ARTICLE

Natural selection and parallel clinal and seasonal changes in *Drosophila melanogaster*

Murillo F. Rodrigues[1,2*], Maria D. Vibranovski[1], Rodrigo Cogni[3]

[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.

Abstract

Spatial and seasonal variation in the environment are ubiquitous. Covariation between a trait and space is often interpreted as the result of selection, but demography can complicate inference. Finding seasonal changes driven by selection is challenging, given effect sizes are small and the environment is subject to stochastic changes within seasons. *Drosophila melanogaster* is known to harbor polymorphisms that change with latitude and seasons. Identifying the role of selection in driving latitudinal and seasonal change is not trivial, but one promising approach is to look for parallel clinal and seasonal variation. 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 linked selection in driving these patterns, and conservatively estimate the proportion of the variants that should be under selection to explain our observations. Allele frequency estimates were obtained from pooled samples from seven different locations along the east coast of the US, and 13 samples collected in the spring and fall in 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 2.8% of the common variants being under selection. Our results highlight the contribution of natural selection in driving allele frequency changes with latitude and seasons in *D. melanogaster*.

Introduction

Many species can be found along environments that gradually change in time and space (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 are called clinal. These clinal traits are usually thought to be a result of

spatially varying selection (Endler 1977), but 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 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) was likely followed by the emergence of adaptations to new climates (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).

68 A classical approach to discern between adaptive and neutral differentiation has been
69 to look at parallel latitudinal variation between continents (Hoffmann et al. 2002; Hoffmann
70 and Weeks 2007; Turner et al. 2008; Paaby et al. 2010; Reinhardt et al. 2014; Schrider et al.
71 2016). If a trait varies clinally in two or more continents, a simple, plausible scenario is
72 latitudinally varying selection acting on ancestral variation (Endler 1977; Barton 1983;
73 Barton 1999). However, some of the clinal patterns observed in North America may be
74 attributed to migration of African flies to low latitude regions and European flies to high
75 latitude regions (Caracristi and Schlotterer 2003; Yukilevich and True 2008; Duchen et al.
76 2013; Kao et al. 2015; Bergland et al. 2016). The main evidence in favor of the secondary
77 contact hypothesis is that there is a cline in the proportion of African ancestry in North
78 America (Kao et al.2015; Bergland et al. 2016), and a weak cline in African ancestry in
79 Australia (only marginally significant; see Bergland et al. 2016).
80 Separating the role of selection and demography in maintaining clinal polymorphisms
81 is not trivial, given that the axis of demography and environmental change are confounded in
82 this system (i.e., migration and the environment are structured along the south-north axis). A
83 simpler, more plausible model would be that admixture introduced "pre-adapted" alleles, that
84 then contributed to clinal adaptation (Flatt 2016). For instance, for proper modelling of
85 selection and demography in this system, one would need more information about rates of
86 migration, population sizes, changes in population size, historical founding of the
87 populations, history of admixture between populations, and sources and nature of selective
88 pressures that vary with latitude. Alternatively, one simple approach that can help disentangle
89 the role of adaptation is to explore signatures of parallel change with latitude and seasons
90 (Cogni et al. 2014; Cogni et al. 2015).
91 *Drosophila* has also been widely used in the study of seasonal adaptation. One of the
92 earliest examples of seasonal genetic variation was observed in chromosomal inversions in *D*.

93 *pseudoobscura* (Dobzhansky 1943). In *D. melanogaster*, it has been shown that flies
94 collected in the spring are more tolerant to stress (Behrman et al. 2015), show higher
95 diapause inducibility (Schmidt and Conde 2006), have increased immune function (Behrman
96 et al. 2018) and have different cuticular hydrocarbon profiles than those collected in the fall
97 (Rajpurohit et al. 2017). Furthermore, genome-wide analyses have identified polymorphisms
98 and inversions that oscillate in seasonal timescales in several localities in the United States
99 and Europe, but the mechanism behind it has not been fully resolved (Bergland et al. 2014;
100 Kapun et al. 2016; Machado et al. 2018; Wittman et al. 2017). Identifying polymorphisms
101 that fluctuate seasonally is challenging, because effect sizes are small and populations are
102 subject to stochastic environmental events that are not tied to seasonality (Machado et al.
103 2018). Indeed, a recent study put in question the interpretation that a set of previously
104 identified seasonal SNPs in *D. melanogaster* is due to seasonally varying selection (Buffalo
105 and Coop 2019).
106 One way to increase power would be to study the relationship between seasonal and
107 latitudinal change. Seasonal and latitudinal variation could be coupled because of two
108 different adaptive mechanisms. First, the environment changes similarly with latitude and
109 through seasons (at least with respect to temperature). Second, the onset of the spring
110 changes with latitude and so seasonal changes in allele frequencies alone could produce clinal
111 variation, a mechanism termed seasonal phase clines (Roff 1980; Rhomberg and Singh
112 1986). Demographic processes that can generate clinal patterns like isolation by distance and
113 secondary contact from diverged populations are independent to variation across seasons,
114 given the short time scale at which the seasonal processes happen (Bergland et al. 2014;
115 Cogni et al. 2015). Therefore, parallel latitudinal and seasonal variation in a trait is strong
116 evidence in favor of natural selection. For instance, it has been observed in *D. melanogaster*
117 that both the prevalence of reproductive diapause and the frequency of a variant in the couch

118 potato gene associated with diapause inducibility vary latitudinally and seasonally (Schmidt
119 et al. 2005; Cogni et al. 2014). The frequency of this diapause-inducing variant drops in the
120 summer and is positively correlated with latitude, as we would expect if it were being driven
121 by natural selection.

122 Besides this classic example, a few studies have looked for signatures of parallel
123 clinal and seasonal variation in *D. melanogaster* (Bergland et al. 2014; Cogni et al. 2015;
124 Kapun et al. 2016; Behrman et al. 2018; Machado et al. 2018). Using a candidate gene
125 approach, Cogni et al. (2015) found an association between clinal and seasonal change in
126 central metabolic genes, that are likely important drivers of climatic adaptation. Kapun et al.
127 (2016) dissected the role of selection in shaping clinal variation in *D. melanogaster*
128 cosmopolitan inversions. By integrating clinal and seasonal variation, information on the
129 stability of inversion clines, and analysis of the effect of climatic variables on inversion
130 frequencies, they concluded that several inversion clines are consistent with natural selection.
131 A recent study also investigated the association between clinal and seasonal variation, and
132 showed the direction of allele frequency change with latitude is concordant with the change
133 with seasons in many polymorphisms across the *D. melanogaster* genome (Machado et al.
134 2018). However, the informative value of parallel clinal and seasonal variation in
135 determining the role of selection was not fully explored.

