Murillo Fernando Rodrigues

# Clinas neutras ou adaptativas? Integrando variação genômica clinal e sazonal para inferir seleção natural em *Drosophila melanogaster*

Adaptive or neutral clines? Integrating genome-wide clinal and seasonal variation to infer natural selection in *Drosophila melanogaster*

São Paulo

2018

Murillo Fernando Rodrigues

# Clinas neutras ou adaptativas? Integrando variação genômica clinal e sazonal para inferir seleção natural em *Drosophila melanogaster*

Adaptive or neutral clines? Integrating genome-wide clinal and seasonal variation to infer natural selection in *Drosophila melanogaster*

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências, na área de Genética e Biologia Evolutiva.
Orientador: Prof. Dr. Rodrigo Cogni
Co-orientadora: Prof. Dra. Maria Vibranovski

São Paulo

2018

# Ficha catalográfica

Rodrigues, Murillo Fernando
Clinas neutras ou adaptativas? Integrando variação genômica clinal e sazonal para inferir seleção natural em *Drosophila melanogaster*

56p

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Adaptação 2. Adaptação sazonal 3. *Drosophila melanogaster* 4. Seleção natural 5. Variação molecular

I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

# Comissão Julgadora

_____ _____

Prof(a). Dr(a). Prof(a). Dr(a).

_____ _____

Prof(a). Dr(a). Prof. Dr. Rodrigo Cogni

# Agradecimentos

Agradeço ao meu orientador, Rodrigo Cogni, por dedicar inúmeras horas ao meu desenvolvimento acadêmico e profissional. Obrigado por propiciar um ambiente acadêmico estimulante, pela paciência e por me ajudar a persistir mesmo em momentos de adversidade. Agradeço à minha co-orientadora, Maria Vibranovski, por sempre estar disposta a me ajudar e por suas sempre pertinentes contribuições.

Aos meus fantásticos colegas de laboratório, Camila, André e Marcos, pelo apoio e ajuda durante os últimos anos. Meu trabalho não teria sido o mesmo sem as contribuições de todos vocês! Obrigado Arthur e Ana, por toda ajuda com as mosquinhas! Agradeço também aos meus colegas de *Journal Club* pelas discussões de excelentes artigos em Evolução.

Agradeço a todos meus amigos e colegas de pós-graduação da Genética e da Ecologia. Sem o ombro amigo de muitos de vocês, meu tempo no mestrado teria sido muito menos divertido e prazeroso. Também, aprendi muito interagindo com todos vocês!

Um especial agradecimento a todos os professores e funcionários do IB/USP. O IB é um ambiente excelente para um biólogo em formação graças a todos vocês! Agradeço aos membros da minha banca de qualificação, Prof. Dr. Carlos Vilela, Dra. Daniela Rossoni e Profa. Dra. Tatiana Torres. Obrigado ao Gustavo e Lenilda, por sempre me ajudar quando precisei.

Thank you to Professor John Pool (UW-Madison), for hosting me for six months in your lab! I learned a lot from you and your group, and I am extremely thankful for you all making me feel welcome. I thank Chris, Jeremy, Quentin, Tiago and Yuheng for being awesome labmates. Thank you for helping me out during my period there. I am

extremely thankful to Ellen, for helping me with my project. Thanks Yuheng for our conversations over lunch, you are a great friend! Special thanks to Tiago and Guilherme, for being great friends and instrumental to my well being while I stayed in Madison! I am also thankful to the other graduate students, postdocs and faculty I interacted with during my time at UW-Madison. Thank you Professor Nicole Perna and Dr. Jeremy Glasner for letting me use the microbiology lab and helping me with basic microbiology techniques. Thank you Brenda and family for welcoming me to Madison, and showing me how great midwesterners can be!

Muito obrigado a todos meus amigos e colegas de graduação, que, de muitas maneiras, me tornaram um biólogo e uma pessoa melhor. Em especial, agradeço aos Queridos da Bio, Bunni, Carol, Gabriel, Isa e Nat, por todos momentos especiais que dividimos ao longo dos últimos seis anos. Tem sido ótimo crescer junto com vocês! Carol e Luiza, muito obrigado também por revisarem a dissertação.

Sou imensamente grato à minha namorada, Luiza Ostrowski, por me dar alegrias, carinho e companhia. Você me faz uma pessoa melhor, e contribuiu muito para que eu conseguisse enfrentar esse desafio!

Agradeço de coração toda minha família pelo apoio durante meus anos de graduação e mestrado. Em especial, agradeço minha mãe e meu pai pelo apoio e amor incondicional! Também, agradeço pelo apoio financeiro. Infelizmente, ciência não é prioridade no nosso país e, em anos recentes, isso tem ficado cada vez mais claro.

Agradeço, por fim, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), pelo financiamento da minha bolsa de mestrado e bolsa de estágio à pesquisa no exterior. Processos nºs 2016/01354-9 e 2017/06374-0.

# Contents

## Resumo

Variação espacial e temporal são ubíquas. Caracteres são chamados clinais quando variam ao longo de um gradiente ambiental, e isso é interpretado como resultado de seleção estruturada no espaço. *Drosophila melanogaster* apresenta clinas em diversos caracteres fenotípicos e genotípicos, as quais são replicadas em diferentes regiões do mundo. Estudos recentes sugeriram que grande parte da variação clinal pode ser atribuída a processos neutros. Como o ambiente varia de maneira similar com a latitude e ao longo do ano, e essa variação sazonal é ortogonal à demografia, uma abordagem possível para inferir seleção natural é integrar variação clinal e sazonal. Neste trabalho, nós testamos se há uma relação entre variação clinal e sazonal ao longo do genoma de *D. melanogaster*. Também, investigamos a proporção de variantes que deveriam estar sob seleção espacial e sazonal para explicar o padrão encontrado. Estimamos a frequência alélica a partir de amostras de *pools* de moscas coletadas em diferentes oito localidades ao longo da costa leste dos Estados Unidos e em diferentes estações do ano na Pensilvânia, EUA. Nós encontramos um padrão genômico de variação clinal refletindo variação sazonal. Esse padrão é mais forte para variantes em regiões exônicas do que intergênicas, consistente com a ação de seleção. A relação entre variação clinal e sazonal encontrada pode ser explicada se 6,6% dos polimorfismos estiverem sob seleção espacial e sazonal. Nossos resultados são consistentes com a hipótese adaptativa de variação clinal e, junto com outras observações, revelam que o papel da demografia na manutenção de clinas em *D. melanogaster* é limitado.

## Abstract

Spatial and temporal variation in the environment are ubiquitous. Traits are called clinal when they vary along an environmental gradient, and this is often interpreted as the result of spatially varying selection. *Drosophila melanogaster* is known to have many phenotypic and genotypic clines, replicated in many regions of the world. Recent studies have suggested that most clinal variation could be attributed to neutral, demographic processes. Because the environment varies in similar ways with latitude and across seasons, and seasonal variation is orthogonal to demography, one promising approach is to integrate clinal and seasonal variation to infer selection. Here, we test whether there is a genome-wide relationship between clinal and seasonal variation, and whether the pattern is consistent with selection. Also, we investigate the proportion of the variants that should be under latitudinal and seasonal selection to explain the pattern we uncovered. We estimate allele frequency from pooled samples of flies from eight different locations along the east coast of the US, and 13 samples collected in the spring and in the fall in Pennsylvania. We show that there is a genome-wide pattern of clinal variation mirroring seasonal variation. This pattern is stronger for exonic when compared to intergenic regions, consistent with natural selection. We find that the genome-wide relationship between clinal and seasonal variation could be explained by about 6.6% of our SNPs being under latitudinal and seasonal selection. Our results are consistent with the adaptive hypothesis of clinal variation and, together with other observations, leave little room for the role of demography in maintaining clines in *D. melanogaster*.

# Introduction

All species are subject to environmental variation structured through space and time. Different environments can impose various selective regimes on populations, resulting in adaptive differentiation. With spatially varying selection, one allele can be beneficial in one environment and disadvantageous in another (Levene 1953). Similarly, selective pressures can cycle through time, and one allele may be alternatively favored and disfavored (Gillespie 1973). Thus, both spatially and temporally varying selection can result in different forms of local adaptation (Ewing 1979).

Spatial heterogeneity generates a pattern of geographical variation in traits that affect fitness (Kawecki and Ebert 2004). Some species can be found along gradually changing environments, and traits that vary with the environment are called clinal. Examples of clinal variation are abundant in many different taxa, such as thale cress (Zuther et al. 2012), monkeyflowers (Kooyers et al. 2015), ivyleaf morning glories (Campitelli and Stinchcombe 2013), Atlantic salmons (Dionne et al. 2007), vervet monkeys (Cardini et al. 2007) and even humans (Hancock et al. 2008).

*Drosophila melanogaster* has been used as a model to understand latitudinal variation because, besides being a genetic model, 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), is thought to have resulted in several climatic adaptations (David and Capy 1988).

Clinal variation has been documented for this species 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 (Schmidt et al. 2005). At the genetic level, latitudinal clines have been identified for allozymes, chromosome inversions 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).

