ARTICLE

Natural selection and parallel clinal and seasonal changes in

*Drosophila melanogaster*

Running title: Parallel clinal and seasonal changes

Murillo F. Rodrigues¹,²*, Maria D. Vibranovski¹, Rodrigo Cogni³

¹ Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of Sao Paulo, Sao Paulo, SP, Brazil
² Present address: Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA
³ Department of Ecology, Institute of Biosciences, University of Sao Paulo, Sao Paulo, SP, Brazil

* Corresponding author:

Murillo F. Rodrigues

Institute of Ecology and Evolution
University of Oregon
272 Onyx Bridge
5289 University of Oregon
Eugene, OR 97403-5289.

Author contributions

M.F.R. and R.C. designed the research. M.F.R. performed the research and analyzed the data. M.F.R., M.D.V. and R.C. discussed the results and conclusions. M.F.R. wrote the manuscript with input from M.D.V and R.C.

Acknowledgements

We would like to thank all the members of the Cogni Lab for input and support throughout the development of this work. We thank Diogo Meyer, Paul Schmidt, Reinaldo Azevedo and Peter Ralph for discussions and three anonymous reviewers for comments on the manuscript.

Funding for this work was provided by São Paulo Research Foundation (FAPESP)
(13/25991-0 and 17/02206-6 to RC, 15/20844-4 to MDV, 16/01354-9 and 17/06374-0 to MFR), and CNPq (307015/2015-7 and 307447/2018-9 to RC), and a Newton Advanced Fellowship from the Royal Society to RC.

Data accessibility

All the data used in this project are available on the NCBI Short Read Archive (BioProject accession numbers PRJNA256231 and PRJNA308584, and NCBI Sequence Read Archive SRA012285.16).
Abstract

Spatial and seasonal variation in the environment are ubiquitous. Environmental heterogeneity can affect natural populations and lead to covariation between environment and allele frequencies. *Drosophila melanogaster* is known to harbor polymorphisms that change both with latitude and seasons. Identifying the role of selection in driving these changes is not trivial, because non-adaptive processes can cause similar patterns. Given the environment changes in similar ways across seasons and along the latitudinal gradient, one promising approach may be to look for parallelism between clinal and seasonal change. Here, we test whether there is a genome-wide relationship between clinal and seasonal variation, and whether the pattern is consistent with selection. We investigate the role of natural selection in driving these allele frequency changes. Allele frequency estimates were obtained from pooled samples from seven different locations along the east coast of the US, and across seasons within Pennsylvania. We show that there is a genome-wide pattern of clinal variation mirroring seasonal variation, which cannot be explained by linked selection alone. This pattern is stronger for coding than intergenic regions, consistent with natural selection. We find that the genome-wide relationship between clinal and seasonal variation could be explained by about 4% of the common autosomal variants being under selection. Our results highlight the contribution of natural selection in driving fluctuations in allele frequencies in *D. melanogaster*.

Introduction

Species occur in environments that vary both in space and time (Ewing 1979; Cardini et al. 2007; Dionne et al. 2007; Hancock et al. 2008; Zuther et al. 2012; Campitelli and Stinchcombe 2013; Kooyers et al. 2015). Traits that covary with the environment in space, also called clinal, are usually thought to be the result of spatially varying selection (Endler 1977). However, non-adaptive processes such as isolation by distance, range expansion and
admixture can produce similar patterns of spatial change (Wright 1943; Vasemägi 2006; Excoffier et al. 2009; Duchen et al. 2013; Bergland et al. 2016). Therefore, disentangling the relative contributions of natural selection and demographic processes in producing clinal variation is challenging.

*Drosophila melanogaster* has been used as a model to study clinal variation because, in addition to being a model species, it is a sub-Saharan fly species that has recently invaded most of the world (David and Capy 1988). These flies migrated into Eurasia approximately 15,000 years ago (Li and Stephan 2006), but the colonization of the Americas and Australia likely happened in a single event within the last several hundred years (Bock and Parsons 1981; Keller 2007). The establishment of populations in dramatically different environments (e.g., temperate regions) was followed by the emergence of adaptations to these new environments (David and Capy 1988; Yukilevich et al. 2010).

Clinal variation has been documented for *D. melanogaster* in many characters both at the phenotypic and genetic level. For instance, at the phenotypic level, flies from higher latitudes are darker (David et al. 1985), bigger (Arthur et al. 2008) and show higher incidence of reproductive diapause than flies from lower latitudes (Schmidt et al. 2005). At the genotypic level, latitudinal clines have been identified for allozymes, chromosome inversions, copy-number variation and single nucleotide polymorphisms (Mettler et al. 1977; Knibb 1982; Oakeshott et al. 1982; Schmidt et al. 2000; de Jong and Bochdanovits 2003; Sezgin et al. 2004; Fabian et al. 2012; Kapun et al. 2016; Schrider et al. 2016). Surprisingly, not much is known about the actual selective pressures in natural fly populations, but temperature is directly or indirectly underlying much of clinal adaptation (Knibb 1982; Krimbas and Powell 1992; Božičević et al. 2016).

A classic approach to discern between adaptive and neutral differentiation has been to look at parallel latitudinal variation between continents (Hoffmann et al. 2002; Hoffmann and
If a trait varies clinally in two or more continents, a plausible scenario is latitudinally varying selection acting on ancestral variation (Endler 1977; Barton 1983; Barton 1999). However, some of the clinal patterns observed in North America may be attributed to migration of African flies to low latitude regions and European flies to high latitude regions (Caracristi and Schlotterer 2003; Yukilevich and True 2008; Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016). The main evidence in favor of the secondary contact hypothesis is that there is a cline in the proportion of African ancestry in North America (Kao et al. 2015; Bergland et al. 2016), and a weak cline in African ancestry in Australia (only marginally significant; see Bergland et al. 2016).

Separating the role of selection and demography in maintaining clinal polymorphisms is not trivial, given that the axis of demography and environmental change are confounded in this system (i.e., migration and the environment are structured along the south-north axis). A simpler, more plausible model would be that admixture introduced “pre-adapted” alleles, that then contributed to clinal adaptation (Flatt 2016). One approach that can help disentangle the role of adaptation is to explore signatures of parallel latitudinal and seasonal changes (Cogni et al. 2014; Cogni et al. 2015).

_Drosophila_ has been widely used in the study of seasonal adaptation. One of the earliest examples of seasonal genetic variation was observed in chromosomal inversions in _D. pseudoobscura_ (Dobzhansky 1943). In _D. melanogaster_, it has been shown that flies collected in the spring are more tolerant to stress (Behrman et al. 2015), show higher diapause inducibility (Schmidt and Conde 2006), have increased immune function (Behrman et al. 2018) and have different cuticular hydrocarbon profiles than those collected in the fall (Rajpurohit et al. 2017). Furthermore, genome-wide analyses have identified polymorphisms and inversions that oscillate in seasonal timescales in several localities in the United States.
and Europe, but the mechanism behind it has not been fully resolved (Bergland et al. 2014; Kapun et al. 2016; Machado et al. 2018; Wittman et al. 2017). Identifying polymorphisms that fluctuate seasonally is challenging, because effect sizes are small, and populations are subject to stochastic environmental events (Machado et al. 2018). A recent analysis suggested seasonal fluctuations in allele frequencies is probably small, and temporal structure independent of seasons may be more important in this system (Buffalo and Coop 2019).