136 Here, we aim to answer whether the clinal and seasonal patterns observed in *D.*
137 *melanogaster* are consistent with natural selection. If natural selection is a key process
138 underlying clinal and seasonal change, flies collected in the north should be more similar to
139 flies collected in early spring, whereas southern flies should be more like flies collected in the
140 fall (Cogni et al. 2014; Bergland et al. 2014; Cogni et al. 2015; Kapun et al. 2016; Behrman
141 et al. 2018; Machado et al. 2018). Thus, we quantified the relationship between clinal and
142 seasonal change along the *D. melanogaster* genome. Given the localized effects of selection,

143 and the expectation that constraint depends to some degree on genic annotations, we also
144 expect the parallelism to vary across genomic annotations. Finally, we estimated how much
145 of the common variants should be under selection to explain the degree of coupling between
146 clinal and seasonal change. By integrating these two independent sources of evidence, we
147 further disentangle the contributions of selection in maintaining the pervasive and long-
148 standing patterns of clinal variation in *D. melanogaster*.

149 Material and Methods

150 Population samples

151 We analyzed 20 samples from seven locations along the United States east coast, collected by
152 (Bergland et al. 2014) (10 samples), and (Machado et al. 2018) (10 samples) (see Table S1).
153 The samples were based on pools of wild-caught individuals. We decided to not include
154 previously collected samples from Maine because they were collected in the fall, whereas all
155 of our other samples were collected in the spring, and we also did not include the DGRP
156 sample from North Carolina, as it is hard to ascertain when they were obtained (Fabian et al.
157 2012; Mackay et al. 2012; Bergland et al. 2014). The Linvilla (Pennsylvania) population was
158 sampled extensively from 2009 to 2015 (six spring, seven fall samples), and was therefore
159 used in our analysis of seasonal variation. We also replicated our results using data from four
160 Australian samples (Anderson et al. 2005; Kolaczkowski et al. 2011). All the data used in this
161 project are available on the NCBI Short Read Archive (BioProject accession numbers
162 PRJNA256231, PRJNA308584 and NCBI Sequence Read Archive SRA012285.16).

163 Mapping and processing of sequencing data

164 Raw, paired-end reads were mapped against the FlyBase *D. melanogaster* (r6.15) and *D.*
165 *simulans* (r2.02) reference genomes (Gramates et al. 2017) using BBSplit from the BBMap
166 suite (https://sourceforge.net/projects/bbmap/; version from February 11, 2019). We removed

167 any reads that preferentially mapped to *D. simulans* to mitigate effects of contamination (the
168 proportion of reads preferentially mapping to *D. simulans* was minimal, never exceeding
169 3%). Then, reads were remapped to *D. melanogaster* reference genome using bwa (MEM
170 algorithm) version 0.7.15 (Li and Durbin 2010). Files were converted from SAM to BAM
171 format using Picard Tools (http://broadinstitute.github.io/picard). PCR duplicates were
172 marked and removed using Picard Tools and local realignment around indels was performed
173 using GATK version 3.7 (McKenna et al. 2010). Single nucleotide polymorphisms (SNPs)
174 and indels were called using CRISP with default parameters (Bansal et al. 2016). We applied
175 several filters to ensure that the identified SNPs were not artifacts. SNPs in repetitive regions,
176 identified using the RepeatMasker library for *D. melanogaster* (obtained from
177 http://www.repeatmasker.org), and SNPs within 5bp of polymorphic indels were removed
178 from our analyses, because these are more likely to be sequencing errors. SNPs with average
179 minor allele frequency in the clinal and seasonal samples less than 10%, with minimum per-
180 population coverage less than 10x (or 4x for the Australian samples) or maximum per-
181 population coverage greater than the 99th quantile were excluded from our analyses. Also,
182 SNPs with multi-allelic states were removed from the analysis. We did not include SNPs in
183 any of the sex chromosomes. SNPs were annotated with their genic classes using SNPeff
184 version 4.3o (Cingolani et al. 2012).

185 **Clinal and seasonal changes in allele frequency**

186 Pool-seq data contains an additional component of error due to sampling (Kofler et al. 2011;
187 Lynch et al. 2014). We decided to model this additional error by incorporating weights
188 proportional to the effective number of chromosomes (N_E):

189 $$N_E = \left(\frac{1}{D} + \frac{1}{N_C}\right)^{-1}$$

190 where N_C is the number of chromosomes in the pool and D is the read depth at that site
191 (Kolaczkowski et al. 2011; Feder et al. 2012; Bergland et al. 2014).
192 To assess latitudinal variation, we used generalized linear models (GLM) to regress
193 allele frequency at each site against latitude according to the form,

194 $$y_i = \alpha + \beta^{clinal} Latitude + \epsilon_i$$

195 Where y_i is the allele frequency in the ith population, α is the intercept, β^{clinal} is the slope,
196 $Latitude$ is the latitude of the population, and ϵ_i is the binomial error given the N_E at that
197 SNP. This kind of regression is suitable for the analysis of clinal variation of allele
198 frequencies estimated from Pool-seq, because it accounts for differences in read depth and
199 number of chromosomes, and the non-normality of allele frequencies (Bergland et al. 2014;
200 Machado et al. 2016).
201 To assess seasonal variation, we followed a similar procedure. We regressed allele
202 frequency at each site against a binary variable corresponding to spring or fall according to
203 the form,

204 $$y_i = \alpha + \beta^{seasonal} Season + \gamma Year + \epsilon_i$$

205 where y_i is the allele frequency in the ith sample, α is the intercept, $\beta^{seasonal}$ is the slope,
206 $Season$ is the season in which the sample was obtained (June and July were encoded as
207 Spring, and September, October and November as Fall), $Year$ is the year, and ϵ_i is the
208 binomial error given the N_E at that SNP.
209 We defined clinal and seasonal SNPs using an outlier approach, because we do not
210 have an adequate genome-wide null distribution to compare our estimates. Thus, we
211 considered that SNPs were outliers if their regression P-value fell in the bottom 1% of the
212 distribution. All statistical analyses were performed in R 3.5.0 (R Core Team 2018).

213 **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:

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

where β_i^{clinal} is the estimated clinal slope for the ith SNP, α is the intercept, β is the slope, $\beta_i^{seasonal}$ is the estimated seasonal slope for the i^{ith} SNP, and ϵ_i is the gaussian error at that SNP. The clinal and seasonal slopes were z-normalized to facilitate interpretation. To confirm our results are robust to potential model misspecifications, we implemented a permutation test in which we rerun the regressions for each SNP using shuffled season and latitude labels 10,000 times.

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:

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

where γ_i is the regression coefficient for the jth genic class, $Class_j$ is the genic class of the ith SNP. Furthermore, we performed a post-hoc Tukey test to find differences in the relationship between clinal and seasonal variation among all genic classes. We also confirmed our results with permuting season and latitude labels to build a null distribution of the expected regression coefficients of each genic class.

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 within inversions. To do so, we performed a regression similar to our previous one, that takes the form:

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

238 where $Inversion_j$ is the inversion status of the SNP, and γ_i is the regression coefficient for
239 the jth status. All statistical analyses were performed in R 3.5.0 (R Core Team 2018).