These clinal traits in *D. melanogaster* are usually thought to be a result of spatially varying selection (Endler 1977). Although a covariation between an environmental variable and a trait is suggestive of natural selection, 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). Because latitudinal clines are often repeated (*e.g.*, across continents), a classical approach to discern between adaptive and neutral differentiation has been to look at parallel clinal variation (Hoffmann et al. 2002; Hoffmann and Weeks 2007; Turner et al. 2008; Paaby et al. 2010; Reinhardt et al. 2014; Schrider et al. 2016). If a trait varies clinally in two or more continents, a simple, plausible scenario is latitudinally varying selection acting on ancestral variation (Endler 1977; Barton 1983; Barton 1999).

A recent study showed that parallel clinal variation between continents could be a result of secondary contact among previously diverged populations of *D. melanogaster* (Bergland et al. 2016). It appears that part of the clinality observed in flies from the east coast of North America and Australia can be attributed to migration of European flies to the high latitude end of the cline and migration of African flies to low latitude locations.

As a consequence, it may be impossible to distinguish adaptive from neutral clines just by looking at patterns of clinal variation.

To reject some of these non-adaptive hypothesis, one would need access to population parameters that are hard to measure or estimate, such as the distribution of fitness effects, landscape of recombination rate, rates of migration, population sizes, changes in population size and historical founding of the populations. Alternatively, one approach to disentangle adaptive from non-adaptive processes is by exploring signatures of parallel variation between latitude and seasons (Cogni et al. 2015).

*Drosophila* has also been widely used in the study of seasonal adaptation. One of the earliest examples of seasonal 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). These studies were either conducted in a common laboratory condition or as a field based mesocosm experiment. Furthermore, genome-wide analysis have identified polymorphisms that oscillate in seasonal timescales in several localities in the United States and Europe (Bergland et al. 2014; Machado et al. 2018).

Many environmental variables that vary along a latitudinal transect vary in a similar fashion through seasons. Processes that can generate clinal patterns like isolation by distance and secondary contact from diverged populations are independent to variation across seasons. Therefore, traits that vary both with latitude and through seasons are likely under natural selection. For instance, 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.

Very few studies have tried to identify parallel clinal and seasonal variation in *D. melanogaster* (Bergland et al. 2014; Cogni et al. 2015; Machado et al. 2018). The association between clinal and seasonal change has been identified in central metabolic genes, which is likely caused by parallel climatic factors driving adaptation in these genes (Cogni et al. 2015). This study focused on a limited number of polymorphisms in metabolic genes. Bergland et al. (2014) found an indicative that clinal variants are more likely to be seasonal. A very recent study observed that seasonal changes mirrors latitudinal changes in *D. melanogaster* with the intent of validating seasonal outliers (Machado et al. 2018), using few seasonal samples from many different locations.

Here, we aim to answer whether the clinal patterns observed in *D. melanogaster* are consistent with the action of natural selection. We do so by looking at parallel clinal and seasonal variation across the genome. Because many environmental variables such as temperature, UV radiation and resource availability vary similarly across latitude and through seasons in a temperate environment, we hypothesize there is a relationship between clinal and seasonal variation. Flies collected in the north should be more similar to flies collected in early spring, whereas southern flies should be more like fall flies. We also tested whether the pattern of clinal mirroring seasonal variation is consistent with selection. Finally, by looking at the relationship between clinal and seasonal variation, we asked how much of the genome is under spatial and seasonal selection. Integrating these two independent sources of evidence can help us disentangle adaptive from non-adaptive processes that contribute to the pervasive and long standing patterns of clinal variation in *D. melanogaster*.

# Material and Methods

## Population samples

We analyzed 21 samples from eight locations along the United States east coast, collected by (Bergland et al. 2014) (10 samples), and (Machado et al. 2018) (11 samples) (see Table S1). Most samples were based on pools of wild-caught individuals. For the Raleigh (North Carolina) sample, one male for each of 92 DGRP inbred strains was pooled and sequenced (Mackay et al. 2012). 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 (Fabian et al. 2012; Bergland et al. 2014). Including these samples would impact our estimates of clinal variation, given that fall flies are more south-like (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. For a more detailed description of sampling, sample preparation and sequencing see Bergland et al. (2014) and Machado et al. (2018). All the data used in this project are available on the NCBI Short Read Archive (BioProject accession numbers PRJNA256231 and PRJNA308584).

## Mapping and processing of sequencing data

Raw, paired-end reads were mapped against the FlyBase *D. melanogaster* reference genome r6.15 (Gramates et al. 2017) 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 10%, with minimum per-population coverage less than 10x or maximum per-population coverage greater than 400x were excluded from our analyses. Also, SNPs with multi-allelic states were removed from the analysis. Because only male flies were sequenced, SNPs identified in the X chromosome had lower coverage, so these SNPs were removed from our analyses. 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 have an additional component of error inherent to them due to the round of additional sampling (Kofler et al. 2011; Lynch et al. 2014). Therefore, it is necessary to model this additional error to avoid biases in our analysis. We decided to do so by incorporating weights proportional to the effective number of chromosomes ($N_E$):

$$N_E = (\frac{1}{R} + \frac{1}{C})^{-}1$$

where $C$ is the number of chromosomes in the pool and $R$ is the read depth at that site (Kolaczkowski et al. 2011; Feder et al. 2012; Bergland et al. 2014; Machado et al. 2016).

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

$$y_i = \alpha + \beta Latitude + \epsilon_i$$

Where $y_i$ is the expected allele frequency in the $i^{th}$ population, $\alpha$ is the intercept, $\beta$ is the slope, $Latitude$ is the latitude of the population, and $\epsilon_i$ is the binomial error given the $N_E$ at that SNP. 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,

$$y_i = \alpha + \beta Season + \epsilon_i$$

where $y_i$ is the observed allele frequency in the $i^{th}$ sample, $\alpha$ is the intercept, $\beta$ is the slope, $Season$ is the season of the sample, 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. Thus, we considered that SNPs were outliers if their regression P-value fell in the bottom 1% of the distribution. All statistical analyses were performed in R 3.5.0 (R 2018).

## Relationship between clinal and seasonal variation

We followed a similar approach to that of Cogni et al. (2015). We reasoned that as clinal variation can be a result of neutral processes, such as isolation-by-distance and admixture (Vasemägi 2006; Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016), an independent parallel evaluation of the action of natural selection is needed. Because environmental factors that vary with latitude also vary with seasons, SNPs that vary both clinally and seasonally are most likely being affected by natural selection. A warm-

favored allele will increase in frequency from spring to fall (positive seasonal slope), whereas it will decrease in frequency with higher latitudes (negative clinal slope). The inverse would be true for a cold-favored allele. Therefore, we predicted that, if selection is pervasive, there will be a negative relationship between clinal and seasonal slopes. We also reasoned that this pattern should be stronger for regions of the genome where sites are more likely to be advantageous, such as exonic regions (Andolfatto 2005).

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

$$\beta_i^{clinal} = \alpha + \beta\beta_i^{seasonal} + \epsilon_i$$

where $\beta_i^{clinal}$ is the estimated clinal slope for the $i^{th}$ SNP, $\alpha$ is the intercept, $\beta$ is the slope, $\beta_i^{seasonal}$ is the estimated seasonal slope for the $i^{th}$ SNP, and $\epsilon_i$ is the gaussian error at that SNP.

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:

$$\beta_i^{clinal} = \alpha + \beta\beta_i^{seasonal} + \gamma_j\beta_i^{seasonal} \times Class_j + \epsilon_i$$

where $\gamma_j$ is the regression coefficient for the $j^{th}$ genic class, $Class_j$ is the genic class of the $i^{th}$ SNP. Furthermore, we performed a Tukey test to find differences in the relationship between clinal and seasonal variation among all genic classes.

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 inside the only common

inversion in our seasonal samples, In(2L)t (Kapun et al. 2016). To do so, we performed a regression similar to our previous one, that takes the form:

$$\beta_i^{clinal} = \alpha + \beta\beta_i^{seasonal} + \gamma_j\beta_i^{seasonal} \times Inversion_j + \epsilon_i$$

where $Inversion_j$ is the inversion status (inside or outside In(2L)t), and $\gamma_j$ is the regression coefficient for the $j^{th}$ status.

## Composite statistic for clinal and seasonal variation

Our approach uses the relationship between clinal and seasonal variation in sets of SNPs to infer selection, and so it is not possible to detect individual SNPs which are under spatial and temporal selection. We developed a composite statistic based on Fisher's method (Fisher 1992) that takes into account clinal and seasonal regression P-values, and the sign of the clinal and seasonal slopes:

$$\chi_i = -2sgn(\beta_i^{clinal} \times \beta_i^{seasonal})(ln(P_i^{clinal}) + ln(P_i^{seasonal}))$$

where $P_i^{clinal}$ is the P-value of the clinal regression and $P_i^{seasonal}$ is the P-value of the seasonal regression. When both P-values are small and the sign of the slopes are opposite, the statistic $\chi$ will be negative in sign and have a large magnitude, suggesting that natural selection is acting on that polymorphism. We defined as outliers those in the bottom 1% of the $\chi$ distribution.

## Enrichment tests

We tested for enrichment of genic classes using our sets of clinal, seasonal, and composite SNPs. We implemented binomial generalized linear mixed models to examine the relationship between being top SNPs for a given statistic and genic class (fixed effect), while controlling for chromosome, recombination rate, and MAF quantile bin (random effects).