Although it is hard to characterize the role of selection in producing latitudinal or seasonal change in allele frequencies, we could gain power by analyzing parallelism between latitudinal and seasonal changes. Two adaptive mechanisms are expected to induce correlations between clinal and seasonal fluctuations in allele frequencies. First, the environment changes similarly with latitude and through seasons (at least with respect to temperature). Second, the onset of spring changes with latitude, and so seasonal changes in polymorphisms alone could produce clinal variation, a mechanism termed seasonal phase clines (Roff 1980; Rhomberg and Singh 1986). Demographic processes that can generate clinal patterns like isolation by distance and secondary contact from diverged populations are independent to variation across seasons, given the short time scale at which the seasonal processes happen (Bergland et al. 2014; Cogni et al. 2015). Therefore, parallel latitudinal and seasonal variation in a trait is strong evidence in favor of natural selection. Indeed, it has been observed in *D. melanogaster* that both the prevalence of reproductive diapause and the frequency of a variant in the couch potato gene associated with diapause inducibility vary latitudinally and seasonally (Schmidt et al. 2005; Cogni et al. 2014). The frequency of this diapause-inducing variant drops in the summer and is positively correlated with latitude, as we would expect if it were being driven by natural selection.

Besides this classic example, a few studies have looked for signatures of parallel clinal and seasonal variation in *D. melanogaster* (Bergland et al. 2014; Cogni et al. 2015;
Kapun et al. 2016; Behrman et al. 2018; Machado et al. 2018). Cogni et al. (2015) found an association between clinal and seasonal change in central metabolic genes, that are likely important drivers of climatic adaptation. Kapun et al. (2016) dissected the role of selection in shaping clinal variation in *D. melanogaster* cosmopolitan inversions. By integrating clinal and seasonal variation, information on the stability of inversion clines, and analysis of the effect of climatic variables on inversion frequencies, they concluded that several inversion clines are consistent with natural selection. A recent study also investigated the association between clinal and seasonal variation and showed the direction of allele frequency change with latitude is concordant with the change with seasons in many polymorphisms across the *D. melanogaster* genome (Machado et al. 2018). Nevertheless, parallel clinal and seasonal variation has not been used to characterize the role of selection in driving genome-wide fluctuations in allele frequency.

Here, we aim to answer whether the clinal and seasonal patterns observed in *D. melanogaster* are consistent with natural selection. If natural selection is a key process underlying clinal and seasonal change, flies collected in the north should be more similar to flies collected in early spring, whereas southern flies should be more like flies collected in the fall (Cogni et al. 2014; Bergland et al. 2014; Cogni et al. 2015; Kapun et al. 2016; Behrman et al. 2018; Machado et al. 2018). We quantified the relationship between clinal and seasonal change along the *D. melanogaster* genome. Given the localized effects of selection, and the expectation that constraint depends to some degree on genic annotations, we also expect the parallelism to vary across genomic annotations. We also assessed if the coupling between clinal and seasonal change stems from just a few selected polymorphisms, or if spatial and temporal selection acts pervasively across the genome. Finally, we estimated how much of the common polymorphisms should be under selection to explain the observed degree of coupling between clinal and seasonal change. By integrating these two independent sources
of evidence, we further disentangle the contributions of selection in maintaining the patterns of clinal and seasonal variation in D. melanogaster populations.

**Material and Methods**

**Population samples**

We analyzed 20 samples from seven locations along the United States east coast, collected by (Bergland et al. 2014) (10 samples), and (Machado et al. 2018) (10 samples) (see Table S1). The samples were based on pools of wild-caught individuals. We decided to not include previously collected samples from Maine because they were collected in the fall, whereas all of our other samples were collected in the spring, and we also did not include the DGRP sample from North Carolina, as it is hard to ascertain when they were obtained (Fabian et al. 2012; Mackay et al. 2012; Bergland et al. 2014). The Linvilla (Pennsylvania) population was sampled extensively from 2009 to 2015 (six spring, seven fall samples), and was therefore used in our analysis of seasonal variation. We also replicated our results using data from four Australian samples (Anderson et al. 2005; Kolaczkowski et al. 2011). All the data used in this project are available on the NCBI Short Read Archive (BioProject accession numbers PRJNA256231, PRJNA308584 and NCBI Sequence Read Archive SRA012285.16).

**Mapping and processing of sequencing data**

Raw, paired-end reads were mapped against the FlyBase D. melanogaster (r6.15) and D. simulans (r2.02) reference genomes (Gramates et al. 2017) using BBSplit from the BBMap suite (https://sourceforge.net/projects/bbmap/; version from February 11, 2019). We removed any reads that preferentially mapped to D. simulans to mitigate effects of contamination (the proportion of reads preferentially mapping to D. simulans was minimal, never exceeding 3%). Then, reads were remapped to D. melanogaster reference genome using bwa (MEM algorithm) version 0.7.15 (Li and Durbin 2010). Files were converted from SAM to BAM
format using Picard Tools (http://broadinstitute.github.io/picard). PCR duplicates were
marked and removed using Picard Tools and local realignment around indels was performed
using GATK version 3.7 (McKenna et al. 2010). Single nucleotide polymorphisms (SNPs)
and indels were called using CRISP with default parameters (Bansal et al. 2016). We applied
several filters to ensure that the identified SNPs were not artifacts. SNPs in repetitive regions,
identified using the RepeatMasker library for D. melanogaster (obtained from
http://www.repeatmasker.org), and SNPs within 5bp of polymorphic indels were removed
from our analyses, because these are more likely to be sequencing errors. SNPs with average
minor allele frequency in the clinal and seasonal samples less than 5%, with minimum per-
population coverage less than 10x (or 4x for the Australian samples) or maximum per-
population coverage greater than the 99th quantile were excluded from our analyses. Also,
SNPs with multi-allelic states were removed from the analysis. We did not include SNPs in
any of the sex chromosomes. SNPs were annotated with their genic classes using SNPeff
version 4.3o (Cingolani et al. 2012).

Clinal and seasonal changes in allele frequency
Pool-seq data contain an additional component of error due to sampling (Kofler et al. 2011;
Lynch et al. 2014). We model the additional error by computing the effective number of
chromosomes ($N_E$) at each site:

$$N_E = \left( \frac{1}{D} + \frac{1}{N_C} \right)^{-1}$$

where $N_C$ is the number of chromosomes in the pool and $D$ is the read depth at that site
(Kolaczkowski et al. 2011; Feder et al. 2012; Bergland et al. 2014).

To assess latitudinal variation, we used generalized linear models (GLM) to regress
allele frequency at each site against latitude according to the form,

$$y_i = \alpha + \beta^{\text{clinal}\text{Latitude}} + \epsilon_i$$
Where \( y_i \) is the allele frequency in the \( i^{th} \) population, \( \alpha \) is the intercept, \( \beta^{\text{cinal}} \) is the regression coefficient for latitude, and \( \epsilon_i \) is the binomial error given the \( N_E \) at that polymorphism. This kind of regression is suitable for the analysis of clinal variation of allele frequencies estimated from Pool-seq, because it accounts for differences in read depth and number of chromosomes, and the non-normality of allele frequencies (Bergland et al. 2014; Machado et al. 2016).

To assess seasonal variation, we followed a similar procedure. We regressed allele frequency at each site against a binary variable corresponding to spring or fall according to the form,

\[
y_i = \alpha + \beta^{\text{seasonal \, Season}} + \gamma \text{Year} + \epsilon_i
\]

where \( y_i \) is the allele frequency in the \( i^{th} \) sample, \( \alpha \) is the intercept, \( \beta^{\text{seasonal}} \) is the seasonal regression coefficient (Season is a dummy variable: June and July were encoded as Spring, and September, October and November as Fall), \( \gamma \) is the regression coefficient for year, and \( \epsilon_i \) is the binomial error given the \( N_E \) at that SNP.

We defined clinal and seasonal SNPs using an outlier approach, because we do not have an adequate genome-wide null distribution to compare our estimates. We considered that SNPs were outliers if their regression P-value fell in the bottom 1% (or 5%) of the distribution. All statistical analyses were performed in R 3.5.0 (R Core Team 2018).