240 **Composite statistic for clinal and seasonal variation**

241 Our approach uses the relationship between clinal and seasonal variation in sets of SNPs to
242 infer selection, and so it is not possible to detect individual SNPs which are under spatial and
243 seasonal selection. We implemented a composite statistic based on Fisher's method (Fisher
244 1992) that considers clinal and seasonal regression P-values, and the sign of the clinal and
245 seasonal slopes:

246 $$\chi_i = -2 \times sgn(\beta_i^{clinal} \times \beta_i^{seasonal})(\ln(P_i^{clinal}) + \ln(P_i^{seasonal}))$$

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

252 **Enrichment tests**

253 We tested for enrichment of genic classes using our sets of clinal, seasonal, and composite
254 SNPs. We tested for enrichment using Fisher's exact test for each genic region and statistic.
255 Because our choice of defining top 1% of outliers is arbitrary, we calculated our P-values
256 based on 10,000 permutations. In each iteration, we shuffled the season and latitude labels
257 and reran the generalized regressions. Using the P-values obtained from regressions in which
258 season and latitude labels were shuffled, we defined for each iteration the top clinal, seasonal
259 and composite SNPs. Then, we calculated the enrichment of each genic class using Fisher's
260 exact test for all iterations. To obtain a P-value for an enrichment of a given genic class, we

261 compared the observed odds ratio in the actual dataset to the distribution of odds ratios
262 observed for datasets in which season and latitude labels were shuffled.
263 We also performed Gene Ontology (GO) enrichment analysis to identify the
264 biological function underlying our sets of outliers (Ashburner et al. 2000). One common issue
265 with GO analysis is that they are biased, as longer genes are likely to harbor more SNPs. We
266 thus used Gowinda (Kofler and Schlötterer 2012), an unbiased (*i.e.*, that accounts for gene
267 length) permutation approach to GO analysis. We obtained annotations from Flybase r6.15
268 and used the following parameters in Gowinda: 10,000 simulations; a minimum of 1 gene per
269 GO category, and considered SNPs 1000bp up and downstream of a gene.
270 Some specific genes (Table 1) have been shown to be strongly involved with clinal or
271 seasonal variation in *D. melanogaster (*Vigue and Johnson 1973; Oakeshott et al. 1982; Berry
272 and Kreitman 1993; Schmidt et al. 2000; Verrelli and Eanes 2001; Duvernell et al. 2003;
273 Sezgin et al. 2004; Schmidt and Paaby 2008; Cogni et al. 2014; Paaby et al. 2014; Behrman
274 et al. 2018). Therefore, we sought to test whether they were enriched among the outliers of
275 our three statistics (clinal P-value, seasonal P-value and composite statistic). We performed
276 permutation tests comparing the proportion of outlier to non-outlier SNPs within a gene to
277 the number expected from a random sample of SNPs (excluding SNPs in intergenic regions).
278 The permutation allowed us to control for the fact that different genes harbor different
279 number of SNPs.

280 **Mitigating the impact of linkage disequilibrium**

281 Selection at one site affects genetic variation at nearby, linked neutral sites (Smith and Haigh
282 1974). Because we assume that sites are independent in our models, the indirect effects of
283 selection can inflate the magnitude of the patterns we investigated. To test the effect of
284 linkage disequilibrium (LD) on our outlier analyses, we computed for all SNPs how their
285 distance to a top SNP affects their P-value or composite statistic. We then smoothed the

286 scatterplot using cubic splines as implemented in ggplot2 (Wickham 2016). To test the effect
287 of linkage on the relationship between clinal and seasonal variation, we implemented a
288 thinning approach. Sampling one SNP per *L* base pairs one thousand times, we constructed
289 sets of SNPs with minimized dependency, where *L* ranged from 0 to 20kb. For each of these
290 sets for a given *L*, we regressed clinal against seasonal slopes, and compared the distribution
291 of the thinned regression coefficients to the coefficients we obtained using all SNPs.

292 **Inferring proportion of selected SNPs**

293 Although a negative, significant relationship between clinal and seasonal variation indicates
294 that a subset of the polymorphisms is under selection, it is difficult to measure how a given
295 slope is related to a biological parameter, such as the proportion of selected SNPs. Most of
296 the observed polymorphisms should be neutral with only a few under direct selection,
297 therefore we expect the relationship between clinal and seasonal change to be minor (i.e., for
298 most SNPs clinal and seasonal change should be unrelated). To get a sense of how the
299 proportion of SNPs with uncoupled clinal and seasonal change would impact our measure of
300 interest, the regression coefficient between clinal and seasonal variation, we performed a set
301 of simple simulations.

302 The polymorphisms most affected by the environment are going to be the ones that
303 change most consistently with seasons and latitude. For the subset of polymorphisms under
304 selection, we expect the relationship between the z-normalized regression coefficients for
305 latitude and seasons to be negative one. This is equivalent to assuming that the rank order of
306 clinal and seasonal change should be the same. Biologically, this would happen if the
307 environment changed in the same ways with latitude and seasons. Although this is probably
308 not the case, we stress that this assumption makes our test conservative, given variants under
309 latitudinal and seasonal selection might not fall along the 1:-1 line, even though we are
310 assuming they should.

311 Under this framework, we simulated datasets with different proportions of points
312 falling along a line ($y = -x$), representing selected SNPs and points without a linear
313 relationship, representing SNPs that are not under selection. For each simulation replicate, we
314 calculated the linear relationship between two variables with varying proportions of points
315 falling on the 1:-1 and points completely unrelated. Each value was drawn from the combined
316 distribution of observed z-normalized regression coefficients for season and latitude. By
317 calculating the observed slope between the regression coefficients of each of our simulated
318 datasets, we could establish a relationship between observed slope and the expected
319 proportion of SNPs that fall along 1:-1 line (or that are under selection). We then used
320 Approximate Bayesian Computation using a sequential sampling scheme (as implemented in
321 the R package EasyABC; Jabot et al. 2013) to estimate the posterior distribution of
322 proportion of selected SNPs that could explain our observed relationship between clinal and
323 seasonal change (using our thinned estimate to alleviate the effects of linkage
324 disequilibrium).

325 **Results**

326 We assembled 21 *D. melanogaster* population samples collected from seven localities across
327 multiple years in the east coast of the United States. All of these samples are the result of a
328 collaborative effort of many researchers from a consortium, the DrosRTEC (Bergland et al.
329 2014; Machado et al. 2018). Seven of our samples span from Florida to Massachusetts and
330 together comprise our clinal set. The seasonal samples were collected in Pennsylvania in the
331 spring (6 samples collected in June or July) and in the fall (7 samples collected in September,
332 October or November). For each sample, a median of 55 individuals (with a range of 33 to
333 116) was pooled and resequenced to an average 75x coverage (ranging from 17 to 215). We
334 also used four clinal samples from the Australia (Anderson et al. 2005; Kolaczkowski et al.
335 2011). More details about the samples can be found on Table S1 (also see Machado et al.

336 2018; Bergland et al. 2014). After all the filtering steps, we identified 572,914 common
337 autosomal SNPs, which were used in our downstream analyses.

338 **Latitude explains more variation than seasons**

339 Latitude explains much of allele frequency variation along the surveyed populations, as there
340 is an excess of low GLM P-value SNPs (Fig. S1A). The mean absolute difference in allele
341 frequency between the ends of the clines is 9.8%. Seasons, on the other hand, explain less of
342 the variation in allele frequency. There is a minor excess of low GLM P-value SNPs (Fig.
343 S1B) and the mean absolute difference in allele frequency between seasons is 2.75%.