We also performed Gene Ontology (GO) enrichment analysis to identify the biological function underlying our sets of outliers (Ashburner et al. 2000). One common issue with GO analysis is that they are biased, as longer genes are likely to harbor more SNPs. We thus used Gowinda (Kofler and Schlötterer 2012), an unbiased (*i.e.*, that accounts for gene length) permutation approach to GO analysis. We obtained annotations from Flybase r6.15 and used the following parameters in Gowinda: 100,000 simulations; a minimum of 1 gene per GO category, and considered SNPs 1000bp up and downstream of a gene.

## Controlling for linked selection

Selection at one site affects genetic variation at linked neutral sites (Smith and Haigh 1974). Because we assume that sites are independent in our models, it is possible that linked selection might explain a great portion of the patterns we investigated. We test the effect of linked selection by implementing a block bootstrap approach. Sampling one SNP per 500bp one thousand times, we constructed sets of SNPs with minimized dependency. For each of these sets, we regressed clinal against seasonal slopes, and compared the distribution of the block-bootstrapped regression coefficients to the coefficients we obtained using all SNPs.

We also investigated whether the relationship between clinal and seasonal variation is affected by local recombination rate. We added an interaction term between seasonal slope and crossover rate to our model, using crossover rates estimated over 100kb windows from Comeron et al. (2012). To help us tease apart the relationship between recombination rate and linked selection, we compared the regression coefficient for the interaction using all SNPs and using our block-bootstrapped sets of SNPs.

## Inferring proportion of selected SNPs

Although a negative, significant relationship between clinal and seasonal variation indicates that a subset of the polymorphisms is being selected, it is difficult to measure how a given slope is related to a biological parameter, such as the proportion of selected SNPs. It is likely that most of the polymorphisms are not under spatial and seasonal selection, and so the noise around our slope estimate is high. To get a sense of how randomness impacts slope estimation, we ran a simple simulation.

We formed three ($X, Y, Z$) vectors drawing numbers from a normal distribution (mean = 0, standard deviation = 1). We created a fourth vector ($W$) as a linear function of $X$ ($W = aX + b$, a = -1 and b = 0). We obtained two pairs of vectors, the independent $Y$ and $Z$ and the linearly dependent pair $X$ and $W$. We regressed the concatenated vector formed from $W$ and $Z$ against the vector consisting of $X$ and $Y$. For each simulation, we varied the relative number of elements in the linear dependent and linear independent vector pairs. Then, we fit a loess curve for the proportion of signal (number of elements in the linear dependent vector pair over the total number of elements) against the estimated regression coefficient.

We assume that for SNPs under spatial and seasonal selection, the regression coefficient for latitude is a linear combination of the regression coefficient for season. For SNPs that are not under spatial and seasonal selection, there should be no relationship between the regression coefficients for latitude and season. By calculating the expected regression between clinal and seasonal variation for datasets with different proportions of points that fall along the $y = -x$ line, we could ask which proportion of selected SNPs is consistent with the regression coefficient of clinal against seasonal variation we observed.

# Results

We assembled 21 *D. melanogaster* population samples collected from eight localities in the east coast of the United States. All of these samples are the result of a collaborative effort of many researchers from two consortiums, the DrosRTEC (Bergland et al. 2014; Machado et al. 2018) and DrosEU (Kapun et al. 2018). Eight 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 216). More details about the samples can be found on Table S1. After all the filtering steps, we identified 610,780 common autosomal SNPs, which we used in our analyses.

## Latitude explains more variation than seasons

Latitude explains much of allele frequency variation along the surveyed populations, as there is an excess of low GLM P-value SNPs (Fig. S1A). At an FDR of 0.01, 41,590 SNPs remain significant, which is approximately 6.8% of the SNPs included in our analysis. At an FDR of 0.05, 97,427 SNPs remain significant (15.9%). We found that the magnitude of the clinal regression coefficients differed among chromosomes, with the rank order being 3R>3L>2L>2R (Table S2). The chromosome arm 3R is covered by three major inversions (In(3R)P, In(3R)Mo, and In(3R)K), what might explain its strong clinality (Fabian et al. 2012; Kapun et al. 2014).

Seasons, on the other hand, explain less of the variation in the Pennsylvania samples. There is a slight excess of low GLM P-value SNPs (Fig. S1B). At an FDR of 0.01 or 0.05, no SNPs remain significant. At an FDR of 0.4, 146 SNPs are significant. The rank

order of seasonal regression coefficients per chromosome was 3L=3R>2L>2R (Table S3). Just one of the major inversions (In(2L)t) segregates at considerable frequency in Pennsylvania (Kapun et al. 2016), and so it is unlikely that inversions account for the seasonal adaptation patterns we have found.

The detection of an excess of statistically significant clinal and seasonal SNPs is subject to numerous assumptions that are most likely violated by our study design. Therefore, we sought to demonstrate that the top significant clinal and seasonal SNPs are enriched for true positive SNPs that underlie clinal and seasonal adaptation. Latitudinal SNPs are more likely to be in exonic, UTR 3', UTR 5', and splice regions (Fig. 1A). Seasonal SNPs are enriched for exonic regions (Fig. 1B). We controlled for recombination rate, minor allele frequency bin and chromosome, and P-values were FDR adjusted.

## Clinal variation mirrors seasonal variation

A clinal pattern can arise as a result of non-adaptive 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, there is much variation within seasons and across years that can impact our inferences, and even though we have samples from different years, many model violations and unknown variables are likely affecting our analysis. By integrating both clinal and seasonal variation, we can infer natural selection with more confidence, because selection is the most likely process to produce a pattern of clinal and seasonal variation in opposite directions (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.

We find a significant negative relationship between clinal and seasonal regression coefficients (Fig. 2A). The relationship is stronger for SNPs within exons and introns when compared to polymorphisms in intergenic, upstream and downstream regions (Fig. 2A). We found that the relationship is not particularly stronger within the only common inversion in Pennsylvania (In(2L)t) (Fig. 2B). The strength of the signal varies among chromosome arms, following the order 3R>3L>2R>2L (Table S4). The relationship between clinal and seasonal coefficients is marginally stronger in high recombination regions ($P = 0.08$; Table S5), suggesting that the efficiency of selection is higher the higher the recombination rate.

Another way of showing that latitudinally varying selective pressures reflect seasonally varying selection in Pennsylvania is by testing if clinal SNPs are more likely to be seasonal. We observed that clinal SNPs are enriched for seasonal SNPs (Fig. 3). The enrichment increases with more stringent P-value quantile cut-offs.

## Top SNPs for clinal and seasonal variation are enriched for metabolic functions

Our approach to detecting selection relies on sets of SNPs, for which we fit a regression of latitudinal against seasonal regression coefficients. It is not possible, however, to single out individual polymorphisms and genes that are being driven by seasonal and spatial selection. Thus, we created a composite statistic that incorporates clinal and seasonal P-values, as well as the difference in sign (direction) of clinal and seasonal regression coefficients. We show that the outlier SNPs for this statistic are enriched among exonic, UTR 3', and UTR 5' (Fig. 1C).

The composite statistic outliers greatly overlap with clinal and seasonal variants (Fig. 4). The overlap is greater among clinal and composite sites, because clinal P-values are generally much smaller than seasonal P-values. However, about one third of the composite outliers are unique.

Although gene ontology enrichment analysis results should be interpreted with caution (Pavlidis et al. 2012), we characterized the biological function of the outlier SNPs for each of these sets: clinal SNPs, seasonal SNPs, and top composite SNPs. At an FDR corrected P-value < 0.05, we found enrichment for 17 GO terms for the clinal SNPs, no enrichment for the seasonal SNPs and five GO terms for the composite statistic SNPs (Tables S6, S7, S8). Most of these enriched terms are involved in metabolism. Because there are many unique SNPs to the composite statistic, we also sought to biologically characterize them (Table S9). Interestingly, we found that SNPs unique to the top composite statistic are enriched for four metabolic functions (macromolecule metabolic process, intramolecular oxidoreductase activity, RNA metabolic process, and nucleic acid metabolic process; Table S9). These results fit nicely with the previous observation that metabolic genes are central to clinal adaptation and vary latitudinally and seasonally (Lavington et al. 2014; Cogni et al. 2015).

## Linked selection contributes little to the relationship between clinal and seasonal variation

Our regression analysis of clinal and seasonal variation assumes that SNPs are independent. However, variation at one site is linked to variation at other sites, and selection can increase this dependency. Therefore, we assessed the impact of linked selection by implementing a block-bootstrap approach. The strength of the genome-wide relationship between clinal and seasonal variation decreased slightly when

randomly sampling one SNP per 500bp ($P < 10^{-4}$; Fig. 5). It did not significantly impact the signal for exonic, intronic, and intergenic regions ($P > 0.05$; Fig. 5). The block-bootstrap estimates for the impact of the recombination rate on the relationship between clinal and seasonal variation are significantly stronger than our initial estimate ($P < 10^{-4}$; Fig. 5). After correcting for dependence, the relationship between recombination rate and our signature of natural selection increases. This supports our previous reasoning that recombination increases the efficiency of selection.

