Broad geographic sampling reveals predictable and pervasive 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], Thomas Flatt[6], Josefa González[7], Talia L. Karasov[8], Iryna Kozeretska[9], Brian P. Lazzaro[10], Thomas JS Merritt[11], John E. Pool[12], Katherine O'Brien[4], Subhash Rajpurohit[4], Paula R. Roy[12], Stephen W. Schaeffer[14], Svitlana Serga[9], Paul Schmidt[4°‡], Dmitri 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 Sanger Institute, Cambridge, 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] Institute of Evolutionary Biology, CSIC- Universitat Pompeu Fabra, Barcelona, Spain
[8] Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA
[9] Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[10] Department of Entomology, Cornell University, Ithaca, NY, 14853
[11] Department of Chemistry & Biochemistry, Laurentian University, Sudbury, ON, Canada
[12] Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI
[13] Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence KS 66045
[14] Department of Biology, The Pennsylvania State University, University Park, PA

Abstract

Local adaptation in response to spatially varying selection pressures is widely recognized as a ubiquitous feature for many organisms. In contrast, our understanding of local adaptation to temporally varying selection pressures is limited. To advance our understanding of local adaptation to temporally varying selection pressures, we studied genomic signatures of seasonal adaptation in *Drosophila melanogaster.* We generated whole-genome estimates of allele frequencies from flies sampled during the spring and fall from 15 localities. We show that seasonal adaptation is a general feature fly populations and that the direction of seasonal adaptation can be predicted by weather conditions in the weeks prior to sampling. We find that seasonal changes in allele frequency are mirrored by spatial variation in allele frequency and that seasonal adaptation affects allele frequencies at ~1.0-2.5% of polymorphisms genome-wide. Our work demonstrates that seasonal adaptation is a major evolutionary force affecting *D. melanogaster* populations living in temperate environments.

Introduction

Organisms living in temperate environments experience dramatic changes in selection pressure across the seasons. Long-lived individuals, as well as species with a limited number of generations per year, exhibit plastic physiological and behavioral strategies that enable them to survive the unfavorable season and exploit the favorable one (Denlinger 2003; Kostál 2006); this is the classical form of seasonal adaptation (Tauber *et al.* 1986). In contrast, for short lived organisms with multiple generations per year, seasonally varying selection can maintain fitness related genetic variation if some genotypes have high fitness in one season but not another (Gillespie 1973). In these organisms, 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 therefore be seen as a form of local adaptation. However, when compared to other forms of local adaptation (Ewing 1979), seasonal adaptation has been considered by some to be uncommon and, when present, unlikely to result in long-term balancing selection (Hedrick 2006). Classic quantitative genetic theory suggests that an optimal, plastic genotype will eventually dominate a population that is exposed to periodically changing environments, particularly when certain environmental cues are reliable indicators of changes in selection pressure(Levins 1968; Via & Lande 1985). Predictions from traditional population genetic models have 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) shows that plasticity is unlikely 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 when the seasonally favored allele is dominant. These recent models, along with others that highlight the importance of population cycles (Bertram & Masel 2017) as well as overlapping generations and age structure (Ellner 1996; Ellner & Sasaki 1996; Ellner & Nelson G Hairston 2015; Bertram & Masel 2017), suggest that seasonal adaptation could be an important feature of short-lived, multivoltine organisms with highly age-structured populations (Behrman *et al.* 2015), such as Drosophila.

In drosophilids there is abundant evidence of seasonal adaptation. Seasonal adaptation was first observed in *D. pseudoobscura* by Dobzhansky and colleagues (Dobzhansky 1948) by tracking allele frequencies of inversions over seasons. Recent studies of other flies in the *obscura* group have demonstrated similar seasonal shifts of inversion frequencies and have linked these seasonal shifts to acute thermal stress (Rodríguez-Trelles *et al.* 2013). In *Drosophila melanogaster*, multiple lines of evidence from phenotypic and genetic analysis demonstrate the presence of seasonal adaptation. When reared in a common laboratory

environment, *D. melanogaster* collected in the spring show higher stress tolerance (Behrman *et al.* 2015), greater propensity to enter diapause (Schmidt & Conde 2009), increased innate immune function (Behrman *et al.* 2018), and modulated cuticular hydrocarbon profiles (Rajpurohit *et al.* 2017) compared to flies collected in the fall. Rapid adaptation over seasonal time scales in these and related phenotypes have also been observed in laboratory (Schmidt & Conde 2009) and field based mesocosm experiments (Rajpurohit *et al.* 2017). Genome-wide analysis identified 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 also 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*, we have a limited understanding of the predictability of allele frequency changes that arise as a consequence of seasonal adaptation and the fraction of polymorphisms genome-wide whose allele frequencies are affected by seasonal adaptation. Furthermore, it is unclear how these patterns of seasonal adaptation vary over spatial gradients, and how spatially vs. temporally varying selection pressures may interact to shape patterns of genetic variation across the genome. To gain insight into these basic questions, it is necessary to have widespread and coordinated sampling efforts at multiple scales; this is one goal of the two Drosophila population genomics consortia DrosRTEC (presented here) and DrosEU (Kapun *et al.* 2018). Here, we focus on seasonal adaptation and estimate allele frequencies genome-wide from *D. melanogaster* collected in the spring and fall at 15 localities in North America and Europe. First, we highlight the genomic signal of seasonal adaptation by examining the general patterns of allele frequency change throughout the genome. Next, using a cross validation approach, we show that allele frequency changes between seasons are predictable when taking into account weather in the weeks immediately prior to sample collection. We demonstrate that aspects of latitudinal variation in allele frequency mirror seasonal fluctuations in allele frequency, even when seasonal and clinal samples were collected on different continents or longitudinal extremes of North America. Finally, using a simulation approach, we show that seasonally varying selection affects allele frequency changes between seasons at more than 1.0 - 2.5% of all common polymorphisms with cumulative seasonal selection coefficients of 10 - 20% per season. Taken together, our work demonstrates that seasonal adaptation is a general feature of *D. melanogaster* populations and has subtle but pervasive effects on patterns of allele frequency genome-wide.

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) each containing pooled DNA sequencing from multiple individual flies. For 15 sampling localities, flies were sampled in the spring and fall over the course of one to six years (Figure 1A). We refer to unique sampling localities or sampling years within a locality as a population. Fly samples were collected through the collaborative effort of the Drosophila Real Time Evolution (DrosRTEC)

consortium and in parallel with similar efforts throughout Europe by the DrosEU consortium (Kapun *et al.* 2018).

We divided the full set of samples into three subsets (Figure 1B). The first 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. The second subset (hereafter 'Core20') is composed of 20 populations consisting of one spring and one fall sample. Populations in the Core20 set are drawn from 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 third subset (hereafter 'ValidationSet') is composed of four populations with one spring and one fall sample. Populations in the ValidationSet are drawn from localities where there are more than two years of sampling. The 'ValidationSet' is used as a part of our cross-validation analysis.

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. For each sample, we resequenced pools of ~75 individuals (range 27-164) to an average of 94x coverage (range 22-220, Supplemental Table 1). All raw sequence data have been deposited to SRA (BioProject Accession #PRJNA308584) and a VCF file with allele frequencies from all populations is available on DataDryad (ACCESSION # TBD).

After filtering, pooled resequencing identified 774,651 common autosomal SNPs which we use in the present analysis (see Materials and methods for details on mapping and filtering criteria). To gain insight into basic patterns of differentiation among our samples, we performed a principal component analysis across all samples. Samples cluster by population and geography (Figure 1C) with the principal component 1 separating European, North American West Coast and East Coast populations while principal component 2 separating the Eastern North America samples by latitude (linear regression $p = 3*10^{-15}$, $R^2=0.84$). No principal component is associated with season following correction for multiple testing.

<u>*Basic genomic signals of seasonal adaptation*</u>. To identify a basic signature of seasonal adaptation, we assessed the per-SNP change in allele frequency between spring and fall among the Core20 set of populations. To perform this test, we used a generalized linear model, regressing allele frequency at each SNP on season (spring or fall) with population specific intercepts and binomial sampling variance. Our estimate of the sampling variance takes into account the number of flies sampled and per-site read depth (i.e., effective coverage (Kolaczkowski *et al.* 2011; Feder *et al.* 2012; Bergland *et al.* 2014; Machado *et al.* 2016)). The genome-wide, observed 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).

The excess of SNPs with low p-values could be a signature of seasonal adaptation or could be a consequence of an incorrect and inflated precision of allele frequency estimates (Supplemental Figure 2) (Machado *et al.* 2016). To resolve whether the genome-wide excess of SNPs with low *p*-values reflects seasonal adaptation or statistical artifact, we performed a permutation test. In our permutation procedure, we randomly shuffled the season label within populations 100 times and refit the generalized linear model. Note that these permutations retain all of the features of the data except for the seasonal labels within populations, and the test is thus robust to any violations of the GLM model assumptions. In addition, we note that our permutation test is conservative, as many permutations necessarily retain much of the seasonal structure among populations. Thus, it is not surprising that all of the permutations are also enriched for low p-values. Still, we find an excess of low p-values in the observed distribution compared even to the conservative permutations (Figure 2A), with a greater proportion of SNPs with $p<0.001$ than 98% of permutations. Our permutation analysis thus suggests that the genome-wide excess of SNPs with low p-values contains a signal of seasonal adaptation.

We examined general patterns of allele frequency change at the top 1% most seasonally variable SNPs identified from the linear models. We find that the mean allele frequency change between spring and fall at these seasonal SNPs is 4-8%, increasing as a function of initial allele frequency (Figure 2B). These new estimates of allele frequency change are smaller than those previously reported from a single orchard (~20%, Bergland *et al* (2014)). Bergland et al (2014) examined the top ~1500 seasonally variable SNPs as opposed to the top ~7500 examined here. However, the average allele frequency change of the top 1500 seasonally variable SNPs identified here is only ~9%. The more modest change in allele frequency between spring and fall that we observe here compared to Bergland et al (2014) is possibly caused by an increased power to detect smaller changes in allele frequency and a reduction of the Beavis effect (Xu 2003), resulting from the larger sample presented in this study. In addition, the broad geographic spread of the populations, the concomitantly large variation in environmental parameters, and our focus on the SNPs showing consistent fluctuations across all populations is likely to reduce the magnitude of the detected fluctuations.

The top 1% most seasonally variable SNPs identified here are significantly enriched for seasonal SNPs identified by Bergland et al (2014) relative to control SNPs matched for several genomic characteristics ($\log_2$ odds ratio = 1.19, $p < 0.001$; see Materials and methods). 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 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 ($\log_2$ odds ratio ± SD = 0.59 ± 0.37, $p = 0.0512$).

