Broad geographic sampling reveals predictable, pervasive, and strong seasonal adaptation in Drosophila

Heather E. Machado[1,2*‡], Alan O. Bergland[1,3*‡], Ryan Taylor[1], Susanne Tilk[1], Emily Behrman[4], Kelly Dyer[5], Daniel K. Fabian[6,7], Thomas Flatt[6,8], Josefa González[9], Talia L. Karasov[10], Iryna Kozeretska[11,12], Brian P. Lazzaro[13], Thomas JS Merritt[14], John E. Pool[15], Katherine O'Brien[4], Subhash Rajpurohit[4], Paula R. Roy[16], Stephen W. Schaeffer[17], Svitlana Serga[9], Paul Schmidt[4○‡], Dmitri A. Petrov[1○‡]

* equal contribution
° equal contribution
‡ to whom correspondence should be addressed: heather.machado@sanger.ac.uk, aob2x@virginia.edu, schmidtp@sas.upenn.edu, dpetrov@stanford.edu

[1] Department of Biology, Stanford University, Stanford, CA
[2] Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK
[3] Department of Biology, University of Virginia, Charlottesville, VA
[4] Department of Biology, University of Pennsylvania, Philadelphia, PA
[5] Department of Genetics, University of Georgia, Athens, GA 30602
[6] Institute of Population Genetics, Vetmeduni Vienna, A-1210 Vienna, Austria
[7] Centre for Pathogen Evolution, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, United Kingdom
[8] Department of Biology, University of Fribourg, CH-1700 Fribourg, Switzerland
[9] Institute of Evolutionary Biology, CSIC- Universitat Pompeu Fabra, Barcelona, Spain
[10] Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA
[11] Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[12] State Institution National Antarctic Center of Ministry of Education and Science of Ukraine, Taras Shevchenko Blvd., 01601, Kyiv, Ukraine
[13] Department of Entomology, Cornell University, Ithaca, NY, 14853
[14] Department of Chemistry & Biochemistry, Laurentian University, Sudbury, ON, Canada
[15] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI
[16] Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence KS 66045
[17] Department of Biology, The Pennsylvania State University, University Park, PA

Abstract

To advance our understanding of adaptation to temporally varying selection pressures, we identified signatures of seasonal adaptation occurring in parallel among *Drosophila melanogaster* populations. To study these evolutionary dynamics, we estimated allele frequencies genome-wide from flies sampled early and late in the growing season from 20 widely dispersed populations. We identify parallel seasonal allele frequency shifts across North America and Europe, demonstrating that seasonal adaptation is a general phenomenon of temperate fly populations. The direction of allele frequency change at seasonally variable polymorphisms can be predicted by weather conditions in the weeks prior to sampling, linking the environment and the genomic response to selection. The extent of allele frequency

fluctuations implies that seasonal evolution drives substantial (5-10%) allele frequency fluctuations at >1% of common polymorphisms across the genome. Our results suggest that fluctuating selection is an important evolutionary force affecting the extent and stability of linked and functional variation.

Introduction

Fluctuations in the environment are an inescapable condition for all organisms. While many of these fluctuations are unpredictable, some are predictable to a degree, including those that occur on diurnal and seasonal time scales. The predictability of cyclic fluctuations is reflected in the fact that many species exhibit plastic physiological and behavioral strategies that enable them to survive the unfavorable season and exploit the favorable one (Denlinger 2003; Kostál 2006); such plastic responses represent the classical form of seasonal adaptation (Tauber *et al.* 1986). However, seasonally varying selection can - in principle - maintain fitness related genetic variation if some genotypes have high fitness in one season but not another (Gillespie 1973). Thus, in the organisms that undergo multiple generations per year (multivoltine), a distinct form of seasonal adaptation occurs when the frequency of alternate genotypes changes in response to seasonal fluctuations in the environment.

Seasonal adaptation can be seen as a form of local adaptation but local adaptation in time – adaptation to temporally varying selection pressures - rather than in space. However, such adaptation in time has been considered by some to be uncommon and, when present, unlikely to result in long-term balancing selection (Ewing 1979; Hedrick 2006). Classic quantitative genetic theory suggests that an optimal, plastic genotype will eventually dominate a population that is exposed to periodically changing environments (Scheiner 1993). This is particularly so when certain environmental cues are reliable indicators of changes in selection pressure (Levins 1968; Via & Lande 1985). Predictions from traditional population genetic models suggested that periodically changing environments will lead to the rapid loss of seasonally favored ecotypes as slight changes in selection pressure from one year to another eventually push allele frequencies at causal alleles to fixation (Hedrick 1976).

Recent theoretical models have called these classical predictions into question. For instance, a quantitative genetic model by Botero et al. (Botero *et al.* 2015) examined whether adaptive tracking, plasticity, or bet-hedging evolve as a consequence of environmental fluctuations. They showed that adaptive tracking rather that plasticity is likely to become a predominant feature of populations living in seasonally fluctuating environments when populations undergo more than three generations per season. Additionally, a population genetic model by Wittmann et al (Wittmann *et al.* 2017) has demonstrated that seasonally varying selection can maintain fitness related genetic variation at many loci throughout the genome provided that dominance shifts from season to season in such a way that, on average, the seasonally favored allele remains even slightly dominant (Curtsinger et al 1994). These recent models, along with others that highlight the importance of population cycles (Bertram & Masel 2019), as well as overlapping generations and age structure (Ellner 1996; Ellner & Sasaki 1996; Ellner & Hairston 2015; Bertram & Masel 2019), suggest that seasonal adaptation and adaptive tracking (Kain *et al.* 2015) could be an important feature of multivoltine organisms such as Drosophila (Behrman *et al.* 2015). More generally, it is possible that adaptive tracking of environmental fluctuations on other time scales might be more common than generally acknowledged.

Despite the lack of theoretical agreement on whether and how seasonal adaptation operates in multivoltine organisms, there is substantial empirical evidence for seasonal adaptation in many organisms including *Drosophila*. Seasonal adaptation was first observed in *D. pseudoobscura* by Dobzhansky and colleagues (e.g., Dobzhansky 1948) by tracking allele frequencies of inversions over seasons. Later studies confirmed and extended these early findings to other species including *D. melanogaster* (Stalker 1976; Stalker 1980; Kapun *et al.* 2016) and *D. subobscura* (Rodríguez-Trelles *et al.* 2013).

In *D. melanogaster*, multiple additional lines of evidence from phenotypic and genetic analyses demonstrate the presence of seasonal adaptation. When reared in a common laboratory environment, flies collected in the spring show higher stress tolerance (Behrman *et al.* 2015), greater propensity to enter reproductive dormancy (Schmidt & Conde 2006), increased innate immune function (Behrman *et al.* 2018), and modulated cuticular hydrocarbon profiles (Rajpurohit *et al.* 2017) as compared to flies collected in the fall. Rapid adaptation over seasonal time scales in these and related phenotypes has also been observed in laboratory (Schmidt & Conde 2006) and field-based mesocosm experiments (Rajpurohit *et al.* 2017; 2018). Genome-wide analysis indicated that a large number of common polymorphisms change in frequency over seasonal time scales in one mid-latitude orchard (Bergland *et al.* 2014), and allele frequency change among seasons has been observed using candidate gene approaches (Cogni *et al.* 2014). In several cases, these adaptively oscillating polymorphisms have been linked to seasonally varying phenotypes (Paaby *et al.* 2014; Behrman *et al.* 2018).

Despite ample evidence of seasonal adaptation in *D. melanogaster*, many aspects of this system remain unexplored. First, we do not know whether seasonal adaptation is a general feature of *D. melanogaster* populations across its range. Previous work (Schmidt and Conde 2006, Bergland *et al.* 2014, Cogni *et al.* 2014, Behrman et al. 2015, Behrman *et al.* 2018) detected seasonal fluctuations in a single locality over the span of one to three years. However, we do not know whether these results of fluctuating selection can be extrapolated to other localities. Second, it is unclear how responsive such adaptation is to specific local environmental conditions. Do seasons generate entirely idiosyncratic selective pressures that are not shared across populations, or does Drosophila experience partly generic selective pressures that are common to temperate populations? A related question is whether seasonal adaptation across the range involves a partially shared set of seasonal polymorphisms. Finally, we do not yet have estimates of the magnitude of effect and the fraction of polymorphisms genome-wide that are affected by seasonal adaptation through direct or linked selection, both in terms of the number of causal variants and the overall stochastic effect on linked variants.

The most straightforward and statistically powerful way to start answering these questions is to assess the extent of parallel shifts in allele frequencies over seasonal time across the *D. melanogaster* range. Should such parallel shifts be detected we can then infer that seasonal adaptation is widespread and at least partly driven by a common set of variants. It will also help us determine the magnitude of the shifts and give us a glimpse into the genetic architecture of seasonal adaptation, for instance whether or not the variants are found genome-wide or clustered into linked blocks or even supergenes.

Examining patterns of allele frequency change on seasonal timescales across a species' range faces two problems. The first problem is logistical: the power of any analysis of parallel seasonal adaptation across dispersed populations requires a large sample size. Such sampling can become prohibitively expensive and logistically difficult for any single lab to carry out, and these difficulties are further exacerbated if heavy constraints are placed on the exact timing of the sampling. The second, and related problem, is the determination of the timing of sampling. We have an intuitive sense of what seasons are; however, it is not clear how to match seasons across locations. While it is possible, albeit logistically difficult, to sample flies according to some matching of seasons using the calendar window (e.g. first week of June for the "spring" and the first week of November for the "fall") or using a pre-determined physiological time (e.g., "spring" is the time after 21 cumulative growing degree days), it is not at all clear that any such matching would be meaningful in terms of the ecology, physiology, and evolution of local populations.

We have elected to solve these two related problems by (i) working as a consortium of Drosophila biologists (DrosRTEC or the Drosophila Real Time Evolution Consortium) and (ii) sampling in a somewhat haphazard manner. A similar effort is being carried out primarily in Europe by the parallel *Drosophila* Population Genomics Consortium DrosEU (Kapun *et al.* 2018). Both consortia began as a result of the 2012 Catalysis Meeting at the National Evolutionary Synthesis Center (NESCent) in Durham, NC, USA. These consortia organize a collective effort of Drosophila biologists who sample at their own locations which then in turn allows for the sampling to be carried out range-wide. We are continuing this effort and expanding the consortium going forward ("*Drosophila Evolution in Space and Time*", DEST) to meet this ambitious goal.

We chose to carry out haphazard sampling for logistical and inferential reasons. In this study, the "spring" sample was generally taken close to the time when the populations of *D. melanogaster* become locally abundant. The "fall" is either simply later in the summer or early fall, generally prior to the time that populations crash at the onset of winter. Beyond the impracticality of requiring that everyone collect flies at a given point in time we simply do not know when, exactly, would be the most appropriate time to collect flies or what the fine-scaled temporal or spatial factors matter for driving changes in allele frequency. Given this lack of knowledge, choosing a specific time point, collection method, or substrate limits the covariance of collection attributes with allele frequencies thereby preventing us from testing for specific gene-environment associations *a posteriori*. Thus, a haphazard approach might in fact have the greatest inferential power provided that the signal of seasonal adaptation can be detected in such a complex dataset.

Using allele frequency data from 20 paired seasonal samples from 15 localities in North America and Europe, we find evidence of parallel, pervasive, and strong seasonal adaptation in *D. melanogaster*. First, we demonstrate that seasonal adaptation is a general phenomenon that occurs in multiple populations of *D. melanogaster* on two continents. We provide evidence suggesting that at least some of the same polymorphisms cycle between seasons in populations sampled across vast geographic distances. Further, we show that allele frequency change between seasons is predictable to some extent when taking into account weather in the weeks immediately prior to sample collection, which hints at the complex and nonlinear nature of

seasonal adaptation. Seasonal alleles tend to show clinal variation, with the alleles that increase in frequency through the summer generally being more frequent in lower latitude locations. Finally, using a simulation approach, we show that the detected seasonal signal is consistent with substantial (5-10%) seasonal allele frequency fluctuations at more than 1% of all common polymorphisms. Taken together, our work demonstrates that seasonal adaptation is a general and predictable feature of *D. melanogaster* populations and has pervasive effects on patterns of allele frequencies genome-wide. More generally we establish that metazoan populations can exhibit adaptive tracking of environmental conditions on extremely short time scales, most likely shorter than a single growing season encompassing ~10 generations, suggesting that such rapid and cyclic evolutionary shifts cannot be discounted *a priori*.

Results

Fly samples and sequence data. We assembled 72 samples of *D. melanogaster* collected from 23 localities in North America and Europe (Supplemental Table 1, Supplemental Figure 1). For 15 sampling localities, flies were collected in the spring and fall over the course of one to six years (Figure 1A). Our sampling and sequencing efforts also complement other genome-wide datasets that investigate genetic variation within and among *D. melanogaster* populations throughout their range (Kolaczkowski *et al.* 2011; Mackay *et al.* 2012; Pool *et al.* 2012; Campo *et al.* 2013; Reinhardt *et al.* 2014; Kim *et al.* 2014; Grenier *et al.* 2015; Svetec *et al.* 2016; Lack *et al.* 2016; Zhao & Begun 2017; Kapun *et al*. 2018).

For our analysis, we divided our samples into three subsets (Figure 1B). The first subset (hereafter 'Core20') is composed of 20 populations consisting of one spring and one fall sample. Populations in the Core20 set are drawn from 15 localities in North America and Europe and we use at most two years of sampling from any one locality. Samples in the Core20 set are used in our basic analyses of seasonal adaptation. The second subset (hereafter 'ValidationSet') is composed of spring and fall samples from four populations. Populations in the ValidationSet are drawn from the two localities where there are more than two years of sampling. The ValidationSet is used as a part of our leave-one-out analysis. The third subset (hereafter the 'clinal' set) of samples was used to examine patterns of clinal variation along the East Coast of North America and consists of four populations sampled in the spring (see Material and Methods and Supplemental Table 1).

To estimate allele frequencies genome-wide we performed pooled DNA sequencing of multiple individual flies per population (Zhu *et al.* 2012; Schlötterer *et al.* 2014). For each sample, we resequenced pools of ~75 individuals (range 27-164) to an average of 94x coverage (range 22-220, Supplemental Table 1). Analyses presented here use a total of 51 samples. Data from the remaining samples, including those that do not constitute paired spring/fall samples, are not included in our analyses here but nonetheless form a part of our larger community-based sampling and resequencing effort and are deposited with the rest of the data. All raw sequence data have been deposited at the NCBI Short Read Archive (SRA; BioProject Accession #PRJNA308584; individual accession numbers for each sample can be found in Supplemental Table 1), a VCF file with allele frequencies from all populations, scripts of the analyses presented here, and genome-wide SNP statistics for seasonality and clinality are available on DataDryad (doi:10.5061/dryad.4r7b826).

Pooled resequencing identified ~1.75M SNPs following quality control and filtering for read depth and average minor allele frequency (minor allele frequency > 0.01). We applied a second filtering to identify ~775,000 SNPs that have observed polymorphism in each population sampled. Unless otherwise noted, we use the smaller set of more common SNPs. Whether this SNP selection process generates bias in our estimates of the strength and magnitude of seasonally variable selection remains to be determined but note that the smaller set still represents approximately half of all detected SNPs.

To gain insight into basic patterns of differentiation among the sampled populations, we performed a principal component (PC) analysis across all samples. Samples cluster by locality and geography (Figure 1C) with PC 1 separating European, North American West Coast and East Coast populations while PC 2 separates the eastern North America samples by latitude (linear regression $p = 3*10^{-15}$, $R^2 = 0.84$). No PC is associated with season, as defined by the time of collection, following correction for multiple testing.

Figure 1:

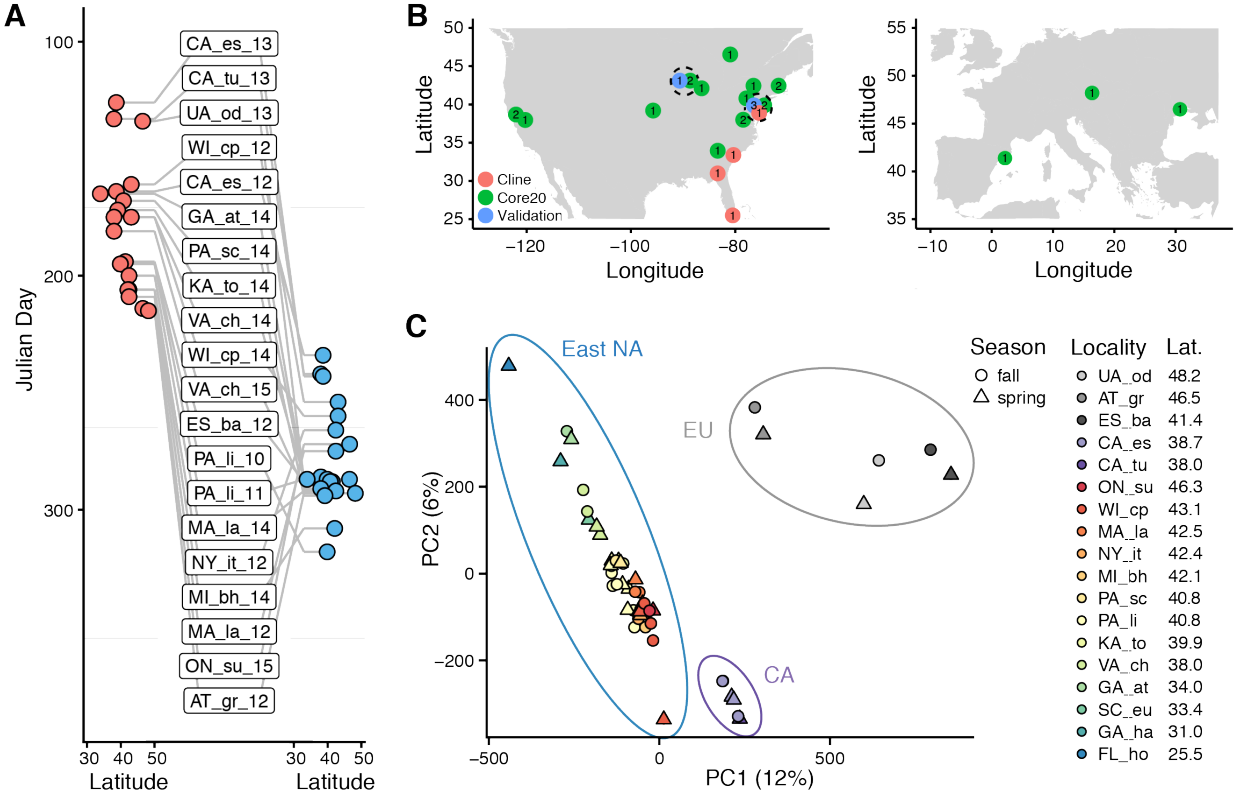


Figure 1. Sampling times, localities, and basic population structure of samples used in this study. A) Distribution of collection times during the spring (red) and fall (blue) in relation to latitude. For samples where the collection month, but not day, was recorded, we specified the 15th of the month to calculate Julian Day. B) Sampling localities for the primary seasonal analysis ('Core20': green), the cross-validation analysis ('ValidationSet': blue) and the latitudinal cline analysis ('Cline': red), distributed across North America and Europe. Numbers

represent the number of years of spring/fall comparisons at that locality. C) Principal components analysis of SNP allele frequencies. Circles are placed around samples from three geographic regions: Europe (EU), California (CA) and Eastern North America (East NA); points are colored by latitude and shapes represent seasons, as defined by collection time.