## Selection accounts for most of the clinal variation

Although we identified a significant negative relationship between clinal and seasonal variation, the magnitude of the regression coefficient appears small ($\beta = -0.013$). 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 a proportion of selected sites. We found that the observed regression coefficient is consistent with approximately 6.6% of the genome being under spatial and seasonal selection (Fig. 6). It is noteworthy that this proportion is similar to 6.8% (clinal SNPs at 0.01 FDR), and not much smaller than 15.9% (clinal SNPs at 0.05 FDR).

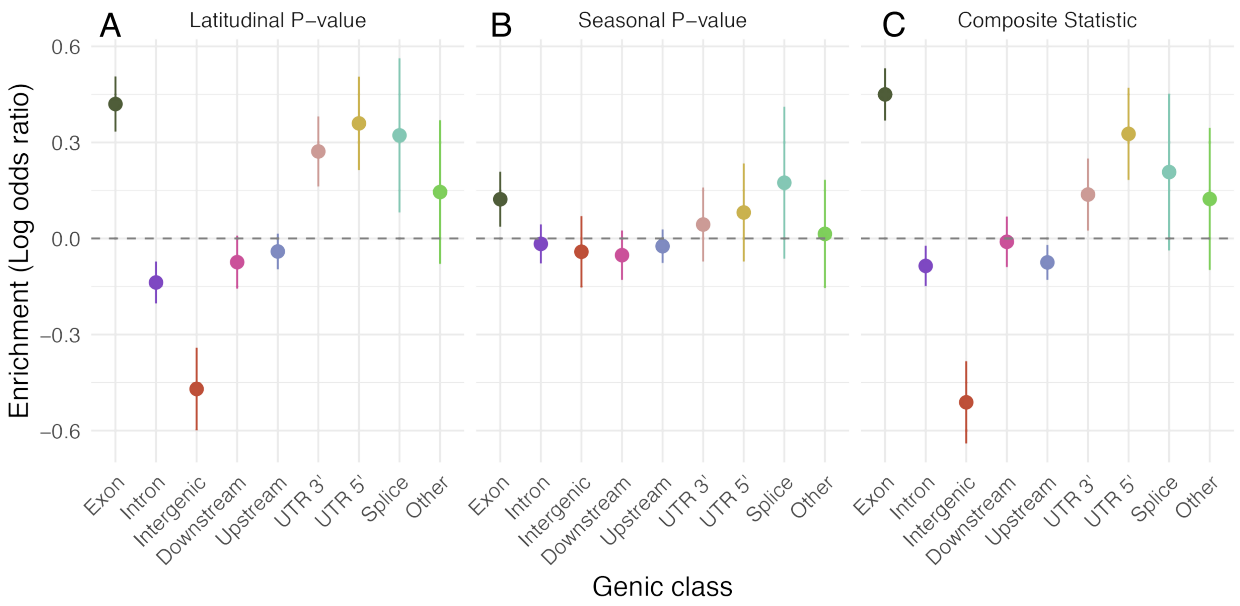


**Figure 1.** Top SNPs are enriched for functionally relevant classes. Enrichment of top 1% SNPs in each genic class. A) for latitudinal P-value, B) seasonal P-value, and C) composite statistic. Composite statistic takes into account the latitudinal and seasonal P-value, and the difference in sign between latitudinal and seasonal regression coefficients. Error bars are 95% CI.

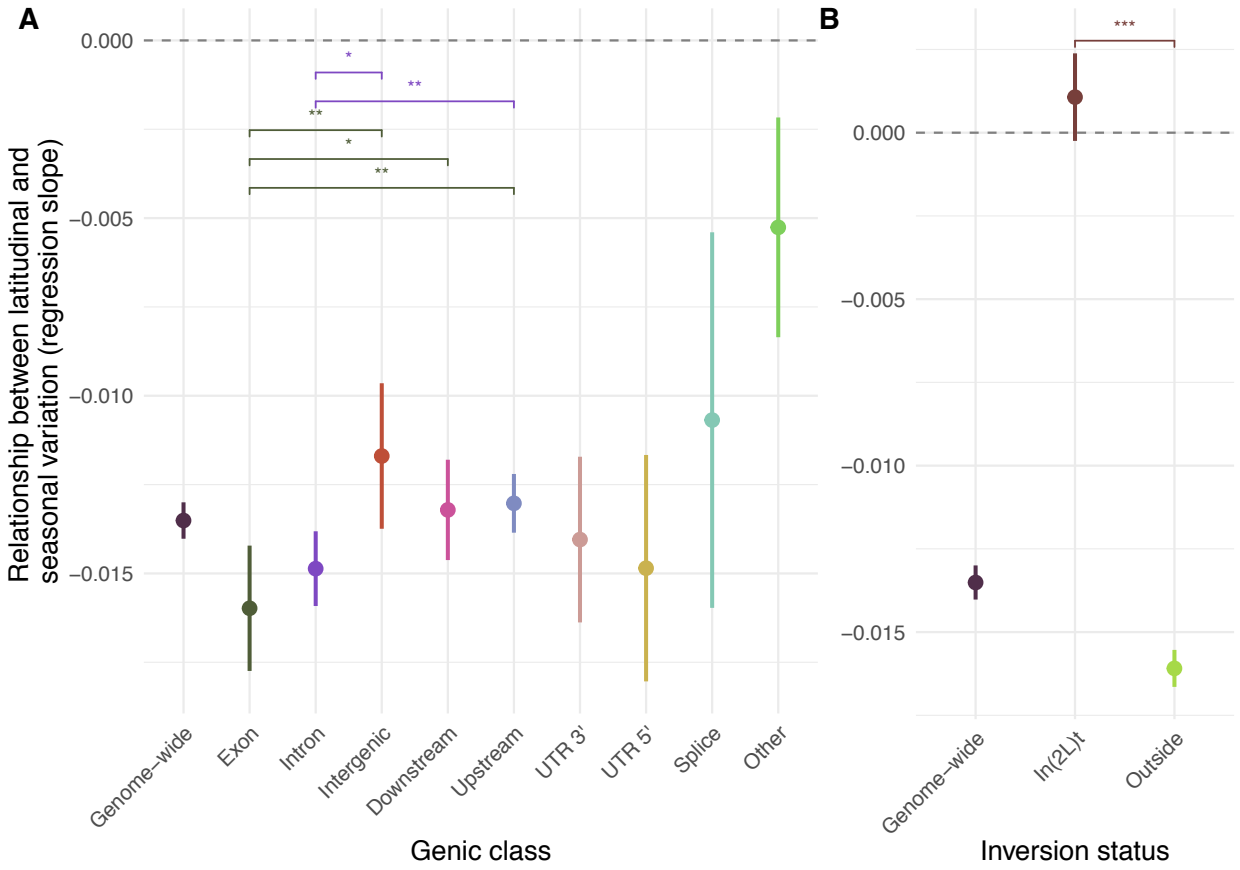


**Figure 2.** Clinal variation reflects seasonal variation. The strength of the relationship varies A) among different genic classes, and B) between inversion statuses. Error bars are 95% CI. Horizontal bars represent post-hoc analysis. * P-value < 0.05, ** P-value < 0.005.