One expectation of rapid changes in allele frequency at seasonally varying SNPs is a localized signal of linked selection. To assess the extent of such linked selection, we calculated the

average, between season F_{ST} surrounding each of the top 1% most seasonally variable SNPs. F_{ST} is indeed elevated surrounding these SNPs but decays to background levels rapidly, on average by 200bp (Figure 2C). The localized signal of linked selection coupled with the broad distribution of seasonally variable SNPs (see below, 'Clustering') suggests that seasonal adaptation affects neutral genetic variation genome-wide.

We further tested whether the seasonally varying polymorphisms that we identify here are enriched among different functional annotation classes. The top 1% most seasonally varying polymorphisms are not significantly enriched for any functional class that we tested, although there is a suggestive enrichment of non-synonymous sites relative to synonymous sites as compared to matched control polymorphisms (Supplemental Figure 3). The top 5% of seasonally varying polymorphisms are significantly enriched for non-synonymous sites as compared to matched control polymorphisms (enrichment = 8%, 95% CI = 2-14% p=0.01). The ratio of non-synonymous to synonymous polymorphisms among the top 5% of seasonal SNPs is also 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.

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 1B). A high degree of concordance in allele frequency change suggests a model wherein seasonal adaptation is, to at least some extent, predictable. To explicitly test this model, we performed a leave-one-out cross validation 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 using a generalized linear model (hereafter, the 'discovery sets'). We identified seasonally variable SNPs from the dropped, 20th, population using a Fisher's exact test with allele counts corrected for double binomial sampling as above. We then tested 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 19 population model and the single population model across a range of significance thresholds. For each population, we calculated a concordance score. This score is based on the relationship between the joint significance threshold for the two models and the proportion of SNPs (genome-wide, or per chromosome) below that threshold that change in a concordant fashion between spring and fall (see Materials and Methods for more details). Note that populations with a positive concordance score are those in which the concordance score is greater than 50% over the bulk of the genome or chromosome. In this cross validation analysis, we also calculated the enrichment of polymorphisms that are identified as seasonal in each single population model and the corresponding 19-population model, regardless of whether allele frequencies changed in a concordant fashion.

Our cross validation analysis yielded several basic results. First, we find that there is only a modest enrichment genome-wide of polymorphisms that are identified as seasonally variable in the single population model and their corresponding discovery sets (Supplemental Figure 4). Nonetheless, we find that the majority of populations show positive genome-wide predictability (Figure 3A). For populations that show the strongest positive genome-wide concordance, the concordance rate reaches ~65% (range 52-69%) among the ~50 (range 35-150) SNPs with the most stringent joint significance threshold. Intriguingly, we found that four of the 20 populations had negative genome-wide predictability (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 sign and magnitude of predictability scores varies among chromosomes, although the genome-wide score is correlated with the per-chromosome scores, particularly for chromosome arms 2L and 2R (Supplemental Figure 5).

We sought to explain why some populations display negative genome-wide predictability. 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 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. The use of a three-week window is somewhat arbitrary but we do note that this duration corresponds to one to two generations and the general period from egg to peak adult reproduction (Schmidt *et al.* 2005; Bergland *et al.* 2012). Thus, a three week window is a biologically meaningful time window. We did not attempt to optimize models based on the duration of the time-window because we have a relatively limited number of populations pairs. We tested spring and fall thermal limits ranging from 0°C to 40°C. For each combination of spring and fall thermal limits, we regressed genome-wide concordance scores on the number of days above and below the thermal limits and recorded the R^2 of each model. We calculated the null distribution of R^2 values by permuting the genome-wide predictability scores with the number of days above and below each thermal limit (n=10,000). For each permutation, we recorded the maximum R^2 and used the distribution of R^2 values as a null distribution to identify which thermal limits are more strongly correlated with genome-wide predictability scores than expected by chance. 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. 3B; $p < 0.05$).

These best fit models predict that populations experiencing warm temperatures prior to the spring sample or cool temperatures prior to the fall sample will have negative genome-wide predictability scores (Supplemental Figure 6A). Spring and fall fly samples were collected over a relatively broad time window (Figure 1A) and so it is possible that the aberrant populations with negative predictability were collected relatively late in the season and are more representative of fly populations sampled from the summer or winter. However, genome-wide predictability scores are not correlated with spring or fall collection date ($p = 0.6$ and 0.9, respectively) or

cumulative growing degrees (p = 0.5, 0.86), latitude (p = 0.9), or the difference in cumulative growing degree days between spring and fall (p = 0.46). The genome-wide concordance score thus appears to be more related to the specific temperature patterns we identified in the spring and the fall.

We tested the predictive power of our thermal limit model by applying it to seasonal allele frequency changes of the four populations in the ValidationSet (Figure 1B). To generate genome-wide predictability scores for these populations, we assessed concordance rates between seasonally variable SNPs identified in each of the ValidationSet populations with seasonally variable SNPs identified among the Core20 set (see Materials and methods for details). We find that populations in the ValidationSet show a similar variability of genome-wide concordance scores as those estimated in the cross validation analysis of the Core20 set and that several populations also appear to be flipped (Linvilla, Pennsylvania – 2012; Cross Plains, Wisconsin – 2013). Next, we used the parameters from the best fit thermal models derived from the Core20 set to predict genome-wide concordance scores of the ValidationSet populations. We find that these predicted genome-wide concordance values are positively correlated with the observed genome-wide concordance score of the ValidationSet populations (Pearson's r = 0.29; Figure 3D). While this correlation of four points is not significantly different from zero (p = 0.7), we nonetheless find that the magnitude of this correlation is greater than 68.8% of the best-fit permuted models. Taken together with the cross validation 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 previous results in other drosophilids (Rodríguez-Trelles *et al.* 2013) 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.

<u>The flipped model</u>. In the analysis presented above, we have used the time of year that flies were collected to designate season. However, our predictability analysis suggests that the designation of season based on time of year may not completely reflect the recent selective history of the populations that we sampled. We therefore reasoned that our ability to detect seasonally variable polymorphisms, along with our ability to assess general signals of seasonal adaptation, will be improved if we flipped the season label of the populations that show negative genome-wide concordance scores.

To evaluate this idea, we flipped the season label for the four populations with negative genome-wide predictability (Figure 3A) 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. Similarly, there are more SNPs with p<0.001 than 100% of 200 permutations for the flipped model, whereas for the original model there are more SNPs with p<0.001 than 98% of the permutations (Supplemental Figure 7). Of course, the increased number of highly seasonal sites in the flipped model as compared to the original model is a foregone conclusion.

To provide a more independent assessment of whether flipping the seasonal label for these four populations improves signals of seasonal adaptation we re-ran other analyses presented above. First, we examined concordance between the top seasonally varying SNPs identified here, using the flipped model, with seasonal SNPs identified in Bergland et al (2014). We carried out this analysis using 18 of the Core20 populations that do not overlap with Bergland et al (2014; hereafter 'reduced Core20'). Note, the shared Pennsylvania populations were not among those populations with season labels altered in our 'flipped model.' The enrichment for the reduced, flipped Core20 model ($\log_2$OR ± SD = 1.02 ± 0.39; p_{perm} < 6e-4) is substantially better than for the reduced, original model ($\log_2$OR ± SD = 0.59 ± 0.37; p_{perm} = 0.0512).

We examined whether strongly seasonally variable SNPs identified in the flipped model are enriched for any functional elements. Enrichment levels for seasonal SNPs identified with the flipped model are qualitatively similar to those from the original model, described above (Supplemental Figure 3).

We re-ran our predictability analysis with the ValidationSet populations. In this re-analysis, we calculated the genome-wide concordance rate for each ValidationSet population, using the flipped, Core20 model as our discovery set. This reanalysis generally increased the magnitude of genome-wide concordance scores but did not modify their sign (Supplemental Figure 7). Next, we used the parameters from the best fit thermal models derived from the (original) Core20 set to predict our re-estimated genome-wide concordance scores (see the Materials and methods for a detailed justification of this approach). We find improved predictions of the genome-wide concordance scores made through the use of the flipped version of the Core20 analysis ('flipped model' Pearson's r = 0.53 'original model': Pearson's r = 0.29). For analyses using the 'flipped model', the correlation coefficient between estimated and observed genome-wide predictability scores are better than 84.5% of the best-fit permuted models (64-95% across the range of best fit thermal limit models). In contrast, the observed, best-fit correlation coefficients for the 'original model' are only better than 68.8% of the best-fit permuted models (range 54.8 - 89%).

Taken together, we conclude that there is some 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.

Spatial differentiation parallels signatures of seasonal adaptation. Next, we sought to assess whether seasonal adaptation reflects spatial patterns of genetic variation across the east coast of North America. Accordingly, we tested for enrichment and parallelism between SNPs that fluctuate seasonally and those that vary in frequency across the East Coast latitudinal cline. To identify clinally varying polymorphisms we fit a generalized linear model, regressing allele frequency on latitude using the clinal set of populations (described above). 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.

Our initial analysis of seasonal and clinal variation yielded two basic results. First, we find that there is an enrichment of SNPs that are both seasonal and clinal, regardless of parallelism, across a range of significance thresholds relative to matched controls for both the flipped and original models (Supplemental Figure 8). The enrichment increases with increasingly stringent seasonal and clinal significance thresholds (Supplemental Figure 8). Rates of enrichment among the most significantly seasonal and clinal sites are higher in the 'original' model than in the 'flipped' model, however at moderate significance levels enrichment rates are higher for the flipped model. Although there is a significant enrichment of seasonal and clinal polymorphisms under either model, we point out that the number of clinal polymorphisms that are seasonal is numerically small and only represents a modest fraction of clinally varying polymorphisms. For examples, under the 'original' model, for the top 1% of clinal SNPs (7739 SNPs), 0.015% (115) are in the top 1% of seasonal sites (expected 0.010%). This represents a 37% enrichment over controls and 50% over the theoretical expectation of 0.01. Thus, the majority of clinal polymorphisms are not seasonal, and vice-a-versa. The mild enrichment that we observe here is consistent with the non-significant, but slightly positive, enrichment reported by Bergland et al (2014).