Seasonal adaptation is a general feature of fly populations. Our first objective was to determine whether or not we find evidence of seasonally cycling genetic variation. To identify such a signal of allele frequency change over time we assessed the per-SNP change in allele frequency between spring and fall among the Core20 set of populations. We first used a generalized linear model (GLM), regressing allele frequency at each SNP on season across all of the populations (see Materials and Methods for details). Indeed, we find that the observed genome-wide distribution of 'seasonal' *p*-values from this model is not uniform and there is a pronounced excess of SNPs with low *p*-values (Figure 2A; compare the red line for the observed data with the grey lines for the shuffled data). We assessed the statistical significance of this result by utilizing the paired spring/fall nature of our dataset. Specifically, we performed a permutation analysis whereby we shuffled the seasonal labels within populations to create 100 permuted datasets and then refit the seasonal GLM model to each such permutation. These permutations have the advantage of retaining all of the features of the data except for the seasonal labels within populations. We do find an excess of low *p*-values in the observed distribution compared to the permutations (Figure 2A), with a greater proportion of SNPs with $p<0.001$ in the unpermuted dataset than in 98% of the permutations. Our permutation analysis thus suggests that the genome-wide excess of SNPs with low *p*-values contains a signal of parallel seasonal adaptation across multiple populations.

To further verify the observed signal of seasonal adaptation, we used a second statistical method for measuring seasonal variation. For this we developed a semi-parametric test, the "rank Fisher's method" (RFM), which is calculated by combining the per-population changes in allele frequency (represented as *p*-values from a Fisher's exact test) between the two seasons normalized by the genome-wide distribution of SNPs with similar attributes (see Materials and Methods for details). This model was motivated by the observation that the GLM model can be affected by an inflated precision of allele frequency estimates (Supplemental Figure 2) or other statistical complexities (Machado *et al* 2016; Wiberg *et al* 2017) and that the observed distribution of the p-values in the GLM is shifted to small values even in the shuffled datasets (Figure 2A). The limitations of the GLM are particularly relevant, as our data are from pooled sequencing, for which we can only estimate an effective sample size (Machado *et al.* 2016). The RFM, by contrast, is relatively insensitive to the sample size estimate (Supplemental Figure 2). Using the RFM we again find a deviation from the expected test statistic distribution indicating an excess of seasonally varying polymorphisms (Figure 2B; compare the red line for the observed data with the grey lines for the shuffled data). Performing a permutation analysis analogous to that used for the GLM, we found the observed enrichment to be greater than in 98.2% of permutations. Together, our GLM and RFM results provide robust evidence that parallel seasonal adaptation is a general feature of these fly populations.

Predictability of seasonal adaptation among populations

The genome-wide signature of seasonal adaptation that we observe indicates that there are consistent changes in allele frequencies between seasons, broadly defined, among populations sampled across multiple years, and localities separated by thousands of miles (Figure 1A,B). However, it is not yet clear whether all populations show signatures of seasonal adaptation at a common set of SNPs and if not, whether such populations are associated with specific environmental effects, such as unusually short summer or an unusually mild winter. To investigate these questions, we performed a leave-one-out analysis. In this analysis, we sequentially removed each of the 20 paired spring-fall populations within the Core20 set and re-identified seasonally variable SNPs among the remaining 19 populations (hereafter, 'discovery set') as well as for the dropped, 20th population ('test set'). We then tested (1) if there is an enrichment of SNPs showing signals of seasonal evolution in both the discovery and test sets, and (2) if allele frequencies changed in the same direction (i.e., concordant allele frequency change) for SNPs that show a significant allele frequency change in both the discovery and test sets (see Materials and Methods for more details). For each population, we then calculated a '*predictability score*'. This score reflects the genome-wide rate of concordance change as a function of the joint significance threshold. Note that populations with a positive predictability score are those in which the concordance rate is greater than 50% over the bulk of the genome or chromosome.

Our predictability analysis yielded several basic results. First, we find that there is an enrichment of polymorphisms identified as strongly seasonal in both the test and discovery sets for a limited number of populations. For instance, at a joint significance quantile of 1%, 4 out of the 20 test sets show a significant enrichment of seasonal SNPs compared to their conjugate discovery set (nominal p-value < 0.05 based on genome-wide Fisher's exact test; Supplemental Figure 3). The populations that show significant enrichment of strongly seasonal sites in the discovery and test sets include: Esparto, CA, USA, 2012 (CA_es); Gross-Enzersdorf, Austria 2012 (AT_gr); Sudbury, Ontario, Canada 2015 (ON_su); and, Charlottesville, VA 2015 (VA_ch). For these populations, enrichment was modest – at most on the order of 25-50%. Although only four populations show nominally significant enrichment at a joint significance quantile of 1%, 14 out of 20 populations show a positive enrichment, and no population shows a significant negative enrichment at a joint significance threshold of 1%. It is unclear whether the extent of this modest, yet sometimes statistically significant, enrichment is due to statistical power, sampling effort, ecology, or the idiosyncratic nature of the evolutionary process (Gould 1989).

Second, we find that the majority of populations show a positive predictability score (Figure 2C). For populations that show the strongest positive predictability score, the concordance rate reaches ~65% (range 52-69%; all populations significantly different from 50% at a nominal p-value < 0.05) among the ~65 (range 36-199) SNPs with the most stringent joint significance threshold. At more modest joint significance thresholds (e.g., the top 10% in both sets), the concordance rate is lower but still statistically different from 50% (maximum nominal p-value at the top 5% of sites is 0.008). Intriguingly, we found that four of the 20 populations had significantly negative predictability scores (Benton Harbor, Michigan 2014; Lancaster, Massachusetts 2012; Topeka, Kansas 2014; and Esparto, California 2012). For these populations, the strength of negative predictability was, in general, as strong as those populations with positive predictability. The predictability score per population does not vary in a systematic way with enrichment (e.g., at a joint significance threshold of 1%, the correlation between enrichment and predictability score is -0.25, $p = 0.41$) and, notably, populations with negative predictability

scores have some of the strongest enrichment in the leave-one-out analysis. The sign and magnitude of predictability scores varies among chromosomes (Supplemental Figure 4), although the genome-wide score is correlated with the per-chromosome scores, particularly for chromosome arms 2L and 2R (Supplemental Figure 5).

Figure 2.

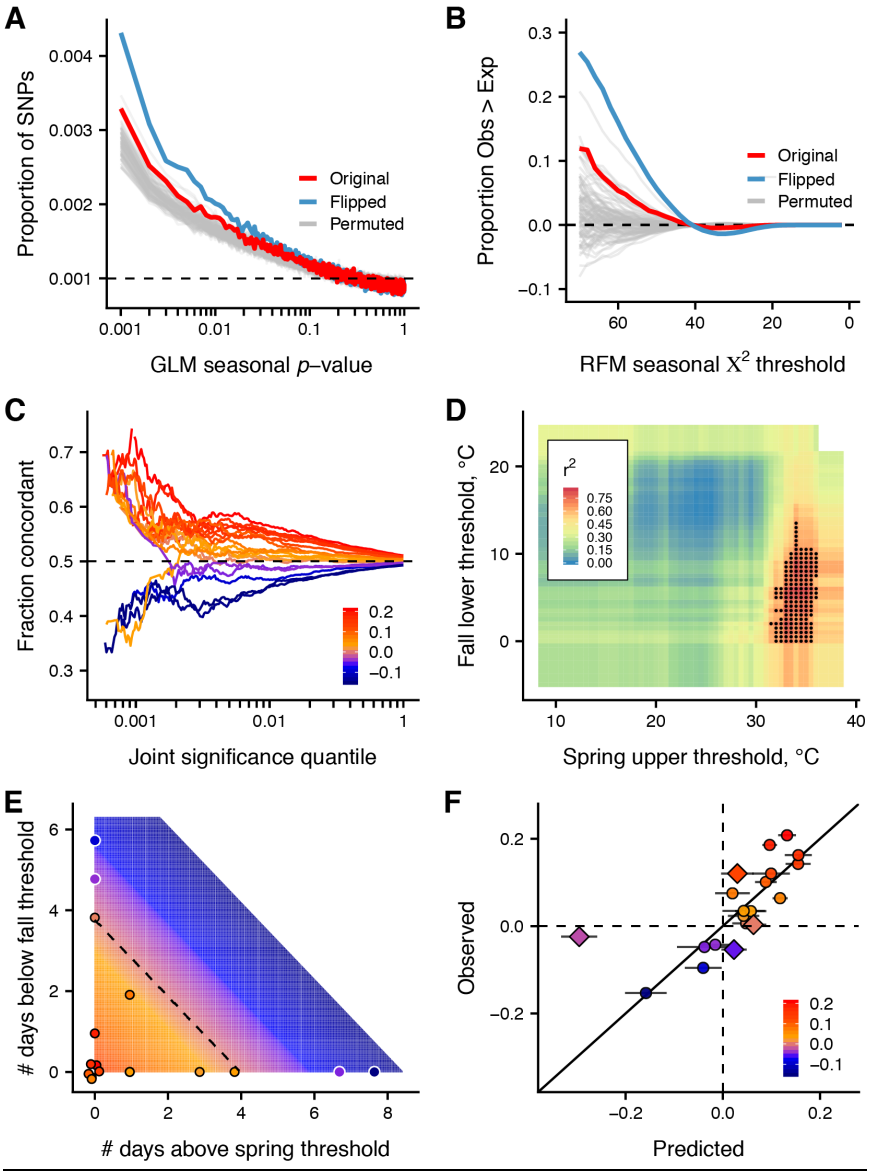


Figure 2. Generality and predictability of seasonal adaptation. A) GLM *p*-value distribution of per-SNP seasonal regressions (red = original dataset in which all seasonal labels are taken at face value; blue = flipped dataset that was generated after detecting that some populations exhibit the opposite pattern of fluctuations compared to the rest of the populations – see the text and Figs. 2CDEF) and permutations (grey)). The y-axis represents the proportion of SNPs within *p*-value bins between 0 and 1 with bin size 0.001. Note, the expected value of a uniform distribution is 0.001, as represented by the dashed line. B) RFM X^2 distribution of per-SNP

seasonal allele frequency change for the original dataset (red), the flipped dataset (blue), permutations (grey), and the null expectation (dashed). Plotted is the proportion of sites with a X^2 value greater than expected per X^2 threshold. C) The fraction of concordant allele frequency changes from the leave-one-out analysis of the Core20 populations. Each line represents one population, colored by the inferred genome-wide predictability score. Value for each population are plotted only when the nominal p-value < 0.05 from a binomial test of difference from the null expectation of a 50% concordance rate (dashed line) with a minimum of 35 SNPs. D) Model fits (r^2) from the thermal limit models. Black points represent models that fit the observed variation in genome-wide predictability scores better than 95% of permutations. E) Surface plot of the best fit thermal model from the leave-one-out analysis; color scheme matches color legend in F; points are colored with their observed genome-wide concordance score. Value on the x- and y-axis are the expected number of days (observed fraction multiplied by 21; some localities had a day or two of weather data missing). F) The best thermal limit models from the Core20 analysis models genome-wide predictability scores of the ValidationSet (diamonds). Circles represent the model-based estimates of Core20 population based on the best fit models identified in Figure 2C.

We sought to explain why some populations display negative predictability by constructing a weather based, thermal-limit model. We hypothesized that populations with negative genome-wide predictability scores were exposed to warm springs or cool falls prior to our sampling. To test this hypothesis, we calculated the number of days in the three weeks prior to spring sampling when maximum temperatures fell above a specified upper thermal limit and the number of cool days in the three weeks prior to fall sampling when minimum temperatures fell below a lower limit. We refer to these counts of days as 'critical days.' The use of a three-week window to assess critical days is somewhat arbitrary but we note that this duration corresponds to one to two generations and the general period from egg to peak adult reproduction (McMillan *et al.* 1970; Schmidt *et al.* 2005; Bergland *et al.* 2012; Klepsatel *et al.* 2013). Thus, a three-week window is a biologically relevant time window. We tabulated the critical days for each population across a range of spring and fall thermal limits (0°C - 40°C). For each combination of spring and fall thermal limits, we regressed predictability scores on the number of spring and fall critical days and assessed model fit in contrast with a null distribution obtained via permutation. We found that the number of days prior to sampling in the spring above ca. 35°C (32-37°C) and number of days in the fall below ca. 5°C (0-10°C) is significantly correlated with genome-wide predictability scores of the 20 populations (Fig. 2D,E; $R^2_{max} = 75.7$, $p_{\text{perm}} = 0.0025$ for the most significant set of spring and fall thermal limits). Inferred model parameters for the best fit models (those with $p_{perm} < 0.05$) and critical days across the range of thermal limits for each population can be found in Supplemental Tables 2 and 3, respectively.

These best fit models indicate that populations experiencing warm temperatures prior to the spring sampling or cool temperatures prior to the fall sampling will have negative genome-wide predictability scores (Figure 2E). This situation would naturally arise if these particular spring and fall fly samples were collected at a tail of the season. However, genome-wide predictability scores are not correlated with spring or fall collection date ($p = 0.6$ and 0.9, respectively), cumulative growing degrees ($p = 0.5$, 0.86), collection locale (residential vs. orchard, $p = 0.6$, 0.4), collection method (aspirator vs. net vs. trap, $p = 0.9$, 0.9), collection substrate (contrasting various types of fruit, compost, $p = 0.8$, 0.9), percent *D. simulans* contamination (0.8, 0.8), latitude ($p = 0.9$), or the difference in cumulative growing degree days between spring and fall (p

= 0.46). The genome-wide predictability score is more related to the specific temperature patterns observed in the spring and the fall than any other factors tested. Although we cannot rule out other explanations such as inadvertent sample swapping, our analysis is generally consistent with a model which suggests that changes in selection pressure within and among seasons may lead to dramatic changes in allele frequencies.

We tested the predictive power of our thermal limit model by applying it to the four populations in the ValidationSet (Figure 1B). We find that populations in the ValidationSet show a similar variability of predictability scores as those estimated in the leave-one-out analysis of the Core20 set and that two populations have negative predictability scores (Supplemental Figure 6B; Media, Pennsylvania 2012; Cross Plains, Wisconsin 2013). Next, we used the parameters from the best fit thermal models derived from the Core20 set (i.e., those where $p_{perm} < 0.05$; Figure 2D,E; Supplemental Table 2,3) to model the genome-wide predictability scores of the ValidationSet populations for which collection dates are known. We find that the modeled genome-wide predictability scores are positively correlated with the observed genome-wide predictability score of the ValidationSet populations (Fig 2F; median Pearson's $r = 0.37$ across the range of best-fit models, minimum Pearson's $r = 0.09$, maximum $r = 0.60$; see Material and Methods). While this correlation of three points is not significantly different from zero ($p = 0.77$, based on the best fit thermal model), we nonetheless find this correlation is greater than 76% of the best-fit permuted models (range 56 - 89% across all models with $p_{perm} < 0.05$). Taken together with the leave-one-out analysis of the Core20 set, we conclude that coarse-grained temperature data of the weeks prior to sampling is sufficient to predict some of the seasonal changes in allele frequencies genome-wide. Our conclusion is consistent with work in other drosophilids which demonstrate that short term changes in temperature, either directly or indirectly, elicits dramatic changes in allele frequencies in the wild and in the laboratory (Rodríguez-Trelles *et al.* 2013; Tobler *et al.* 2014; Mallard *et al.* 2018; Barghi et al 2019).

The flipped model. In the analysis presented above, we used the time of year that flies were collected to designate season. However, our leave-one-out analysis and thermal limit model suggest that the designation of season based on time of year may not completely reflect the recent selective history of the populations that we sampled. Notably, our analysis suggests that some populations experienced a particularly warm spring or a cold fall prior to sampling, and that such temperature extremes elicit dramatic changes in allele frequencies that are opposite from what would be expected from the calendar time season designation. We therefore reasoned that our ability to detect seasonally variable polymorphisms, along with our ability to assess general signals of seasonal adaptation, would improve if we flip the season label of the populations that show negative genome-wide predictability scores.

To evaluate this idea, we flipped the season label for the four populations with negative genome-wide predictability (Figure 2C) and calculated seasonal *p*-values using a GLM (hereafter, the 'flipped model'). As expected, the genome-wide distribution of *p*-values from the flipped model is more heavily enriched for low *p*-values than the original model. The flipped model out-performs the original model in the permutation analysis, where there are more SNPs with $p<0.001$ than 100% of 100 permutations for the flipped model (compared to 98% of the permutations for the original model; Figure 2A). We also re-calculated the RFM statistic after flipping the season labels on these four populations and again find a stronger signal than for the

original model and greater than 100% of the permutations (Figure 2B). Of course, the stronger signal of seasonal allele frequency change in the flipped model is a foregone conclusion given the particular way we are changing the seasonal labels.