Relationship between clinal and seasonal variation

We examined the relationship between clinal and seasonal variation by regressing the clinal against seasonal slopes for all SNPs:

\[
\beta_i^{\text{cinal}} = \alpha + \beta \beta_i^{\text{seasonal}} + \epsilon_i
\]

where \( \beta_i^{\text{cinal}} \) is the estimated clinal slope for the \( i^{th} \) SNP, \( \alpha \) is the intercept, \( \beta \) is the slope, \( \beta_i^{\text{seasonal}} \) is the estimated seasonal slope for the \( i^{th} \) SNP, and \( \epsilon_i \) is the gaussian error at that
SNP. The regression line was fit using Huber’s M estimator to improve robustness to outliers. The clinal and seasonal slopes were z-normalized to facilitate interpretation of the magnitude; this way, the regression coefficient can be interpreted as a correlation coefficient. To confirm our results are robust to potential model misspecifications, we implemented a permutation test in which we rerun the generalized linear regressions for each SNP using shuffled season and latitude labels 2,000 times. The same procedure was implemented for most of the statistical tests, except where indicated otherwise.

To assess differences in the relationship between clinal and seasonal variation across different genic classes (exon, intron, 5' UTR, 3' UTR, upstream, downstream intergenic, splice), we added the following interaction term:

\[ \beta_i^{clinal} = \alpha + \gamma_i \beta_i^{seasonal} \text{Class}_j + \epsilon_i \]

where \( \gamma_i \) is the regression coefficient for the \( j^{th} \) genic class, \( \text{Class}_j \) is the genic class of the \( i^{th} \) SNP.

There are some chromosomal inversions segregating in the populations we studied, and they are known to contribute to adaptation (Wright and Dobzhansky 1946; García-Vázquez and Sánchez-Refusta 1988; Kapun et al. 2014). We tested whether the relationship between clinal and seasonal variation is stronger at SNPs that were identified to be markers of inversions in Kapun et al. (2014). To do so, we performed a regression that takes the form:

\[ \beta_i^{clinal} = \alpha + \gamma_i \beta_i^{seasonal} \times \text{Inversion}_j + \epsilon_i \]

where \( \text{Inversion}_j \) is the inversion status of the SNP, and \( \gamma_i \) is the regression coefficient for the \( j^{th} \) status. We considered SNPs surrounding the inversion breakpoints in the same manner (Corbett-Detig and Hartl 2012). All statistical analyses were performed in R 3.5.0 (R Core Team 2018) and can be found at gitlab.com/mufernando/clinal_sea.git.
We tested for enrichment of genic classes using our sets of clinal and seasonal SNPs using Fisher’s exact test for each genic region and statistic. To control for confounders, such as read depth variation, we computed P-values based on 2,000 permutations. In each iteration, we shuffled the season and latitude labels and reran the generalized regressions. Using the P-values obtained from regressions in which season and latitude labels were shuffled, we defined for each iteration the top clinal and seasonal SNPs. Then, we calculated the enrichment of each genic class using Fisher’s exact test. To obtain a P-value for an enrichment of a given genic class, we compared the observed odds ratio in the actual dataset to the distribution of odds ratios observed for datasets in which season and latitude labels were shuffled.

Mitigating the impact of linkage disequilibrium

Selection at one site affects genetic variation at nearby, linked neutral sites (Smith and Haigh 1974). Because we assume that sites are independent in our models, the indirect effects of selection can inflate the magnitude of the patterns we investigated. To test the effect of linkage disequilibrium (LD) in our outlier analyses, we computed for all SNPs how their distance to a top SNP affects their P-value. We then smoothed the scatterplot using cubic splines as implemented in ggplot2 (Wickham 2016). To test the effect of linkage on the relationship between clinal and seasonal variation, we implemented a thinning approach. Sampling one SNP per L base pairs one thousand times, we constructed sets of SNPs with minimized dependency, where L ranged from 1 to 20kb. For each of these sets for a given L, we regressed clinal against seasonal slopes, and compared the distribution of the thinned regression coefficients to the coefficients we obtained using all SNPs.

Inferring proportion of selected SNPs

Although a negative, significant relationship between clinal and seasonal variation indicates that a subset of the polymorphisms is under selection, it is difficult to measure how a given
slope is related to a biological parameter, such as the proportion of SNPs that are driven by selection. Most of the observed polymorphisms should be neutral and only a few under direct selection, therefore we expect the relationship between clinal and seasonal change to be minor (i.e., for most SNPs clinal and seasonal change should be unrelated). To gain some intuition of how the proportion of SNPs with uncoupled clinal and seasonal change (e.g., neutral polymorphisms) would impact our measure of interest, the regression coefficient between clinal and seasonal variation, we performed a set of simple simulations.

The polymorphisms most affected by the environment are going to be the ones that change most consistently with seasons and latitude. In this thought experiment, we will assume that the rank order of clinal and seasonal change should be the same (e.g., for SNPs under selection, the linear relationship between clinal and seasonal change will be assumed to be exactly -1). Biologically, this would happen if the environment changed in the same ways with latitude and seasons. This assumption is most likely not true, but it renders our thought experiment conservative (Fig. S4). Note the coefficient cannot be less than -1, because the two variables being regressed were z-normalized and thus have same variance.

We simulated datasets with different proportions of points falling along a line ($y = -x$), representing selected SNPs and points without a linear relationship, representing SNPs that are not under spatial and seasonal selection. For each simulation replicate, we calculated the linear relationship between two variables with varying proportions of points falling on the 1:-1 line to independent points. Each value was drawn from the combined distribution of observed z-normalized regression coefficients for season and latitude. By calculating the observed slope between the regression coefficients of each of our simulated datasets, we could establish a relationship between observed slope and the expected proportion of SNPs that fall along 1:-1 line (or that are under selection). We then used Approximate Bayesian Computation using a sequential sampling scheme (as implemented in the R package
We assembled 21 *D. melanogaster* population samples collected from seven localities across multiple years in the east coast of the United States. All of these samples are the result of a collaborative effort of many researchers from a consortium, the DrosRTEC (Bergland et al. 2014; Machado et al. 2018). Seven of our samples span from Florida to Massachusetts and together comprise our clinal set. The seasonal samples were collected in Pennsylvania in the spring (6 samples collected in June or July) and in the fall (7 samples collected in September, October or November). For each sample, a median of 55 individuals (with a range of 33 to 116) was pooled and resequenced to an average 75x coverage (ranging from 17 to 215). We also used four clinal samples from the Australia (Anderson et al. 2005; Kolaczkowski et al. 2011). More details about the samples can be found on Table S1 (also see Machado et al. 2018; Bergland et al. 2014). After all the filtering steps, we identified 797,792 common autosomal SNPs, which were used in our downstream analyses.

Allele frequency changes with latitude, seasons and years

Latitude explains much of allele frequency variation along the surveyed populations, as there is an excess of low GLM P-value SNPs (Fig. 1A). The mean absolute difference in allele frequency between the ends of the clines is 9.2%. Seasons, on the other hand, explain less of the variation in allele frequency. There is only a minor excess of low GLM P-value SNPs (Fig. 1B) and the mean absolute difference in allele frequency between seasons is 2.6%. We also found that year of sampling is a good predictor of allele frequency change (in
Pennsylvania), more so than seasons, given there is a huge excess of low GLM P-values (Fig. 1C).

Figure 1. Distribution of P-values from the generalized linear models of allele frequency and latitude, and allele frequency and seasons/years.