Second, we tested whether SNPs that are clinal and seasonal below a range of significance thresholds change in allele frequency in a parallel fashion. We define parallel changes in allele frequency across time and space to correspond to the situation when the high frequency allele in the 'spring' is the same as the high frequency allele in the northern US as well as high frequency 'fall' alleles being found predominantly in the southern US. We find that the rate of parallelism increases with increasingly stringent seasonal and clinal significance thresholds (Figure 4A). Parallel changes in allele frequency between seasons and clines is similar to previously published genome-wide and locus specific results (Cogni *et al.* 2014; Bergland *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'. The increase in parallelism rates using the flipped model can be taken as strong, and orthogonal, evidence that the flipped model generates a higher confidence set of seasonal SNPs.

Parallel changes in phenotype and genotype in temperate environments could be influenced by seasonal migration of southern flies into northern locales through the growing season. Seasonal migration would not necessarily invalidate our conclusion of seasonal adaptation because one would still need to invoke a model of selection on a resident, overwintering population for allele frequencies to oscillate from year to year. Nonetheless, we sought to test whether seasonal migration strongly influences the signal of seasonal adaptation by identifying seasonal SNPs using populations that are geographically and genetically isolated from the East Coast (Campo *et al.* 2013; Bergland *et al.* 2015). We assessed seasonal changes in allele frequency separately for Californian (n=3) and European (n=3) populations using the GLM approach, employing both the 'original' and 'flipped' seasonal labels, and tested for parallelism with

polymorphisms that vary along the East Coast of North America. We found a strong signal of parallelism between seasonal and latitudinal variation in both the California and the Europe comparisons (Figure 4B, 4D). For the Californian populations, parallelism was greater for the 'flipped model' than for the 'original model' (Figure 4D). These results suggest that parallel changes in allele frequency between seasons and across the latitudinal cline is not exclusively driven by seasonal migration in the spring from neighboring southern populations. Thus, the identified seasonal SNPs are not merely those that mark local clines along the east coast of North America

The strength and genomic extent of seasonal adaptation. Our results suggest that seasonal adaptation is driven, at least in part, by predictable changes in allele frequency at polymorphisms spread throughout the genome. Accordingly, we sought to quantify the fraction of polymorphisms whose frequencies are affected, both directly and indirectly, by seasonal adaptation and to estimate the average strength of seasonally varying selection.

To robustly estimate the strength and extent of seasonally varying selection, we developed a semi-parametric test statistic based on the per-population change in allele frequency among seasons. In this method, we first calculated the per-population change in allele frequency for each SNP using two one-tailed Fisher's exact tests. Specifically, first we test whether the SNP increases in frequency from spring to fall and then we test whether it decreases in frequency from spring to fall. Next, we rank normalized both one-tailed p-values for each SNP relative to other SNPs with similar sequencing depth and allele frequency. Finally, we combined the one-tailed, rank-normalized p-values for each SNP using Fisher's Method (Fisher 1925), producing two χ^2 distributions representing the magnitude and concordance of allele frequency change among populations. Our method is limited in power by the number of SNPs used for rank normalization (median 277 per bin). However, our method is robust to an inflation in sampling precision, in contrast to the GLM method (Supplemental Figure 2). We performed these analyses using the seasonal labels from the 'original' and 'flipped' models. The genome-wide distribution of χ^2 statistics was also calculated for 5000 permuted datasets wherein spring and fall labels were randomly flipped within populations. Consistent with our analysis of seasonal changes in allele frequency using generalized linear models, we find an excess of polymorphisms with elevated signals of consistent seasonal allele frequency change relative to the null expectation of no allele frequency change using nominal *p*-values as well as relative to the permuted datasets (Figure 5AB). The excess of seasonally varying polymorphisms relative to permutations was greater for the 'flipped model' ($p < 0.002$) than for the 'original model' ($p = 0.018$), as expected.

We estimated the fraction of the genome affected by seasonal adaptation and the average strength of seasonal selection. To estimate these parameters, we compared the observed distributions of χ^2 statistics 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 detail). Using an ABC-style approach, we identified which parameter estimates were most likely to produce the

observed genome-wide distribution of χ^2 statistics. The model we employ for the ABC analysis assumes one generation of selection per season and assumes that every site in the genome changes in frequency independently and is subject to the same 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 just the proportion of causal sites. We find a ridge of best fit parameters, with a trade-off between the strength of selection and the proportion of sites under selection (Figure 5CD). Our simulations suggest that the frequency of at least 2.5% of common polymorphisms in the fly genome are affected by seasonal adaptation of the strength of ~20% per season under the flipped model, or 1% with strength of selection of 10% per season for the original model. The estimated strength of selection and fraction of the genome affected by seasonally varying selection are greater for the 'flipped model' than the 'original model,' as expected (Figure 5CD). 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 9). We found no substantial shift in our parameter estimates across subsampled datasets (mean of 100 samples).

The distribution of seasonally variable SNPs throughout the genome. Analyses presented above suggest that seasonal adaptation affects patterns of allele frequency genome-wide. Accordingly, we characterized several aspects of the genomic distribution 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, typically segregating at less than 15% frequency in most populations (Supplemental Figure 10). Using a simple sign test, we find that no inversion consistently changed in frequency using season labels as defined in either the 'original or the 'flipped' model. We note that *In(2R)Ns* is 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), reflecting previously reported signals of seasonal change (Kapun *et al.* 2016). However, *In(2R)Ns* segregates at a much lower frequency than many of the most strongly seasonal SNPs, and seasonal changes at it are relatively mild compared to other seasonally varying polymorphisms. 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, we examined the abundance of the top 1% of seasonally variable SNPs as defined by the GLM model (n=7,748) among bins of 100 consecutive SNPs (average bin size of 12Kb). We compared the observed distribution of seasonal SNPs per bin to the nominal expectation based on binomial sampling, as well as to the empirical distribution of matched control SNPs (Figure 6A). We found

significant over-dispersion in the distribution of seasonal SNPs per bin compared to both the theoretical and the matched control distributions (KS test $p<10^{-16}$), indicating that seasonal SNPs are more clustered than expected. We cannot determine to what extent this clustering is due to linkage of the causal SNPs to neighboring SNPs or to the clustering of causal seasonal variants. The distribution of control polymorphisms was indistinguishable from the binomial expectation (KS 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 seasonally variable SNPs, 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 among all autosomes but are enriched on chromosome 2L (Figure 6B). Across the window sizes tested, the probability of observing at least one highly seasonal SNP in any given window is less than expected by chance given a Poisson distribution, consistent with clustering on both large and small genomic scales (Figure 6B). Nonetheless, the observed clustering is modest considering that there is a >75% chance of identifying a seasonally variable SNP in any randomly selected 50Kb window (Figure 6B). The general signal of overdispersion is similar for both the 'flipped' and 'original' models (Figure 6A,B) and consistent with similar patterns previously observed in Bergland *et al* (2014). The broad genomic distribution and large number of seasonally variable SNPs suggests that seasonal adaptation is highly polygenic.

We performed a sliding-window analysis to identify regions of the genome with an excess of seasonally varying polymorphisms. We found approximately 40 regions of the genome with a significant excess of seasonally varying polymorphisms. The locations of these regions change substantially between the 'original' and 'flipped' models (Figure 6C). For simplicity, we focus the remainder of our analysis here on the regions identified using the 'flipped' model. Regions of the genome enriched for seasonally variable SNPs are enriched on both arms of chromosome 2 and deficient along chromosome 3 (χ^2 = 22.9, p = 4e-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 variable gene (Cogni *et al.* 2014) found to be associated with diapause (Schmidt *et al.* 2008), has 7 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 lack power to discover many many causal genes involved in seasonal adaptation. There was no excess of seasonal SNPs surrounding the insulin receptor (*InR*), wherein seasonally variable indels and SNPs had been previously reported(Paaby *et al.* 2014).

Gene-ontology analysis (Huang *et al.* 2009a; b) of the ~330 genes within the 40 regions of significant excess of seasonal polymorphisms did not identify any significant enrichment for different biological processes, cellular components, or molecular function following multiple testing correction. However, we find that genes associated with several molecular function categories including inositol metabolism and phosphorylation are marginally enriched following multiple testing correction ($p_{adjusted} < 0.08$). 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 (Pavlidis *et al.* 2012) should be taken when interpreting this result as it is driven by a set of three, closely linked genes (CG17026, CG17028, CG17029).

Discussion

Herein, we have studied the genomic signatures of seasonal adaptation in the fruit fly, *D. melanogaster*. Our work focused on addressing the basic questions of whether seasonal adaptation is general, predictable, and pervasive. Rigorously addressing these questions requires accurate allele frequency estimates (Zhu *et al.* 2012) genome-wide from paired spring-fall samples collected across a large portion of *D. melanogaster*'s range. Generating such samples can be prohibitively expensive and time-consuming for a single lab and therefore we formed a consortium, DrosRTEC, to facilitate seasonal sampling across broad spatial scales (Figure 1). Our consortium's work is mirrored by similar efforts to assess the patterns of spatial differentiation throughout Europe by the DrosEU consortium (Kapun *et al.* 2018).

Our survey of allele frequency changes between spring and fall among over 20 populations demonstrates that seasonal adaptation, defined here as a form of local adaptation, is a general feature of *D. melanogaster* living in temperate locales. Previous analyses of seasonal changes in allele frequency in *D. melanogaster* have focused on a single mid-latitude population (Linvilla, Pennsylvania; (Cogni *et al.* 2014; Bergland *et al.* 2014)). The generality of that work is necessarily limited because it is restricted to a single locale. By demonstrating consistent seasonal changes in allele frequencies among multiple populations (Figure 2A), we argue here suggests that seasonal adaptation is a general phenomenon of temperate *D. melanogaster* populations, and not restricted to a single orchard, biogeographic region, or continent. Consequently, we suggest that some aspects of natural selection vary between seasons consistently across a large portion of *D. melanogaster*'s range and there is some common genetic basis for seasonal adaptation.

Although we identified consistent patterns of evolution between spring to fall, our work highlights the intrinsic difficulty of the very definition of spring and fall. In general, we define spring as the time of year close to when *D. melanogaster* first becomes locally abundant. Fall is the time that is close to the end of the growing season, prior to the seasonal collapse, but not extirpation (Ives 1970; Shpak *et al.* 2010; Bergland *et al.* 2014; Machado *et al.* 2016), of *D. melanogaster* populations. The fact that we detect abundant parallel changes in allele frequency among seasons using this sampling scheme (Figure 2A, Figure 3A) suggests that for a large number of populations, there is a biological meaning to spring and fall defined this way. While we find clear

evidence that many samples show consistent changes in allele frequency between spring and fall (Figure 3AC), other populations show strong signals of reversed allele frequency change. For these populations, the data looks as if the spring and fall labels are switched (Figure 3AC). These 'flipped' populations are geographically dispersed (California, Kansas, Michigan, and Massachussetts) with samples collected over two non-consecutive years (2012, 2014). We do not believe it is likely that the labels were flipped as a result of sample handling. Rather, we provide evidence consistent with the model that idiosyncratic aspects of the environment prior to sampling drove opposing allele frequency changes in these populations. Notably, an ecological model that we built using information about local temperature extremes allowed us to predict the sign and magnitude of allele frequency changes genome-wide with relatively high accuracy (Figure 3D, Supplemental Figure 7B) for additional populations, which would not be likely had the labels were in fact switched during sample handling.