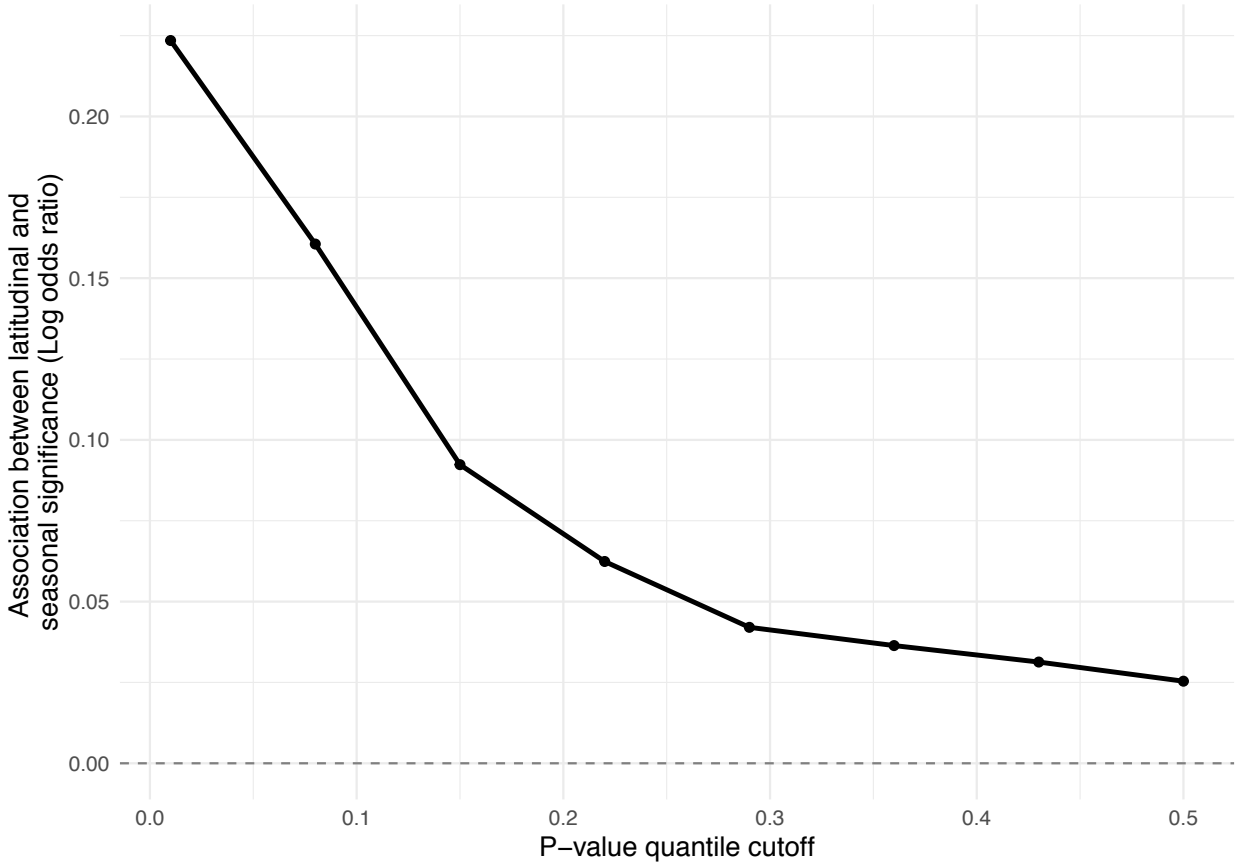


**Figure 3.** Clinal SNPs are more likely to be seasonal. Enrichment of seasonal SNPs among clinal SNPs using different P-value cutoffs.

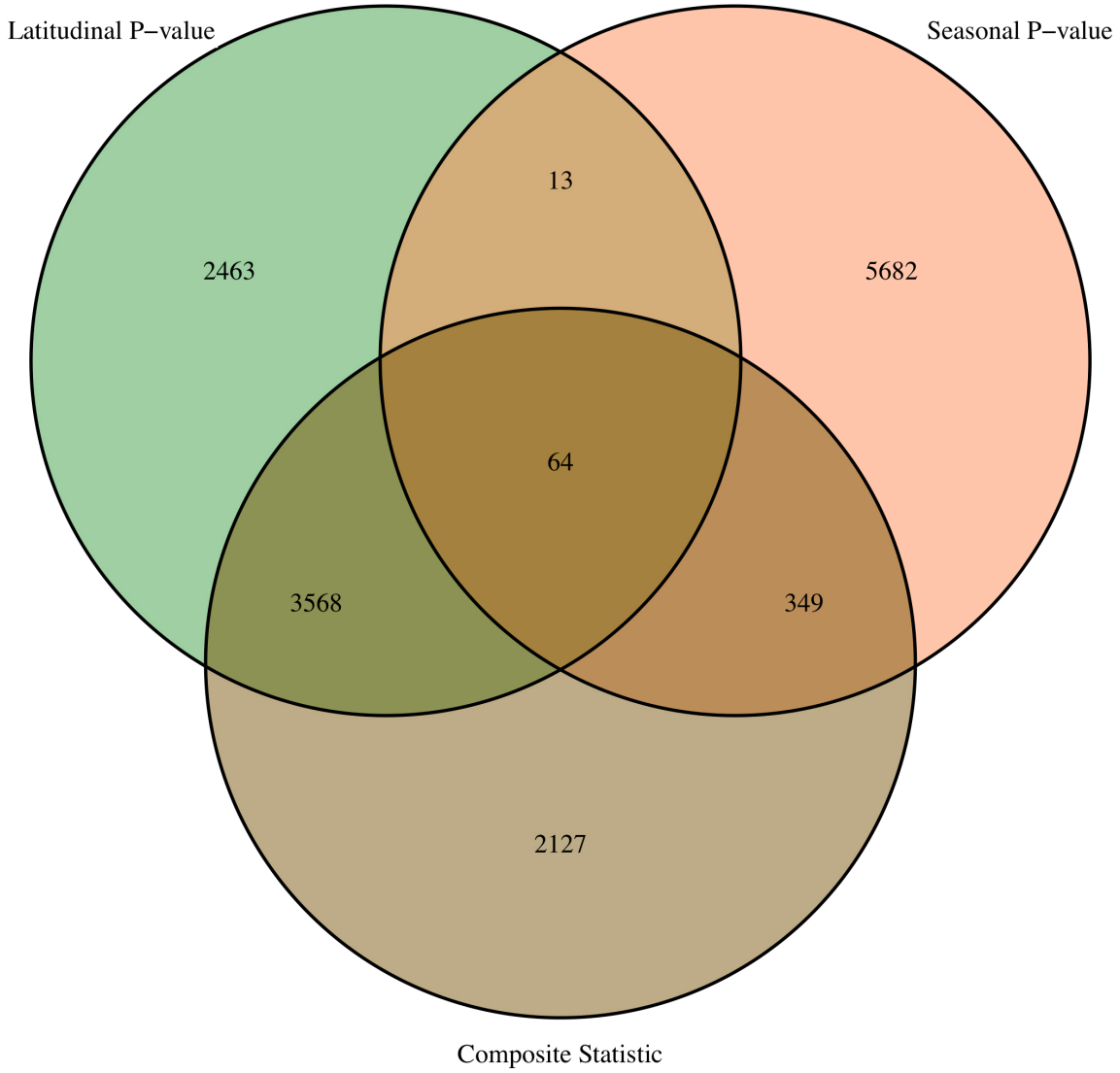


**Figure 4.** Overlap among 1% top SNPs for latitudinal P-value, seasonal P-value, and composite statistic.

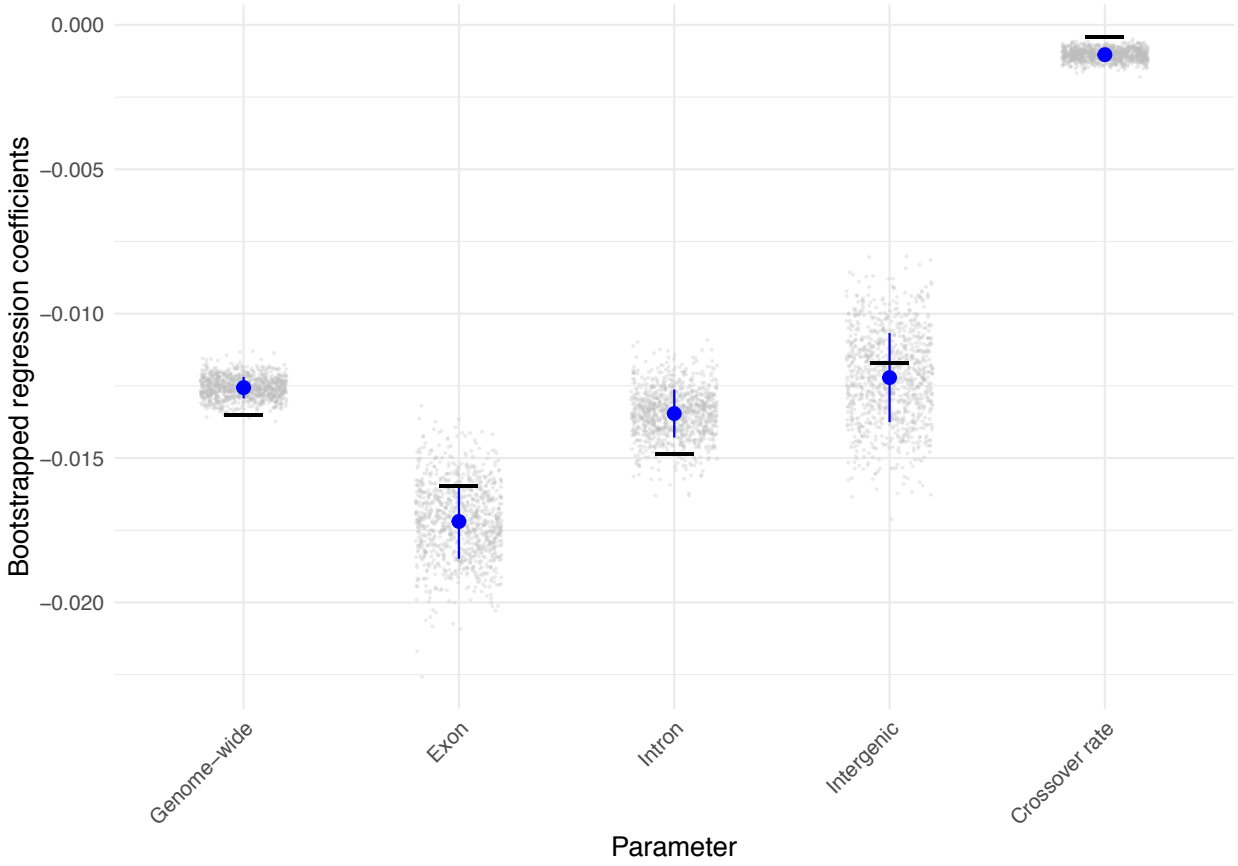


**Figure 5.** Block-bootstrapped estimates (one SNP per 500bp). Blue dots are the mean across simulations, blue lines standard deviation and gray dots the values for each simulation. Black horizontal lines are the original estimates.