Our generalized linear models do not account for dependency between samples, which can be a problem when regressing allele frequency on seasons. To investigate whether this could be an issue, we performed Durbin-Watson tests for autocorrelation in the residuals of the seasonal regressions using Julian days as the time variable. We found a minor excess of low P-values (Fig. S1), and the season P-values are weakly, but negatively correlated with the Durbin-Watson test P-values (Spearman $\rho = -0.004, P = 0.0001$). This indicates that the assumption of independency is being met for most variants, and that autocorrelation is not artificially creating patterns of seasonality in allele frequency.

Given we do not have enough information to build an appropriate null distribution to calibrate our P-values, we sought to demonstrate that top significant clinal and seasonal SNPs are enriched for functional variants, which are more likely to contribute to adaptation. Latitudinal SNPs are more likely to be in exonic and UTR 3' regions (Fig. 2A), whereas seasonal SNPs are enriched for exonic, UTR 3' and UTR 5' regions (Fig. 2B). Further, top
latitudinal and seasonal SNPs seem to be underrepresented within upstream and downstream regions (Fig. 2). The enrichment of top latitudinal SNPs among functional genic classes has been reported by previous studies which used different latitudinal samples (Fabian et al. 2012; Kolaczowski et al. 2011, Machado et al. 2016). Enrichment of top seasonal SNPs among some functional regions have been observed before (Bergland et al. 2014; Machado et al. 2018). Using a 5% cutoff, our results with respect to enrichments is largely replicated (Fig. S2).

**Figure 2.** Top SNPs are enriched for functionally relevant classes. Enrichment of top 1% SNPs in each genic class for A) latitudinal P-value and B) seasonal P-value. The histograms show the distribution of odds ratios when latitude and season labels were permuted, and the vertical bars show the observed odds ratios.

Clinal variation is related to seasonal variation

A clinal pattern can arise solely as a result of demographic processes, such as isolation by distance and admixture (Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016). Although seasonality is less affected by such processes, seasonal change is less pronounced and more subject to stochastic changes in the environment, making it harder to detect seasonal change with precision. Here, we integrate both clinal and seasonal change estimates across a large number of SNPs in the genome (about 790,000 SNPs). We expect the overall pattern that
emerges to be informative of the relative role of natural selection, because selection is a plausible process to produce a pattern of clinal variation mirroring seasonal variation (Cogni et al. 2015). With higher latitudes, cold-favored alleles rise in frequency, whereas from spring to fall, the frequency of cold-favored alleles decreases. Conversely, with higher latitudes, warm-favored alleles decrease in frequency, whereas from spring to fall, the frequency of warm-favored alleles increases.

To assess if clinal variation and seasonal variation are related, we performed a linear regression of the slopes of clinal and seasonal allele frequency change (see Methods). We find a significant negative linear relationship between clinal and seasonal regression coefficients (Fig. 3A, Table S2). The relationship is strongest for SNPs within exons, and the weakest for unclassified SNPs and those within intergenic regions (Fig. 3A; Table S3).

Nonetheless, the relationship is different than zero for all classes (except for the unclassified), what would be consistent either with pervasive linked selection or widespread distribution of variants that are important for adaptation, even within non-coding regions. Our results are robust to modelling assumptions, as they are significant when compared to the datasets with permuted of latitude and season labels (Fig. 3C). Qualitatively similar results were replicated using a different minor allele frequency cutoff and using samples from Maine, which were sampled in the summer (Fig. S3A-B).
Figure 2. Clinal change parallels seasonal change. A) Linear relationship between latitudinal change and seasonal change for each genic class. B) Association between top latitudinal and seasonal variants for different P-value cutoffs. C) Null distribution of the regression coefficient between latitudinal and seasonal change for each genic class (after permuting the latitude and season labels). Vertical bars represent the observed values.

Given that previous studies have demonstrated the importance of cosmopolitan inversions in climatic adaptation (Kapun et al. 2016), we looked at how clinal and seasonal variation are related within common cosmopolitan inversions. In our dataset, we found only 16 out of hundreds of marker SNPs identified in Kapun et al. (2014). It is unclear whether this is due to lifting over their annotations to a more recent version of the reference genome or if these variants are truly absent within the samples used in our study. We found that the relationship between clinal and seasonal change for those markers is high but did not reach significance due to small number of variants (Table S4). Most of the divergence (and differentiation) between flies with and without inversions should be located around the
breakpoints (Kapun et al. 2014, Kapun et al. 2018). Thus, we checked whether the pattern of parallelism between clinal and seasonal variation is stronger surrounding common cosmopolitan inversions (considering SNPs within 2Mb of the breakpoints identified in Corbett-Detig and Hartl, 2012). We found that clinal and seasonal change are strongly coupled surrounding the inversions (Table S5). Nevertheless, the pattern is still strong outside these regions, indicating our main results are not purely driven by frequency changes of inversions.

Another way of testing for parallelism between clinal and seasonal change is by testing if clinal SNPs are more likely to be seasonal (and vice-versa). We observed that clinal SNPs are enriched for seasonal SNPs (Fig. 2B). The enrichment increases with more stringent lower P-value quantile cut-offs, as we would expect if even strictly non-significant variants were informative of the role of selection.

We also confirmed our main finding, that clinal and seasonal change are related using samples from Australia. To measure clinal change in Australia, we only used four low coverage samples in Australia, two for each of low and high latitude locations. There is a genome-wide relationship between clinal variation in Australia and seasonal variation in Pennsylvania that, although minor, is significant (Fig. S3C). We also found a relationship between latitudinal regression significance and seasonal significance using the Australian samples (Fig. S3C).

The effects of linkage disequilibrium on clinal and seasonal variation

Variation at one site is linked to variation at other sites, and selection will increase this dependency (Smith and Haigh 1974). First, we assessed if latitudinal and seasonal P-values, were dependent on how distant a SNP was from our top 1% SNPs. We show that both statistics are dependent on distance from the outlier SNPs (Fig. 3A-B), but the effect virtually disappears after 5kb.
Our regression analysis of clinal and seasonal variation assumes that SNPs are independent. Thus, we assessed the impact of linkage on the relationship between clinal and seasonal change by implementing a thinning approach. First, we tested how the genome-wide regression estimate varied with changing window sizes. The effect of non-independency of variants on the relationship between clinal and seasonal change is rather small (Fig. 3C). Nevertheless, we chose a conservative window length of 5kb for downstream analyses. The strength of the genome-wide relationship between clinal and seasonal variation decreased when randomly sampling one SNP per 5kb, but it still remained significantly different from zero (P= 0.001; Fig. 3D). Thus, it seems the signal is widespread across many loci. The thinning did not significantly impact the signal for many regions, but the strength of the signal within splicing, UTR 5’, upstream, downstream and intergenic regions decreased and is not significantly different from 0 (Fig. 3D). It may be that most of the signal coming from those regions are due to the linked effects of selection.
Selection accounts for a substantial proportion of genome-wide clines

Although we identified a significant negative relationship between clinal and seasonal variation, the magnitude of the regression coefficient appears small (5kb bootstrap estimate $\beta = 0.037$). It is hard to measure the biological relevance of such coefficient, and so we designed a simple simulation approach that translates our statistic of interest into a biological parameter, that is the proportion of selected sites. We assume that: (i) for the subset of variants that are under selection, we would expect to see a perfectly linear relationship between z-normalized clinal and seasonal regression coefficients; (ii) for the variants that are
not contributing to adaptation, then we would expect absolutely no relationship between clinal and seasonal variation; (iii) variants are not linked. Assumption (i) renders our test conservative, as the relationship would not necessarily be negative one because some variants may be strongly seasonal but not necessarily strongly latitudinal and vice-versa. In Figure S4, we show that if the true slope for selected SNPs is not -1, we would be severely underestimating the proportion of selected SNPs. Accounting for linkage disequilibrium (assumption iii) is notoriously complicated, especially because we cannot accurately measure LD from pooled sequencing (Feder et al. 2012). However, we show that most of the effects of autocorrelation can be accounted for with our thinning approach, thus we used the observed relationship between clinal and seasonal change after 5kb thinning.