To provide an orthogonal assessment of whether flipping the seasonal label for these four populations improves signals of seasonal adaptation we re-ran our predictability analysis. In this re-analysis, we calculated the predictability score for each ValidationSet population using the flipped Core20 model as our discovery set. The observed predictability scores estimated using the flipped Core20 model is more strongly correlated with the model-based estimates from our thermal limit model: the median correlation between model-estimated and observed predictability scores of $r = 0.51$ (original model: $r = 0.37$). This correlation coefficient is greater than 83% of the best-fit permuted models (compared with 68.8% for the original model; minimum 0.25, maximum 0.73; Supplemental Figure 6ABC). Thus, the use of the flipped model as a discovery set increased the predictive power of our thermal limit model in an independent dataset.

Another way to assess the increased predictive power of the flipped model is to examine the overlap between seasonal SNPs identified here with those identified by Bergland *et al.* (2014). Two populations are shared between those included in the Core20 and those analyzed in Bergland *et al.* (2014) and these shared populations could drive the signal of enrichment that we potentially observe. Accordingly, we identified the top 1% most seasonally variable SNPs from a model fit with 18 of the Core20 populations that do not overlap with those used in the earlier Bergland *et al.* (2014) study. These top SNPs identified here with 18 populations are marginally enriched for those identified by Bergland *et al.* (2014) relative to control SNPs matched for chromosome, heterozygosity, and average recombination rate ($\log_2$ odds ratio ± SD = 0.59 ± 0.37, $p_{perm} = 0.0512$). Applying the flipped season labels for this reduced Core20 set substantially improves the overlap ($\log_2$OR ± SD = 1.02 ± 0.39; p_{perm} = 6e-4). Note that the Pennsylvanian populations studied by Bergland *et al.* (2014) were not among those populations with season labels altered in our flipped model.

Taken together, we conclude that there is support for the idea that the flipped model generates a higher confidence set of seasonally varying SNPs. For the remainder of our analyses, presented below, we used seasonal SNPs identified by both the original and flipped models. In the results that follow we find that the use of the flipped model generally increases the strength of the signals that we observe, suggesting improved biological relevance of the flipped model.

Latitudinal differentiation parallels signatures of seasonal adaptation. Phenotypic and genetic analyses have demonstrated local adaptation among populations of *D. melanogaster* sampled across latitudinal clines (see Adrion *et al.* 2015 and references therein). Signals of local adaptation across latitudinal gradients have even been suggested to result from clines in seasonality (Rhomberg & Singh 1989): higher latitudes are more stressful as a consequence of severe winters and shorter growing seasons whereas more benign conditions are found in tropical and sub-tropical areas. Thus, spatial heterogeneity in selection pressures associated with temperate environments may be paralleled by seasonal fluctuations in similar selection pressures.

Figure 3.

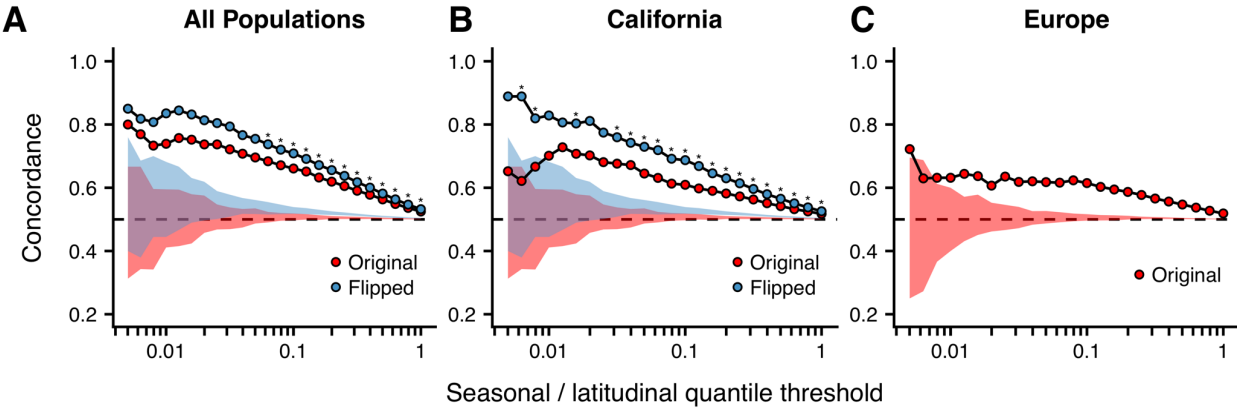


Figure 3. Concordance rates of seasonal and clinal polymorphisms. A) Seasonal sites were identified using 18 of the Core20 populations to avoid a shared collection locale (Linvilla, Pennsylvania) with populations used for clinal analysis. The y-axis represents the concordance rate of allele frequency change across time and space, assuming that the winter favored allele is the same as the allele favored in high latitudes. Seasonal sites were also identified in B) California populations (n=3) and C) Europe populations (n=3) and contrasted with clinal polymorphisms identified along the East Coast of North America. Shaded areas are the 96% confidence intervals of 100 matched control datasets. Stars represent significant differences between the original and flipped concordance rates based on a Fisher's exact test (nominal $p < 0.05$).

Parallel changes in phenotype and genotype across time and space have been observed in *D. melanogaster* (Cogni *et al.* 2014; Bergland *et al.* 2014; Paaby *et al.* 2014; Cogni *et al.* 2015; Kapun *et al.* 2016; Rajpurohit *et al.* 2017; 2018; Kapun & Flatt 2019). To assess whether we observe similar general patterns in our data, we tested whether SNPs that are clinal and seasonal below a range of significance thresholds change in allele frequency in a parallel fashion. Parallel changes in allele frequency across time and space occur when the sign of allele frequency change between spring and fall is the same as between high- and low-latitude populations. We re-identified seasonally varying SNPs using 18 of the Core20 populations to avoid a shared collection locale with the clinal set (see Supplemental Table 1). Seasonally varying SNPs were identified among these 18 populations using both the original and flipped seasonal labels. Clinally varying polymorphisms were identified using single-locus models (GLM, with allele frequency regressed against latitude) as well as models that account for population structure (BayEnv2; Günther & Coop 2013).

We find that the rate of parallelism increases with increasingly stringent seasonal and clinal significance thresholds (Figure 3A) using either model of clinality (Supplemental Figure 7). Parallel changes in allele frequency between seasons and clines is similar to previously published genome-wide (Bergland *et al.* 2014) and locus specific results (Stalker 1976; Stalker 1980; Cogni *et al.* 2014; Kapun *et al* 2016). Parallelism rates are higher when assessing seasonal variation using the flipped model as compared to when we use the original model (Figure 3A).

533 This increase in parallelism rates using the flipped model can be taken as yet another strong, and
534 orthogonal, evidence that the flipped model generates a higher confidence set of seasonal SNPs.
535
536 Based on these analyses, there is clear evidence that parallel changes in allele frequency occur
537 across time and space. Are these parallel changes driven by large scale seasonal migration? To
538 address this question, we identified seasonal SNPs using populations that are geographically and
539 genetically isolated from the East Coast (Campo *et al.* 2013; and see Figure 1A). We assessed
540 seasonal changes in allele frequency separately for Californian (n=3) and European (n=3)
541 populations and tested for parallelism with polymorphisms that vary along the East Coast of
542 North America. One of the Californian populations contained a 'flipped' sample and so we also
543 re-assessed seasonal changes in allele frequency for those populations using a flipped model. We
544 found a strong signal of parallelism between seasonal and latitudinal variation in both the
545 California and the Europe comparisons (Figure 3B,C). For the Californian populations,
546 parallelism was greater for the flipped model than for the original model (Figure 3B).
547
548 These results suggest that parallel changes in allele frequency between seasons and across the
549 latitudinal cline is not exclusively driven by seasonal migration in the spring from neighboring
550 southern populations. Our conclusion is consistent with patterns of enrichment of jointly
551 seasonal and clinal polymorphisms (Supplemental Figure 8). Notably, for the flipped seasonal
552 model, there is weak evidence of an enrichment of highly seasonal and clinal polymorphisms, a
553 pattern otherwise expected if seasonal migration was a strong determinant of seasonal evolution.
554 Indeed, strongly seasonal sites identified from the flipped model are enriched for moderately
555 clinal polymorphism, and *vice versa*, a pattern ostensibly consistent with independent seasonal
556 evolution among populations arrayed along a spatial gradient. In contrast, seasonal sites
557 identified with the original seasonal model, in which seasons are defined by calendar time and
558 not temperature, are significantly enriched for strongly clinal polymorphisms suggesting that
559 seasonal migration may contribute to certain aspects of seasonal evolution. We note, however,
560 that the observed enrichment of seasonal and clinal polymorphisms only represents a small
561 fraction of all seasonal sites: for instance, using the original model at a joint quantile of 1%, we
562 expect ~84 (95% CI based on bootstrap resampling of matched control polymorphism: 63-104)
563 of the top 1% (n=~7500) seasonally varying SNPs to be among the top 1% of clinal sites;
564 however, we observe ~115 to be clinal. Thus, while we observe a significant overlap of seasonal
565 and clinal polymorphisms the vast majority of strongly seasonal polymorphisms are not strongly
566 clinal, again suggesting that clinal and seasonal evolution in *D. melanogaster* are largely
567 decoupled. Additionally, our analysis presented here assumes that patterns of clinal variation
568 along the West Coast of North America and Europe are distinct from patterns of clinal variation
569 along the East Coast; whether or not this assumption is valid remains to an open question (but
570 see Oakeshott *et al*. 1982).
571
572 *The strength and genomic extent of seasonal adaptation*. Our results suggest that seasonal
573 adaptation is driven, at least in part, by changes in allele frequency at a shared set of
574 polymorphisms spread throughout the genome. Here, we investigate the magnitude of change
575 and the number of sites that vary seasonally.
576

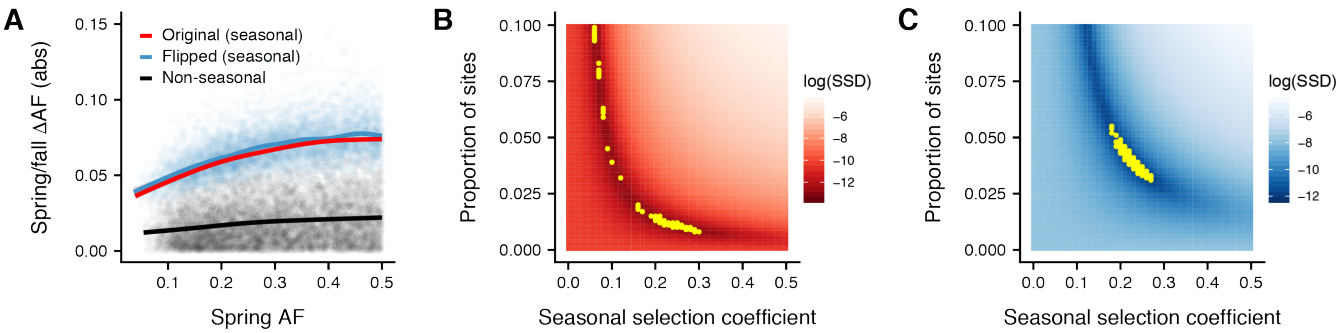


Figure 4. The strength and magnitude of seasonal adaptation. A) Average spring/fall allele frequency change for each of the top 1% of seasonally varying SNPs (red/blue) and non-seasonal, matched control SNPs (black), as a function of the folded spring allele frequency. Lines represent a moving average across SNPs with a given spring allele frequency. BC) ABC estimates for the proportion of sites affected by seasonal selection and the associated *s* for the original (A) and flipped model (B), showing the sum of squared distances (SSD) between the observed and simulation RFM X^2 test statistic distributions (see Materials and Methods for details), over varying strengths of selection and proportion of sites under selection. Yellow points represent the top 1% of simulated parameter value combinations, reflecting an ABC acceptance rate of 0.01).

To gain basic insight into the magnitude of seasonally variable changes in allele frequency, we first examined the difference in allele frequencies between seasons for the most significantly seasonal polymorphisms. The top 1% most seasonally variable SNPs identified by the GLM change in frequency by a difference of 4 and 8%, on average (Figure 4A). This change is not different between the original and flipped models (Figure 4A), but is substantially smaller than estimates of ~20% reported by Bergland *et al.* (Bergland *et al.* 2014).

To robustly estimate the strength and extent of seasonally varying selection, we used our semi-parametric RFM test in an Approximate Bayesian Computation (ABC) analysis. This analysis allows us to estimate the fraction of the genome affected by seasonal adaptation and the average strength of seasonal selection by comparing the observed distributions of the test statistic from both the flipped and original models to simulations where a fraction of polymorphisms was randomly selected to change in frequency between seasons with a given selection coefficient (see Materials and Methods for more details). Using the sum of the square distances between the observed and simulated test statistic as our metric, we identified which parameter estimates were most likely to produce the observed genome-wide distribution of the test statistic. The model we employ for this analysis is based on one generation per season and assumes that every site in the genome changes in frequency independently and is subject to an equal strength of selection. Note that this model is appropriate here as we aim to estimate the proportion of both causal and linked SNPs that change in frequency due to seasonal selection and not to assess the number of causal sites. We note that our analysis is potentially biased because we are conditioning on polymorphisms that are present in all populations tested, including North American and European ones with different long-term dynamics and histories of colonization, expansion, and admixture (Keller 2007; Bergland *et al.* 2016; Kapopoulou *et al.* 2018). The filtering requirement that we impose may therefore bias our SNP set to those which are subject to balancing selection,

causing an upward bias in our estimates. We note, however, there are no qualitative differences found as a function of filtering for common SNPs (compare Figure 4B to Supplemental Figure 9).

We find a ridge of best fit parameters, with a trade-off between the proportion of sites under selection and strength of selection (Figure 4B,C). For the original model, the data are consistent with a minimum of ~1% of sites subject to an effective cumulative seasonal selection coefficient ("*s*", for simplicity) of ~0.3, or a maximum of ~10% of sites at a *s* of ~0.05. For the flipped model, the data are consistent with a minimum of ~2.5% at an *s* of ~0.3 and a maximum of ~5% at an *s* of ~0.2. As expected, the estimated strength of selection and the fraction of the genome affected by seasonally varying selection is greater for the flipped model than the original model (Figure 4B vs. 4C). In order to assess whether and to what extent our estimates are confounded by clustering of seasonally varying polymorphisms, we performed the same analyses using datasets of subsampled polymorphisms every 1Kb and 5Kb (Supplemental Figure 10). We found no substantial shift in our parameter estimates across subsampled datasets.

The distribution of seasonally variable SNPs throughout the genome. The analyses presented above suggest that seasonal adaptation affects patterns of allele frequency genome-wide. We therefore aimed to characterize several aspects of the genomic distribution and functional enrichment of seasonally varying polymorphisms.

First, we examined seasonal changes in the frequency of the major, cosmopolitan inversions that segregate in wild *D. melanogaster* populations. We estimated inversion frequency from pooled allele frequency data by calculating the average frequency of SNPs previously identified as being closely linked to these inversions (Kapun *et al.* 2014). With the exception of *In(2L)t*, we find that most inversions are rare in the populations that we examined, segregating at less than 15% frequency in most populations (Supplemental Figure 11). Using a simple sign test, we find that no inversion consistently changes in frequency using season labels as defined in either the original or the flipped model. We note that *In(2R)Ns* is at a higher frequency in the fall compared to the spring in a quarter of the Core20 populations ($p_{binomial\text{-}test}$ = 0.048; Bonferroni adjusted p = 0.28), consistent with previously reported signals of seasonal change (Stalker 1976; Stalker 1980; Kapun *et al.* 2016; Kapun & Flatt 2019). However, *In(2R)Ns* segregates at a lower frequency (<15% across populations surveyed) than many of the most strongly seasonal SNPs on chromosome 2R (over 85% of the top seasonally varying SNPs have minor allele frequency greater than 15%), and seasonal changes of *In(2R)Ns* are relatively mild compared to other seasonally varying polymorphisms (cf. Figure 4A Supplemental Figure 11). We conclude that inversions are not the exclusive drivers of seasonal adaptation in *D. melanogaster*, although we cannot rule out their role in contributing to seasonal adaptation.

To determine if seasonally variable SNPs are non-randomly distributed throughout the genome, we examined the abundance of the top 1% of seasonally variable SNPs (GLM model; n=7,748) among bins of 100 consecutive SNPs. We found significant over-dispersion in the distribution of seasonal SNPs per bin compared to both the theoretical and the matched control distributions (Figure 5A; Kolmogorov-Smirnov test $p=10^{-16}$), indicating that seasonal SNPs are more clustered than expected, consistent with an elevated seasonal F_{st} surrounding the most strongly

seasonal sites (Figure 5B). We cannot determine to what extent this clustering is due to linkage to causal SNPs or to the clustering of causal seasonal variants. The distribution of control polymorphisms was indistinguishable from the binomial expectation in our analysis (Kolmogorov-Smirnov test p=1).

Although seasonally variable SNPs are non-randomly distributed throughout the genome, they are nonetheless broadly distributed. To further quantify the distribution of these sites, we calculated the probability of observing at least one of the top 1% most seasonally variable SNPs among 1000 randomly selected genomic windows ranging in size from 100bp to 100Kb. We find that seasonally variable SNPs are distributed across all autosomes but are enriched on chromosome 2L (Figure 5C,D). Across the window sizes tested, the probability of observing at least one top seasonal SNP in any given window is less than expected by chance given a Poisson distribution, consistent with clustering at both large and small genomic scales (Figure 5C). Despite significant clustering, seasonally variable SNPs show broad genomic distribution, as evidenced by a >75% chance of identifying a seasonally variable SNP in any randomly selected 50Kb window (Figure 5C). The general signal of over-dispersion is similar for both the flipped and original models (Figure 5A,C). The broad genomic distribution and large number of seasonally variable SNPs suggests that seasonal adaptation is highly polygenic. A polygenic model is consistent with the view that many of the traits underlying seasonal adaptation (Schmidt & Conde 2006; Behrman *et al.* 2015; Betini *et al.* 2017; Rajpurohit *et al.* 2017; Noh *et al.* 2017; Behrman *et al.* 2018; Rajpurohit *et al.* 2018), and adaptation to temperate environments in general (Rand *et al.* 2010; Levine *et al.* 2011; Chen *et al.* 2012; Rajpurohit *et al.* 2013; Cockerell *et al.* 2014; Svetec *et al.* 2015; Chakraborty & Fry 2016; Juneja *et al.* 2016; Bozicevic *et al.* 2016; Svetec *et al.* 2016), are themselves quantitative traits (Mackay & Huang 2018).