344 Given we do not have enough information on the selective pressures and migration
345 patterns in nature to build an appropriate null distribution to calibrate our P-values, we sought
346 to demonstrate that top significant clinal and seasonal SNPs are enriched for true positive
347 SNPs that underlie adaptation. Latitudinal SNPs are more likely to be in exonic, UTR 3' and
348 UTR 5' regions (Fig. 1A), whereas seasonal SNPs are enriched for exonic and UTR 5'
349 regions (Fig. 1B). The enrichment of top latitudinal SNPs among functional genic classes has
350 been reported by previous studies which used different latitudinal samples (Fabian et al.
351 2012; Kolaczowski et al. 2011, Machado et al. 2016). Further, enrichment of top seasonal
352 SNPs among functional regions have been observed before (Bergland et al. 2014; Machado et
353 al. 2018). The distribution of the top SNPs along the genome is shown in Fig. S8 and see Fig.
354 S2 for the enrichment test using a 5% cut-off. Our P-values were obtained from a
355 permutation test in which we shuffled season and latitude labels to build our null distribution.

356 **Clinal variation is related to seasonal variation**

357 A clinal pattern can arise solely as a result of demographic processes, such as isolation by
358 distance and admixture (Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016). Although
359 seasonality is less affected by such processes, seasonal change is less pronounced and more
360 subject to stochastic changes in the environment, making it harder to detect seasonal change

361 with precision. Here, we integrate both clinal and seasonal change estimates across a large
362 number of SNPs in the genome (about 580,000 SNPs). We expect the overall pattern that
363 emerges to be informative of the relative role of natural selection, because selection is a
364 plausible process to produce a pattern of clinal variation mirroring seasonal variation (Cogni
365 et al. 2015). With higher latitudes, cold-favored alleles rise in frequency, whereas from spring
366 to fall, the frequency of cold-favored alleles decreases. Conversely, with higher latitudes,
367 warm-favored alleles decrease in frequency, whereas from spring to fall, the frequency of
368 warm-favored alleles increases.

369 To assess if clinal variation and seasonal variation are related, we performed a linear
370 regression of the slopes of clinal and seasonal allele frequency change (see Methods). We
371 find a significant negative linear relationship between clinal and seasonal regression
372 coefficients (Fig. 2A, S3, S4 and S5; Table S2). The relationship is strongest for SNPs within
373 exons, and the weakest for unclassified SNPs and those within intergenic regions (Fig. 2A;
374 Table S3 and S5). Nonetheless, the relationship is different than zero for all classes (except
375 for the unclassified), what would be consistent either with pervasive linked selection or
376 widespread distribution of variants that are important for adaptation, even within non-coding
377 regions. We stress that our results are robust to modelling assumptions, as they are significant
378 when compared to the datasets with permuted of latitude and season labels (Fig. 2A).

379 Given that previous studies have demonstrated the importance of cosmopolitan
380 inversions in climatic adaptation (Kapun et al. 2016), we looked at how clinal and seasonal
381 variation are related within common cosmopolitan inversions. Our results are in concordance
382 with previous findings, in that SNPs within the region In(3R)P spans have a strikingly strong
383 negative relationship between clinal and seasonal variation (Fig. 2B; Table S4). It is
384 noteworthy that the difference between the genome-wide estimate and the estimate using

385 only SNPs outside inversions is small, suggesting that inversions alone do not explain our
386 main result, and variants relevant for adaptation are widespread across the genome.
387 Another way of testing for parallelism between clinal and seasonal change is by
388 testing if clinal SNPs are more likely to be seasonal. We observed that clinal SNPs are
389 enriched for seasonal SNPs (Fig. 2C). The enrichment increases with more stringent lower P-
390 value quantile cut-offs, as we would expect if even strictly non-significant variants were
391 informative of the role of selection.
392 We also confirmed our main finding, that clinal and seasonal change are related using
393 samples from Australia. There is a genome-wide relationship between clinal variation in
394 Australia and seasonal variation in Pennsylvania that, although minor, is significant (Fig.
395 S6A). We also found a relationship between latitudinal regression significance and seasonal
396 significance using the Australian samples (Fig. S6B). To measure clinal change in Australia,
397 we only used four low coverage samples in Australia, two for each of low and high latitude
398 locations, thus our power was tremendously reduced.

399 **Top SNPs for clinal and seasonal variation are enriched for metabolic functions**

400 Our approach to detecting selection relies on a large set of SNPs, for which we fit a
401 regression of latitudinal against seasonal regression coefficients. It is not possible, however,
402 to single out individual polymorphisms and genes that are being driven by spatial and
403 seasonal selection. Thus, we implemented a composite statistic that incorporates clinal and
404 seasonal P-values, as well as the difference in sign (direction) of clinal and seasonal
405 regression coefficients.
406 The composite statistic outliers greatly overlap with clinal and seasonal variants (Fig.
407 S8). The overlap is greater among clinal and composite sites, as expected given clinal P-
408 values are generally much smaller than seasonal P-values. Nevertheless, the statistic
409 incorporates the difference in sign that can be helpful in distinguishing spurious from causal

410 variants involved with adaptation. Consistent with this reasoning, even though about one
411 third of the composite outliers are unique, outlier SNPs for the composite statistic are
412 enriched among exonic, UTR 3', and UTR 5' regions and depleted within intergenic regions
413 (Fig. 1C). The enrichment among genic regions seen for the composite statistic is comparable
414 to those for latitudinal P-values, suggesting that incorporating both clinal and seasonal
415 variation can inform us adaptive processes in natural populations.

416 We characterized the biological function of the outlier SNPs for each of these sets:
417 clinal SNPs, seasonal SNPs, and top composite SNPs. At an FDR corrected P-value < 0.05,
418 we found enrichment for 84 GO terms for the clinal SNPs, no enrichment for the seasonal
419 SNPs and five GO terms for the composite statistic SNPs (Tables S6, S7, S8). Although
420 caution is needed when interpreting GO analysis results (Pavlidis et al. 2012), some of these
421 enriched terms are involved in metabolism. These results agree with the previous observation
422 that metabolic genes are central to clinal adaptation and vary latitudinally and seasonally
423 (Lavington et al. 2014; Cogni et al. 2015).

424 There are many examples of genes that were empirically implicated with clinal and
425 seasonal adaptation in *D. melanogaster* (Vigue and Johnson 1973; Oakeshott et al. 1982;
426 Berry and Kreitman 1993; Schmidt et al. 2000; Verrelli and Eanes 2001; Duvernell et al.
427 2003; Sezgin et al. 2004; Schmidt and Paaby 2008; Cogni et al. 2014; Paaby et al. 2014;
428 Behrman et al. 2018). We thus tested whether those genes were enriched for top clinal,
429 seasonal and composite statistic SNPs (i.e., SNPs within introns and exons of a gene). Out of
430 10 genes with strong support from previous studies, four had fewer than ten SNPs in our
431 dataset, and thus we had no power to look for enrichment. Two out of the other six genes
432 were enriched for top clinal SNPs, one for top seasonal SNPs, and three for the composite
433 statistic (Table 1). We found an excess of top latitudinal, seasonal and composite SNPs in the
434 couch potato gene (*cpo*), which is associated with diapause and is known to harbor SNPs that

435 vary clinally and seasonally (Schmidt and Paaby 2008; Cogni et al. 2014). The insulin
436 receptor (*InR*) gene, which was shown to have seasonal and clinal variants (Paaby et al.
437 2014), is enriched with both latitudinal and composite SNPs (Table 1). Adh, a metabolic gene
438 implicated in climatic adaptation, shows an excess of top composite SNPs (Vigue and
439 Johnson 1973; Berry and Kreitman 1993; Sezgin et al. 2004). The genes *Tep2* and *Tep3*,
440 which have been associated with seasonal variation in immunity (Behrman et al. 2018), do
441 not show an excess of outlier SNPs for any of our statistics (Table 1). These results further
442 highlight the utility of the composite statistic, and could be used to guide future functional
443 studies.