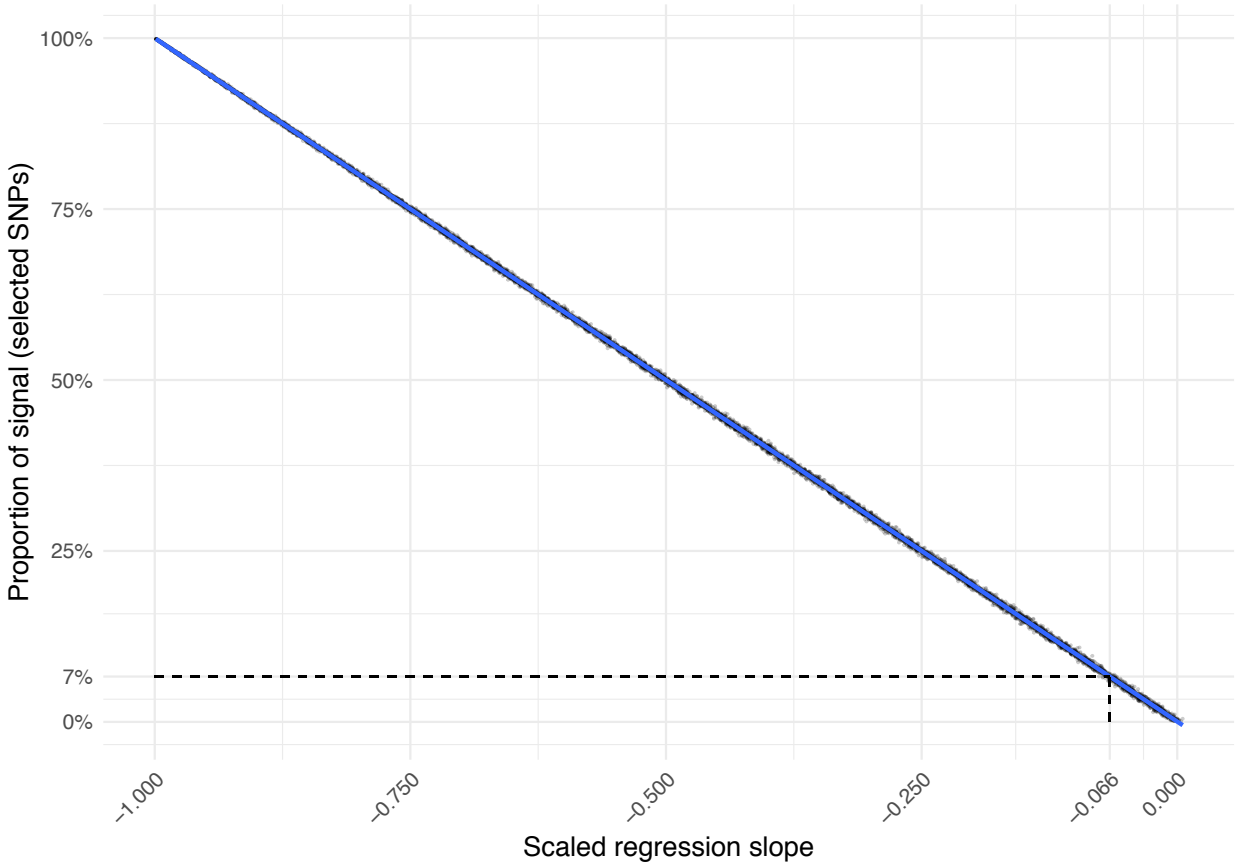


**Figure 6.** Relationship between the regression slope between two variables and the proportion of linearly dependent elements. Black points represent each simulation, and blue line is the fitted loess curve.

## 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). Although clinal variation is often taken as evidence of spatially varying selection, many studies have shown that demographic processes can generate similar patterns (Vasemägi 2006; Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016). Disentangling adaptive from non-adaptive processes is not trivial, and ideally one would need access to population parameters that are hard to estimate. One alternative lies in looking for patterns of clinal and seasonal variation (Cogni et al. 2015), since variation over short periods should be orthogonal to most confounding demographic processes (Bergland et al. 2014). Because we expect the environment to change in similar ways along a latitudinal transect and within a temperate location across seasons, integrating both clinal and seasonal variation can be informative of natural selection.

Here, using *D. melanogaster* samples from the east coast of the US, 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) most of the observed clinal variation can be explained by natural selection acting both spatially and seasonally. We show that clinal variation recapitulates seasonal variation, 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. 3). The relationship between clinal and seasonal regression is stronger for genic when compared to intergenic regions (Fig. 2A). Because variants in genic regions are more likely to impact fitness when compared to intergenic regions (Andolfatto 2005), this result can be explained by selection.

We sought to estimate how much of the variants should be under latitudinal and seasonal selection to explain the genome-wide pattern of clinal and seasonal concordance we found. Our results suggest that the observed strength of the relationship between clinal and seasonal variation is explained by about 6.6% of the SNPs being under spatial and seasonal selection (Fig. 6). We estimate that approximately 6.8% of the SNPs are clinal at 1% FDR and 15.9% at 5% FDR. A previous study has obtained a slightly smaller number (4.3% at 1% FDR), but they included samples from Maine that were collected in the fall (Machado et al. 2016). We do not have an appropriate null distribution that takes into account simultaneously selective and demographic processes, therefore our choice of 1% or 5% FDR is somewhat arbitrary. We should emphasize the exact genome-wide proportions are likely to be smaller, because we applied some stringent filters that may have removed a significant number of real variants. Nonetheless, our simple analysis showed that most clinal variation in common, autosomal SNPs is consistent with selection that varies spatially and seasonally. This is an important result, as more recent studies have favored non-adaptive hypothesis to explain clinal patterns in *D. melanogaster* (Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016).

One such hypothesis is the secondary contact hypothesis, which states that clinal patterns could be attributed to migration from Europe to high latitude regions and from Africa to low latitude regions (Bergland et al. 2016). The interpretation that this could generate and maintain most of the clinal variation observed in *D. melanogaster* is highly unlikely. Clinal variation has been observed in this and related species for a long time (David and Bocquet 1975a; David and Bocquet 1975b; Coyne and Beecham 1987; Singh and Rhomberg 1987a). Given the high levels of gene flow estimated between the ends of the cline both in the US and Australia (Singh and Rhomberg 1987b; Agis and Schlötterer

2001; Kennington et al. 2003; Schmidt et al. 2005), differentiation at neutral sites should quickly dissipate, unless there is a considerable amount of ongoing migration from Africa and Europe into both the US and Australia. Furthermore, secondary contact cannot account for any of the patterns we uncovered here, as variation over very short timescales (*e.g.*, seasonal variation) is orthogonal to admixture.

Demography and linkage disequilibrium could impact our estimates, and even explain some of the patterns we found. Demography could potentially mimic some of the seasonal patterns we have observed here. First, rural populations of *D. melanogaster* in temperate regions collapse during the winter and recover from spring to autumn. If early spring fly populations are recolonized from local urban refugia, then the seasonal variation we observed is a result of migration of urban genotypes into rural areas. In fact, reproductive diapause cycles in orchards and reaches high frequencies early in the spring. However, its frequency in urban fruit markets in Philadelphia is much lower (Schmidt and Conde 2006), making it unlikely that seasonal variation is a result of migration of urban flies to rural areas. Flies have been shown to survive and reproduce during winter season in temperate regions, supporting our argument that rural flies can withstand a harsh winter season (Mitrovski and Hoffmann 2001; Hoffmann et al. 2003). Furthermore, (Bergland et al. 2014) used simulations and found that local refugia recolonization is unlikely promoting the seasonal oscillation of hundreds of SNPs. Second, the same observed seasonal pattern could be produced by migration of flies from the south in the summer, and from the north in the winter. Northward migration of flies in the summer seems biologically plausible, as populations reach their peak during this season and migration rates are high in *D. melanogaster*. On the other hand, southward migration during the winter seems unlikely, because the winter starts earlier in the north, fly populations are dramatically reduced during this season, and climate

would additionally hinder their long-range migration. We know that the seasonal variation is not exclusively caused by selection in the summer, but also in the winter. The incidence of diapause and the allele associated with diapause incidence, for instance, increases in frequency during the winter (Schmidt and Conde 2006; Cogni et al. 2014). Therefore, alternating long-range migration from the ends of the cline cannot explain the patterns we observe.

Linkage disequilibrium could affect our estimates in two ways. First, chromosomal inversions segregate along the cline and seasonally in Pennsylvania (Kapun et al. 2016). Although inversions do contribute to adaptation (Fabian et al. 2012; Kapun et al. 2016), we demonstrate that the genome-wide pattern of clinal reflecting seasonal variation cannot be explained by sites within inversions (Fig. 2B). This result is in concordance with previous studies, which have shown that not all clinal and seasonal variation can be attributed to inversions (Fabian et al. 2012; Bergland et al. 2014; Machado et al. 2018). More generally, because we expect rapid changes in allele frequency across seasons to intensify linkage disequilibrium, linked selection could bias our estimates. However, we show that this effect is rather small (Fig. 5). Previous studies have shown that linked selection is important in this short time scale, but it decays rapidly, and returns to background levels within 200 bp (Bergland et al. 2014; Machado et al. 2018). Therefore, we can conclude the genome-wide pattern of clinal variation mirroring seasonal variation is a result of selection acting on many sites across the genome.