Next, we tested whether the seasonally varying polymorphisms that we identify here using either the original or flipped model are enriched among different functional annotation classes (Supplemental Figure 12). The top 1% most seasonally varying polymorphisms are not significantly enriched for any functional class that we tested. The top 5% of seasonally varying polymorphisms are weakly enriched for non-synonymous sites as compared to matched control polymorphisms (enrichment = 8%, 95% CI = 2-14%, nominal p=0.01). The ratio of non-synonymous to synonymous polymorphisms among the top 5% of seasonal SNPs is also marginally higher than that for matched controls (enrichment = 7%, 95% CI = 1 − 15%, p=0.03). We found no significant enrichment or under-enrichment for intergenic regions, introns, UTR's, or coding regions.

Figure 5.

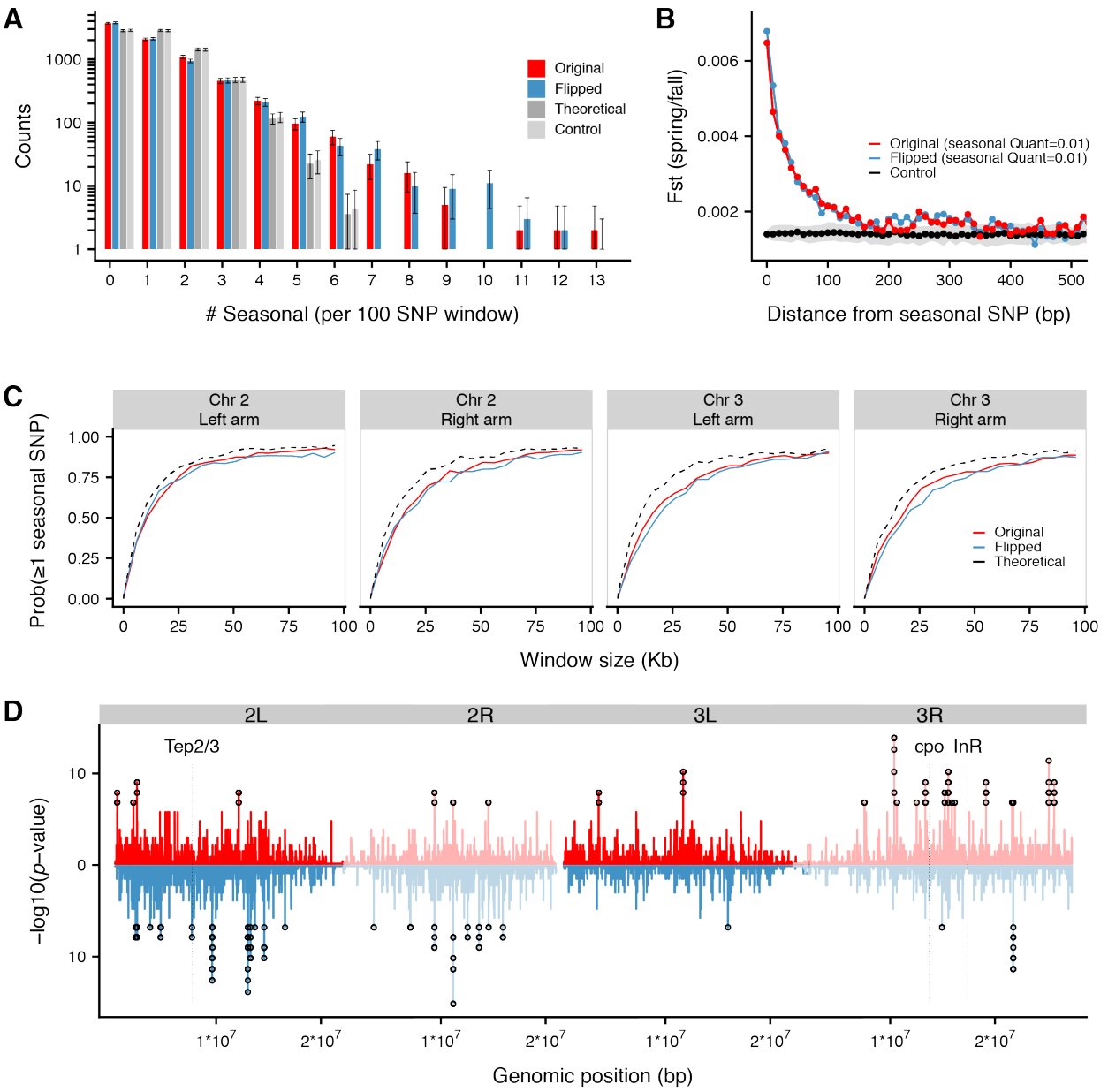


Figure 5. Genomic distribution of seasonal SNPs. A) Distribution of seasonal sites (quantile: 0.01) per 100 SNP window. The observed distribution (red and blue) is over-disbursed compared to both the theoretical distribution (dark grey) and that of matched controls (light grey). Error bars are 2 standard deviations. B) Spring/fall genetic differentiation (F_{st}) at highly seasonal SNPs (distance = 0) and for SNPs with increasing distance from a seasonal SNP. Also plotted is the mean (black) and 96% confidence interval (grey area) of 100 matched controls. C) Proportion of windows with at least one seasonal SNP, with increasing window size (Kb). The observed proportions (solid lines) and expected theoretical proportion (dashed line) is plotted for each chromosome arm. D) Sliding window analysis of regions with high densities of seasonal SNPs. Open circles denote regions of the genome with an excess of seasonally variable SNPs following multiple testing correction. Red lines, above the x-axis, represent the strength of enrichment of

strongly seasonal sites as identified by the original model. Blue lines, below the x-axis, represent the enrichment of strongly seasonal sites identified in the flipped model.

We performed a sliding-window analysis to identify regions of the genome with an excess of seasonally varying polymorphisms. We found 25 regions of the genome with a significant excess of seasonally varying polymorphisms for either model. The locations of these regions change substantially between the 'original' and 'flipped' models (Figure 5D). For simplicity, we focus the remainder of our analysis here on the regions identified using the 'flipped' model, however the coordinates of all identified regions of both the 'original' and 'flipped' are reported in Supplemental Table 4. Regions of the genome enriched for seasonally variable SNPs are found on both arms of chromosome 2 and deficient along chromosome 3 ($\chi^2 = 22.9$, $p = 4e\text{-}5$). Intriguingly, one of 40 most seasonally variable regions identified with the 'flipped model' contains the genes *Tep2* and *Tep3,* which we have previously associated with seasonal variation in immune function in flies (Behrman *et al.* 2018). The *Tep2/3* region contains 11 of the top seasonally variable SNPs, ranking this region among the top 0.1% genomic windows. The genomic window surrounding *couch-potato* (*cpo*), a seasonally varying gene (Cogni *et al.* 2014) found to be associated with reproductive dormancy (Schmidt *et al.* 2008), has seven of the top seasonally variable SNPs ranking it among the top 1% of windows genome-wide. Note that the enrichment of seasonal SNPs surrounding *cpo* is not significant after conservative, multiple testing correction ($p_{adjusted} = 1$) which would be appropriate for *de novo* discovery of such regions. If we treat *cpo* as the true seasonal locus, this result implies that we currently lack power to discover many causal genes involved in seasonal adaptation. There was no excess of seasonal SNPs surrounding the insulin-like receptor gene (*InR*), wherein seasonally variable indels and SNPs have been previously reported and linked to variation in life-history traits (Paaby *et al.* 2014).

Gene-ontology analysis (Huang *et al.* 2009a; b) of the 198 genes within the 25 regions we identified show significant enrichment for different biological processes, cellular components, and molecular function following multiple testing correction (Supplemental Table 5). For instance, we find that genes associated with several molecular function categories including inositol metabolism and phosphorylation are enriched following multiple testing correction ($p_{adjusted} < 0.013$). This result is *prima facie* consistent with a role of inositol related compounds in drosophilid overwintering and cold-tolerance (Vesala *et al.* 2012). However, extreme caution should be taken when interpreting this result, as in any GO analysis (Pavlidis *et al.* 2012), as it is driven by a set of three, closely linked genes (CG17026, CG17028, CG17029).

Discussion

In this study we identify genomic signatures of seasonal adaptation in *D. melanogaster*. Our approach relies on the detection of parallel changes in allele frequency from spring to fall across 20 populations. We use coarse definitions of spring and fall, with spring as the time of year close to when *D. melanogaster* first becomes locally abundant and fall as the time close to the end of the growing season but before populations become too scarce to sample (Ives 1970; Behrman et al. 2015). We sampled populations across a wide geographic range, such that the length of growing season as well as other environmental conditions vary substantially. Despite this

environmental heterogeneity and the coarseness of the definitions of spring and fall, we detect a clear signal of parallel seasonal allele frequency shifts across these populations.

Our analysis revealing pervasive seasonal adaptation would not be possible without accurate allele frequency estimates (Zhu *et al.* 2012), genome-wide, from paired spring-fall samples collected across *D. melanogaster*'s range. Our approach used less expensive pooled sample sequencing which were collected by the DrosRTEC consortium, to facilitate seasonal sampling across broad spatial scales, which is mirrored in Europe by the DrosEU consortium (Kapun *et al.* 2018).

By detecting consistent seasonal changes in allele frequencies among multiple populations in North America and Europe, we demonstrate that seasonal adaptation is a general phenomenon of temperate *D. melanogaster* populations, and not restricted to a single orchard (Linvilla Orchard, Media, Pennsylvania; Schmidt & Conde 2006; Cogni *et al.* 2014; Bergland *et al.* 2014; Behrman *et al* 2015), biogeographic region (East Coast of North America; Behrman *et al* 2018), or even a continent. Furthermore, the parallel frequency shifts across multiple populations that we describe here suggests that seasonal adaptation across the whole range is at least partly driven by a consistent set of loci. Note that we generally lack power to detect alleles that cycle in a small subset of populations or are private to a single population. Therefore, we focus only on polymorphisms that segregate in all populations – and which therefore have an overall high population frequency - and only look at the signal of parallel cycling. Moreover, we did not attempt to investigate cycling of structural variants such as indels and transposable elements that are known to generate adaptive mutations at an appreciable rate (reviewed in Barrón *et al.* 2014). Thus, our conclusions are limited to the shared genetic basis of seasonal adaptation as revealed by the cycling of common SNPs.

Our ability to detect a shared genetic basis of seasonal adaptation across these populations implies that our admittedly crude definitions of spring and fall nonetheless carry biological meaning. It also implies that there are generic features in the environment that exert consistent selective pressures despite the environmental heterogeneity across sampling times and locations. It is possible that such generic selective pressures relate to temperature, humidity, population density, resource availability, or other biotic and abiotic factors. However, what these selective pressures are specifically and how they act across seasons remains to be determined. Note, however, that our results here (Figure 2) and elsewhere (Bergland *et al.* 2014) suggest that temperature extremes may be a strong proximate or ultimate selective agent.

Although we identified consistent patterns of evolution between spring and fall, not all populations fit the general pattern, with some showing strong signals of reversed allele frequency change (Figure 2C). These 'flipped' populations are geographically dispersed (California, Kansas, Michigan, and Massachusetts) with samples collected over two non-consecutive years (2012, 2014). We demonstrated that this flipping is associated with particularly warm springs and particularly cool falls prior to sampling. Indeed, the ecological model we built using information about local temperature extremes allowed us to predict the sign and magnitude of allele frequency changes genome-wide for additional populations not used to build the model (Figure 2F, Supplemental Figure 6). We thus hypothesize that, at the time of sampling, some of our spring samples had already experienced sufficiently strong selection in the generically

summer way and that the fall samples had experienced selective pressures in the generically winter way. This would result in seasonal alleles in the spring already showing fall-like patterns and the ones in the fall showing spring-like patterns. More generally, it is likely that natural selection does not push populations in the same consistent direction across the growing season and that there are phases within the growing season with distinct ecological and selection pressures. In the future, more dense temporal sampling across the growing season might help map the action of natural selection to more specific environmental factors. Taken at face value the implication is that metazoan populations are capable of adaptive tracking on extremely short temporal scales on the order of less than 10 generations. These data thus further argue that evolution and ecology do often operate on similar timescales (Hendry 2009) and evolutionary adaptation cannot be generally ignored *a priori* in demographic or ecological investigations (Rudman *et al*. 2017).

Our analysis allowed us to estimate the proportion of the genome affected by seasonal cycling and the magnitude of seasonal fluctuations. Our best guess is that the frequency of ~4% of common polymorphisms are affected by seasonally variable selection coefficients, cumulatively on the order of ~20% per season selection. However, there is a large range of plausible values that tradeoff the number of sites (1-10% of all sites) with the strength of selection (10-30% per season). Note that we do not claim that these 1-10% of common polymorphisms are causal nor that they cycle independently. In fact, we speculate that the vast majority of the seasonal sites are only linked to causal loci and many cycle together by fluctuating linked selection. Indeed, the number of seasonal loci is implausibly large for all of them to be individually causal (Bergland *et al.* 2014) under any of these scenarios. The lack of strong enrichment for specifically functional sites such as non-synonymous compared to synonymous loci (Supplemental Figure 12) is also consistent with the vast majority not being causal. At the same time, our model of unlinked loci captures the breadth and magnitude of such linked selection and demonstrates that seasonal adaptation is affecting, either directly or indirectly, a large number of variants across a range of populations. Given that we can only detect consistent pattern of seasonal cycling we are likely to underestimate the impact of linked seasonal selection. Seasonal adaptation therefore constitutes a profound stochastic evolutionary force acting on linked variation that is likely to rival in strength genetic draft due to selective sweeps and should be much stronger over seasonal timescales than random genetic drift in populations as large as *D. melanogaster* (Karasov *et al.* 2010).

While we cannot determine the precise number and location of causal sites in the present study, the the broad distribution of the seasonal SNPs across the genome implies that seasonal adaptation is highly polygenic. Indeed, there is greater than a 75% chance that a random 50Kb region of the genome will have at least one strongly seasonal site, reminiscent of the data on the density of GWAS hits in human populations (Boyle *et al.* 2017). Therefore, the cycling we observe cannot be due to a single causal region such as an inversion or a region of particularly low recombination. At the same time, seasonal SNPs are significantly clustered, suggesting that some genomic regions are more likely to contain one or more causal sites. We conducted a screen for regions of the genome with an excess of highly seasonal SNPs (Figure 5D). We found some outlier regions contain genes previously linked to seasonally variable phenotypes (immunity and *Tep2*/*Tep3* - Behrman *et al.* 2018; reproductive dormancy - *cpo* (Cogni *et al.* 2014; Schmidt *et al.* 2008). Additional sampling may reduce the false positive rate and should help us map individual seasonally causal loci.

Our observation that seasonal adaptation is polygenic and rapid suggests that there exists a substantial number of common polymorphisms that are subject to strong fluctuating selection and which are nonetheless maintained across the range of *D. melanogaster*. The most straightforward explanation of these results is that these polymorphisms are subject to some form of balancing selection (Levene 1953; Wittmann *et al.* 2017; Bertram & Masel 2019). Pervasive balancing selection is consistent with the recent realization that simple models of mutation-selection balance are inconsistent with the extent of quantitative genetic variation in a variety of fitness related traits (Charlesworth 2015). A full understanding of the genomic consequences of balancing selection caused by strong fluctuating selection will require the identification of individual causal loci. Doing so will require substantial additional sampling across seasons as well as from closely and broadly distributed locales throughout the whole geographic range of *D. melanogaster* in combination with the genetic mapping of seasonally fluctuating phenotypes (Schmidt & Conde 2006; Behrman *et al.* 2015; Rajpurohit *et al.* 2017; Behrman *et al.* 2018; Rajpurohit *et al.* 2018). Finally, such efforts can be strongly augmented by manipulative field experiments and functional work to investigate the molecular effects of putatively causal genetic variants. All of this work will need to be done in the context of other processes that underlie persistence of populations in the face of environmental heterogeneity such as phenotypic plasticity and bet hedging (Via & Lande 1985; David *et al.* 1997; Ayrinhac *et al.* 2004; Bergland *et al.* 2008; MacMillan *et al.* 2015; Overgaard *et al.* 2015; Kain *et al.* 2015) as it is becoming clear that all of these processes are occurring contemporaneously with seasonal adaptation.

Materials and methods

Population sampling and sequence data. We assembled 72 samples of *D. melanogaster*, 60 representing newly collected and sequenced samples and 26 representing previously published samples (Bergland *et al.* 2014; Kapun *et al.* 2016). Locations, collection dates, number of individuals sampled, and depth of sequencing for all samples are listed in Supplemental Table 1. For each sample, members of the Drosophila Real-Time Evolution Consortium collected an average of 75 male flies using direct aspiration from substrate, netting, or trapping at orchards and residential areas. Flies were confirmed to be *D. melanogaster* by examination of the male genital arch. We extracted DNA by first pooling all individuals from a sample, grinding the tissue together in extraction buffer, and using a lithium chloride – potassium acetate extraction protocol (see Bergland *et al.* 2014 for details on buffers and solutions). We prepared sequencing libraries using a modified Illumina protocol (Bergland *et al.* 2014) and Illumina TrueSeq adapters. Paired-end 125bp libraries were sequenced to an average of 94x coverage either at the Stanford Sequencing Service Center on an Illumina HiSeq 2000, or at the Stanford Functional Genomics facility on an Illumina HiSeq 4000.