444 **The effects of linkage disequilibrium on clinal and seasonal variation**

445 Variation at one site is linked to variation at other sites, and selection will increase this
446 dependency. To mitigate this effect, we first assessed if latitudinal and seasonal P-values, and
447 our composite statistic were dependent on how distance a SNP was from our top 1% SNPs.
448 We show that all statistics are dependent on distance from our outlier SNPs (Fig. 3A-C), but
449 the effect virtually disappears after 5kb.

450 Our regression analysis of clinal and seasonal variation assumes that SNPs are
451 independent. Thus, we assessed the impact of linkage on the relationship between clinal and
452 seasonal change by implementing a thinning approach. First, we tested how the genome-wide
453 regression estimate varied with changing window sizes. The dependency among sites affects
454 our estimates, but levels off after 5kb (Fig. 3D). Nevertheless, we chose a conservative
455 window length of 10kb for downstream analyses. The strength of the genome-wide
456 relationship between clinal and seasonal variation decreased from -0.041 to -0.028 when
457 randomly sampling one SNP per 10kb (P= 0.0449), but it still remained significantly
458 different from zero (P= 0.0029; Fig. 3E). The thinning did not significantly impact the signal
459 for exonic regions, but the strength of the signal within intronic regions significantly

460 decreased (P= 0.0209; Fig. 3E). For many of the thinned datasets, the regression coefficient
461 within intronic and intergenic regions reached values close to zero, as we would expect if
462 most of the signal within these regions were caused by linked selection. This suggests that the
463 approach is efficient in mitigating the impact of linked selection on our estimates.

464 **Selection accounts for a substantial proportion of genome-wide clines**

465 Although we identified a significant negative relationship between clinal and seasonal
466 variation, the magnitude of the regression coefficient appears small (10kb bootstrap estimate
467 $\beta = 0.0277$). It is hard to measure the biological relevance of such coefficient, and so we
468 designed a simple simulation approach that translates our statistic of interest into a biological
469 parameter, that is the proportion of selected sites. We assume that: (i) for the subset of
470 variants that are under selection, we would expect to see a perfectly linear relationship
471 between z-normalized clinal and seasonal regression coefficients; (ii) for the variants that are
472 not contributing to adaptation, then we would expect absolutely no relationship between
473 clinal and seasonal variation; (iii) variants are not linked. Assumption (i) renders our test
474 conservative, as the relationship would not necessarily be 1:-1 because some variants may be
475 strongly seasonal but not necessarily strongly latitudinal and vice-versa. Accounting for
476 linkage disequilibrium (assumption iii) is notoriously complicated, especially because we
477 cannot accurately measure LD from pooled sequencing (Feder et al. 2012). However, we
478 show that most of the effects of autocorrelation can be accounted for with our thinning
479 approach, thus we used the observed relationship between clinal and seasonal change with
480 10kb thinning.

481 We asked what the expected relationship between clinal and seasonal regression
482 coefficients would be if some of the variants were subject to spatially and seasonally varying
483 selection but most were neutral. To assess this, we built datasets with varying proportion of
484 adaptive variants (with linear relationship between clinal and seasonal variation) to neutral

485 variants (no relationship between clinal and seasonal variation) and calculated the expected
486 linear relationship between the two variables.
487 We found that the observed regression coefficient is consistent with approximately
488 2.8% of the common variants being under selection (Fig. 4), with 95% of the posterior
489 estimates ranging from 2.5% to 3%. Thus, our analysis suggests that a sizable fraction of
490 common, autosomal SNPs changes due to selection.

491 **Discussion**

492 Clinal patterns have been observed in both phenotypic and genotypic traits in many different
493 species (Hancock et al. 2008; Baxter et al. 2010; Adrion et al. 2015). Especially in systems in
494 which there is collinearity between the axis of gene flow and environmental heterogeneity,
495 disentangling the contribution of selection and demography to producing clines is not trivial.
496 Detecting seasonal cycling in allele frequencies is also challenging, mostly because the effect
497 size is likely to be small and the environment changes unpredictably within seasons. We
498 relied on integrating patterns of clinal and seasonal variation (Cogni et al. 2015), since
499 variation over short periods is independent of most confounding demographic processes
500 (Bergland et al. 2014). Parallel environmental change and the differences in the onset of
501 seasons with latitude can create parallelism between clinal and seasonal variation. Natural
502 selection would be the process driving the coupling between latitudinal and seasonal change
503 in both scenarios, thus investigating the relationship between clinal and seasonal change can
504 help discern the role of natural selection. Here, using *D. melanogaster* samples from the east
505 coast of the US, we show that (i) there is a genome-wide pattern of clinal variation mirroring
506 seasonal variation, (ii) this is consistent with the action of natural selection, and (iii) a
507 significant number of common variants are being driven by selection.
508 We show that clinal and seasonal change are related, as there is a significant negative
509 relationship between clinal and seasonal regression coefficients in the *D. melanogaster*

510 genome (Fig. 2A), and there is an enrichment for seasonal SNPs among clinal SNPs (Fig.
511 2C). The relationship between clinal and seasonal regression is stronger for genic when
512 compared to intergenic regions (Fig. 2A). Demographic processes would impact the genome
513 as a whole, but variants in genic regions are more likely to impact fitness (Andolfatto 2005).
514 We estimate that about 2.8% of common, autosomal SNPs should be under selection to
515 explain the genome-wide pattern of clinal and seasonal concordance uncovered (Fig. 4).
516 Further, we show that some genes previously known for their role in local adaptation are
517 enriched for variants that vary seasonally and latitudinally (Table 1).

518 Chromosomal inversions segregate along the cline and seasonally in Pennsylvania,
519 and thus could impact our estimates (Kapun et al. 2016). Although inversions do contribute
520 to adaptation (Fabian et al. 2012; Kapun et al. 2016), we demonstrate that the genome-wide
521 pattern of clinal reflecting seasonal variation is not being driven solely by inversions (Fig.
522 2B) and previous studies have also shown that not all clinal and seasonal variation can be
523 attributed to inversions (Fabian et al. 2012; Bergland et al. 2014; Machado et al. 2018). More
524 generally, because we expect selection to intensify linkage disequilibrium, the relationship
525 between clinal and seasonal variation could be mostly driven by a few large effect loci.
526 However, we show that our results cannot be fully explained by linked selection (Fig. 3).
527 When accounting for autocorrelation along the chromosomes, the relationship between clinal
528 and seasonal change disappears for variants within intergenic regions, but not for exons (Fig.
529 3E). The effects of linkage disequilibrium are important in this short time scale, but it decays
530 rapidly, and returns to background levels after 5kb (Fig. 3A-D). Therefore, our results
531 support a scenario where the genome-wide pattern of clinal variation mirroring seasonal
532 variation is a result of selection acting on many sites across the genome.

