dover Publications., 1980).

2. A. Charmantier, R. H. McCleery, L. R. Cole, C. Perrins, L. E. B. Kruuk, B. C. Sheldon, Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*. **320**, 800–803 (2008).

3. A. M. Simons, Modes of response to environmental change and the elusive empirical evidence for bet hedging. *Proc. Biol. Sci.* **278**, 1601–1609 (2011).

4. D. N. Reznick, F. H. Shaw, F. H. Rodd, R. G. Shaw, Evaluation of the Rate of Evolution in Natural Populations of Guppies (Poecilia reticulata). *Science*. **275**, 1934–1937 (1997).

5. P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches. *Science*. **296**, 707–711 (2002).

6. N. G. Hairston, S. P. Ellner, M. A. Geber, T. Yoshida, J. A. Fox, Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.* **8**, 1114–1127 (2005).

7. R. D. H. Barrett, S. M. Rogers, D. Schluter, Natural selection on a major armor gene in threespine stickleback. *Science*. **322**, 255–257 (2008).

8. R. D. H. Barrett, S. Laurent, R. Mallarino, S. P. Pfeifer, C. C. Y. Xu, M. Foll, K. Wakamatsu, J. S. Duke-Cohan, J. D. Jensen, H. E. Hoekstra, Linking a mutation to survival in wild mice. *Science*. **363**, 499–504 (2019).

9. D. J. Rennison, S. M. Rudman, D. Schluter, Genetics of adaptation: Experimental test of a biotic mechanism driving divergence in traits and genes. *Evolution Letters*. **0** (2019), doi:10.1002/evl3.135.

10. S. M. Rudman, S. Greenblum, R. C. Hughes, S. Rajpurohit, O. Kiratli, D. B. Lowder, S. G. Lemmon, D. A. Petrov, J. M. Chaston, P. Schmidt, Microbiome composition shapes rapid genomic adaptation of Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20025–20032 (2019).

11. J. B. S. Haldane, The cost of natural selection. *Journal of Genetics*. **55**, 511–524 (1957).

12. E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. **169**, 1177–1186 (2017).

13. P. D. Gingerich, Rates of Evolution - Effects of Time and Temporal Scaling. *Science*. **222**, 159–161 (1983).

14. M. Turelli, N. H. Barton, Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and G x E interactions. *Genetics*. **166**, 1053–1079 (2004).

15. R. C. Lewontin, *The genetic basis of evolutionary change* (Columbia University Press New York, 1974), vol. 560.

16. M. T. Kinnison, N. G. Hairston Jr, A. P. Hendry, Cryptic eco-evolutionary dynamics. *Ann. N. Y. Acad. Sci.* **1360**, 120–144 (2015).

17. D. Schluter, L. Nagel, Parallel speciation by natural selection. *American Naturalist*. **146**, 292–301 (1995).

18. S. Rajpurohit, E. Gefen, A. O. Bergland, D. A. Petrov, A. G. Gibbs, P. S. Schmidt, Spatiotemporal dynamics and genome-wide association genome-wide association analysis of desiccation tolerance in Drosophila melanogaster. *Mol. Ecol.* **27**, 3525–3540 (2018).

19. T. N. Grainger, S. M. Rudman, P. Schmidt, J. M. Levine, Competitive history shapes rapid evolution in a seasonal climate. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2015772118.

20. S. Tilk, A. Bergland, A. Goodman, P. Schmidt, D. Petrov, S. Greenblum, Accurate Allele Frequencies from Ultra-low Coverage Pool-Seq Samples in Evolve-and-Resequence Experiments. *G3* . **9**, 4159–4168 (2019).

21. E. L. Behrman, S. S. Watson, K. R. O'Brien, M. S. Heschel, P. S. Schmidt, Seasonal variation in life history traits in two Drosophila species. *J. Evol. Biol.* **28**, 1691–1704 (2015).

22. T. Flatt, Life-History Evolution and the Genetics of Fitness Components in Drosophila melanogaster. *Genetics*. **214**, 3–48 (2020).

23. J. B. S. Haldane, Suggestions as to quantitative measurement of rates of evolution. *Evolution*. **3**, 51–56 (1949).

24. P. D. Gingerich, Quantification and comparison of evolutionary rates. *Am. J. Sci.* **293**, 453–478 (1993).

25. A. P. Hendry, T. J. Farrugia, M. T. Kinnison, Human influences on rates of phenotypic change in wild animal populations. *Mol. Ecol.* **17**, 20–29 (2008).

26. T. F. C. Mackay, S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. H. Anholt, M. Barrón, C. Bess, K. P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L. Perales, L.-L. Pu, C. Qu, M. Ràmia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley, Y.-Q. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman, R. A. Gibbs, The Drosophila melanogaster Genetic Reference Panel. *Nature*. **482**, 173–178 (2012).

27. J. Stapley, J. Reger, P. G. D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A. D. Ball, A. P. Beckerman, J. Slate, Adaptation genomics: the next generation. *Trends Ecol. Evol.* **25**, 705–712 (2010).

28. T. Dobzhansky, Genetics of natural populations; altitudinal and seasonal changes produced by natural selection in certain populations of Drosophila persimilis. *Genetics*. **33**, 158–176 (1948).

29. M. Kapun, T. Flatt, The adaptive significance of chromosomal inversion polymorphisms in Drosophila melanogaster. *Mol. Ecol.* **28**, 1263–1282 (2019).

30. G. Bell, Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 87–97 (2010).

31. A. P. Hendry, M. T. Kinnison, Perspective: The Pace of Modern Life: Measuring Rates of Contemporary Microevolution. *Evolution*. **53**, 1637–1653 (1999).

32. P. S. Schmidt, D. R. Conde, Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in Drosophila melanogaster. *Evolution*. **60**, 1602–1611 (2006).

33. A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, D. A. Petrov, Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLoS Genet.* **10**, e1004775 (2014).

34. H. Machado, A. O. Bergland, R. Taylor, S. Tilk, E. Behrman, K. Dyer, D. Fabian, T. Flatt, J. Gonzalez, T. Karasov, I. Kozeretska, B. Lazzaro, T. Merritt, J. Pool, K. O'Brien, S. Rajpurohit, P. Roy, S. Schaeffer, S. Serga, P. Schmidt, D. Petrov, Broad geographic sampling reveals predictable and pervasive seasonal adaptation in Drosophila. *bioRxiv* (2018), p. 337543.

35. B. Charlesworth, Causes of natural variation in fitness: evidence from studies of Drosophila populations. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1662–1669 (2015).

36. M. J. Wittmann, A. O. Bergland, M. W. Feldman, P. S. Schmidt, D. A. Petrov, Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. *Proc. Natl. Acad. Sci. U. S. A*. **114**, E9932–E9941 (2017).

37. J. Bertram, J. Masel, Different mechanisms drive the maintenance of polymorphism at loci subject to strong versus weak fluctuating selection. *Evolution*. **73**, 883–896 (2019).

38. M. Baym, S. Kryazhimskiy, T. D. Lieberman, H. Chung, M. M. Desai, R. Kishony, Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. *PLoS One*. **10**, e0128036 (2015).

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