The following sequence data processing was performed on both the new and the previously published data. We trimmed low-quality 3' and 5' read ends (sequence quality < 20) using the program *cutadapt* v1.8.1 (Martin 2011). We mapped the raw reads to the *D. melanogaster* genome v5.5 (and for *D. simulans* genome v2.01, flybase.org) using *bwa* v0.7.12 *mem* algorithms, with default parameters (Li & Durbin 2009), and used the program *SAMtools* v1.2 for bam file manipulation (functions *index*, *sort*, and *mpileup*) (Li *et al.* 2009). We used the program *picard* v2.0.1 to remove PCR duplicates (http://picard.sourceforge.net) and the program

GATK v3.2-2 for indel realignment (McKenna *et al.* 2010). We called SNPs and indels using the program *VarScan* v2.3.8 using a *p*-value of 0.05, minimum variant frequency of 0.005, minimum average quality of 20, and minimum coverage of 10 (Koboldt *et al.* 2012). We filtered out SNPs within 10bp of an indel (they are more likely to be spurious), variants in repetitive regions (identified by *RepeatMasker* and downloaded from the UCSC Genome browser), SNPs with a median frequency of less than 1% across populations, regions with low recombination rates (~0 cM/Mb; Comeron *et al.* 2012), and nucleotides with more than two alleles. Because we sequenced only male individuals, the X chromosome had lower coverage and was not used in our analysis. After filtering, we had a total of 1,763,522 autosomal SNPs. This set was further filtered to include only SNPs found polymorphic in all samples ("common polymorphisms"), resulting in 774,651 SNPs that represent the core set used in our analyses.

Due to the phenotypic similarity of the species *D. melanogaster* and *D. simulans*, we tested for *D. simulans* contamination by competitively mapping reads to both genomes. We omitted any of our pooled samples with greater than 5% of reads mapping preferentially to *D. simulans* (Charlottesville, VA 2012, fall; Media, PA 2013 fall). For the remaining samples reads that mapped better to the *D. simulans* were removed from the analysis. For the populations used in this analysis, the average proportion of reads mapping preferentially to *D. simulans* was less than 1% (see Supplemental Table 1), and there was no significant difference in the level of *D. simulans* contamination between spring and fall samples (t-test p=0.90).

In silico simulation analysis shows that competitive mapping accurately estimates the fraction of *D. simulans* contamination. To conduct these simulations, we used a single *D. simulans* genome from an inbred line derived from an individual caught California (Signor *et al*. 2017) and a single *D. melanogaster* genome from a DGRP line (Mackay *et al.* 2012). We used *wgsim* to generate simulated short-reads from each genome and mixed these short reads together in various proportions. We then mapped the short reads back to the composite genome for competitive mapping, as described above. We calculated contamination rate as the number of total reads mapping to the *D. simulans* reference genome divided by the number of reads mapping to both *D. simulans* and *D. melanogaster*. We also calculated the locus specific residual cross mapping rate by calculating the number of *D. simulans* reads still mapping to the *D. melanogaster* genome even after competitive mapping. These simulations demonstrate that the estimation of the cross-species mapping is precise (Pearson's $r = 0.9999$, $p = 1.6 \times 10^{-24}$, Supplemental Figure 13A), but underestimates the true contamination rate by ~2%. The level of residual *D. simulans* contamination does not correlate with contamination rate and is roughly 9% (Supplemental Figure 13B). Additionally, the proportion of seasonal sites (top 1%) is not correlated with the rate of residual mapping of *D. simulans* reads to the *D. melanogaster* genome as inferred through our simulation (r = -0.0026, p=0.93).

To assess seasonal variation we analyzed population genomic sequence data from 20 spring and 20 fall samples ('Core20'). These samples represent a subset of the sequenced samples. We used samples that had both a spring and a fall samples taken from the same locality in the same year. We also used a maximum of two years of samples for a given locality to prevent the analysis from being dominated by characteristics of a single population. When there was more than two years of samples for a given population, we chose to use the two years with the earliest spring

collection time. This decision was made on the assumption that the earlier spring collection would better represent allele frequencies following overwintering selection. This left 20 paired spring/fall samples, taken from 12 North American localities spread across 6 years and 3 European localities across 2 years (Supplemental Table 1). The localities and years of sampling are as follows: Esparto, California 2012 and 2013; Tuolumne, California 2013; Lancaster, Massachusetts 2012 and 2014; Media, Pennsylvania 2010 and 2011; Ithaca, New York 2012; Cross Plains, Wisconsin 2012 and 2014; Athens, Georgia 2014; Charlottesville, Virginia 2014 and 2015, State College, Pennsylvania 2014; Topeka, Kansas 2014; Sudbury, Ontario, Canada 2015; Benton Harbor, Michigan 2014, Barcelona, Spain 2012; Gross-Enzersdorf, Vienna 2012; Odesa, Ukraine 2013. The four sets of paired spring/fall samples that were not included in the Core20 set were used for cross-validation ('ValidationSet': Media, Pennsylvania 2009, 2012, 2014, Cross Plains, Wisconsin 2013). For comparison of seasonal with latitudinal variation, we used sequence data from four spring samples along the east coast of the United States (Homestead, Florida 2010; Hahia, Georgia 2008; Eutawville, South Carolina 2010, Linvilla, Pennsylvania 2009). We performed a principal components analysis of allele frequency per SNP per sample using the *prcomp* function in *R* (frequency data scaled by SNP).

GLM identification of seasonal sites. We identified seasonal sites using two separate methods, a general linear regression model (GLM) and a rank Fisher's method (RFM). All statistical analyses were performed in *R* (R Core Team 2014). To perform the GLM we used the *glm* function with binomial error, weighted by the "effective coverage" (N_c)- a measure of the number of chromosomes sampled, adjusted by the read depth:

$$N_c = (1/N + 1/R)^{-1}$$

where N 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). This adjusts for the additional error introduced by sampling of the pool at the time of sequencing. The seasonal GLM is a regression of allele frequency by season (e.g., spring versus fall):

$$y_i = \text{season} + \text{population} + e_i$$

where y_i is the allele frequency at the i^{th} SNP, and e_i is the binomial error at the i^{th} SNP. Although the GLM is a powerful test, the corrected binomial sampling variance (N_c) is likely an anti-conservative estimate of the true sampling variance associated with pooled sequencing (Machado *et al.* 2016; Spitzer et al 2019). Therefore, we use the results of this test as a seasonal outlier test, rather than an absolute measure of the deviation from genome-wide null expectation of no allele frequency change between seasons.

RFM identification of seasonal sites.

As a secondary measure of enrichment of seasonally varying sites above neutrality, as well as to estimate the proportion and average selection coefficient of seasonally varying sites, we used a method in which calculated combined *p*-values based on class-based, ranked *p*-values from a Fisher's exact test estimated for each population. In this "rank Fisher's method" (RFM), we calculated a per-SNP *p*-value for a single spring/fall population comparison using a Fisher's

exact test. Then, *p*-values for either a spring to fall increase in allele frequency or spring to fall decrease in allele frequency (i.e., each tail tested separately) were ranked among all other SNPs with the same total number of reads and same total number of alternate reads. The class-based rank of each SNP becomes its new *p*-value, providing uniform sets of *p*-values across the genome. These rank *p*-value calculations were made for each of the 20 CoreSet populations, separately. As the power (minimum *p*-value) is relative to the number of SNPs being ranked per bin, and as the bin sizes are equivalent across populations, each population has equivalent power. We then combine ranked *p*-values using Fisher's method for each SNP, taking two times the negative sum of logged *p*-values across the 20 spring/fall comparisons (each tail tested separately). The null distribution for this statistic (X^2) is a Chi-squared distribution with degrees of freedom equal to 40 (two times the number of comparisons). We used the distribution of these per-SNP X^2 test statistics to 1) test for a seasonal signal above noise and 2) perform an ABC estimation of selection parameters (see below, *The strength …*).

Matched controls. With the assumption that the majority of the genome is not seasonal, we can use matched genomic controls as a null distribution to test for enrichment of different features of seasonal polymorphisms. The use of matched control SNP sets was employed for tests of enrichment of clinal polymorphism, enrichment with seasonal SNPs identified by Bergland *et al.* (Bergland *et al.* 2014), and in tests of genomic clustering. We matched each SNP identified as significantly seasonal (at a range *p*-values) to another SNP, matched for chromosome, effective coverage, median spring allele frequency, inversion status either within or outside of the major inversions *In(2L)t, In(2R)NS, In(3L)P, In(3R)K, In(3R)Mo*, and *In(3R)P*, and recombination rate. We used the same *D. melanogaster* inversion breakpoints used in Corbett-Detig & Hartl (2012) and the recombination rates from Comeron *et al.* (2012). We randomly sampled 100 of the possible matches per SNP (excluding the focal SNP) to produce 100 matched control sets. Any SNPs with fewer than 10 possible matches were discarded from the matched control analyses. We defined 95% confidence intervals from the matched controls as the 3rd and 98th ranked values for the quantity being tested (e.g., percent concordance or proportion of genic classes).

Leave-one-out analysis. To test the general predictability of seasonal change in our dataset, we performed a leave-one-out analysis. In this analysis, we performed seasonal regressions for subsets of the data, dropping one paired sample and comparing it to a seasonal test of the remaining 19. We then measured the percent concordance of spring/fall allele frequency change, defined as the proportion of SNPs that agree in the direction of allele frequency change and sign of regression coefficient. This was performed 20 times, once for each paired sample.

To estimate genome-wide (or chromosome specific) predictability scores, we calculated the rate of change of the concordance score as a function of quantile threshold. To estimate this rate, we used the generalized linear model with binomial error,

$$concordance_i = quantile_i + e,$$

where 'concordance' is the fraction of SNPs falling below the i^{th} quantile threshold of both the target (19 population) and test (1 population) models that changed in frequency in a concordance fashion and e is the binomial error with weights corresponding to the total number of SNPs falling below that threshold. Thus, the genome-wide (or chromosome specific) predictability

score is heavily influenced by concordance rates of higher quantiles because that is where the bulk of SNPs reside.

The thermal limit model. To test whether heterogeneity in genome-wide predictability scores can be explained by aspects of weather, we obtained climate data (daily minimum and maximum temperatures) from the Global Historical Climatology Network-Daily Database (Menne *et al* 2012). We matched each locality to a weather station based on geographic distance. For the Core20 set of populations, three of the collections did not have precise collection dates for one or both seasonal collections (Athens, Georgia 2014; Media, Pennsylvania 2010, 2011), or were not associated with any climate data from the GHCND database (Odesa, Ukraine); for the ValidationSet populations, one of the populations did not have precise collection dates for one of both seasonal collections (Media, Pennsylvania 2009). These populations were removed from our weather model analysis.

To fit the thermal limit model for the Core20 and ValidationSet populations, we regressed the observed predictability score on the number of days above and below the thermal limits using a simple linear regression of the form,

$$\text{PredictabilityScore} = \text{CriticalDays}_{\text{Spring}} + \text{CriticalDays}_{\text{Fall}}$$

and recorded model fit (R^2). We calculated the null distribution of R^2 values by permuting the genome-wide predictability scores across among populations (n=10,000). For each permutation, we recorded the maximum R^2 across the full range of thermal limits and used the distribution of R^2 values as a null distribution. Using this null distribution, we then identified which thermal limits are more strongly correlated with genome-wide predictability scores than expected by chance (p_{perm}).

Predicted predictability scores were calculated using the best fit thermal limit models ($p_{perm} < 0.05$). We calculated the median predicted score across the range of models. Confidence intervals were calculated using one standard deviation across the range of best fit models.

Latitudinal cline concordance. To identify SNPs that changed consistently in allele frequency with latitude (clinal), we first identified SNPs that varied in allele frequency along a 14.4° latitudinal transect up the east coast of the United States. We used one spring sample from each of the following populations (identified as "Region_City_Year"): PA_li_2011 (39.9°N), SC_eu_2010 (33.4°N), GA_ha_2008 (30.1°N), and FL_ho_2010 (25.5°N).

First, we regressed allele frequency with population latitude in a general linear model (*glm*: *R* v3.1.0), using a binomial error model and weights proportional to the effective coverage (N_c):

$$y_i = \text{latitude} + e_i$$

where y_i is the allele frequency at the ith SNP, and e^i is the binomial error given the ith at the SNP. To confirm the robustness of the clinal GLM results, clinality also was assessed using a second, more complex model, that accounts for existing covariance in the dataset (bayenv2: Coop *et al.* 2010; Günther & Coop 2013). Results of the seasonal and clinal concordance analysis using the *bayenv2* model were comparable to that of the GLM model (Supplementary Figure 11) and are not presented in the main text.

We then tested the concordance in allele frequency change by season with the allele frequency change by latitude. We performed three separate seasonal regressions (see above) for comparison with the latitudinal regression: spring verses fall for the 18 non-Pennsylvania paired samples, spring versus fall for the three California paired samples, and spring versus fall for the three Europe paired samples. With the removal of the Pennsylvania samples, none of these three seasonal regressions contained samples from any of the four populations used for the latitudinal regression. Taking sets of increasingly clinal and increasingly seasonal SNPs, we assessed the proportion of sites that both increase in frequency from spring to fall and increase in frequency from north to south or that decrease in frequency from spring to fall and decrease in frequency from north to south. We compared this concordance with the concordance of 100 sets of matched controls.

The strength and genomic extent of seasonal adaptation. We estimated the proportion of sites under selection and the selection strength using the RFM in an ABC analysis (R package *abc*; Csillery *et al.* 2012) using simulated datasets of neutral and selected sites. To simulate neutral SNPs we used the observed spring allele frequency for a given SNP in a given paired sample as the "true" allele frequency and performed a binomial draw of the size equal to the spring and of the fall samples. This was done across all SNPs for each paired sample, producing a simulated, null dataset the same size as the observed dataset, with the same allele frequency distributions as the observed data. This simulated dataset will not necessarily reflect the same sampling (i.e., drift) process as observed in nature from spring to fall, and as neither the local census nor variance effective population size are known, such processes cannot be accurately modeled. However, as our rank *p*-value Fisher's method relies only on the rank within a paired sample, and the consistency of the rank across paired samples, there is not likely to be an artificially inflated seasonal signal due to incorrectly estimated sampling error (as there can be for a regression *p*-value; Supplemental Figure 2).

Selected sites were drawn from the same spring allele frequencies and read depths as the neutral sites, with the exception that the fall allele frequency was drawn from a new allele frequency, following a specified shift from the observed spring frequency by a factor of *s*:

$$AF_{fall} = AF_{spring} + AF_{spring} * (1\text{-}AF_{spring}) * s$$

where *AF* is the allele frequency and *s* is the cumulative "seasonal selection coefficient" from the spring to the fall sample. We use this cumulative *s* since the true number of generations between the spring and fall samples is not known. We simulated datasets with seasonal selection coefficients ranging from 0 to 0.5 (in 0.01 increments), and proportions of sites under selection ranging from 0 to 0.1 (in 0.001 increments). Each simulated dataset was then analyzed in the same manner as the observed data, producing per-dataset distributions of X^2 test statistics. For

each X^2 bin of each dataset, we calculated the proportion of sites greater than expected. To avoid double counting, we excluded the bottom half of the distribution (X^2<20), which consists of negative observed minus expected values and is largely symmetric with the top half. We calculated the square difference between the simulation and the observed proportion of sites per bin, resulting in 28 summary statistics. We also calculated the sum of squared differences across bins, which we refer to as the SSD. ABC parameter estimates were robust to using either all 28 per-bin summary statistics or the SSD, therefore we reduced our analysis to the single SSD summary statistic. We used the rejection method in the *abc* function, choosing the tolerance level of 0.01 based on lowest error in a cross-validation (*cv4abc* function; Supplemental Figure 14).

Seasonal changes in inversion frequencies. To test whether large cosmopolitan inversions change in frequency between spring and fall we calculated the average frequency of SNPs that are closely linked to inversion karyotype (Kapun *et al.* 2014).

Genomic characteristics of seasonal SNPs. We assessed the uniformity of the genomic distribution of highly seasonal SNPs by comparing the observed dataset to that of matched controls and theoretical null distributions. For this analysis we examined the genomic distribution of the top 1% of seasonally varying polymorphisms (n=7,748). We first measured the number of seasonal SNPs per bin 100 SNP bin (average of 1 SNP per bin) and compared this distribution to that of the mean of 100 sets of matched controls to the expectation from a binomial distribution. A distribution that is more strongly peaked around 1 than expected is an under-dispersed genomic distribution of seasonal SNPs (more evenly distributed than expected) whereas a distribution less strongly peaked around 1 is over-dispersed and indicates clustering of seasonal SNPs. We assessed the similarity of the observed and expected distributions using a Kolmogorov-Smirnov test. We also assessed this distribution across genomic length scales. We calculated the probability of observing at least one highly seasonal SNP among 1000 randomly selected genomic windows ranging in size from 100bp to 100Kb. We compared this to the expected probability given a Poisson distribution.

We calculated a "seasonal F_{st}" for each SNP by taking the median of the spring/fall F_{st} (Weir & Cockerham 1984, equations 1:4) values across paired samples.

Data availability. All raw sequence data have been deposited to SRA (BioProject Accession #PRJNA308584; individual accession numbers for each sample can be found in Supplemental Table 1), a VCF file with allele frequencies from all populations, scripts to conduct analyses presented here, and genome-wide SNP statistics for seasonality and clinality are available on DataDryad (doi:10.5061/dryad.4r7b826).

Acknowledgements

We thank the National Evolutionary Synthesis Center (NESCent) for sponsoring the 2012 Catalysis meeting that initiated the Drosophila Real Time Evolution Consortium. The meeting was attended by Alan Bergland, Alisa Sedghifar, Brian Helmuth, Brian Lazzarro, Chau-Ti Ting, David Kidd, Dmitri Petrov, Fabian Staubach, Hannah Burrack, Jim Fry, John Lessard, John Coulbourne, John Pool, Josefa Gonzalez, Julien Ayroles, Kelly Dyer, Kim Hughes, Maaria

Kankare, Nadia Singh, Paul Schmidt, Regan Early, Stephen Porder, Subhash Rajpurohit, Sui Fai Lee, and Thomas Flatt and we kindly thank all participants for their participation. We also thank Jamie Blundell and all members of the Schmidt and Petrov labs who provided exceptionally valuable feedback. This work was funded by the NIH grants R01GM100366 to PS and DAP, R35GM118165 to DAP, R01GM100366 to PS, F32GM097837 and R35GM119686 to AOB, European Commission grant H2020-ERC-2014-CoG-647900 to JG, a NSERC Discovery Grant RGPIN-2018-05551 and Canada Research Chair 950-230113 to TJSM.