533 An important mechanism that can cause clinal patterns has been neglected from
534 recent discussions of clinal variation in *Drosophila*. The latitudinal variation on the onset of

535 seasons can produce clines, a phenomenon termed “seasonal phase clines” (Roff 1980;
536 Rhomberg and Singh 1986). Under this model, a correlation between clinal and seasonal
537 change is expected. Our latitudinal samples were all collected within one month of difference
538 (during the spring), and so our observations could be partially explained by differences in the
539 seasonal phase. Seasonality may be causing some of the patterns we uncovered, as we found
540 the relationship between clinal and seasonal change is strong when considering only the top
541 1% clinal SNPs (Fig. S5B), though the pattern is not replicated for the top 1% seasonal SNPs
542 (Fig. S5C). Our data does not allow for proper disentangling of seasonal phase clines from
543 parallel environmental change, but change on the onset of seasons alone cannot explain our
544 results. We found that latitude is usually a much better predictor of allele frequency
545 differences (Fig. S1), and the magnitude of change along the cline is much greater than what
546 we found within a population across seasons (9.8% vs. 2.75%). We show that including
547 Maine samples (which were sampled in the fall) in our analyses does not meaningfully
548 change our main results (Fig. S4). However, the magnitude of change in allele frequency
549 between Florida and Maine is smaller (mean change of 9.1% vs. 9.8% from Florida to
550 Massachusetts), which may be due to the difference in seasonal phase, as the samples from
551 Maine were collected in the Fall, but the sample from Massachusetts was collected in the
552 Spring.

553 Importantly, our results should be robust to differential admixture from Europe and
554 Africa to the ends of the clines (Duchen et al. 2013; Kao et al. 2015; Bergland et al. 2016).
555 Given that variation over seasonal time scales are less affected by broader scale migration
556 patterns, it is not plausible that secondary contact can create parallel clinal and seasonal
557 variation. Furthermore, the evidence for secondary contact in Australia is quite weak (see
558 Bergland et al. 2016), but we show that clinal variation in Australia is related to seasonal
559 variation in Pennsylvania (Fig. S6). Secondary contact may have contributed ancestral

560 variation, that have since been selectively sorted along the cline (Flatt 2016). Consistent with
561 this interpretation, the proportion of African ancestry is lower in low recombination regions,
562 suggesting selection is mediating admixture in *D. melanogaster* (Pool 2015).

563 Population substructure and migration could cause seasonal variation in allele
564 frequency in *D. melanogaster*. One possibility is that rural populations of *D. melanogaster* in
565 temperate regions collapse during the winter and recover from spring to autumn. However,
566 reproductive diapause cycles in orchards and reaches high frequencies early in the spring,
567 whereas its frequency in urban fruit markets in Philadelphia is much lower (Schmidt and
568 Conde 2006). Another possibility is that seasonal variation is produced by migration of flies
569 from the south in the summer, and from the north in the winter. There is little evidence that
570 long range migration produces clines in *D. melanogaster*, though this process seems
571 important in *D. simulans* (Bergland et al. 2014; Machado et al. 2016). Importantly, flies have
572 been shown to survive and reproduce during winter season in temperate regions, so rural flies
573 can withstand a harsh winter season and be subject to selection (Mitrovski and Hoffmann
574 2001; Hoffmann et al. 2003). More recent studies have also shown that fly populations adapt
575 to seasonal variation in outdoor cages (e.g., Rudman et al. 2019). These seasonal patterns
576 have been replicated in many populations across North America and Europe (Machado et al.
577 2018), bolstering the argument for seasonal selection. Regardless of the main process behind
578 seasonal change, comparing clinal and seasonal variation would still be worthwhile, as it
579 provides another way of independently assessing change in allele frequencies.

580 A recent study suggested the temporal change in allele frequencies reported in
581 Bergland et al. (2014) is not consistent with seasonal fluctuations driven by temporal
582 selection (Buffalo et al. 2019). Though we use a different set of samples, we show that the
583 distribution of seasonal P-values is close to uniform (Fig. S1), and there is only a slight
584 enrichment for coding SNPs. Finding truly seasonal SNPs is not trivial because the change in

585 allele frequency is generally much smaller and the populations are subject to stochastic
586 changes in the environment not related to seasons. For instance, Machado et al. 2018 found
587 that when they flipped the season labels of some samples the seasonal model fit was greatly
588 improved. On the other hand, finding SNPs driven by selection in space is hard because of
589 demography as a confounder. The main rationale behind our work is that integrating
590 signatures of clinal change with seasonal change can increase power to detect signatures of
591 selection. We show that the relationship between seasonal and latitudinal change disappears
592 when season labels are permuted (Fig. 2D). Thus, our results suggest that there is at least
593 some seasonal effect in our samples that is not pure noise.

594 A few previous studies have investigated the relationship between clinal and seasonal
595 variation (Bergland et al. 2014; Cogni et al. 2015; Kapun et al. 2016; Machado et al. 2018),
596 but here we further explore parallelism between clinal and seasonal variation to tease apart
597 the role of selection. Cogni et al. (2015) focused on a few SNPs in metabolic genes, and so
598 they could not draw robust conclusions about processes that impact most of the genome. Our
599 work is also fundamentally different from that of Machado et al. (2018), which revolves
600 around identifying seasonal variation in *D. melanogaster* by aggregating data from many
601 locations in North America and Europe. They also investigated the relationship between
602 clinal and seasonal variation, but do not interpret how parallel clinal and seasonal change can
603 help us identify the role of selection. In addition, because our interest is to ultimately
604 understand processes driving clinal variation in the east coast of the US, we only included
605 seasonal samples from Pennsylvania. This location had the greatest number of years surveyed
606 (7 years, as opposed to 2 years for most of the other locations; see Machado et al. 2018), and
607 it is the population for which we have the most information about how phenotypes vary
608 seasonally (Schmidt and Conde 2006; Behrman et al. 2015; Rajpurohit et al. 2017; Behrman
609 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 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).

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.

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.

Anderson, A. R., Hoffmann, A. A., McKechnie, S. W., Umina, P. A., & Weeks, A. R. 2005. The latitudinal cline in the In (3R) Payne inversion polymorphism has shifted in the last 20 years in Australian Drosophila melanogaster populations. Molecular Ecology. 14(3), 851-858.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature. 437(7062), 1149.

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.

Berry A, Kreitman M. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in Drosophila melanogaster on the east coast of North America. Genetics 134(3):869-93.

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

Božičević V, Hutter S, Stephan W, Wollstein A. 2016. Population genetic evidence for cold adaptation in European Drosophila melanogaster populations. Molecular ecology 25(5):1175-91.

Buffalo V, Coop G. 2019. Estimating the genome-wide contribution of selection to temporal allele frequency change. BioRxiv, 798595.