We asked what the expected relationship between clinal and seasonal regression coefficients would be if some of the variants were subject to spatially and seasonally varying selection but most were neutral. To assess this, we built datasets with varying proportion of adaptive variants (with linear relationship between clinal and seasonal variation) to neutral variants (no relationship between clinal and seasonal variation) and calculated the expected linear relationship between the two variables.

We found that the observed regression coefficient is consistent with approximately 3.9% of the common variants being under selection (Fig. 4), with 95% of the posterior estimates ranging from 3.7% to 4.2%. Thus, our analysis suggests that a sizable fraction of common, autosomal SNPs changes due to selection. Note this proportion is in line with previous estimates for the proportion of clinal SNPs (3.7% in Machado et al. 2015) and seasonal variants (~4% in Machado et al. 2018).
Figure 4. Proportion of SNPs that should be under selection to explain the genome-wide relationship between clinal and seasonal variation. A) background points represent each simulation, and the line is the fitted loess curve. Dashed lines relate the observed regression coefficient after thinning (5kb) to the expected proportion of selected SNPs. B) posterior distribution of the proportion of selected SNPs, given the observed relationship between clinal and seasonal variation (5kb thinned), with horizontal line highlighting the mean.

Discussion

Clinal patterns have been observed in both phenotypic and genotypic traits in many different species (Hancock et al. 2008; Baxter et al. 2010; Adrion et al. 2015). Especially in systems in which there is collinearity between the axis of gene flow and environmental heterogeneity, disentangling the contribution of selection and demography in producing clines is not trivial. Detecting seasonal cycling in allele frequencies is also challenging, mostly because the effect size is likely to be small and the environment may change unpredictably within seasons. We relied on integrating patterns of clinal and seasonal variation, since variation over short periods is independent of most confounding demographic processes. Parallel environmental change and the differences in the onset of seasons with latitude can create parallelism between clinal and seasonal variation. Natural selection would be the process driving the coupling between latitudinal and seasonal change in both scenarios, thus investigating it can
help discern the role of natural selection. Here, using *D. melanogaster* samples from the east coast of the United States and Australia, we show that (i) there is a genome-wide pattern of clinal variation mirroring seasonal variation, (ii) this is consistent with the action of natural selection, and (iii) a significant number of common variants may be being driven by selection.

We show that clinal and seasonal change are related, as there is a significant negative relationship between clinal and seasonal regression coefficients in the *D. melanogaster* genome (Fig. 2A), and there is an enrichment for seasonal SNPs among clinal SNPs (Fig. 2B). The relationship between clinal and seasonal regression is stronger for genic when compared to intergenic regions (Fig. 2A,C). Demographic processes would impact the genome as a whole, but variants in genic regions are more likely to impact fitness (Andolfatto 2005). Further, we estimate that about 3.9% of common, autosomal SNPs should be under selection to explain the genome-wide pattern of clinal and seasonal concordance uncovered (Fig. 4).

Because we expect selection to intensify linkage disequilibrium, the relationship between clinal and seasonal variation could be mostly driven by a few large effect loci. However, we show that our results cannot be fully explained by linked selection (Fig. 3). When accounting for autocorrelation along the chromosomes, the relationship between clinal and seasonal change disappears for variants within intergenic regions, but not for exons (Fig. 3D). The effects of linkage disequilibrium are important in this short time scale, but it decays rapidly, and returns to background levels after 5kb (Fig. 3A-C). We also investigated whether our results could be due to differences in inversion frequencies, which are known to underlie much of climatic adaptation in *D. melanogaster* (Kapun et al. 2016). The coupling between clinal and seasonal change surrounding common inversions is generally high, but in isolation it cannot explain the genome-wide patterns we observed. Therefore, our results support a
scenario in which the genome-wide pattern of clinal variation mirroring seasonal variation is a result of selection acting on many sites across the genome.

An important mechanism that can cause clinal patterns has been neglected from recent discussions of clinal variation in Drosophila. The latitudinal variation on the onset of seasons can produce clines, a phenomenon termed “seasonal phase clines” (Roff 1980; Rhomberg and Singh 1986). Under this model, a correlation between clinal and seasonal change is expected. Our latitudinal samples were all collected within one month of difference (during the spring), and so our observations could be partially explained by differences in the seasonal phase. Our data does not allow for proper disentangling of seasonal phase clines from parallel environmental change but change on the onset of seasons alone cannot explain our results. We found that latitude is usually a much better predictor of allele frequency differences (Fig. 1A-B), and the magnitude of change along the cline is much greater than what we found within a population across seasons (9.2% vs. 2.64%). We show that including Maine samples (which were sampled in the fall, in contrast to all other samples that were sampled in the spring) in our analyses does not meaningfully change our main results (Fig. S3). However, the rate of change (per degree of latitude) in allele frequency from Florida to Maine is smaller than the rate from Florida to Massachusetts. This could be due to differences in sampling year, but we found that the rate of change in frequency between Virginia and Maine is also smaller than the rate between Virginia and Massachusetts (FL and ME were sampled in 2009, whereas VA and MA were sampled in 2012; Fig. S5). We believe these differences are due to the shift in seasonal phase, as the samples from Maine were collected in the fall, but all other samples were collected in the spring.

Importantly, our results should be robust to differential admixture from Europe and Africa to the ends of the clines (Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016). Given variation over seasonal time scales are less affected by broader scale migration
patterns, it is not plausible that secondary contact can create parallel clinal and seasonal variation. Furthermore, the evidence for secondary contact in Australia is quite weak (see Bergland et al. 2016), but we show that clinal variation in Australia is related to seasonal variation in Pennsylvania (Fig. S3). Secondary contact may have contributed ancestral variation, that have since been selectively sorted along the cline (Flatt 2016). Consistent with this interpretation, the proportion of African ancestry is lower in low recombination regions, suggesting selection is mediating admixture in *D. melanogaster* (Pool 2015).

Population substructure and migration could cause seasonal variation in allele frequency in *D. melanogaster*. One possibility is that rural populations of *D. melanogaster* in temperate regions collapse during the winter and recover from spring to fall. However, reproductive diapause cycles in orchards and reaches high frequencies early in the spring, whereas its frequency in urban fruit markets in Philadelphia is much lower (Schmidt and Conde 2006). Another possibility is that seasonal variation is produced by migration of flies from the south in the summer, and from the north in the winter. There is little evidence of long-range migration in *D. melanogaster*, though this process seems important in *D. simulans* (Bergland et al. 2014; Machado et al. 2016). Importantly, flies have been shown to survive and reproduce during winter season in temperate regions, so rural flies can withstand a harsh winter season and be subject to selection (Mitrovski and Hoffmann 2001; Hoffmann et al. 2003, Rudman et al. 2019). These seasonal patterns have been replicated in many populations across North America and Europe (Machado et al. 2018), bolstering the argument for seasonal selection. Given the patterns we uncovered here are the result of subtle, but repeatable changes across multiple seasons, it is hard to imagine that selection is not the main causing force, even if it’s acting to maintain cryptic population structure within each location.

A recent study suggested the temporal changes in allele frequency reported in Bergland et al. (2014) is only weakly consistent with seasonal selection (Buffalo et al. 2019).
Consecutive spring-fall pairs showed some signal of adaptation, but the effect was small and disappeared at larger timescales (same season but across different years). Similarly, Machado et al. 2018 found that when they flipped the season labels of some samples the seasonal model fit was greatly improved. Consistent with these observations, we found there is strong temporal structure across years (Fig. 1C). Though we use a different set of samples, we show that the distribution of seasonal P-values is only slightly enriched for low P-values (Fig. 1B). Importantly, top seasonal SNPs are enriched for functional genic classes, when compared to datasets in which the season labels were permuted (Fig. 2B). These results highlight how difficult it is to find truly seasonal SNPs, likely because change in allele frequency is generally much smaller and natural populations are subject to stochastic changes in the environment not related to seasons. Once we have larger time series data, we will not need to rely on seasons as a proxy for environmental heterogeneity, and instead it will be possible to perform a proper causal analysis.