Author contributions

Heather E. Machado: Designed the study, performed analysis, wrote manuscript, edited manuscript, collected flies; Alan O. Bergland: Designed the study, performed analysis, wrote manuscript, edited manuscript, collected flies; Ryan Taylor: Performed analysis, edited manuscript; Susanne Tilk: Prepared sequencing libraries; Emily Behrman: Collected flies; Kelly Dyer: Collected flies; Daniel K. Fabian: Collected flies; Thomas Flatt: Collected flies, edited manuscript; Josefa González: Collected flies, edited manuscript; Talia L. Karasov: Collected flies; Iryna Kozeretska: Collected flies; Brian P. Lazzaro: Collected flies, edited manuscript; Thomas JS Merritt: Collected flies; John E. Pool: Collected flies, edited manuscript; Katherine O'Brien: Collected flies; Subhash Rajpurohit: Collected flies; Paula R. Roy: Collected flies; Stephen W. Schaeffer: Collected flies, edited manuscript; Svitlana Serga: Collected flies; Paul Schmidt: Designed the study, wrote manuscript, edited manuscript, collected flies; Dmitri Petrov: Designed the study, wrote manuscript, edited manuscript.

Supplemental Tables

Supplemental Table 1: Detailed information for each sample, including time and location of sample, collector, and SRA accession number.

Supplemental Table 2: Model coefficients for best-fit thermal limit models ($p_{perm} < 0.05$).

Supplemental Table 3: Critical days across a range of thermal limits for each population.

Supplemental Table 4: Regions of the genome with elevated density of seasonally variable SNPs for the original and flipped models

Supplemental Table 5: Results from Gene Ontology analysis of regions harboring an excess of seasonal SNPs using the flipped model.

1205
1206
1207

References

Ayrinhac A, Debat V, Gibert P *et al.* (2004) Cold adaptation in geographical populations of Drosophila melanogaster: phenotypic plasticity is more important than genetic variability. *Functional Ecology*, **18**, 700–706.

Adrion JR, Hahn MW, Cooper BS (2015) Revisiting classic clines in Drosophila melanogaster in the age of genomics. *Trends in Genetics*, **31**, 434-444.

Barghi N, Tobler R, Nolte V, Jakšić AM, Mallard F, Otte KA, et al. (2019) Genetic redundancy Fuels polygenic adaptation in *Drosophila*. *PloS Biology*, **17**, e3000128.

Barron MG, Fiston-Lavier AS, Petrov DA, Gonzalez J. (2014) Population genomics of transposable elements in Drosophila. *Annu Rev Genet*, **48** , 561-81.

Behrman EL, Howick VM, Kapun M, Staubach F, Bergland AO, Petrov DA, Lazzarro BP, Schmidt PS (2018) Rapid seasonal evolution in innate immunity of wild Drosophila melanogaster. *Proceedings of the Royal Society B: Biological Sciences*, **285**, 20172599.

Behrman EL, Watson SS, O'Brien KR, Heschel MS, Schmidt PS (2015) Seasonal variation in life history traits in two Drosophila species. *Journal of Evolutionary Biology*, **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 Genetics*, **10**, e1004775.

Bergland AO, Chae H-S, Kim Y-J, Tatar M (2012) Fine-scale mapping of natural variation in fly fecundity identifies neuronal domain of expression and function of an aquaporin. *PLoS Genetics*, **8**, e1002631.

Bergland AO, Genissel A, Nuzhdin SV, Tatar M (2008) Quantitative trait loci affecting phenotypic plasticity and the allometric relationship of ovariole number and thorax length in Drosophila melanogaster. *Genetics*, **180**, 567–582.

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*. *Molecular Ecology*, **25**, 1157-1174.

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

Betini GS, McAdam AG, Griswold CK, Norris DR (2017) A fitness trade-off between seasons causes multigenerational cycles in phenotype and population size. *eLife*, **6**, e18770.

Botero CA, Weissing FJ, Wright J, Rubenstein DR (2015) Evolutionary tipping points in the capacity to adapt to environmental change. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 184–189.

Boyle EA, Li YI, Pritchard JK. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, **169**, 1177–1186.

Bozicevic V, Hutter S, Stephan W, Wollstein A (2016) Population genetic evidence for cold adaptation in European *Drosophila melanogaster* populations. *Molecular Ecology*, **25**, 1175–1191.

Campo D, Lehmann K, Fjeldsted C *et al.* (2013) Whole-genome sequencing of two North American *Drosophila melanogaster* populations reveals genetic differentiation and positive selection. *Molecular Ecology*, **22**, 5084–5097.

Carson HL, Stalker HD (1948) Reproductive Diapause in *Drosophila robusta*. *Proceedings of the National Academy of Sciences of the United States of America*, **34**, 124–129.

1254 Chakraborty M, Fry JD (2016) Evidence that Environmental Heterogeneity Maintains a
1255 Detoxifying Enzyme Polymorphism in *Drosophila melanogaster*. *Current Biology*, **26**, 219–
1256 223.
1257 Charlesworth B. (2015) Causes of natural variation in fitness: Evidence from studies of
1258 Drosophila populations. *PNAS*, **112** ,1662-1669
1259 Chen Y, Lee SF, Blanc E *et al.* (2012) Genome-Wide Transcription Analysis of Clinal Genetic
1260 Variation in Drosophila. *PloS one*, **7**, e34620.
1261 Cockerell FE, Sgró CM, McKechnie SW (2014) Latitudinal clines in heat tolerance, protein
1262 synthesis rate and transcript level of a candidate gene in *Drosophila melanogaster*. *Journal*
1263 *of Insect Physiology*, **60**, 136–144.
1264 Cogni R, Kuczynski C, Koury S *et al.* (2014) The intensity of selection acting on the *couch*
1265 *potato* gene -- spatial–temporal variation in a diapause cline. *Evolution*, **68**, 538–548.
1266 Cogni R, Kuczynski K, Lavington E *et al.* (2015) Variation in *Drosophila melanogaster* central
1267 metabolic genes appears driven by natural selection both within and between populations.
1268 *Proceedings. Biological sciences / The Royal Society*, **282**, 20142688–20142688.
1269 Comeron JM, Ratnappan R, Bailin S (2012) The many landscapes of recombination in
1270 *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1002905.
1271 Coop G, Witonsky D, Di Rienzo A, Pritchard J.K. (2010). Using Environmental Correlations to
1272 Identify Loci Underlying Local Adaptation. *Genetics*, **185**, 1411-1423.
1273 Corbett-Detig RB, Hartl DL (2013) Correction: Population Genomics of Inversion
1274 Polymorphisms in *Drosophila melanogaster*. PLOS Genetics, 9, 10.1371
1275 Csillery K, Francois O, Blum MGB. (2012), abc: an R package for approximate Bayesian
1276 computation (ABC). *Methods in Ecology and Evolution,* **3**, 475-479.
1277 Curtsinger JW, Service PM, Prout T. (1994) Antagonistic Pleiotropy, Reversal of Dominance,
1278 and Genetic Polymorphism. *The American Naturalist,* **144**, 210-228.
1279 David JR, Gibert P, Gravot E *et al.* (1997) Phenotypic plasticity and developmental temperature
1280 in Drosophila: Analysis and significance of reaction norms of morphometrical traits. *Journal*
1281 *of Thermal Biology*, **22**, 441–451.
1282 Denlinger DL (2003) Regulation of diapause. *Annual Review of Entomology*, **47**, 93–122.
1283 Dobzhansky T (1948) Genetics of natural populations; altitudinal and seasonal changes produced
1284 by natural selection in certain populations of *Drosophila persimilis*. *Genetics*, **33**, 158–176.
1285 Ellner S (1996) Environmental fluctuations and the maintenance of genetic diversity in age or
1286 stage-structured populations. *Bulletin of Mathematical Biology*, **58**, 103–127.
1287 Ellner S, Nelson G Hairston J (2015) Role of Overlapping Generations in Maintaining Genetic
1288 Variation in a Fluctuating Environment. *The American Naturalist*, **143**, 403–417.
1289 Ellner S, Sasaki A (1996) Patterns of genetic polymorphism maintained by fluctuating selection
1290 with overlapping generations. *Theoretical Population Biology*, **50**, 31–65.
1291 Ewing EP (1979) Genetic variation in a heterogeneous environment VII. Temporal and spatial
1292 heterogeneity in infinite populations. *American Naturalist*, **114**, 197–212.
1293 Feder AF, Petrov DA, Bergland AO (2012) LDx: Estimation of Linkage Disequilibrium from
1294 High-Throughput Pooled Resequencing Data (R Wu, Ed,). *PloS one*, **7**, e48588.
1295 Gould, SJ. (1989). *Wonderful life: the Burgess Shale and the nature of history.* W.W. Norton &
1296 Company, Inc. New York, NY.
1297 Gillespie J (1973) Polymorphism in random environments. *Theoretical Population Biology*, **4**,
1298 193–195.

1299 Grenier JK, Arguello JR, Moreira MC *et al.* (2015) Global diversity lines - a five-continent
1300 reference panel of sequenced *Drosophila melanogaster* strains. *G3:*
1301 *Genes|Genomes|Genetics*, **5**, 593–603.
1302 Günther T, Coop G. (2013) Robust identification of local adaptation from allele frequencies.
1303 *Genetics*, **195**: 205-220.
1304 Hedrick PW (1976) Genetic variation in a heterogeneous environment. II. Temporal
1305 heterogeneity and directional selection. *Genetics*, **84**, 145–157.
1306 Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: The age of
1307 genomics. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 67–93.
1308 Hendry AP. (2017) *Eco-evolutionary dyanmics*. Princeton University Press, Princeton, NJ.
1309 Huang DW, Sherman BT, Lempicki RA (2009a) Systematic and integrative analysis of large
1310 gene lists using DAVID bioinformatics resources. *Nature protocols*, **4**, 44–57.
1311 Huang DW, Sherman BT, Lempicki RA (2009b) Bioinformatics enrichment tools: paths toward
1312 the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, **37**, 1–13.
1313 Ives PT (1970) Further genetic studies of the south Amherst population of *Drosophila*
1314 *melanogaster*. *Evolution*, **24**, 507–518.
1315 Juneja P, Quinn A, Jiggins FM (2016) Latitudinal clines in gene expression and cis-regulatory
1316 element variation in *Drosophila melanogaster*. *BMC Genomics*, **17**, 981.
1317 Kain JS, Zhang S, Zade JA *et al.* (2015) Variability in thermal and phototactic preferences in
1318 Drosophila may reflect an adaptive bet-hedging strategy. *Evolution*, **69**, 3171–3185.
1319 Kapopoulou A, Pfeifer SP, Jensen JD, Laurent S. (2018) The demographic history of African
1320 *Drosophila melanogaster*. *Genome Biology and Evolution,* **10**, 2338-2342.
1321 Kapun M, Flatt T (2019) The adaptive significance of chromosomal inversion polymorphisms in
1322 *Drosophila melanogaster*. *Molecular Ecology*, **28**, 1263-1282.
1323 Kapun M, Aduriz MGB, Staubach F *et al.* (2018) Genomic analysis of European *Drosophila*
1324 *melanogaster* populations on a dense spatial scale reveals longitudinal population structure
1325 and continent-wide selection. *bioRxiv*, 313759.
1326 Kapun M, Fabian DK, Goudet J, Flatt T (2016) Genomic Evidence for Adaptive Inversion Clines
1327 in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **33**, 1317–1336.
1328 Kapun M, van Schalkwyk H, McAllister B, Flatt T, Schlötterer C (2014) Inference of
1329 chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations
1330 of *Drosophila melanogaster*. *Molecular Ecology*, **23**, 1813–1827.
1331 Karasov T, Messer P, Petrov DA (2010) Evidence that Adaptation in *Drosophila* Is Not Limited
1332 by Mutation at Single Sites. *PLoS Genetics*, **6**, e1000924.
1333 Keller A. (2007) *Drosophila melanogaster*'s history as a human commensal. *Current Biology*,
1334 **17**, R77-81.
1335 Kim YB, Oh JH, McIver LJ *et al.* (2014) Divergence of *Drosophila melanogaster* repeatomes in
1336 response to a sharp microclimate contrast in Evolution Canyon, Israel. *Proceedings of the*
1337 *National Academy of Sciences of the United States of America*, **111**, 201410372–10635.
1338 Klepsatel P, Gáliková M, De Maio N *et al.* (2013) Reproductive and post-reproductive life
1339 history of wild-caught *Drosophila melanogaster* under laboratory conditions. *Journal of*
1340 *Evolutionary Biology*, **26**, 1508–1520.
1341 Koboldt DC, Zhang Q, Larson DE *et al.* (2012) VarScan 2: somatic mutation and copy number
1342 alteration discovery in cancer by exome sequencing. *Genome Research*, **22**, 568–576.

1343 Kolaczkowski B, Kern AD, Holloway AK, Begun DJ (2011) Genomic differentiation between
1344 temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*, **187**,
1345 245–260.
1346 Kostál V (2006) Eco-physiological phases of insect diapause. *Journal of Insect Physiology*, **52**,
1347 113–127.
1348 Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE (2016) A Thousand Fly Genomes: An
1349 Expanded Drosophila Genome Nexus. *Molecular Biology and Evolution*, **33**, 3308–3313.
1350 Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *The*
1351 *American Naturalist*, **87**, 331–333.
1352 Levine MT, Eckert ML, Begun DJ (2011) Whole-Genome Expression Plasticity across Tropical
1353 and Temperate *Drosophila melanogaster* Populations from Eastern Australia. *Molecular*
1354 *Biology and Evolution*, **28**, 249–256.
1355 Levins R (1968) *Evolution in changing environments: some theoretical explorations*. Princeton
1356 University Press, Princeton, NJ.
1357 Lynch M, Bost D, Wilson S, Maruki T, Harrision S (2014) Population-genetic inference from
1358 pooled-sequencing data. *Genome Biology and Evolution,* **6**, 1210-1218.
1359 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform.
1360 *Bioinformatics*, **25**, 1754–1760.
1361 Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and
1362 SAMtools. *Bioinformatics*, **25**, 2078–2079.
1363 Machado HE, Bergland AO, O'Brien KR *et al.* (2016) Comparative population genomics of
1364 latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Molecular*
1365 *Ecology*, **25**, 723–740.
1366 Mackay TFC, Huang W (2018) Charting the genotype-phenotype map: lessons from the
1367 *Drosophila melanogaster* Genetic Reference Panel. *WIREs Dev Biol,* **7**, e289.
1368 Mackay TFC, Richards S, Stone EA *et al.* (2012) The *Drosophila melanogaster* Genetic
1369 Reference Panel. *Nature*, **482**, 173–178.
1370 MacMillan HA, Ferguson LV, Nicolai A *et al.* (2015) Parallel ionoregulatory adjustments
1371 underlie phenotypic plasticity and evolution of Drosophila cold tolerance. *Journal of*
1372 *Experimental Biology*, **218**, 423–432.
1373 Mallard F, Nolte V, Tobler R, Kapun M, Schlötterer C (2018) A simple genetic basis of
1374 adaptation to a novel thermal environment results in complex metabolic rewiring in
1375 Drosophila. *Genome Biology*, **19**, 119.
1376 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
1377 *EMBnet.journal*, **17**, pp. 10–12.
1378 McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce
1379 framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**,
1380 1297–1303.
1381 McMillan I, Fitz-Earle M, Robson DS (1970) Quantitative genetics of fertility. II. Lifetime egg
1382 production of *Drosophila melanogaster*--experimental. *Genetics*, **65**, 355–369.
1383 Noh S, Everman ER, Berger CM, Morgan TJ (2017) Seasonal variation in basal and plastic cold
1384 tolerance: Adaptation is influenced by both long- and short-term phenotypic plasticity.
1385 *Ecology and Evolution*, **7**, 5248–5257.
1386

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

Overgaard J, Kristensen TN, Mitchell KA, Hoffmann AA (2015) Thermal Tolerance in Widespread and Tropical Drosophila Species: Does Phenotypic Plasticity Increase with Latitude? *The American Naturalist*, **178**, S80–S96.

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.

Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A Critical Assessment of Storytelling: Gene Ontology Categories and the Importance of Validating Genomic Scans. *Molecular Biology and Evolution*, **29**, 3237–3248.

Pool JE, Corbett-Detig RB, Sugino RP *et al.* (2012) Population genomics of Sub-Saharan *Drosophila melanogaster*: African diversity and Non-African admixture. *PLoS Genetics*, **8**, e1003080.

R Core Team. (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rajpurohit S, Gefen E, Bergland AO *et al.* (2018) Spatiotemporal dynamics and genome-wide association analysis of desiccation tolerance in *Drosophila melanogaster*. *Molecular Ecology*, **27**, 3525-3540.

Rajpurohit S, Hanus R, Vrkoslav V *et al.* (2017) Adaptive dynamics of cuticular hydrocarbons in Drosophila. *Journal of Evolutionary Biology*, **30**, 66–80.

Rajpurohit S, Nedved O, Gibbs AG (2013) Meta-analysis of geographical clines in desiccation tolerance of Indian drosophilids. *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology*, **164**, 391–398.

Rand DM, Weinreich DM, Lerman D, Folk D, Gilchrist GW (2010) Three Selections Are Better Than One: Clinal Variation of Thermal QTL From Independent Selection Experiments in Drosophila. *Evolution*, **64**, 2921–2934.

Reinhardt JA, Kolaczkowski B, Jones CD, Begun DJ, Kern AD (2014) Parallel geographic variation in *Drosophila melanogaster*. *Genetics*, **197**, 361–373.

Rhomberg, L.R. and R.S. Singh. (1989) Evidence for a link between local and seasonal cycles in gene frequencies and latitudinal clines in a cyclic parthenogen. *Genetica*, **78**, 73-79.

Rodríguez-Trelles F, Tarrío R, Santos M (2013) Genome-wide evolutionary response to a heat wave in Drosophila. *Biology Letters*, **9**, 20130228.

Rudman SM, Barbour M, Csillery K, Gienapp P, Guillaume F, Hairston N, Hendry A, Lasky JR, Rafajlović M, Räsänen K, Schmidt PS, Seehausen O, Therkildsen NO, Turcotte MM, Levine J. (2017). What genomic data can reveal about eco-evolutionary dynamics. *Nature Ecology & Evolution,* **2**, 9-15.