Although a few previous studies have investigated the relationship between clinal and seasonal variation (Cogni et al. 2015; Machado et al. 2018), we are the first to thoroughly investigate this pattern in a genome-wide context and to demonstrate that this signature is consistent with the action of spatial and seasonal 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 many locations in North America and Europe. They investigated the relationship between clinal and seasonal variation as a way to validate their outlier seasonal SNPs. In addition, because our interest is to ultimately understand processes driving clinal variation in the east coast of the US, 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; Machado et al. 2018), and it is the population for which we have the most information about how phenotypes vary seasonally (Schmidt and Conde 2006; Behrman et al. 2015; Rajpurohit et al. 2017; Behrman et al. 2018). Also, *D. melanogaster* populations are structured, and so there are differences in polymorphisms segregating in the US and Europe.

Many species occur along spatially structured environments, and show clinal variation in many traits, so a question that remains open is whether these patterns are the result of natural selection or neutral demographic processes. Temporal variation is also ubiquitous, especially in temperate environments, so seasonal adaptation could be an important feature of organisms that have multiple generations each year (also called multivoltine) (Behrman et al. 2015). Here, we demonstrate that by integrating clinal and seasonal variation, it is possible to discern adaptive and non-adaptive hypothesis for clinal variation. This approach could potentially be applied to other multivoltine species that occur along environmental gradients. Notably, this approach can be applied in invasive species, which are known to often have short-generation times and to reproduce quickly (Sakai et al. 2001). Our data is consistent with the adaptive

hypothesis for clinal variation in *D. melanogaster*, and our results could hold for many of the other species which show clinal variation in natural populations.

# References


Adrion JR, Hahn MW, Cooper BS. 2015. Revisiting classic clines in Drosophila melanogaster in the age of genomics. Trends Genet. 31:434–444.

Agis M, Schlötterer C. 2001. Microsatellite variation in natural Drosophila melanogaster populations from New South Wales (Australia) and Tasmania. Mol. Ecol. 10:1197–1205.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature 437:1149–1152.

Arthur AL, Weeks AR, Sgrò CM. 2008. Investigating latitudinal clines for life history and stress resistance traits in Drosophila simulans from eastern Australia. J. Evol. Biol. 21:1470–1479.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25:25–29.

Bansal V, Bansal V, Libiger O. 2016. A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics 32:3213.

Barton NH. 1983. Multilocus clines. Evolution 37:454–471.

Barton NH. 1999. Clines in polygenic traits. Genet. Res. 74:223–236.

Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M, et al. 2010. A Coastal Cline in Sodium Accumulation in Arabidopsis thaliana Is Driven by Natural Variation of the Sodium Transporter AtHKT1;1. PLoS Genet. 6:e1001193.

Behrman EL, Howick VM, Kapun M, Staubach F, Bergland AO, Petrov DA, Lazzaro BP, Schmidt PS. 2018. Rapid seasonal evolution in innate immunity of wild Drosophila melanogaster. Proc. Biol. Sci. [Internet] 285. Available from: http://dx.doi.org/10.1098/rspb.2017.2599

Behrman EL, Watson SS, O'Brien KR, Heschel MS, Schmidt PS. 2015. Seasonal variation in life history traits in two Drosophila species. J. Evol. Biol. 28:1691–1704.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. PLoS Genet. 10:e1004775.

Bergland AO, Tobler R, González J, Schmidt P, Petrov D. 2016. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in Drosophila melanogaster. Mol. Ecol. 25:1157–1174.

Bock IR, Parsons PA. 1981. Species of Australia and New Zealand. Genetics and biology of Drosophila 3:291–308.

Campitelli BE, Stinchcombe JR. 2013. Natural selection maintains a single-locus leaf shape cline in Ivyleaf morning glory, Ipomoea hederacea. Mol. Ecol. 22:552–564.

Cardini A, Jansson A-U, Elton S. 2007. A geometric morphometric approach to the study of ecogeographical and clinal variation in vervet monkeys. J. Biogeogr. 34:1663–1678.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6:80–92.

Cogni R, Kuczynski C, Koury S, Lavington E, Behrman EL, O'Brien KR, Schmidt PS, Eanes WF. 2014. The intensity of selection acting on the couch potato gene--spatial-temporal variation in a diapause cline. Evolution 68:538–548.

Cogni R, Kuczynski K, Lavington E, Koury S, Behrman EL, O'Brien KR, Schmidt PS, Eanes WF. 2015. Variation in Drosophila melanogaster central metabolic genes appears driven by natural selection both within and between populations. Proceedings of the Royal Society of London B: Biological Sciences 282:20142688.

Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in Drosophila melanogaster. PLoS Genet. 8:e1002905.

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 inDrosophila 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.

Duchen 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.

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.

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 inDrosophila melanogaster: Body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. J. Genet. 82:207–223.

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, Aduriz MGB, Staubach F, Vieira J, Obbard D, Goubert C, Stabelli OR, Kankare M, Haudry A, Wiberg RAW, et al. 2018. Genomic analysis of European Drosophila populations reveals longitudinal structure and continent-wide selection. bioRxiv [Internet]:313759. Available from: https://www.biorxiv.org/content/early/2018/06/04/313759

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.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. Ecol. Lett. 7:1225–1241.

Keller A. 2007. Drosophila melanogaster's history as a human commensal. Curr. Biol. 17:R77–R81.

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.

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.

Levene H. 1953. Genetic Equilibrium When More Than One Ecological Niche is Available. Am. Nat. 87:331–333.

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 glycerol-3-phosphate dehydrogenase clines in Drosophila melanogaster on different continents. Evolution 36:86–96.

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.

Rajpurohit S, Hanus R, Vrkoslav V, Behrman EL, Bergland AO, Petrov D, Cvačka 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.

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. Annu. 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.

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.

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.

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 material

**Table S1.** Information of the samples used in this study.

| Accession # | Sample name | Location | Latitude | Collection date | # Flies | Median depth | Month | Season | Seasonal set | Clinal set |
|---|---|---|---|---|---|---|---|---|---|---|
| SRR1525685 | FL1 | Homestead, FL | 25.47 | Jul-08 | 39 | 59 | 7 | Spring | 0 | 1 |
| SRR1525694 | FL2 | Homestead, FL | 25.47 | Jul-08 | 48 | 38 | 7 | Spring | 0 | 1 |
| SRR1525695 | GA | Hahira, GA | 30.99 | Jul-08 | 51 | 102 | 7 | Spring | 0 | 1 |
| SRR1525696 | SC | Euatwville, SC | 33.40 | Jul-08 | 48 | 83 | 7 | Spring | 0 | 1 |
| SRR3590551 | VA_07_2012 | Charlotttesville, VA | 38.03 | Jul-12 | 95 | 70 | 7 | Spring | 0 | 1 |
| SRR1525697 | NC | Raleigh, NC | 35.5 | 2003 | 98 | 33 | NA | NA | 0 | 1 |
| SRR3939095 | PA_06_2013 | Linvilla, PA | 39.88 | Jun-13 | 54 | 71 | 6 | Spring | 0 | 1 |
| SRR3590557 | MA_07_2012 | Lancaster, MA | 42.46 | Jul-12 | 90 | 51 | 7 | Spring | 0 | 1 |
| SRR1525768 | PA_07_2009 | Linvilla, PA | 39.53 | Jul-09 | 55 | 187 | 7 | Spring | 1 | 0 |
| SRR1525769 | PA_11_2009 | Linvilla, PA | 39.53 | Nov-09 | 74 | 66 | 11 | Fall | 1 | 0 |
| SRR1525770 | PA_07_2010 | Linvilla, PA | 39.53 | Jul-10 | 116 | 17 | 7 | Spring | 1 | 0 |
| SRR1525771 | PA_11_2010 | Linvilla, PA | 39.53 | Nov-10 | 33 | 76 | 11 | Fall | 1 | 0 |
| SRR1525772 | PA_07_2011 | Linvilla, PA | 39.53 | Jul-11 | 75 | 53 | 7 | Spring | 1 | 0 |
| SRR1525773 | PA_10_2011 | Linvilla, PA | 39.53 | Oct-10 | 47 | 74 | 10 | Fall | 1 | 0 |
| SRR3590560 | PA_10_2012 | Linvilla, PA | 39.53 | Oct-12 | 100 | 26 | 10 | Fall | 1 | 0 |
| SRR3590561 | PA_07_2012 | Linvilla, PA | 39.53 | Jul-12 | 115 | 60 | 7 | Spring | 1 | 0 |
| SRR3590563 | PA_9_2012 | Linvilla, PA | 39.53 | Sep-12 | 50 | 55 | 9 | Fall | 1 | 0 |
| SRR3939096 | PA_10_2014 | Linvilla, PA | 39.88 | Oct-14 | 50 | 109 | 10 | Fall | 1 | 0 |
| SRR3939097 | PA_06_2014 | Linvilla, PA | 39.88 | Jun-14 | 92 | 68 | 6 | Spring | 1 | 0 |
| SRR3939098 | PA_10_2015 | Linvilla, PA | 39.88 | Oct-15 | 52 | 102 | 10 | Fall | 1 | 0 |
| SRR3939099 | PA_07_2015 | Linvilla, PA | 39.88 | Jul-15 | 74 | 216 | 7 | Spring | 1 | 0 |