There are several implications of our predictability analysis. The first, and most general, is that relatively fine-scaled variation in selection pressures, on the order of generation time, can promote rapid adaptive evolution and drive observable fluctuations in allele frequency. Finely scaled (sensu (Levene)) variation in selection pressures is generally thought to lead to the adaptive evolution of phenotypic plasticity (Via & Lande 1985). Flies clearly exhibit extensive plastic responses to a variety of environmental factors (David *et al.* 1997; Bergland *et al.* 2008), many of which vary across seasonal time scales, and these plastic responses could conceivably account for population persistence and growth across seasons. Indeed, behavioral and physiological changes (phenotypic plasticity) account for population persistence overwinter in a variety of other organisms. Why, then, is seasonal adaptation, defined here as allele frequency change over seasonal time-scales, such a prominent feature of temperate fly populations? One possibility is that many aspects of plasticity are relatively costly (Schou *et al.* 2015), or that the environment is fundamentally unpredictable, thus promoting the existence of alternate genotypes that are favored in one season and not another.

We argue that our results are consistent with polygenic response to rapid changes in selection pressure. Our simulations show that the overall, genome-wide, signal is consistent with seasonal adaptation affecting allele frequencies at ~2.5% of common polymorphisms genome-wide and changing allele frequencies by upwards of 10% between spring and fall. Note that the strength of selection reported here is lower than in Bergland et al (2014) but we are also looking specifically at parallel changes across a much larger portion of the species range. To the extent that some SNPs shift only in some populations this would tend to lower the estimate of strength of consistent selection.

The analysis of the top seasonal SNPs reveals interesting patterns. As before (Bergland *et al.* 2014) seasonally variable SNPs are spread throughout the genome, with some overdispersion. We see little enrichment in terms of classic annotations for nucleotide function (Supplemental Figure 3), suggesting perhaps that many of the seasonal SNPs that we identify are regulatory in nature and/or only tightly linked to the truly causal genetic variants. In addition, we provide evidence that some of top seasonally variable SNPs also vary across a latitudinal cline, often in

a parallel fashion (Figure 4), however note that most clinal SNPs do not show seasonal fluctuations. The mirroring of seasonal and clinal variation provides an independent confirmation that our analysis likely picked a high proportion of causal seasonal SNPs. These SNPs will provide a rich substrate for functional analysis.

Our work demonstrates that seasonal adaptation is a pervasive process taking place across a large fraction of *D. melanogaster*'s range. The results are tantalizing but full understanding of the system will require additional sampling across seasons as well as from close and broadly distributed locales across the whole geographic range. Broad scale sampling will enable us to identify other aspects of the environment that elicit rapid adaptive evolution and further quantify the fraction of polymorphisms in the genome that underly rapid adaptation. Broad sampling can be coupled with additional work using controlled population cages that expose the flies to the seasonal environments while minimizing migration, and functional work to try to understand the molecular effects of individual SNPs and variants.

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 flies using direct aspiration from substrate, netting, or trapping with banana and yeast bait. 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 and Illumina TrueSeq adapters. Paired-end 125bp libraries were sequenced to an average of 94x coverage either at the Stanford Sequencing Service Center on a 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). Due to the phenotypic similarity of the species *D. melanogaster* and *D. simulans*, we tested for *D. simulans* contamination by

competitively mapping to both genomes. Reads that mapped better to the *D. simulans* were removed from the analysis. Any samples with greater than 5% of reads mapping preferentially to *D. simulans* was omitted from the dataset (see Supplemental Table 1). The average proportion of reads mapping preferentially to *D. simulans* was less than 1%. 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, resulting in 774,651 SNPs that represent the core set used in our analyses.

To asses 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; Linvilla, 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; Odessa, Ukraine 2013. The five sets of paired spring/fall samples that were not included in the Core20 set were used for cross-validation ('ValidationSet': Linvilla, Pennsylvania 2009, 2012, 2014, 2015, 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).

<u>*Seasonal sites*</u>. We identified seasonal sites using two separate methods, a general linear regression model (GLM) and a rank *p*-value Fisher's method. All statistical analyses were performed in *R* v3.1.0 (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 ith SNP, and e_i is the binomial error at the ith SNP. Although the GLM is a powerful test, the binomial error assumption is likely an underestimate of the pool-seq error (Machado *et al.* 2016). 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.

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 looked at the top 1% most seasonal SNPs, as assessed by GLM *p*-value (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 underdispersed genomic distribution of seasonal SNPs (more evenly distributed than expected) whereas a distribution less strongly peaked around 1 is overdispersed and indicates clustering of seasonal SNPs. We assessed the similarity of two distributions using a Kolmogorov-Smirnov (KS) test, where a low p-value indicates significantly different distributions. 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.

Matched controls. With the assumption that the majority of the genome is not seasonal, we can use matched genomic controls as neutral references. 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).

Predictability analysis. To test the general predictability of seasonal change in our dataset, we performed a leave-one-out cross validation 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) concordance 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,

$$concordance = quantile + e,$$

where 'concordance' was the fraction of SNPs falling below the quantile threshold 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) concordance score is heavily influenced by concordance rates of higher quantiles because that is where the bulk of SNPs reside.

To test whether heterogeneity in genome-wide concordance 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:INOJh_Os}. We matched each locality to a weather station based on geographic distance. Three of the collections did not have precise collection dates for one or both seasonal collections (Linvilla, Pennsylvania 2010, 2011), or were not associated with any climate data from the GHCND database (Odessa, Ukraine). These populations were removed from our weather model analysis.

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 along 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). 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. This type of regression is particularly appropriate for the analysis of clinal variation of allele frequencies, as it takes into account precision (number of chromosomes sampled per population) and the curve-linear behavior at low allele frequencies.

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 matched controls.

The strength and genomic extent of seasonal adaptation. To test for an enrichment of seasonally varying sites above neutrality and to estimate the proportion and selection coefficient of seasonally varying sites, we designed a "rank p-value Fisher's method" test. In this test, 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 proportional 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. We then combine ranked p-values using Fisher's method for each SNP, taking two times the negative log sum of the p-values across the 20 spring/fall comparisons (each tail tested separately). The null distribution for this statistic is a Chi-squared distribution with degrees of freedom equal to 40 (two times the number of comparisons).

We estimated the proportion of sites under selection and the selection strength 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 neutral dataset the same size as the observed dataset, with the same allele frequency distributions as the observed data. This neutral dataset will not necessarily reflect the same sampling error as observed in nature from spring to fall, and as this is not known it 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 no way for there to be an artificially inflated seasonal signal due to incorrectly estimated sampling error (as there can be for a regression p-value).

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 seasonal selection allele frequency:

$$AF_{fall} = AF_{spring} + AF_{spring} * (1-AF_{spring}) * s_{sf}$$

where *AF* is the allele frequency and s_{sf} is the cumulative "seasonal selection coefficient" from the spring to the fall sample. We use this cumulative s_{sf} since the true number of generations between the spring and fall samples is not known. We simulated datasets with seasonal selection coefficients ranging from 0.01 to 0.5, and proportions of sites under selection ranging from 0.001 to 0.1.

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. We performed a principal components analysis of allele frequency per SNP per sample using the *prcomp* function in *R* (frequency data scaled by SNP).

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).

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 all members of the Schmidt and Petrov labs who provided exceptionally valuable feedback. This work was funded by the NIH grants RO1GM1000366 to PS and DAP, R35GM118165 to DAP, and RO1GMXXXX to PS.

References

Behrman EL, Howick VM, Kapun M *et al.* (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. (SV Nuzhdin, Ed,). *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 (2015) Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in Drosophila melanogaster. *Molecular Ecology*, n/a–n/a.