A few previous studies have investigated the relationship between clinal and seasonal variation (Bergland et al. 2014; Cogni et al. 2015; Kapun et al. 2016; Machado et al. 2018), but here we further explore parallelism between clinal and seasonal variation to tease apart the role of selection. Cogni et al. (2015) focused on a few SNPs in metabolic genes, and so they could not draw robust conclusions about processes that impact most of the genome. Our work is also fundamentally different from that of Machado et al. (2018), which revolves around identifying seasonal variation in *D. melanogaster* by aggregating data from many locations in North America and Europe. Here, we only included seasonal samples from Pennsylvania. This location had the greatest number of years surveyed (7 years, as opposed to 2 years for most of the other locations; see Machado et al. 2018), and it is the population for which we have the most information about seasonal variation in relevant phenotypes,
such as diapause (Schmidt and Conde 2006; Behrman et al. 2015; Rajpurohit et al. 2017; Behrman et al. 2018).

Many species occur along spatially structured environments and show clinal variation in traits, so a question that remains open is: what is the role of selection in producing and maintaining these patterns? Seasonal variation is also ubiquitous, especially in temperate environments, so seasonal change could be an important feature of organisms that have multiple generations each year (Behrman et al. 2015). Here, we demonstrate that by integrating clinal and seasonal variation, we can discern the contributions of selection in driving allele frequency changes with the environment. This approach could potentially be applied to other multivoltine species that occur along environmental gradients, including invasive species, which are known to often have short generation times and to reproduce quickly (Sakai et al. 2001).
Coyne JA, Beecham E. 1987. Heritability of two morphological characters within and among natural populations of Drosophila melanogaster. Genetics 117:727–737.

David J, Capy P, Payant V, Tsakas S. 1985. Thoracic trident pigmentation in Drosophila melanogaster: Differentiation of geographical populations. Genet. Sel. Evol. 17:211–224.

David JR, Bocquet C. 1975a. Evolution in a cosmopolitan species: Genetic latitudinal clines in Drosophila melanogaster wild populations. Experientia 31:164–166.

David JR, Bocquet C. 1975b. Similarities and differences in latitudinal adaptation of two Drosophila sibling species. Nature 257:588–590.

David JR, Capy P. 1988. Genetic variation of Drosophila melanogaster natural populations. Trends Genet. 4:106–111.

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L. 2007. Clinal Variation In Mhc Diversity With Temperature: Evidence For The Role Of Host? Pathogen Interaction On Local Adaptation In Atlantic Salmon. Evolution 61:2154-2164.

Dobzhansky T. 1943. Genetics of Natural Populations IX. Temporal Changes in the Composition of Populations of Drosophila Pseudoobscura. Genetics 28:162–186.

Duchêne P, Zivkovic D, Hutter S, Stephan W, Laurent S. 2013. Demographic inference reveals African and European admixture in the North American Drosophila melanogaster population. Genetics 193:291–301.

Duvernell DD, Schmidt PS, Eanes WF. 2003. Clines and adaptive evolution in the methuselah gene region in Drosophila melanogaster. Molecular ecology 12(5):1277-85.

Endler JA. 1977. Geographic variation, speciation, and clines. Monogr. Popul. Biol. 10:1–246.

Ewing EP. 1979. Genetic Variation in a Heterogeneous Environment VII. Temporal and Spatial Heterogeneity in Infinite Populations. Am. Nat. 114:197–212.

Excoffier L, Foll M, Petit RJ. 2009. Genetic Consequences of Range Expansions. Annu. Rev. Ecol. Evol. Syst. 40:481–501.

Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, Flatt T. 2012. Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America. Mol. Ecol. 21:4748–4769.

Feder AF, Petrov DA, Bergland AO. 2012. Ldx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. PLoS One 7:e48588.

Fisher RA. 1992. Statistical Methods for Research Workers. In: Kotz S, Johnson NL, editors. Breakthroughs in Statistics: Methodology and Distribution. New York, NY: Springer New York. p. 66–70.

Flatt T. 2016. Genomics of clinal variation in Drosophila: disentangling the interactions of selection and demography. Molecular ecology 25(5):1023-6.

García-Vázquez E, Sánchez-Refusta F. 1988. Chromosomal polymorphism and extra bristles of Drosophila melanogaster: joint variation under selection in isofemale lines. Genetica 78:91–96.

Gillespie J. 1973. Polymorphism in random environments. Theor. Popul. Biol. 4:193–195.

Gramates LS, Marygold SJ, Santos GD, Urbano J-M, Antonazzo G, Matthews BB, Rey AJ, Tabone CJ, Crosby MA, Emmert DB, et al. 2017. FlyBase at 25: looking to the future. Nucleic Acids Res. 45:D663–D671.

Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. 2008. Adaptations to climate in candidate genes for common metabolic disorders. PLoS Genet. 4:e32.

Hoffmann AA, Anderson A, Hallas R. 2002. Opposing clines for high and low temperature resistance in Drosophila melanogaster. Ecol. Lett. 5:614–618.

Hoffmann AA, Scott M, Partridge L, Hallas R. 2003. Overwintering in Drosophila melanogaster: outdoor field cage experiments on clinal and laboratory selected populations help to elucidate traits under selection. J. Evol. Biol. 16:614–623.

Hoffmann AA, Weeks AR. 2007. Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. Genetica 129:133–147.

de Jong G, Bochdanovits Z. 2003. Latitudinal clines Drosophila melanogaster: Body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. J. Genet. 82:207–223.

Jabot, F., Faure, T., & Dumoulin, N. 2013. Easy ABC: performing efficient approximate Bayesian computation. Coopresampling schemes using R. Methods in Ecology and Evolution. 4(7), 684-687.

Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. 2015. Population genomic analysis uncovers African and European admixture in Drosophila melanogaster populations from the south-eastern United States and Caribbean Islands. Mol. Ecol. 24:1499–1509.

Kapun M, Fabian DK, Goudet J, Flatt T. 2016. Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. Mol. Biol. Evol. 33:1317–1336.

Kapun M, van 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.
Kapun, Martin, and Thomas Flatt. The Adaptive Significance of Chromosomal Inversion Polymorphisms in Drosophila Melanogaster. Molecular Ecology 28, no. 6 (2019): 1263–82.

Keller A. 2007. Drosophila melanogaster’s history as a human commensal. Curr. Biol. 17:R77–81.

Kennington WJ, Killeen JR, Goldstein DB, Partridge L. 2003. Rapid laboratory evolution of adult wing area in Drosophila melanogaster in response to humidity. Evolution 57:932–936.

Knibb WR. 1982. Chromosome inversion polymorphisms in Drosophila melanogaster II. Geographic clines and climatic associations in Australasia, North America and Asia. Genetica 58:213–221.

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, Schlötterer C. 2012. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. Bioinformatics 28:2084–2085.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. 2011. Genomic differentiation between temperate and tropical Australian populations of Drosophila melanogaster. Genetics 187:245–260.

Kooyers NJ, Greenlee AB, Colicchio JM, Oh M, Blackman BK. 2015. Replicate altitudinal clines reveal that evolutionary flexibility underlies adaptation to drought stress in annual Mimulus guttatus. New Phytol. 206:152–165.

Krimbas CB, Powell JR. 1992. Drosophila inversion polymorphism. CRC press.