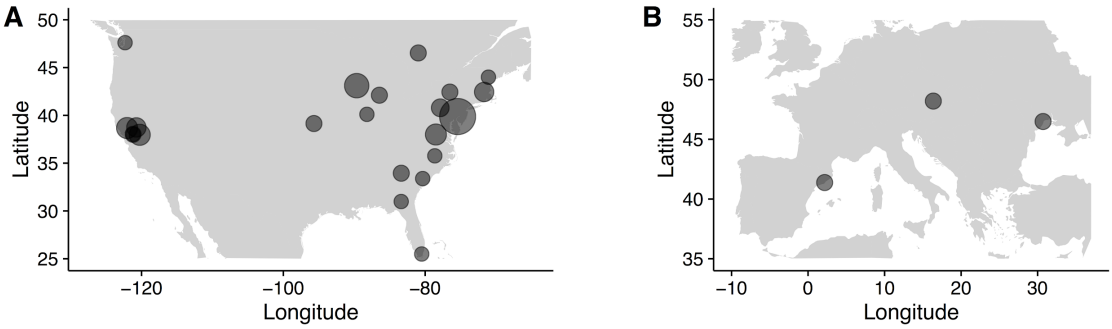
Scheiner SM (1993) Genetics and Evolution of Phenotypic Plasticity. *Annual Review of Ecology and Systematics,* **24**, 35-68.

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, **15**, 749–763.

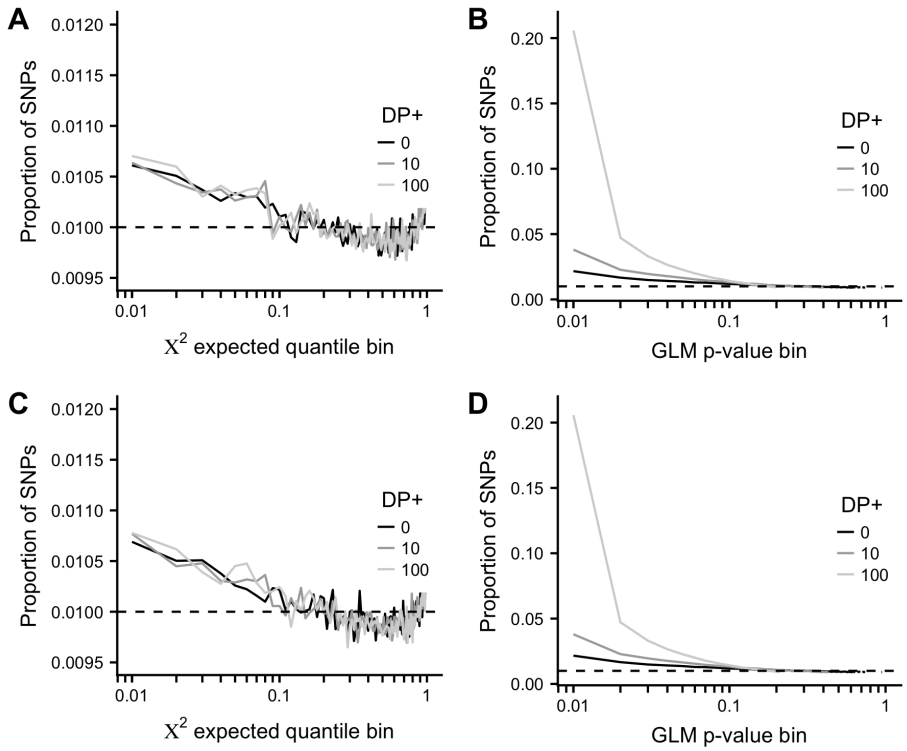
Schmidt PS, Conde DR (2006) Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in *Drosophila melanogaster*. *Evolution*, **60**, 1602-1611.

1432 Schmidt PS, Matzkin L, Ippolito M, Eanes WF (2005) Geographic variation in diapause
1433 incidence, life-history traits, and climatic adaptation in *Drosophila melanogaster*. *Evolution*,
1434 **59**, 1721–1732.
1435 Schmidt PS, Zhu C-T, Das J *et al.* (2008) An amino acid polymorphism in the couch potato gene
1436 forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proceedings of the*
1437 *National Academy of Sciences of the United States of America*, **105**, 16207–16211.
1438 Signor SA, New FN, Nuzhdin S. A large panel of *Drosophila simulans* reveals an abundance of
1439 common variants. *Genome Biology and Evolution*, **10**, 189–206.
1440 Spitzer K, Pelizzola M, Futschik A. (2019). Modifying the Chi-square and the CMH test for
1441 population genetic inference: adapting to over-dispersion. http://arxiv.org/abs/1902.08127.
1442 Stalker HD. (1976). Chromosome studies in wild populations of *Drosophila*
1443 *melanogaster*. *Genetics,* **82**: 323-347.
1444 Stalker HD (1980). Chromosome Studies in Wild Populations of *Drosophila melanogaster*. II.
1445 Relationship of Inversion Frequencies to Latitude, Season, Wing-Loading and Flight
1446 Activity. *Genetics*, **95**, 211-223.
1447 Svetec N, Cridland JM, Zhao L, Begun DJ (2016) The Adaptive Significance of Natural Genetic
1448 Variation in the DNA Damage Response of *Drosophila melanogaster*. *PLoS Genetics*, **12**,
1449 e1005869.
1450 Svetec N, Zhao L, Saelao P, Chiu JC, Begun DJ (2015) Evidence that natural selection maintains
1451 genetic variation for sleep in *Drosophila melanogaster*. *BMC evolutionary biology*, **15**, 41.
1452 Tauber MJ, Tauber CA, Masaki S (1986) *Seasonal Adaptations of Insects*. Oxford University
1453 Press, Oxford, UK.
1454 Tobler R, Franssen SU, Kofler R *et al.* (2014) Massive habitat-specific genomic response in *D.*
1455 *melanogaster* populations during experimental evolution in hot and cold environments.
1456 *Molecular Biology and Evolution*, **31**, 364–375.
1457 Vesala L, Salminen TS, Kostál V, Zahradnickova H, Hoikkala A (2012) Myo-inositol as a main
1458 metabolite in overwintering flies: seasonal metabolomic profiles and cold stress tolerance in
1459 a northern drosophilid fly. *Journal of Experimental Biology*, **215**, 2891–2897.
1460 Via S, Lande R (1985) Genotype-environment interaction and the evolution of phenotypic
1461 plasticity. *Evolution*, **39**, 505–522.
1462 Weir BS, Cockerham CC. (1984) Estimating F-Statistics for the Analysis of Population
1463 Structure. *Evolution*, **38**, 1358-1370
1464 Wiberg RAW, Gaggiotti OE, Morrissey MB, Ritchie MG (2017) Identifying consistent allele
1465 frequency differences in studies of stratified populations. (L Johnson, Ed,). *Methods in*
1466 *Ecology and Evolution*, **8**, 1899–1909.
1467 Wittmann MJ, Bergland AO, Feldman MW, Schmidt PS, Petrov DA (2017) Seasonally
1468 fluctuating selection can maintain polymorphism at many loci via segregation lift.
1469 *Proceedings of the National Academy of Sciences of the United States of America*, **114**,
1470 E9932–E9941.
1471 Zhao L, Begun DJ (2017) Genomics of parallel adaptation at two timescales in Drosophila. *PLoS*
1472 *Genetics*, **13**, e1007016.
1473 Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical Validation of Pooled Whole
1474 Genome Population Re-Sequencing in Drosophila melanogaster. *PloS one*, **7**.
1475

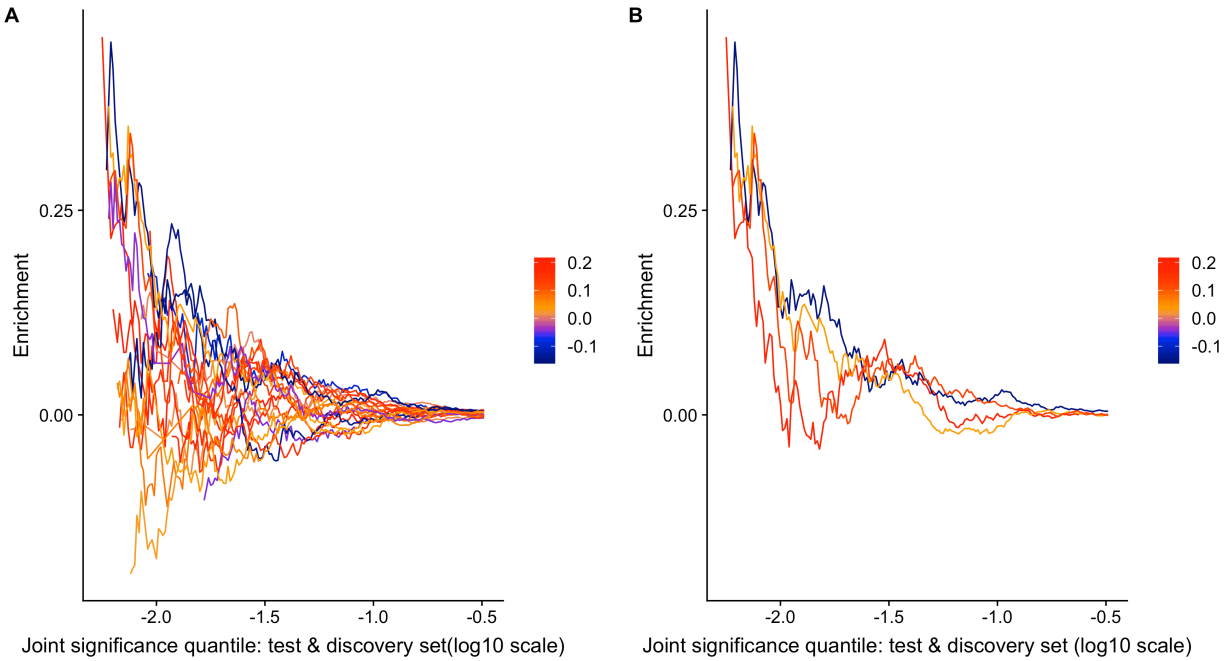
Supplemental Figures



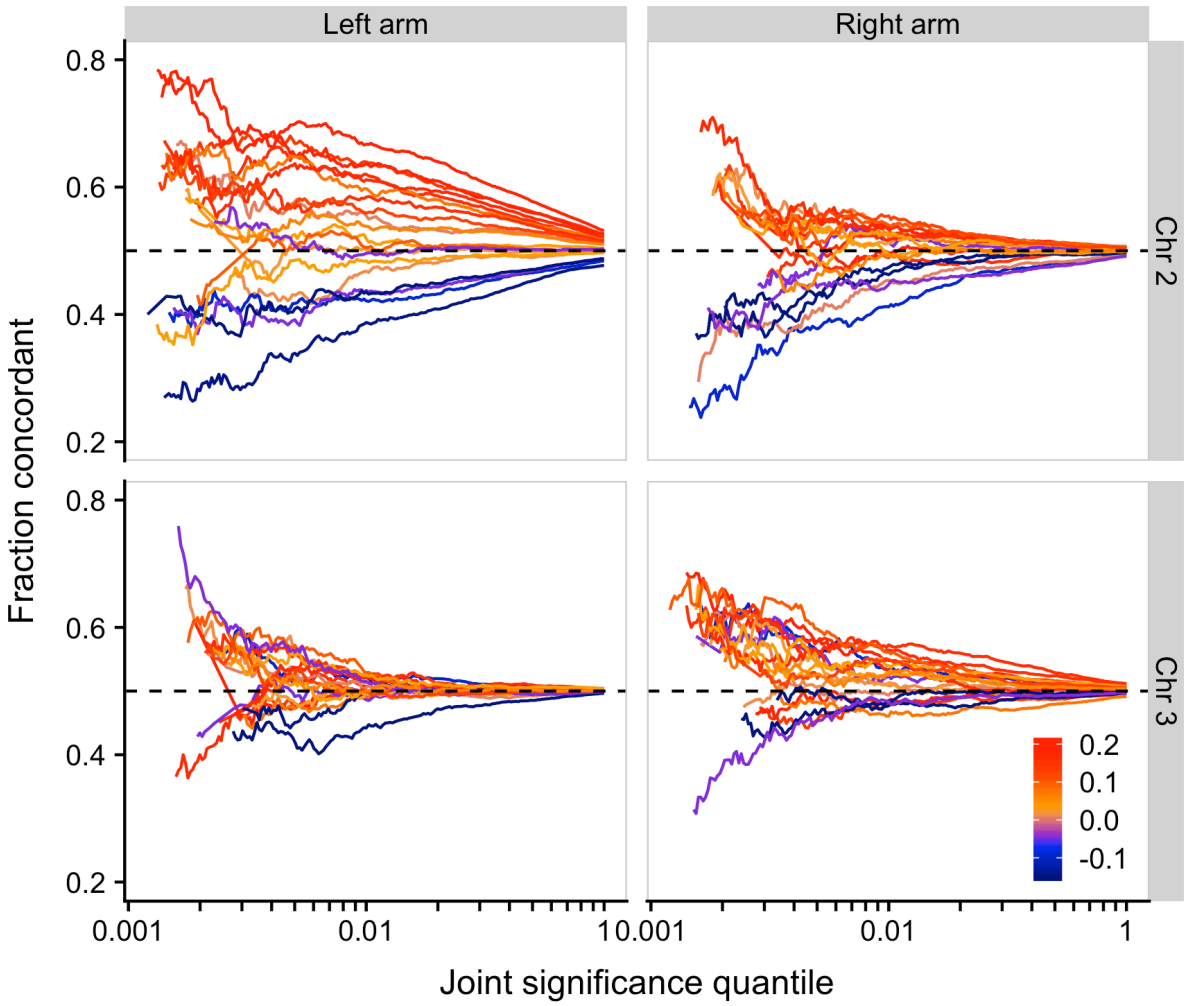
Supplemental Figure 1. Sampling localities in A) North America and B) Europe for all samples collected, regardless of use in the present analysis. Sizes of circles reflect number of samples per locality.



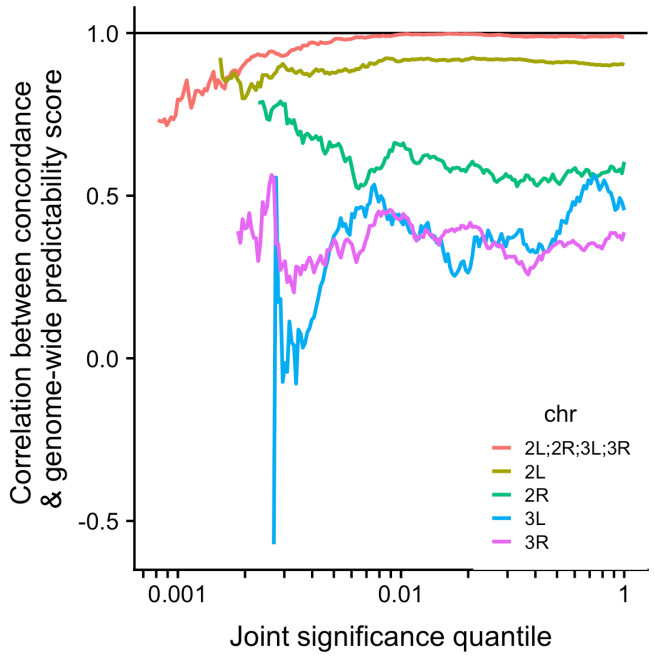
Supplemental Figure 2. The effect of artificial increases of sample size on the Core20 seasonal analysis using A,C) the rank Fisher's method and B,D) the GLM analyses. Top row: sample size inflated by a fixed amount. Bottom row: sample size inflated by a Poisson sampled amount. Shown is the proportion of SNPs in each bin (solid lines) compared with the null expectation (dashed line). Black: original dataset. Dark grey: 10 reads added to each sample (same allele frequencies). Light grey: 100 reads added to each sample (same allele frequencies).



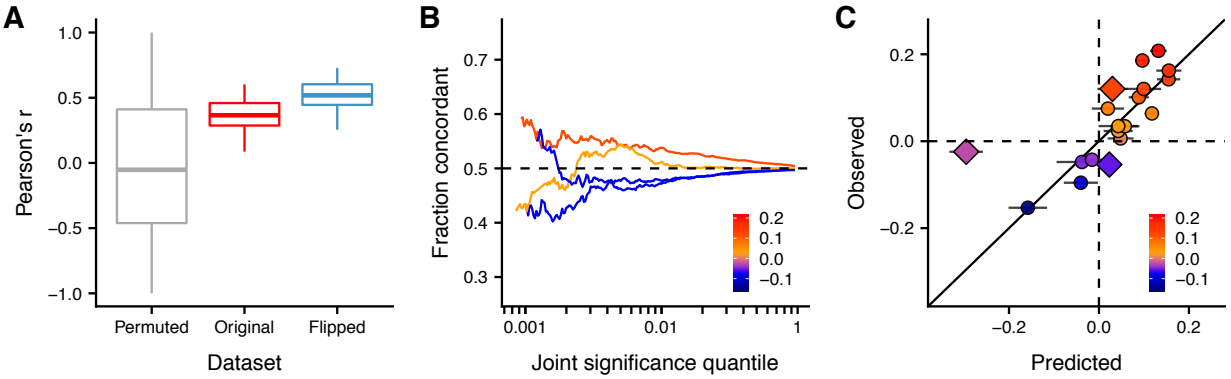
Supplemental Figure 3. The enrichment of shared seasonal sites from the leave-one-out analysis of the Core20 populations. Each line represents one population, colored by the inferred genome-wide predictability score. (A) Enrichment scores for all populations; (B) Enrichment scores for populations with significant enrichment (nominal $p < 0.05$) at a joint significance threshold less than 1% (-2).