Bertram J, Masel J (2017) A central role for boom-bust cycles in maintaining genetic variation in fluctuating environments. *bioRxiv*, 164723.
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.
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.
Cogni R, Kuczynski C, Koury S *et al.* (2014) The intensity of selection acting on the *couch potato* gene -- spatial–temporal variation in a diapause cline. *Evolution*, **68**, 538–548.
David JR, Gibert P, Gravot E *et al.* (1997) Phenotypic plasticity and developmental temperature in Drosophila: Analysis and significance of reaction norms of morphometrical traits. *Journal of Thermal Biology*, **22**, 441–451.
Denlinger DL (2003) Regulation of diapause. *Annual Review of Entomology*, **47**, 93–122.
Dobzhansky T (1948) Genetics of natural populations; altitudinal and seasonal changes produced by natural selection in certain populations of Drosophila persimilis. *Genetics*, **33**, 158–176.
Ellner S (1996) Environmental fluctuations and the maintenance of genetic diversity in age or stage-structured populations. *Bulletin of Mathematical Biology*, **58**, 103–127.
Ellner S, Nelson G Hairston J (2015) Role of Overlapping Generations in Maintaining Genetic Variation in a Fluctuating Environment. *The American Naturalist*, **143**, 403–417.
Ellner S, Sasaki A (1996) Patterns of genetic polymorphism maintained by fluctuating selection with overlapping generations. *Theoretical Population Biology*, **50**, 31–65.
Ewing EP (1979) Genetic variation in a heterogeneous environment VII. Temporal and spatial heterogeneity in infinite populations. *American Naturalist*, **114**, 197–212.
Feder AF, Petrov DA, Bergland AO (2012) LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data (R Wu, Ed,). *PloS one*, **7**, e48588.
Fisher RA (1925) *Statistical methods for research workers*.
Gillespie J (1973) Polymorphism in random environments. *Theoretical Population Biology*, **4**, 193–195.
Hedrick PW (1976) Genetic variation in a heterogeneous environment. II. Temporal heterogeneity and directional selection. *Genetics*, **84**, 145–157.
Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: The age of genomics. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 67–93.
Huang DW, Sherman BT, Lempicki RA (2009a) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, **4**, 44–57.
Huang DW, Sherman BT, Lempicki RA (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research*, **37**, 1–13.
Ives PT (1970) Further genetic studies of the south Amherst population of *Drosophila melanogaster*. *Evolution*, **24**, 507–518.
Kapun M, Aduriz MGB, Staubach F *et al.* (2018) Genomic analysis of European Drosophila melanogaster populations on a dense spatial scale reveals longitudinal population structure and continent-wide selection. *bioRxiv*, 313759.
Kapun M, Fabian DK, Goudet J, Flatt T (2016) Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. *Molecular Biology and Evolution*, **33**, 1317–1336.
Kapun M, van Schalkwyk H, McAllister B, Flatt T, Schlötterer C (2014) Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of Drosophila melanogaster. *Molecular Ecology*, **23**, 1813–1827.
Koboldt DC, Zhang Q, Larson DE *et al.* (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, **22**, 568–576.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ (2011) Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*, **187**, 245–260.
Kostál V (2006) Eco-physiological phases of insect diapause. *Journal of Insect Physiology*, **52**, 113–127.
Levene H Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, **87**, 331–333.
Levins R (1968) *Evolution in changing environments: some theoretical explorations*.
Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
Machado HE, Bergland AO, O'Brien KR *et al.* (2016) Comparative population genomics of latitudinal variation in Drosophila simulans and Drosophila melanogaster. *Molecular Ecology*, **25**, 723–740.
Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, pp. 10–12.
McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.
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.
Rajpurohit S, Hanus R, Vrkoslav V *et al.* (2017) Adaptive dynamics of cuticular hydrocarbons in Drosophila. *Journal of Evolutionary Biology*, **30**, 66–80.
Rodríguez-Trelles F, Tarrío R, Santos M (2013) Genome-wide evolutionary response to a heat wave in.
Schmidt PS, Conde DR (2009) Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in *Drosophila melanogaster*. *dx.doi.org*, **60**, 1602.
Schmidt PS, Matzkin L, Ippolito M, Eanes WF (2005) Geographic variation in diapause incidence, life-history traits, and climatic adaptation in *Drosophila melanogaster*. *Evolution*, **59**, 1721–1732.
Schmidt PS, Zhu C-T, Das J *et al.* (2008) An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 16207–16211.
Schou MF, Loeschcke V, Kristensen TN (2015) Strong Costs and Benefits of Winter Acclimatization in Drosophila melanogaster. (M Sears, Ed,). *PloS one*, **10**, e0130307.
Shpak M, Wakeley J, Garrigan D, Lewontin RC (2010) A Structured Coalescent Process for Seasonally Fluctuating Populations. *Evolution*, **64**, 1395–1409.
Tauber MJ, Tauber CA, Masaki S (1986) *Seasonal Adaptations of Insects*. Oxford University Press on Demand.
Vesala L, Salminen TS, Kostál V, Zahradnickova H, Hoikkala A (2012) Myo-inositol as a main metabolite in overwintering flies: seasonal metabolomic profiles and cold stress tolerance in a northern drosophilid fly. *Journal of Experimental Biology*, **215**, 2891–2897.
Via S, Lande R (1985) Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, **39**, 505–522.
Wittmann MJ, Bergland AO, Feldman MW, Schmidt PS, Petrov DA (2017) Seasonally fluctuating selection can maintain polymorphism at many loci via segregation lift.

Proceedings of the National Academy of Sciences of the United States of America, **114**, E9932–E9941.
Xu S (2003) Theoretical Basis of the Beavis Effect. *Genetics*, **165**, 2259–2268.
Zhu Y, Bergland AO, González J, Petrov DA (2012) Empirical Validation of Pooled Whole Genome Population Re-Sequencing in Drosophila melanogaster (J-S Seo, Ed,). *PloS one*, **7**.

Figures

Figure 1:

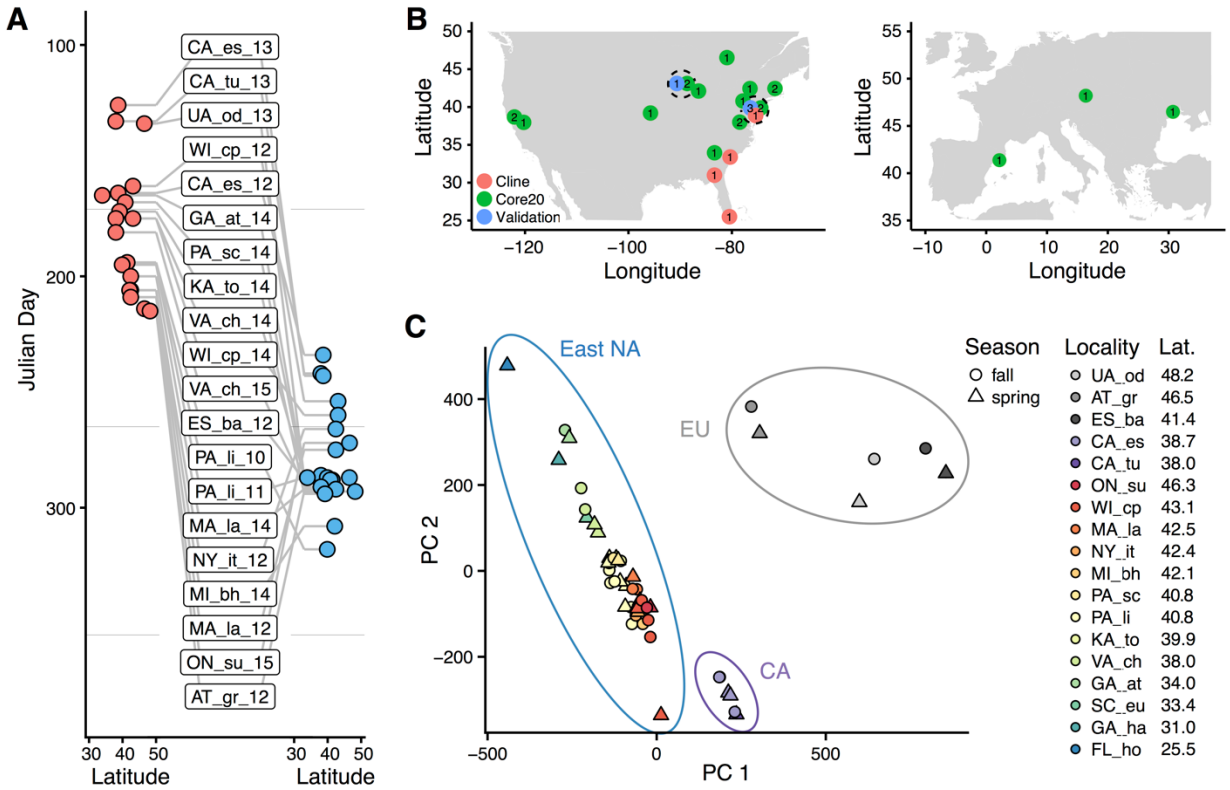


Figure 1. A) Distribution of collection times during the spring (red) and fall (blue) in relation to latitude. B) Sampling localities for the primary seasonal analysis ('Core20': green), the cross-validation analysis ('ValidationSet': blue) and the latitudinal cline analysis ('Cline': red), distributed 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 per sample. Circles are placed around samples from three geographic regions: Europe (EU), California (CA) and Eastern North America (East NA).

Figure 2

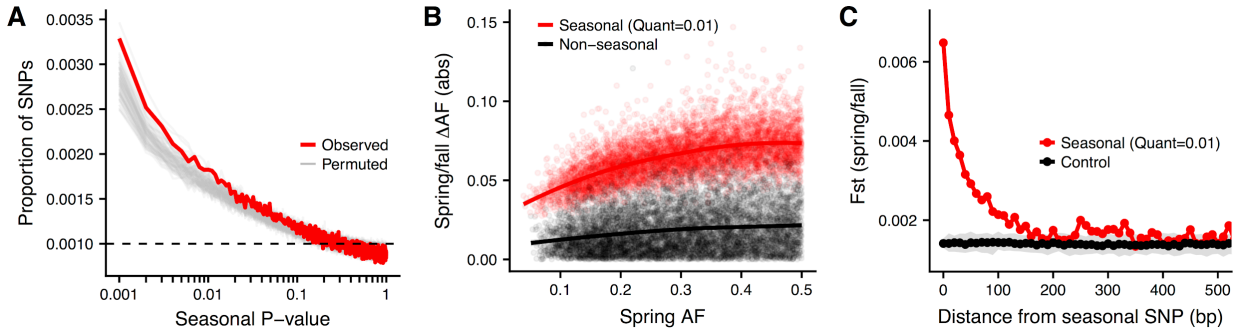


Figure 2. Genomic characteristic of seasonal variants. A) P-value distribution of per-SNP seasonal regressions (red line) and permutations (grey lines). P-value bins are of size 0.001, with a corresponding expected uniform distribution at 0.001 (dashed line). B) Average spring/fall allele frequency change for each of the top 1% of seasonally varying SNPs (red) and non-seasonal, matched control SNPs (black), as a function of spring allele frequency. Lines represent a moving average across SNPs with a given spring allele frequency. C) 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.

Figure 3

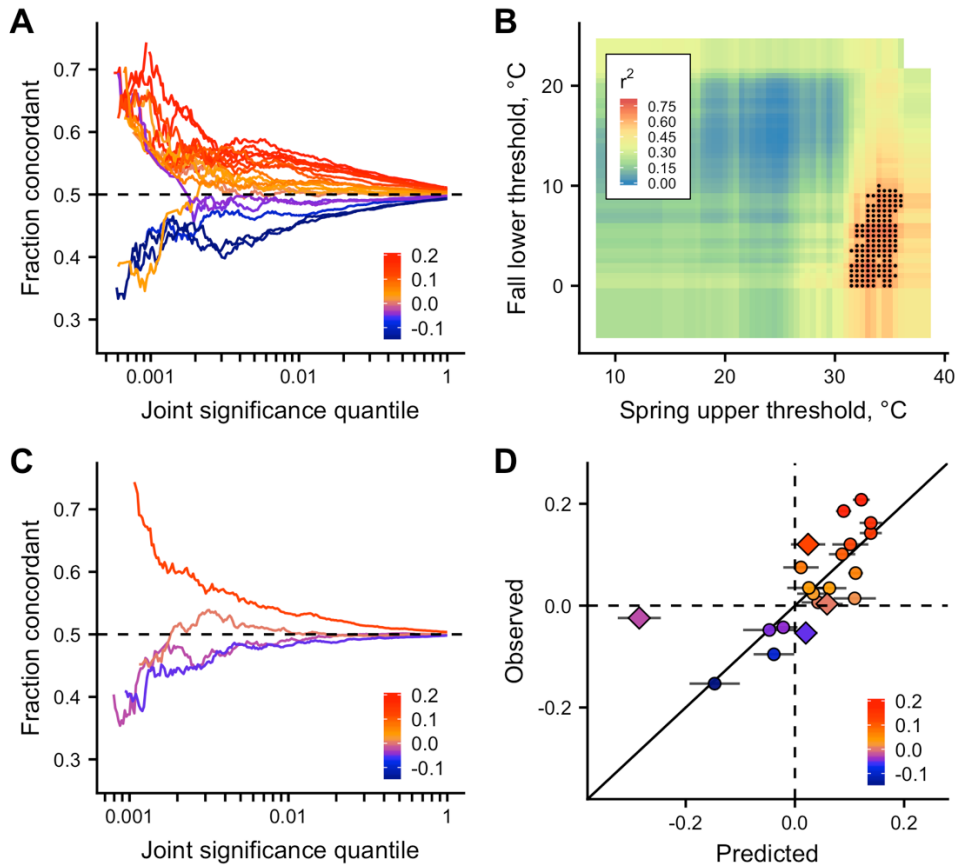


Figure 3. Predictability of seasonal adaptation and the thermal limit model. A) The fraction of concordant allele frequency changes from the leave-one-out cross validation analysis of the Core20 populations. Each line represents one population, colored by the inferred genome-wide predictability score. B) 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. C) Genome-wide predictability scores from the ValidationSet of populations in relation to the Core20 model. D) The best thermal limit models from the Core20 analysis predict genome-wide predictability scores of the ValidationSet (diamonds).