Lavington E, Cogni R, Kuczynski C, Koury S, Behrman EL, O’Brien KR, Schmidt PS, Eanes WF. 2014. A small system—high-resolution study of metabolic adaptation in the central metabolic pathway to temperate climates in Drosophila melanogaster. Mol. Biol. Evol. 31:2032–2041.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595.

Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet. 2:e166.

Lynch M, Bost D, Wilson S, Maruki T, Harrison S. 2014. Population-genetic inference from pooled-sequencing data. Genome Biol. Evol. 6:1210–1218.

Machado HE, Bergland AO, O’Brien KR, Behrman EL, Schmidt PS, Petrov DA. 2016. Comparative population genomics of latitudinal variation in Drosophila simulans and Drosophila melanogaster. Mol. Ecol. 25:723–740.

Machado HE, Bergland AO, Taylor R, Tilk S, Behrman E, Dyer K, Fabian DK, Flatt T, González J, Karasov TL, et al. 2018. Broad geographic sampling reveals predictable and pervasive seasonal adaptation in Drosophila. Available from: http://dx.doi.org/10.1101/337543

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The Drosophila melanogaster Genetic Reference Panel. Nature 482:173–178.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Mettler LE, Voelker RA, Mukai T. 1977. Inversion Clines in Populations of DROSOPHILA MELANOGASTER. Genetics 87:169–176.

Mitrovski P, Hoffmann AA. 2001. Postponed reproduction as an adaptation to winter conditions in Drosophila melanogaster: evidence for clinal variation under semi-natural conditions. Proc. Biol. Sci. 268:2163–2168.

Oakeshott JG, Gibson JB, Anderson PR, Knibb WR, Anderson DG, Chambers GK. 1982. Alcohol dehydrogenase and glycero-3-phosphate dehydrogenase clines in Drosophila melanogaster on different continents. Evolution 36:86–96.

Paaby AB, Bergland AO, Behrman EL, Schmidt PS. 2014. A highly pleiotropic amino acid polymorphism in the Drosophila insulin receptor contributes to life-history adaptation. Evolution 68:3395–3409.

Paaby AB, Blacket MJ, Hoffmann AA, Schmidt PS. 2010. Identification of a candidate adaptive polymorphism for Drosophila life history by parallel independent clines on two continents. Mol. Ecol. 19:760–774.

Pavlidis P, Jensen JD, Stephan W, Stamatakis A. 2012. A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Mol. Biol. Evol. 29:3237–3248.

Pool JE. The mosaic ancestry of the Drosophila genetic reference panel and the D. melanogaster reference genome reveals a network of epistatic fitness interactions. 2015. Molecular biology and evolution. 32(12):3236-51.

Rajpurohit S, Hanus R, Vrkošlav V, Behrman EL, Bergland AO, Petrov D, Čvačka J, Schmidt PS. 2017. Adaptive dynamics of cuticular hydrocarbons in Drosophila. J. Evol. Biol. 30:66–80.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. Available from: https://www.R-project.org/

Reinhardt JA, Kolaczkowski B, Jones CD, Begun DJ, Kern AD. 2014. Parallel Geographic Variation in Drosophila melanogaster. Genetics 197:361–373.
Rhomberg LR, Singh RS. 1986. Evidence for a link between local and seasonal cycles in gene frequencies and latitudinal gene clines in a cyclic parthenogen. Genetics 87(1):73-9.

Roff D. 1980. Optimizing development time in a seasonal environment: the ‘ups and downs’ of clinal variation. Oecologia 45(2):208-2.

Rudman, Seth M., Sharon Greenblum, Rachel C. Hughes, Subhash Rajpurohit, Ozan Kiratli, Dallin B. Lowder, Skyler G. Lemmon, Dmitri A. Petrov, John M. Chaston, and Paul Schmidt. 2019. “Microbiome Composition Shapes Rapid Genomic Adaptation of Drosophila Melanogaster.” BioRxiv, 632257.

Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J, With KA, Baughman S, Cabin RJ, Cohen JE, Ellstrand NC, et al. 2001. The Population Biology of Invasive Species. Ann. Rev. Ecol. Syst. 32:305–332.

Schmidt PS, Conde DR. 2006. Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in Drosophila melanogaster. Evolution 60:1602–1611.

Schmidt PS, Duvernell DD, Eanes WF. 2000. Adaptive evolution of a candidate gene for aging in Drosophila. Proc. Natl. Acad. Sci. U. S. A. 97:10861–10865.

Schmidt PS, Matzkin L, Ippolito M, Eanes WF, Hey J. 2005. Geographic variation in diapause incidence, life-history traits, and climatic adaptation in Drosophila melanogaster. Evolution 59:1721–1732.

Schmidt PS, Paaby AB. 2008. Reproductive diapause and life-history clines in North American populations of Drosophila melanogaster. Evolution 62:1204–1215.

Schrider DR, Hahn MW, Begun DJ. 2016. Parallel evolution of copy-number variation across continents in Drosophila melanogaster. Mol. Biol. Evol. 33:1308–1316.

Sezgin E, Duvernell DD, Matzkin LM, Duan Y, Zhu C-T, Verrelli BC, Eanes WF. 2004. Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in Drosophila melanogaster. Genetics 168:923–931.

Singh RS, Rhomberg LR. 1987a. A Comprehensive Study of Genic Variation in Natural Populations of Drosophila melanogaster. II. Estimates of Heterozygosity and Patterns of Geographic Differentiation. Genetics 117:255–271.

Singh RS, Rhomberg LR. 1987b. A Comprehensive Study of Genic Variation in Natural Populations of Drosophila melanogaster. I. Estimates of Gene Flow from Rare Alleles. Genetics 115:313–322.

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

Turner TL, Levine MT, Eckert ML, Begun DJ. 2008. Genomic analysis of adaptive differentiation in Drosophila melanogaster. Genetics 179:455–473.

Vasemägi A. 2006. The adaptive hypothesis of clinal variation revisited: single-locus clines as a result of spatially restricted gene flow. Genetics 173:2411–2414.

Verrelli BC, Eanes WF. 2001. Clinal variation for amino acid polymorphisms at the Pgm locus in Drosophila melanogaster. Genetics 157(4):1649-63.

Vigue CL, Johnson FM. 1973. Isozyme variability in species of the genus Drosophila. VI. Frequency-property-environment relationships of allelic alcohol dehydrogenases in D. melanogaster. Biochemical genetics 9(3):213-27.

H. Wickham. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wittmann MJ, Bergland AO, Feldman MW, Schmidt PS, Petrov DA. 2017. Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift. Proceedings of the National Academy of Sciences 114(46): E9932-41.

Wright S. 1943. Isolation by distance. Genetics 28:114–138.

Wright S, Dobzhansky T. 1946. Genetics of natural populations; experimental reproduction of some of the changes caused by natural selection in certain populations of Drosophila pseudoobscura. Genetics 31:125–156.

Yukilevich, Roman, and John R. True. 2008. “African Morphology, Behavior and Phermones Underlie Incipient Sexual Isolation between Us and Caribbean Drosophila Melanogaster.” Evolution; International Journal of Organic Evolution 62 (11): 2807–28.

Yukilevich R, Turner TL, Aoki F, Nuzhdin SV, True JR. 2010. Patterns and processes of genome-wide divergence between North American and African Drosophila melanogaster. Genetics 186:219–239.

Zuther E, Schulz E, Childs LH, Hincha DK. 2012. Clinal variation in the non-acclimated and cold-acclimated freezing tolerance of Arabidopsis thaliana accessions. Plant Cell Environ. 35:1860–1878.
Supplementary information

Figure S1. Distribution of Durbin-Watson P-values. We tested whether there is autocorrelation in the residuals of the seasonal generalized linear models.