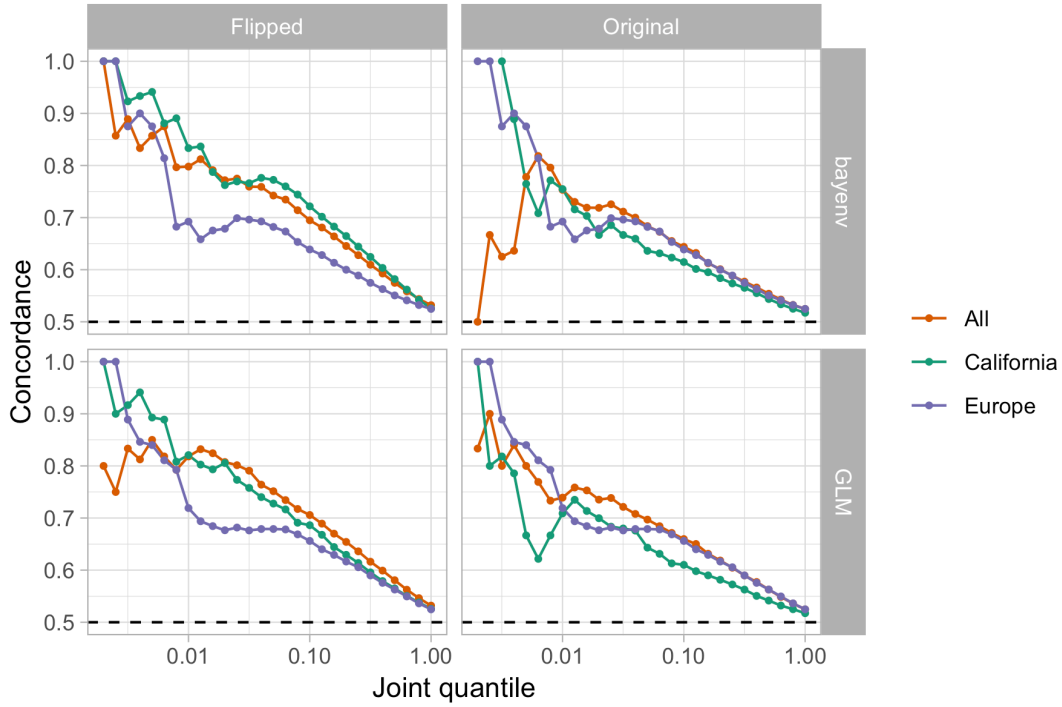
Supplemental Figure 4. Predictability of seasonal adaptation, by chromosome. The fraction of concordant allele frequency changes from the leave-one-out cross validation analysis of the Core20 populations. Each line represents one population, colored by the inferred genome-wide predictability score.



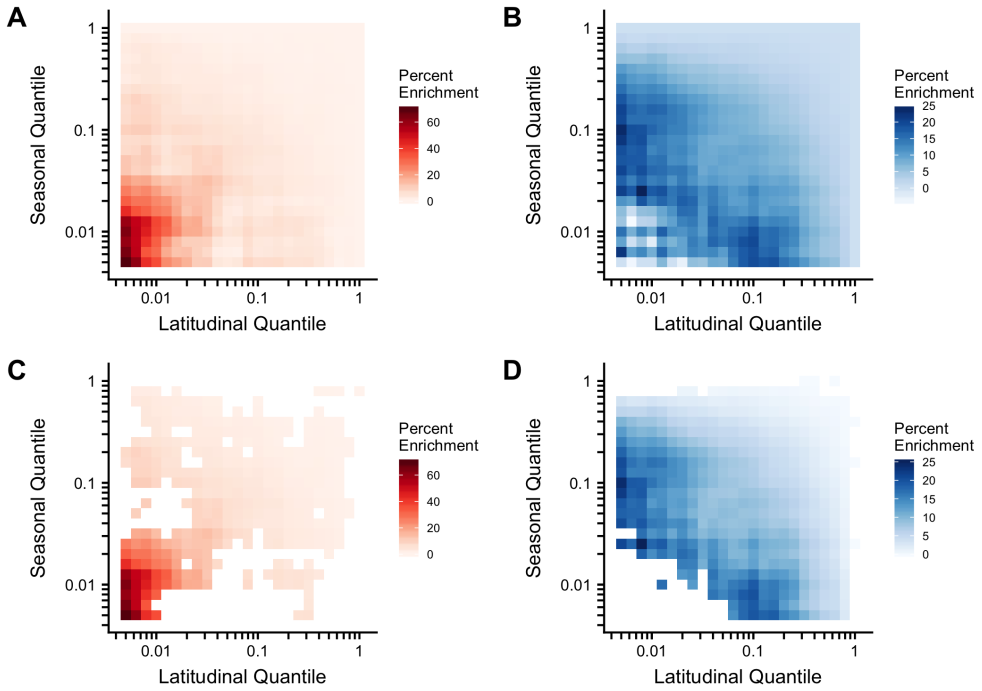
Supplemental Figure 5. Correlation between genome-wide predictability scores and per-chromosome predictability scores.



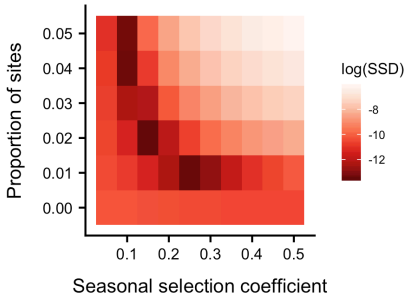
Supplemental Figure 6. Additional plots related to the predictability analysis. A) Pearson's correlation coefficient between the ValidationSet's observed genome-wide concordance score and their predicted scores for the best-fit permuted models (null) versus observed scores generated from the original Core20 model versus the flipped Core20 model. B) Genome-wide concordance scores from the ValidationSet of populations in relation to the flipped Core20 model. C) The best thermal limit models from the Core20 analysis predict genome-wide concordance scores of the ValidationSet (diamonds).



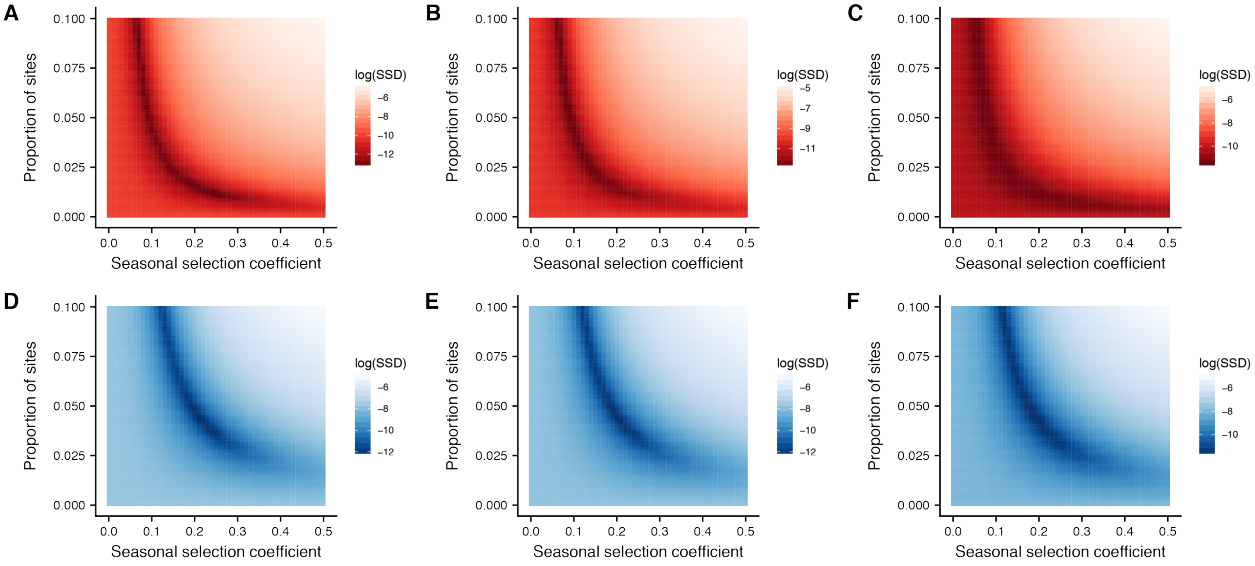
Supplemental Figure 7. Comparison of the general linear model (GLM: main text) and the bayenv model for assessing clinal and seasonal concordance. The joint quantile is the quantile threshold for both the seasonal and the clinal analysis (plot is cumulative).



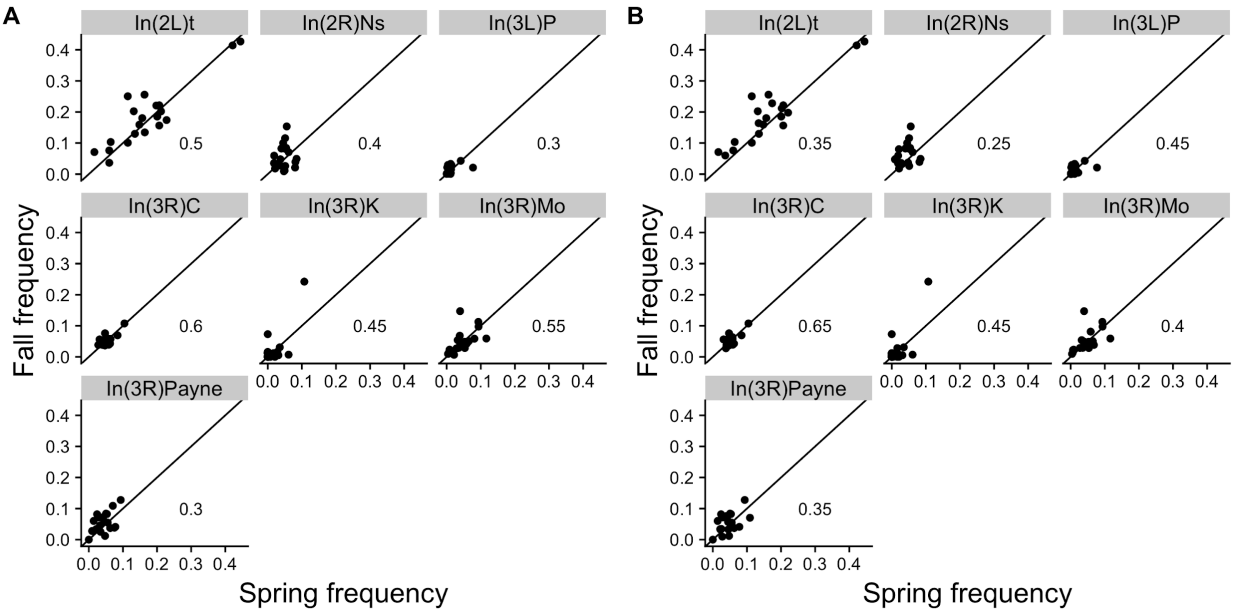
Supplemental Figure 8. The enrichment of SNPs found to be both seasonal and latitudinal, regardless of concordance, as a function of *p*-value quantile thresholds. Darker hues indicate a greater proportion of concordant SNPs. The percent enrichment is calculated as the percent increase in SNPs at a joint quantile threshold over matched controls. A,C) Original dataset. B,D) Flipped dataset. A,B) All values, regardless of a significant enrichment. C,D) Only values greater than 95% of controls (otherwise, colored white).



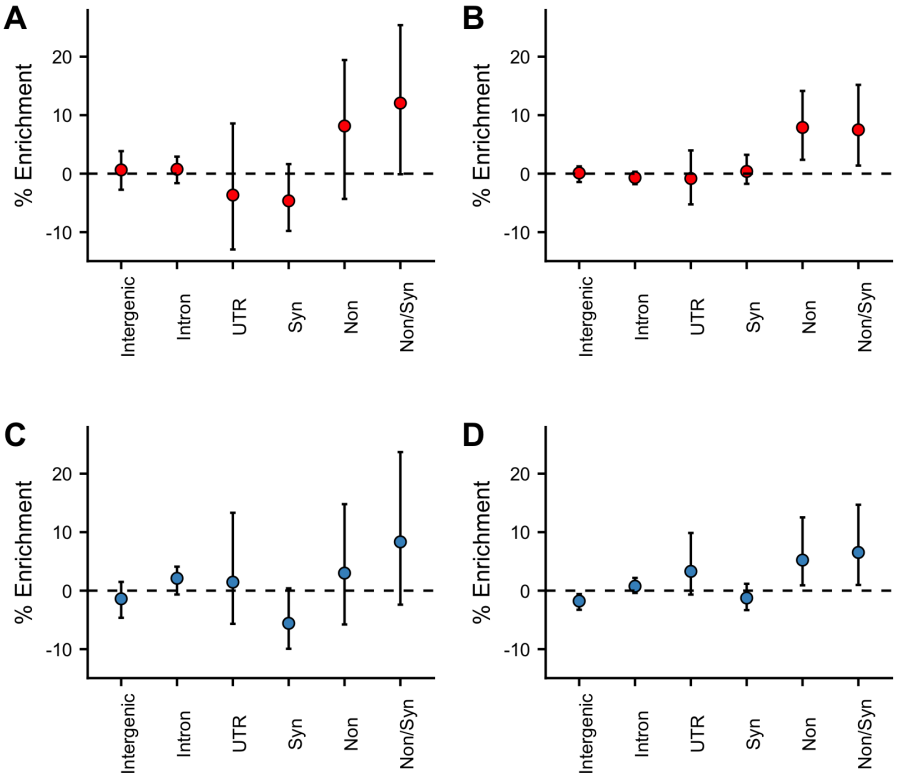
Supplemental Figure 9. ABC estimates for the proportion of sites affected by seasonal selection and the associated *s* for the original model *without* filtering sites found to be monomorphic in one or more populations. Plotted is the sum of squared distances (SSD) between the observed and simulation RFM X^2 test statistic distributions, for varying strengths of selection and proportion of sites under selection.



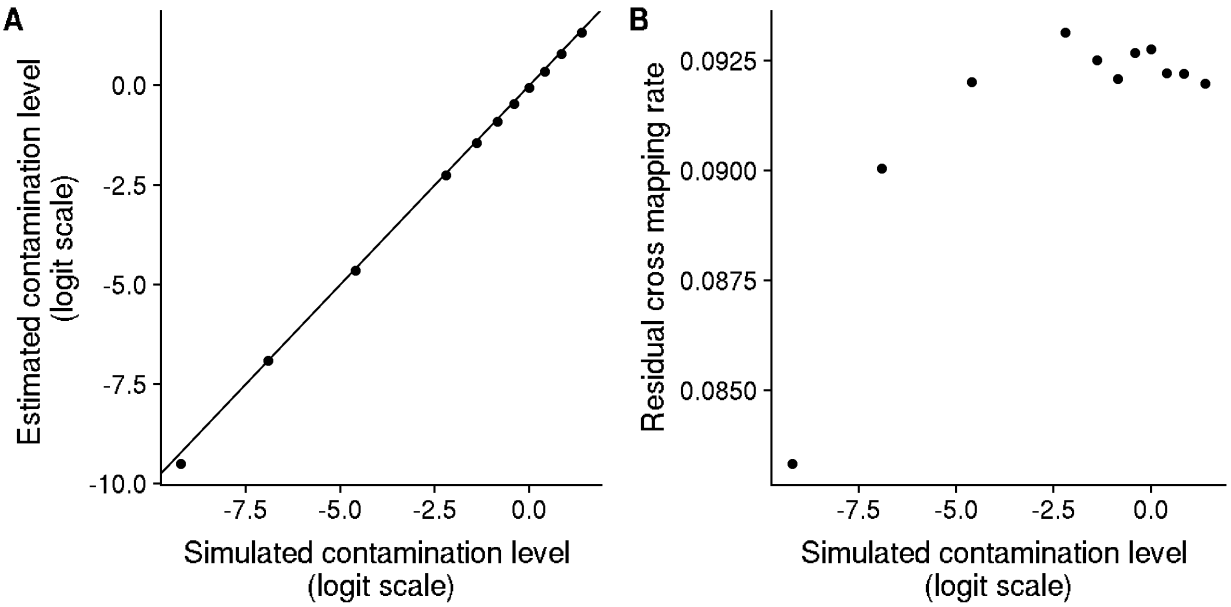
Supplemental Figure 10. Effect of subsampling on ABC parameter estimates. Top row: original dataset. Bottom row: flipped dataset. A,D) No subsampling. B,E) Sampling one SNP per 1Kb. C,F) Sampling one SNP per 5Kb.



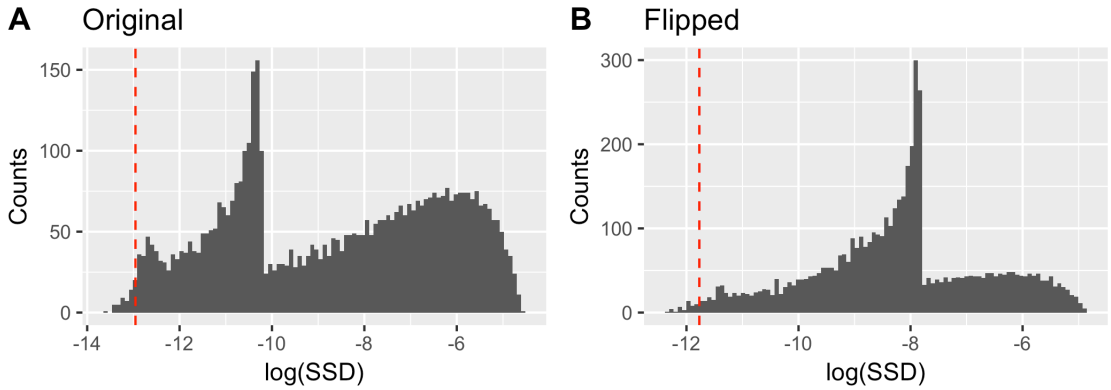
Supplemental Figure 11. Inversion frequencies. Each datapoint is a separate population. Numbers are the fraction of populations where spring is greater than fall. A) Original dataset. B) Flipped dataset.



Supplemental Figure 12. Enrichment of genic classes for A) the top 1% and B) top 5% of seasonal sites for the original analysis, and C) the top 1% and D) top 5% for the flipped analysis. Points represent means and error bars represent 95% CI across100 bootstrapped sets of matched control SNPs.



Supplemental Figure 13. Competitive mapping effectively reduces cross species mapping between *D. simulans* and *D. melanogaster*. (A) The estimated rate of *D. simulans* contamination can be precisely estimated by calculating the number of reads mapping preferentially to the *D. simulans* genome. Estimated contamination rates are ~2% lower than the expected rate. (B) The residual cross species mapping rate, i.e. the fraction of *D. simulans* reads mapping to the *D. melanogaster* genome is roughly consistent at ~9%.



Supplemental Figure 14. Distribution of the summary statistic SSD used in the ABC parameter estimation. Red dashed line indicates the cutoff for the top 1% of simulations (corresponds to the ABC tolerance level). A) Original dataset. B) Flipped dataset.