Figure 4

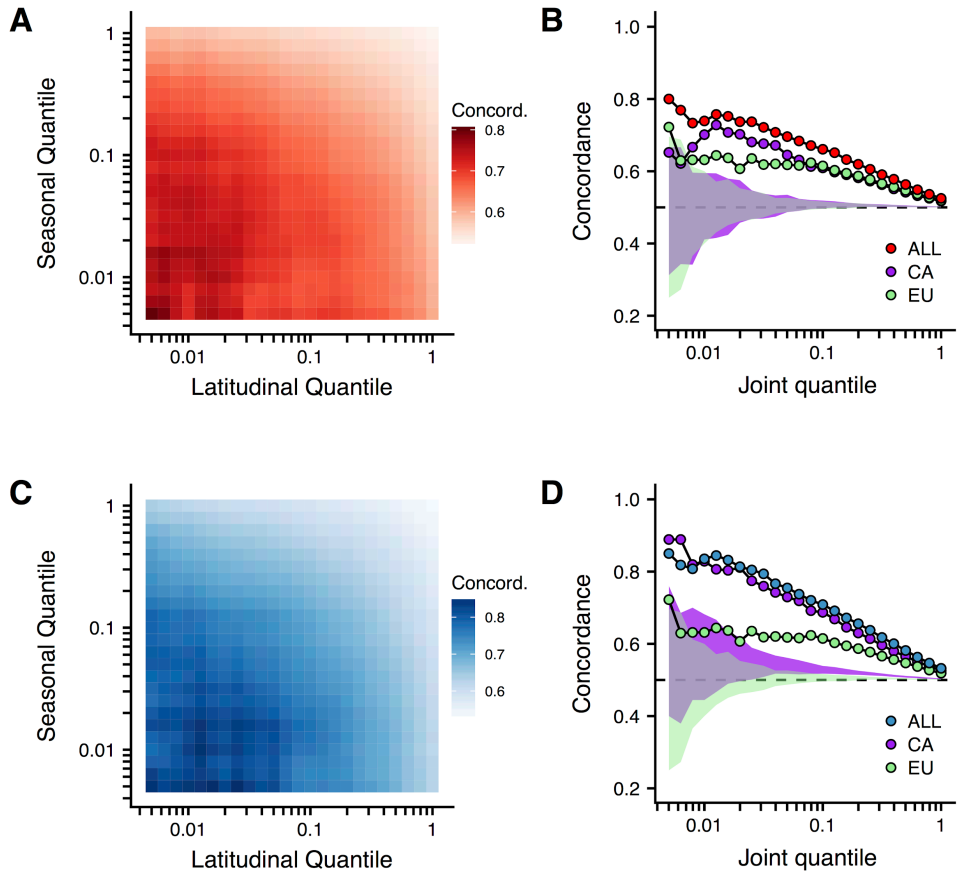


Figure 4. Seasonal and latitudinal allele frequency concordance (spring is matched to north). A,C) Percent enrichment in the number of SNPs significant in both the latitudinal and seasonal regressions (compared to the median of 100 matched controls), for the original (A) and flipped (C) datasets. All non-white hues represent a significant enrichment compared to matched controls. B,D) Seasonal/latitudinal concordance for all 18 samples (red, blue), 3 California samples (CA: purple), and the 3 Europe samples (EU: green), for the original (B) and flipped (D) datasets. Purple and green areas are the 96% confidence intervals of 100 matched control datasets for CA and EU, respectively. Latitudinal regression: spring FL_ho, SC_eu, GA_ha, PA_li. Seasonal regression: all 18 paired samples from populations not used in the latitudinal regression (omitted PA_li 2010 and 2011).

Figure 5

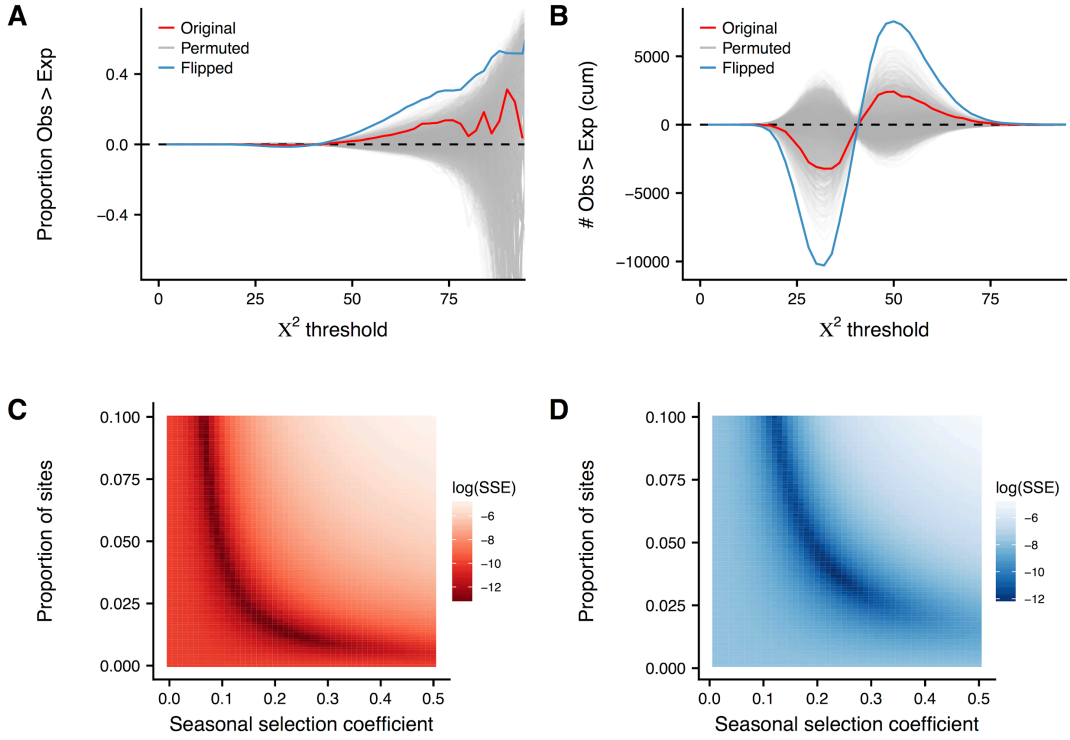


Figure 5. The rank P-value fisher's method enrichment of seasonal SNPs. A) The proportion of sites and (C) the cumulative distribution of sites observed minus expected across X^2 test statistic values. Plotted is the observed distribution (red), the distribution for the 'flipped' dataset (blue), permutations (grey), and the neutral expectation (dashed). B,D) Parameter estimates (selection coefficient and proportion of SNPs under selection) from simulated datasets for B) the original analysis and D) the 'flipped' analysis.