Figure S2. Enrichment of top 5% SNPs in each genic class for A) latitudinal P-value, B) seasonal P-value. Histograms show the distribution of odds ratios when season labels were permuted, and the vertical bars indicated the observed odds ratio.
Figure S3. Relationship between clinal and seasonal change using A) a more stringent minor allele frequency cutoff, B) latitudinal samples from Maine, which were collected in the summer and C) samples from Australia to calculate latitudinal change.

Figure S4. Relationship between proportion of points linearly related (in comparison to points that are not related) and the expected relationship between two z-normalized variables. Each color represents a different assumed degree of linear relatedness (ranging from -1 to -0.1). Note how for a given estimated relationship between two variables, the slope we assume...
(-1) results in the smallest possible proportion of points linearly related, demonstrating that our test is conservative.

Figure S5. Rate of allele frequency differences between two populations (normalized by difference in latitude). Note the rates are smaller for comparisons involving the samples from Maine. FL: Florida (July 2008 and 2010), ME: Maine (October 2009), MA: Massachusetts (July 2012), VA: Virginia (July 2012).
Table S1. Information of the samples used in this study.

| Accession # | Sample name | Location          | Latitude | Collection date | # Flies | Median depth | Month | Season | Seasonal set | Clinical set |
|-------------|-------------|-------------------|----------|-----------------|---------|--------------|-------|--------|--------------|--------------|
| SRR11 77951 | AU_LO1      | Queensland, Australia | 16.90    | 2004            | 17      | 10           |       |        |              |              |
| SRR11 77952 | AU_LO2      | Queensland, Australia | 16.90    | 2004            | 17      | 12           |       |        |              |              |
| SRR11 77953 | AU_HI1      | Tasmania, Australia  | 42.77    | 2004            | 15      | 10           |       |        |              |              |
| SRR11 77955 | AU_HI2      | Tasmania, Australia  | 42.77    | 2004            | 15      | 14           |       |        |              |              |
| SRR15 25685 | FL1         | Homestead, FL       | 25.47    | Jul-08          | 39      | 59           | 7     | Spring | 0            | 1            |
| SRR15 25694 | FL2         | Homestead, FL       | 25.47    | Jul-10          | 48      | 37           | 7     | Spring | 0            | 1            |
| SRR15 25698 | ME1         | Bowdoinham, ME      | 44.02    | Oct-09          | 75      | 86           | 9     | Fall   | 0            | 0            |
| SRR20 60283 | ME2         | Bowdoinham, ME      | 44.02    | Oct-09          | 75      | 22           | 9     | Fall   | 0            | 0            |
| SRR15 25695 | GA          | Hahira, GA          | 30.99    | Jul-10          | 51      | 101          | 7     | Spring | 0            | 1            |
| SRR15 25696 | SC          | Eutawville, SC      | 33.40    | Jul-10          | 48      | 83           | 7     | Spring | 0            | 1            |
| SRR35 90551 | VA_07_012   | Charlottesville, VA  | 38.03    | Jul-12          | 69      | 70           | 7     | Spring | 0            | 1            |
| SRR39 39095 | PA_06_013   | Linvilla, PA        | 39.88    | Jun-13          | 54      | 37           | 6     | Spring | 0            | 1            |
| SRR35 90557 | MA_07_2012  | Lancaster, MA       | 42.46    | Jul-12          | 90      | 51           | 7     | Spring | 0            | 1            |
| SRR15 25768 | PA_07_009   | Linvilla, PA        | 39.53    | Jul-09          | 55      | 186          | 7     | Spring | 1            | 0            |
| SRR15 25769 | PA_11_009   | Linvilla, PA        | 39.53    | Nov-09          | 74      | 66           | 11    | Fall   | 1            | 0            |
| SRR15 25770 | PA_07_010   | Linvilla, PA        | 39.53    | Jul-10          | 11      | 17           | 7     | Spring | 1            | 0            |
| SRR15 25771 | PA_11_010   | Linvilla, PA        | 39.53    | Nov-10          | 33      | 76           | 11    | Fall   | 1            | 0            |
| SRR15 25772 | PA_07_011   | Linvilla, PA        | 39.53    | Jul-11          | 75      | 53           | 7     | Spring | 1            | 0            |
| SRR15 25773 | PA_10_011   | Linvilla, PA        | 39.53    | Oct-10          | 47      | 74           | 10    | Fall   | 1            | 0            |
| SRR35 90560 | PA_10_012   | Linvilla, PA        | 39.53    | Oct-12          | 10      | 25           | 10    | Fall   | 1            | 0            |
| SRR35 90561 | PA_07_012   | Linvilla, PA        | 39.53    | Jul-12          | 11      | 59           | 7     | Spring | 1            | 0            |
| SRR35 90563 | PA_9_20     | Linvilla, PA        | 39.53    | Sep-12          | 50      | 55           | 9     | Fall   | 1            | 0            |
### Table S2. Summary of a regression of z-normalized latitude regression coefficients against z-normalized season regression coefficients genome-wide. CI stands for 95% confidence interval.

| Predictors | Estimates | CI       | p   |
|------------|-----------|----------|-----|
| Intercept  | 0.013     | 0.011 – 0.015 | <0.001 |
| Seasonal slope | -0.039 | -0.041 – -0.037 | <0.001 |

Observations 797792

### Table S3. Summary of a regression of z-normalized latitude regression coefficients against z-normalized season regression coefficients for each genic class. CI stands for 95% confidence interval.

| Predictors         | Estimates | CI       | p   |
|--------------------|-----------|----------|-----|
| Intercept          | 0.013     | 0.011 – 0.015 | <0.001 |
| Seasonal slope: Intergenic | -0.029 | -0.037 – -0.021 | <0.001 |
| Seasonal slope: Exon | -0.050 | -0.057 – -0.043 | <0.001 |
| Seasonal slope: Intron | -0.045 | -0.050 – -0.041 | <0.001 |
| Seasonal slope: Downstream | -0.036 | -0.042 – -0.031 | <0.001 |
| Seasonal slope: Upstream | -0.035 | -0.039 – -0.032 | <0.001 |
| Seasonal slope: UTR_3 | -0.045 | -0.055 – -0.035 | <0.001 |
| Seasonal slope: UTR_5 | -0.044 | -0.057 – -0.031 | <0.001 |
Table S4. Summary of a regression of z-normalized latitude regression coefficients against z-normalized season regression coefficients for each inversion status. CI stands for 95% confidence interval.

| Latitude slope | Predictors   | Estimates | CI           | p     |
|----------------|--------------|-----------|--------------|-------|
|                | Intercept    | 0.013     | 0.011 – 0.015| <0.001|
|                | Seasonal slope:Outside | -0.039 | -0.041 – -0.036 | <0.001|
|                | Seasonal slope:Inside | -0.174 | -0.607 – 0.258 | 0.429 |
| Observations   | 797792       |           |              |       |

Table S5. Summary of a regression of z-normalized latitude regression coefficients against z-normalized season regression coefficients for each SNPs surrounding inversion breakpoints. CI stands for 95% confidence interval.

| Latitude slope | Predictors   | Estimates | CI           | p     |
|----------------|--------------|-----------|--------------|-------|
|                | Intercept    | 0.013     | 0.011 – 0.015| <0.001|
|                | Seasonal slope:In(2Lt) | 0.004 | -0.003 – 0.010 | 0.248 |
|                | Seasonal slope:In(2R)NS | -0.019 | -0.026 – -0.012 | <0.001|
|                | Seasonal slope:In(3L)P | -0.068 | -0.074 – -0.061 | <0.001|
|                | Seasonal slope:In(3R)P | -0.170 | -0.177 – -0.163 | <0.001|
|                | Seasonal slope:Outside | -0.030 | -0.033 – -0.028 | <0.001|
| Observations   | 797792       |           |              |       |