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.

Caracristi, G., and C. Schlötterer. 2003. "Genetic Differentiation between American and European Drosophila Melanogaster Populations Could Be Attributed to Admixture of African Alleles." Molecular Biology and Evolution 20 (5): 792–99.

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.

Cogni R, Kuczynski K, Koury S, Lavington E, Behrman EL, O'Brien KR, Schmidt PS, Eanes WF. 2017. On the long-term stability of clines in some metabolic genes in Drosophila melanogaster. Scientific reports 7:42766.

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 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.
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.
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 inDrosophila 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 B ayesian computation sampling 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.
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.
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 glycerol-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, 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.
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. Genetica78(1):73-9.
Roff D. 1980. Optimizing development time in a seasonal environment: the 'ups and downs' of clinal variation. Oecologia 45(2):202-8.
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. 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.

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.

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).

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.

Table 1. Enrichment for outlier SNPs in genes involved with clinal and seasonal adaptation.

		Top Clinal		Top Seasonal		Top Composite	
Gene	SNPs	Top SNPs	P	Top SNPs	P	Top SNPs	P
Adh	11	1	0.1033	0	1.0000	2	**0.0059**
cpo	821	32	**0.0001**	14	**0.0439**	25	**0.0001**
Gpdh	10	0	1.0000	0	1.0000	0	1.0000
Hex-C	2	0	-	0	-	0	-
InR	167	25	**0.0001**	2	0.4928	40	**0.0001**
mth	7	0	-	0	-	0	-
Pgm	1	0	-	0	-	0	-
Tep2/3	190	0	1.0000	0	1.0000	0	1.0000
Treh	28	1	0.2409	0	1.0000	1	0.2518
UGP	1	0	-	0	-	0	-

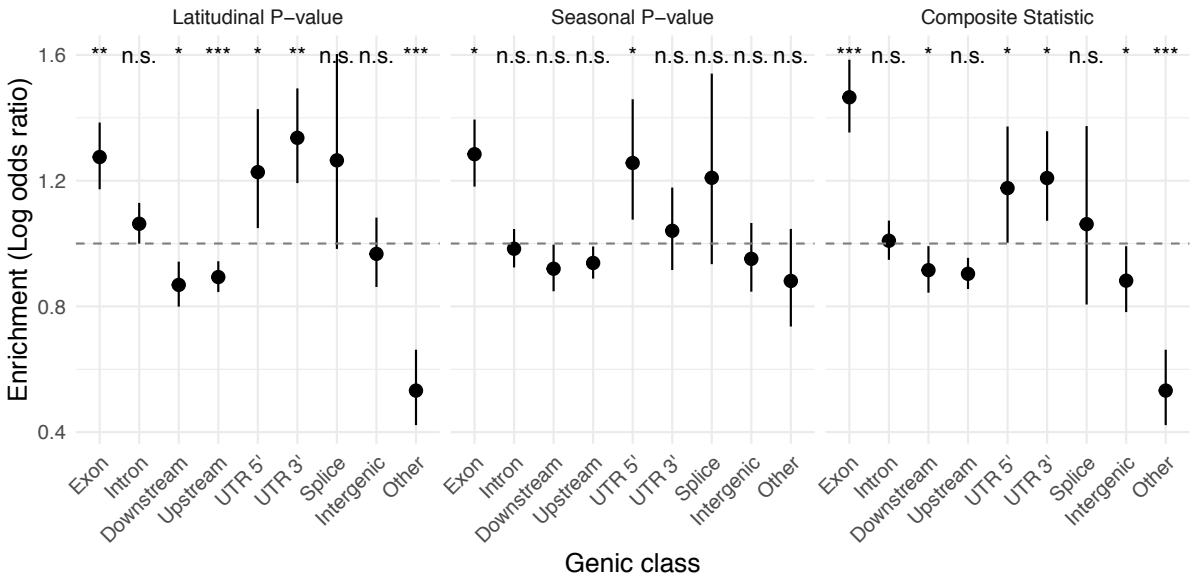

Figure 1. Top SNPs are enriched for functionally relevant classes. Enrichment of top 1% SNPs in each genic class for A) latitudinal P-value, B) seasonal P-value, and C) composite statistic. The composite statistic takes into account the latitudinal and seasonal P-value, and the difference in sign between latitudinal and seasonal regression coefficients. Error bars are 95% CI. * P-value < 0.05, ** P-value < 0.005, *** P-value < 0.0005, **** P-value < 0.00005.

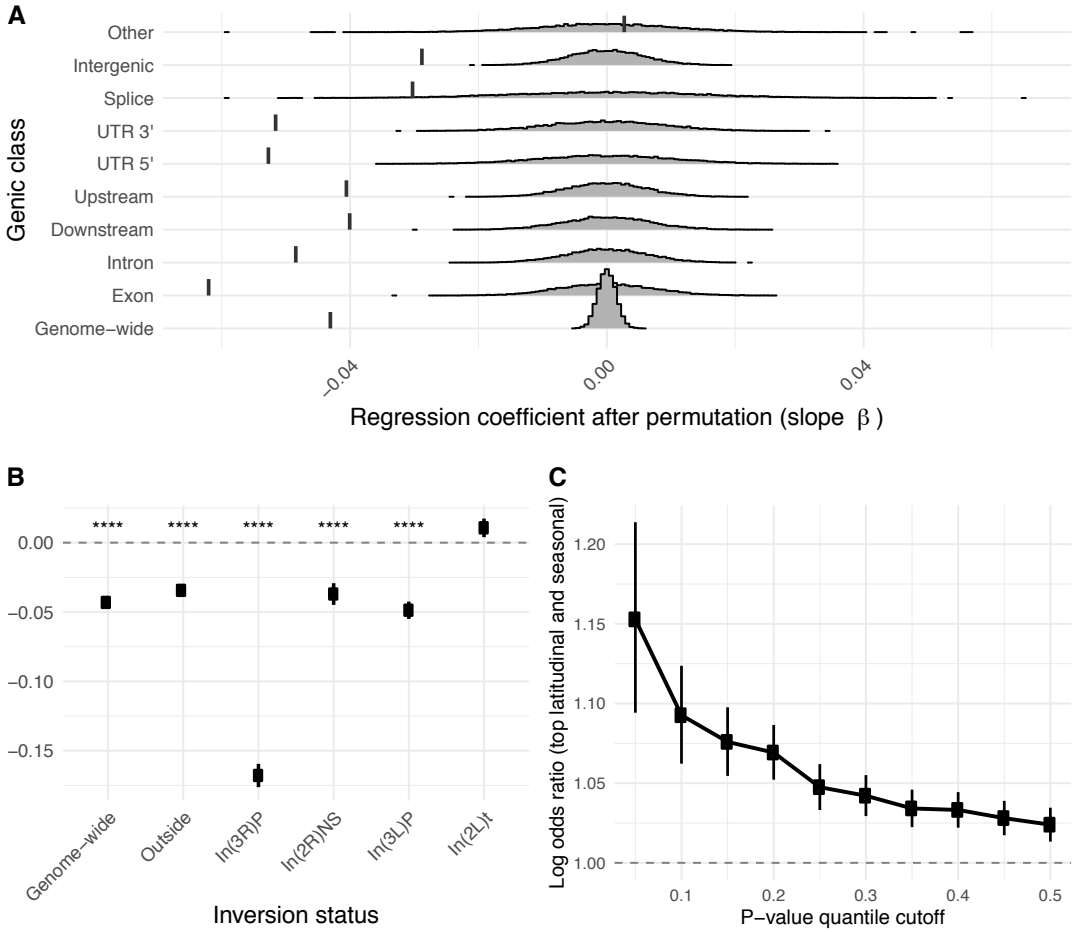

Figure 2. Clinal variation mirrors seasonal variation. A) Null distribution of the regression coefficients between latitudinal and seasonal change for each genic class (after permuting the latitude and season labels). Vertical bars represent the observed values. B) The relationship between clinal and seasonal change depends on inversion statuses. C) Latitudinal and seasonal significance are associated. Error bars are 95% CI. * P-value < 0.05, ** P-value < 0.005, *** P-value < 0.0005, **** P-value < 0.00005.