Figure 6

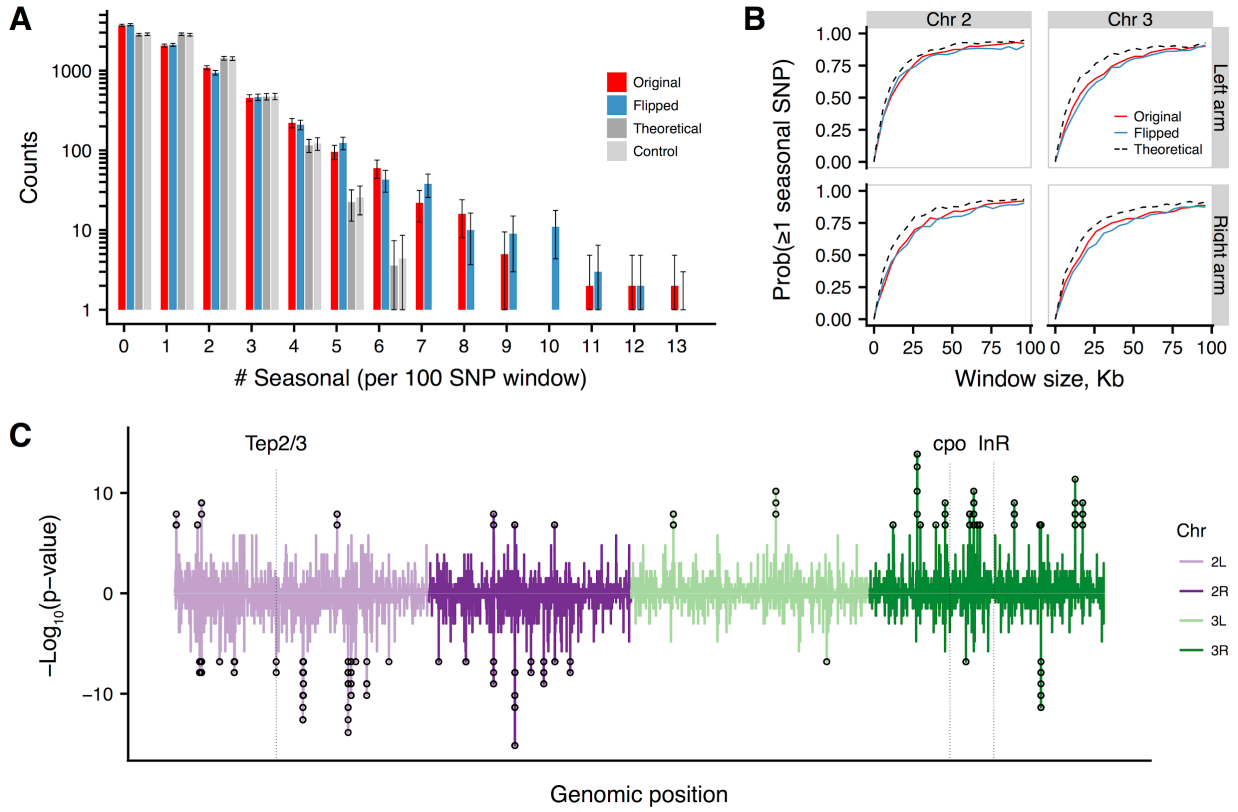
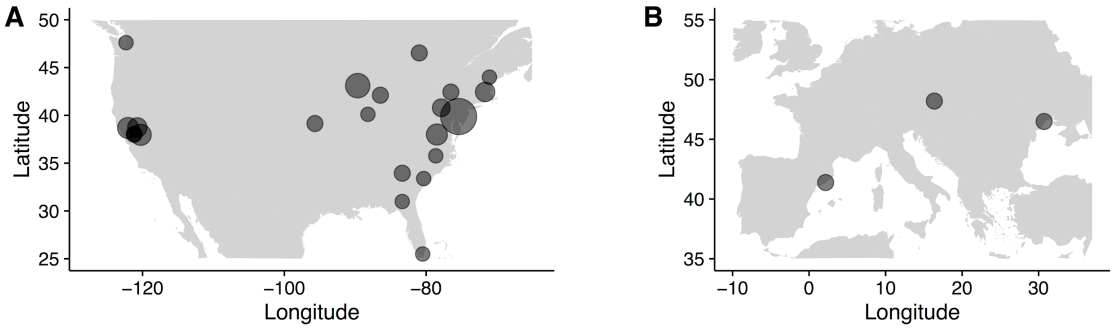
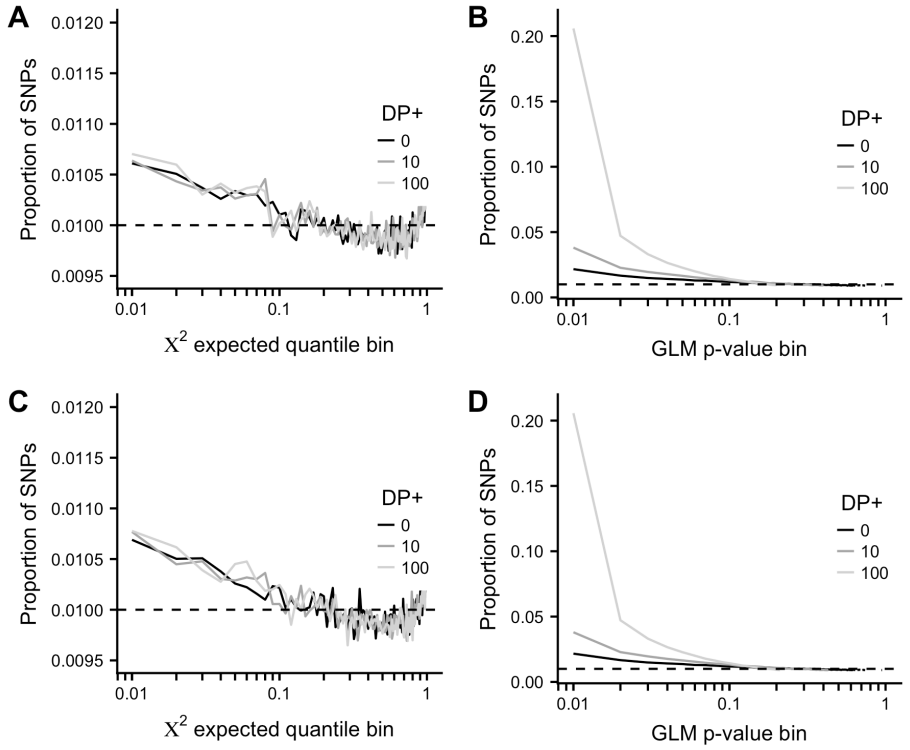


Figure 6. Genomic distribution of seasonal SNPs. (A) Genomic over-dispersion of seasonal SNPs, plotted as the distribution of the number of seasonal sites (quantile: 0.01) per 100 SNP window. The observed distribution (red) is over-disbursed compared to both the theoretical distribution (dark grey) and that of matched controls (light grey). Error bars are 2 standard deviations. B) Genomic over-dispersion of seasonal SNPs, plotted as the proportion of windows with at least one seasonal SNP, with increasing window size (bp). The observed proportion (solid line) and expected theoretical proportion (dashed line) is plotted for each chromosome arm. C) sliding window analysis of regions with high densities of seasonal SNPs.

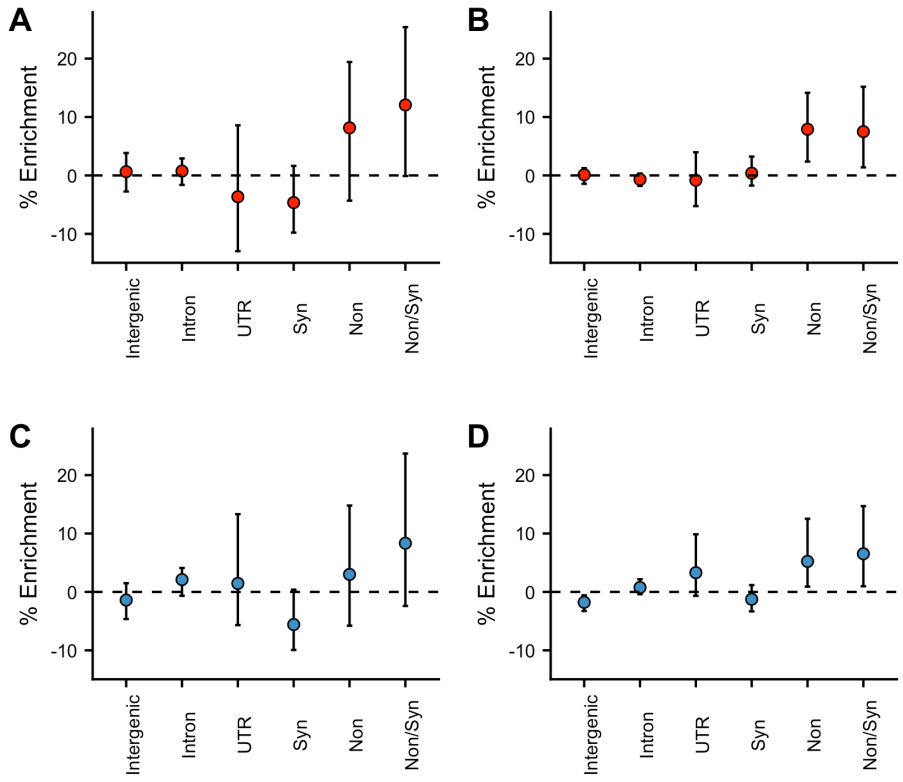
Supplemental Figures



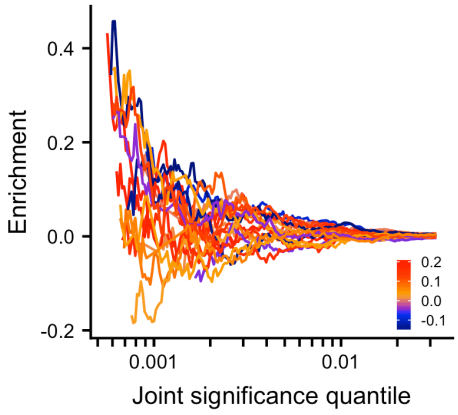
Supplemental Figure 1. Sampling localities for all samples collected, regardless of use in the present analysis. Sizes of circles reflect number of samples per locality.



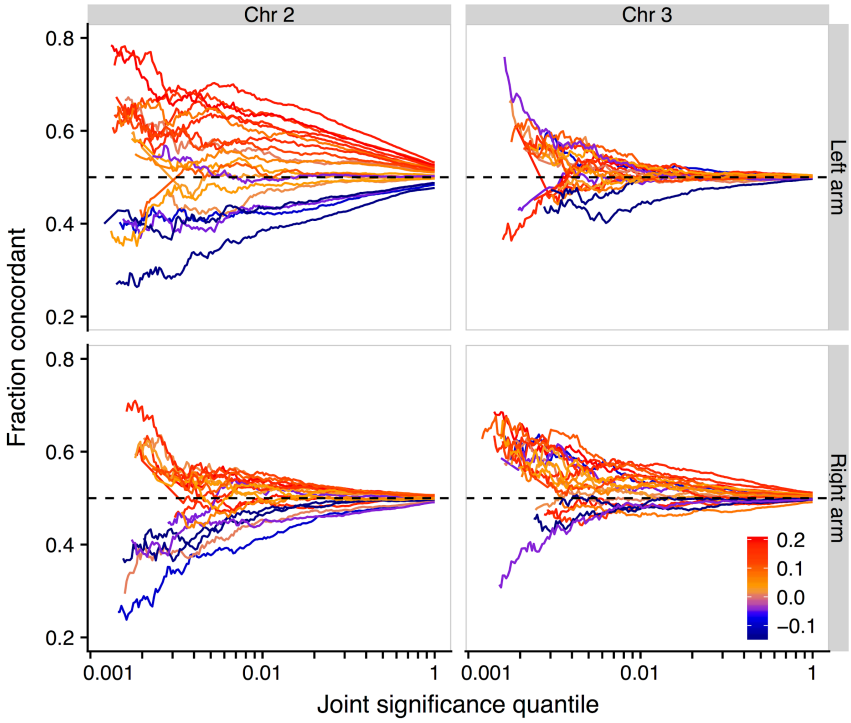
Supplemental Figure 2. The effect of artificial increases of sample size on the CoreSet seasonal analysis using A,C) the rank Fisher's method and B,D) the GLM analyses. Top row: sample size inflated by a fixed amount. Bottom row: sample size inflated by a Poisson sampled amount. Shown is the proportion 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).