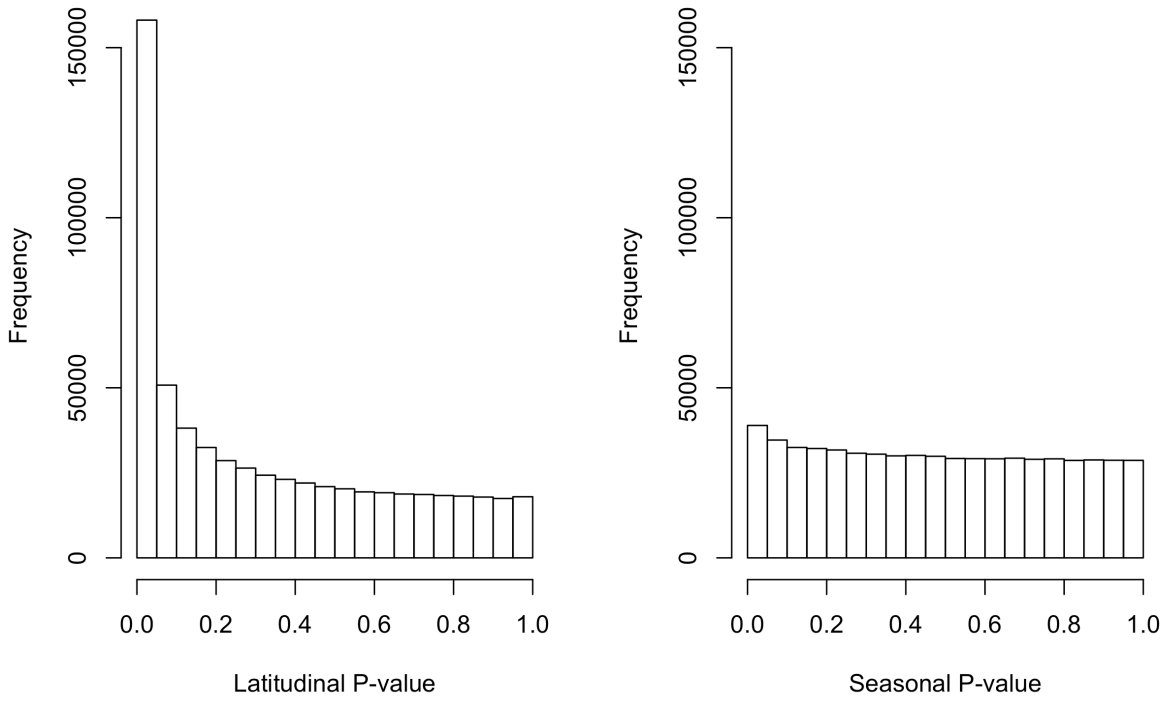


**Figure S1**. Distribution of seasonal and latitudinal GLM P-values.

**Table S2.** Summary of a regression of absolute latitude regression coefficients against chromosomes. CI stands for 95% confidence interval.

| | **Absolute latitude regression coefficient** | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.73 | 0.73 – 0.74 | **<0.001** |
| CHROM2R | -0.09 | -0.10 – -0.09 | **<0.001** |
| CHROM3L | 0.01 | 0.00 – 0.01 | **0.007** |
| CHROM3R | 0.24 | 0.24 – 0.25 | **<0.001** |
| Observations | 610780 | | |
| $R^2$ / adjusted $R^2$ | 0.038 / 0.038 | | |

**Table S3.** Summary of a regression of absolute seasonal regression coefficients against chromosomes. CI stands for 95% confidence interval.

| | **Absolute seasonal regression coefficient** | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.78 | 0.78 – 0.78 | **<0.001** |
| CHROM2R | -0.01 | -0.01 – -0.00 | **0.002** |
| CHROM3L | 0.02 | 0.01 – 0.02 | **<0.001** |
| CHROM3R | 0.02 | 0.01 – 0.02 | **<0.001** |
| Observations | 610780 | | |
| $R^2$ / adjusted $R^2$ | 0.000 / 0.000 | | |

**Table S4.** Summary of a regression of latitude regression coefficients against seasonal regression coefficients, with an interaction term between seasonal regression and chromosome. CI stands for 95% confidence interval.

| | | reg lat | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | -0.0001 | -0.0002 – -0.0000 | **0.039** |
| reg sea | -0.0014 | -0.0024 – -0.0004 | **0.006** |
| reg_sea:CHROM2R | -0.0112 | -0.0127 – -0.0097 | **<0.001** |
| reg_sea:CHROM3L | -0.0133 | -0.0147 – -0.0119 | **<0.001** |
| reg_sea:CHROM3R | -0.0240 | -0.0253 – -0.0226 | **<0.001** |
| Observations | 610780 | | |
| $R^2$ / adjusted $R^2$ | 0.006 / 0.006 | | |

**Table S5.** Summary of a regression of latitude regression coefficients against seasonal regression coefficients, with an interaction term between seasonal regression and crossover rate (Comeron et al. 2012). CI stands for 95% confidence interval.

| | **Regression coefficient for latitude** | | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 0.0000 | -0.0001 – 0.0001 | 0.396 |
| Regression coefficient for seasons | -0.0126 | -0.0134 – -0.0117 | **<0.001** |
| regression_sea:CO_rate | -0.0004 | -0.0007 – -0.0002 | **0.001** |
| Observations | 608395 | | |
| $R^2$ / adjusted $R^2$ | 0.005 / 0.005 | | |

**Table S6.** Gene ontology hits using 1% top latitudinal SNPs. Only terms with a FDR corrected P-value less than 0.05 are shown.

| GO category | # candidate SNPs within genes | FDR corrected P-value | # genes in this GO term | Description |
|---|---|---|---|---|
| GO:0010171 | 11 | 0.013785 | 18 | body morphogenesis |
| GO:0016623 | 4 | 0.013785 | 4 | oxidoreductase activity, acting on the aldehyde or oxo group of donors, oxygen as acceptor |
| GO:0032991 | 317 | 0.013785 | 1792 | macromolecular complex |
| GO:0007288 | 10 | 0.013785 | 18 | sperm axoneme assembly |
| GO:0005214 | 32 | 0.013785 | 123 | structural constituent of chitin-based cuticle |
| GO:0042302 | 34 | 0.013785 | 129 | structural constituent of cuticle |
| GO:0046483 | 160 | 0.013785 | 908 | heterocycle metabolic process |
| GO:1901360 | 173 | 0.013785 | 973 | organic cyclic compound metabolic process |
| GO:0035082 | 10 | 0.020846667 | 21 | axoneme assembly |
| GO:0006139 | 152 | 0.020846667 | 863 | nucleobase-containing compound metabolic process |
| GO:0034641 | 214 | 0.020846667 | 1403 | cellular nitrogen compound metabolic process |
| GO:0090304 | 118 | 0.020846667 | 707 | nucleic acid metabolic process |
| GO:0005578 | 17 | 0.02549 | 35 | proteinaceous extracellular matrix |
| GO:0006725 | 165 | 0.029770714 | 942 | cellular aromatic compound metabolic process |
| GO:0006807 | 248 | 0.044773333 | 1605 | nitrogen compound metabolic process |
| GO:0006979 | 35 | 0.048411875 | 94 | response to oxidative stress |
| GO:0048079 | 7 | 0.048671177 | 12 | regulation of cuticle pigmentation |

**Table S7.** Gene ontology hits using 1% top seasonal SNPs. Only terms with a FDR corrected P-value less than 0.05 are shown.

| GO category | # candidate SNPs within genes | FDR corrected P-value | # genes in this GO term | Description |
|---|---|---|---|---|
| no terms with FDR P-value < 0.05 | | | | |

**Table S8.** Gene ontology hits using 1% top composite statistic SNPs. Only terms with a FDR corrected P-value less than 0.05 are shown.

| GO category | # candidate SNPs within genes | FDR corrected P-value | # genes in this GO term | Description |
|---|---|---|---|---|
| GO:0010171 | 10 | 0.035683333 | 18 | body morphogenesis |
| GO:0016623 | 4 | 0.035683333 | 4 | oxidoreductase activity, acting on the aldehyde or oxo group of donors, oxygen as acceptor |
| GO:1901360 | 194 | 0.035683333 | 973 | organic cyclic compound metabolic process |
| GO:0046483 | 178 | 0.036645 | 908 | heterocycle metabolic process |
| GO:0006979 | 40 | 0.03731 | 94 | response to oxidative stress |

**Table S9.** Gene ontology hits using SNPs unique to 1% top composite statistic. Only terms with a FDR corrected P-value less than 0.05 are shown.

| GO category | # candidate SNPs within genes | FDR corrected P-value | # genes in this GO term | Description |
|---|---|---|---|---|
| GO:0043170 | 292 | 0.038865 | 2486 | macromolecule metabolic process |
| GO:0016863 | 3 | 0.038865 | 3 | intramolecular oxidoreductase activity, transposing C=C bonds |
| GO:0016070 | 64 | 0.038865 | 541 | RNA metabolic process |
| GO:0090304 | 82 | 0.038865 | 707 | nucleic acid metabolic process |