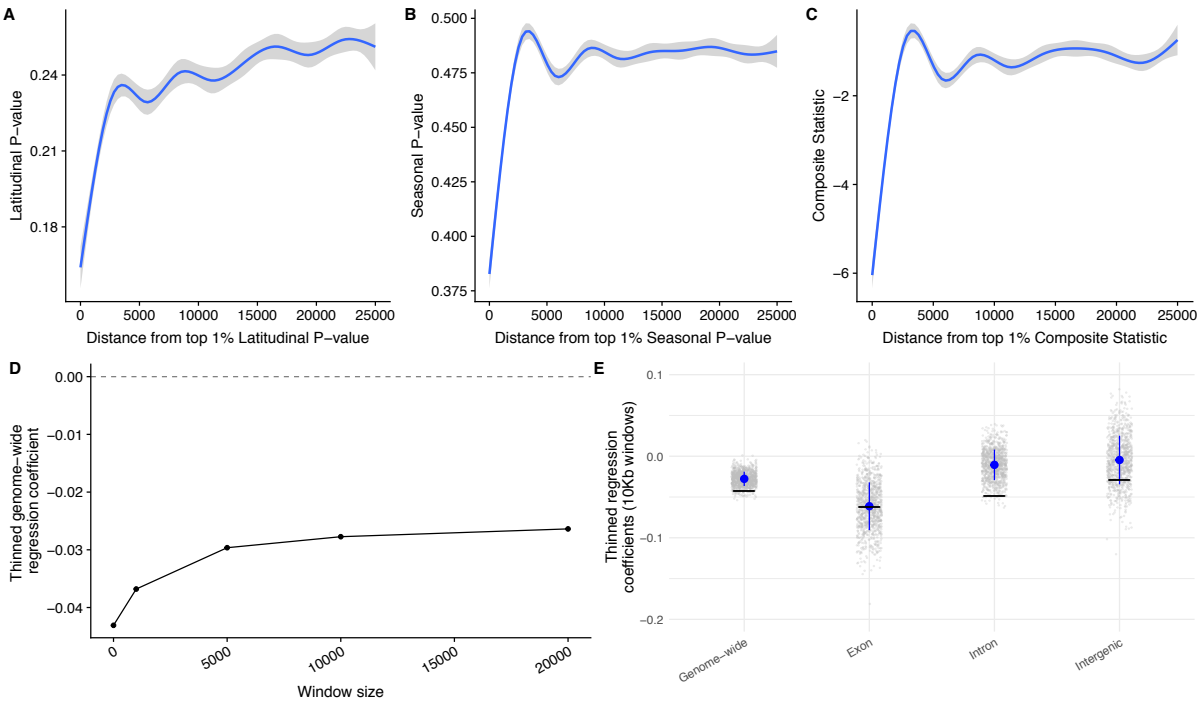

Figure 3. Effects of linkage disequilibrium. The A) latitudinal P-value, B) seasonal P-value and C) composite statistics are dependent on the distance to the respective top 1% outliers. The relationship between clinal and seasonal variation is affected by dependency among SNPs. D) the strength relationship between clinal and seasonal variation decreases with the size of the thinning window. E) comparison among original estimates (bar) and values obtained after thinning using a window size of 10kb (point-range). Point-ranges show the mean and standard deviation across simulations and gray dots in the background the values for each simulation.

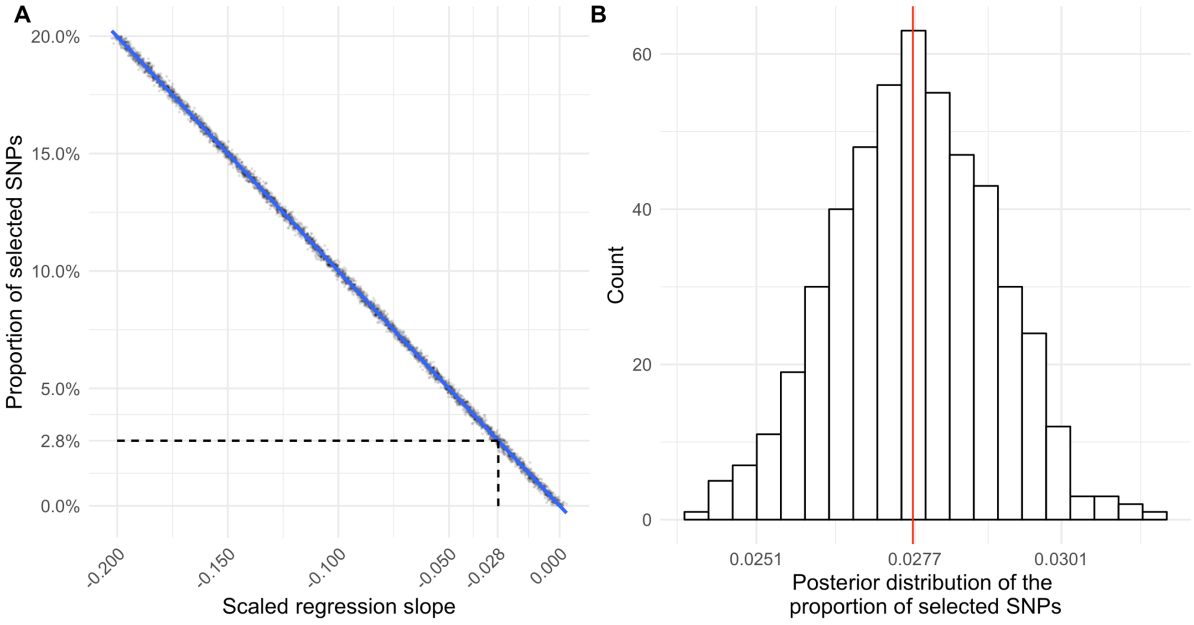

Figure 4. Proportion of SNPs that should be under selection to explain the genome-wide relationship between clinal and seasonal variation. A) background points represent each simulation, and the line is the fitted loess curve. Dashed lines relate the observed scaled (or z-normalized) regression coefficient after thinning (10kb) of our data 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 (10kb thinned), with horizontal line highlighting the mean.