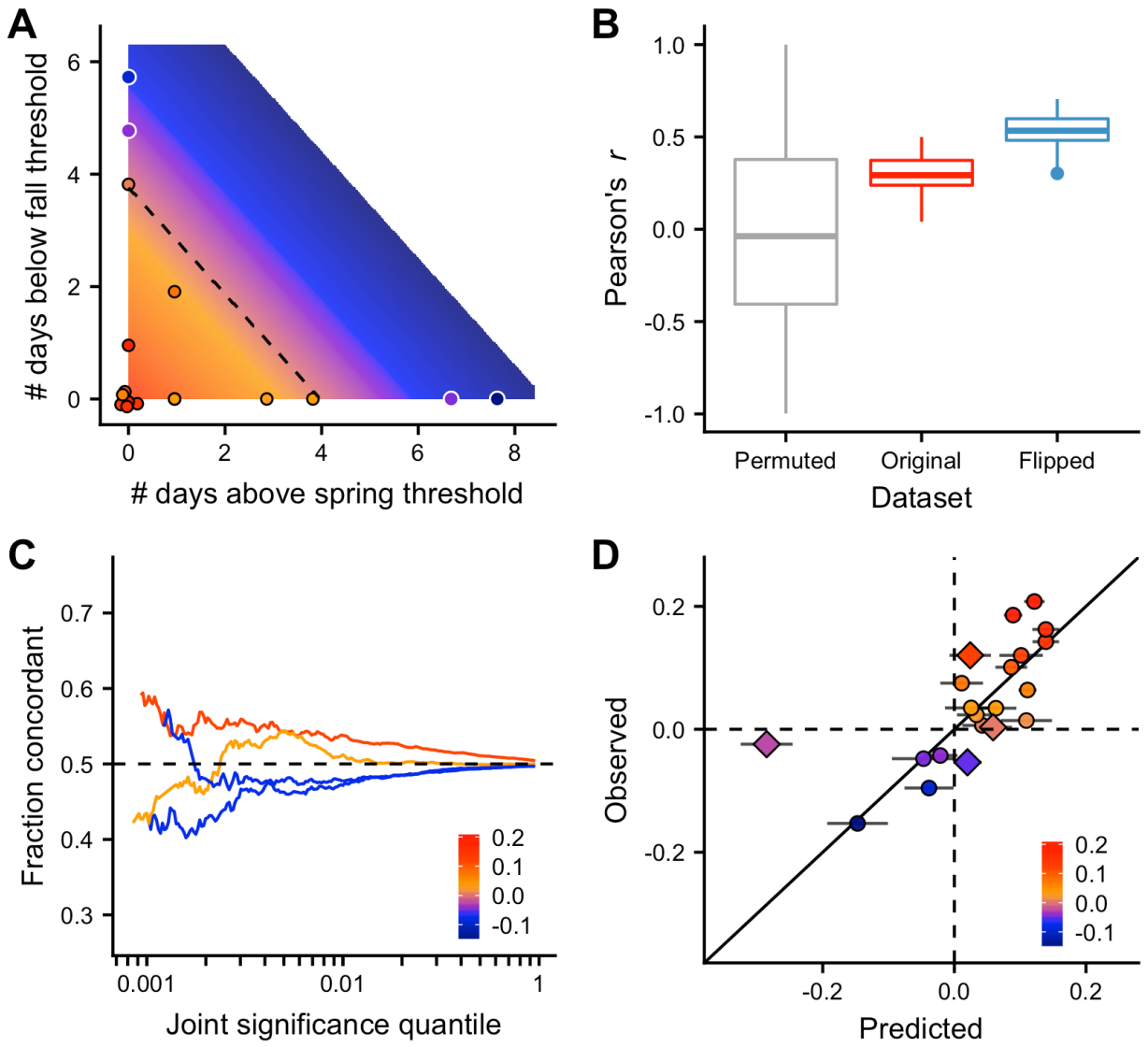
Supplemental Figure 3. Enrichment of genic classes for A) the top 1% and B) top 5% of seasonal sites for the original analysis, and C) the top 1% and D) top 5% for the flipped analysis.



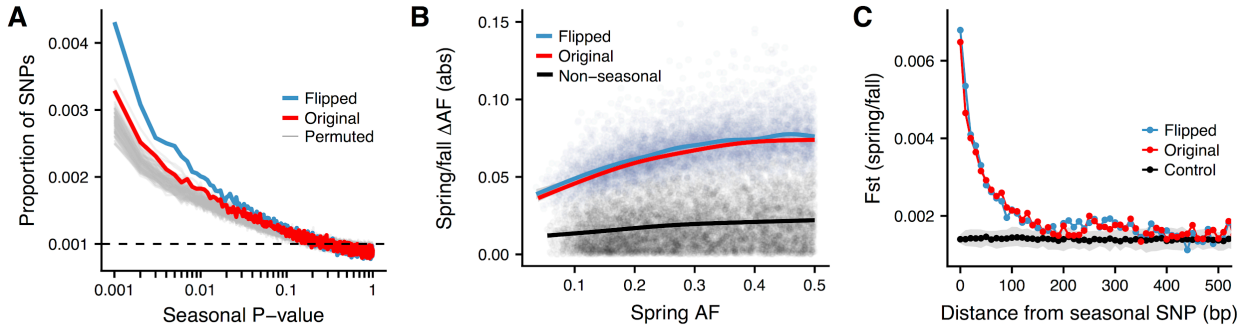
Supplemental Figure 4. The enrichment of shared seasonal sites from the leave-one-out cross validation analysis of the Core20 populations. Each line represents one population, colored by the inferred genome-wide predictability score.



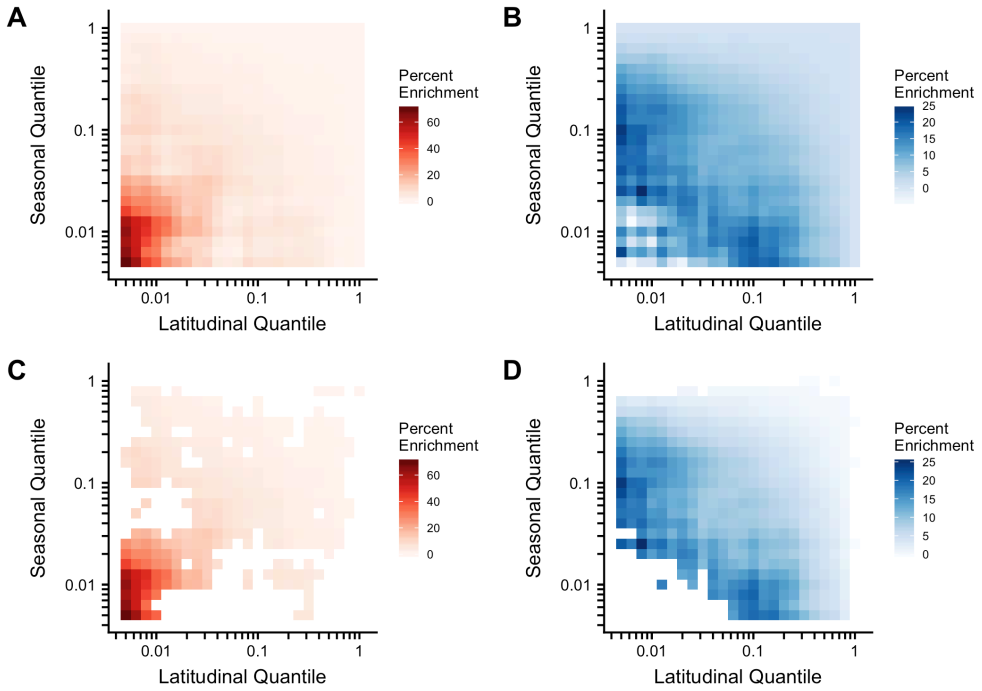
Supplemental Figure 5. Predictability of seasonal adaptation, by chromosome. The fraction of concordant allele frequency changes from the leave-one-out cross validation analysis of the Core20 populations. Each line represents one population, colored by the inferred genome-wide predictability score.



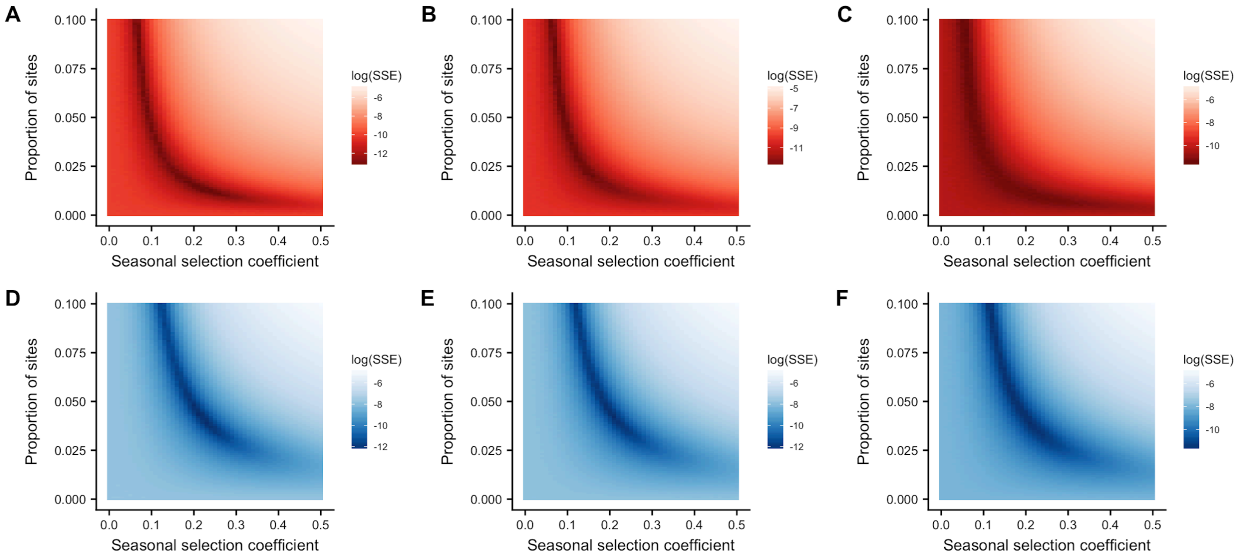
Supplemental Figure 6. Additional plots related to the predictability analysis. A) Surface plot of the best fit thermal model from the leave-one-out cross validation analysis; color scheme matches color legend in C & D; points are colored with their observed genome-wide concordance score. B) Pearson's correlation coefficient between the ValidationSet's obseved genome-wide concordance score and their predicted scores for the best-fit permuted models (null) versus obseved scores generated from the original Core20 model versus the flipped Core20 model. C) Genome-wide concordance scores from the ValidationSet of populations in relation to the flipped Core20 model. D) The best thermal limit models from the Core20 analysis predict genome-wide concordance scores of the ValidationSet (diamonds).



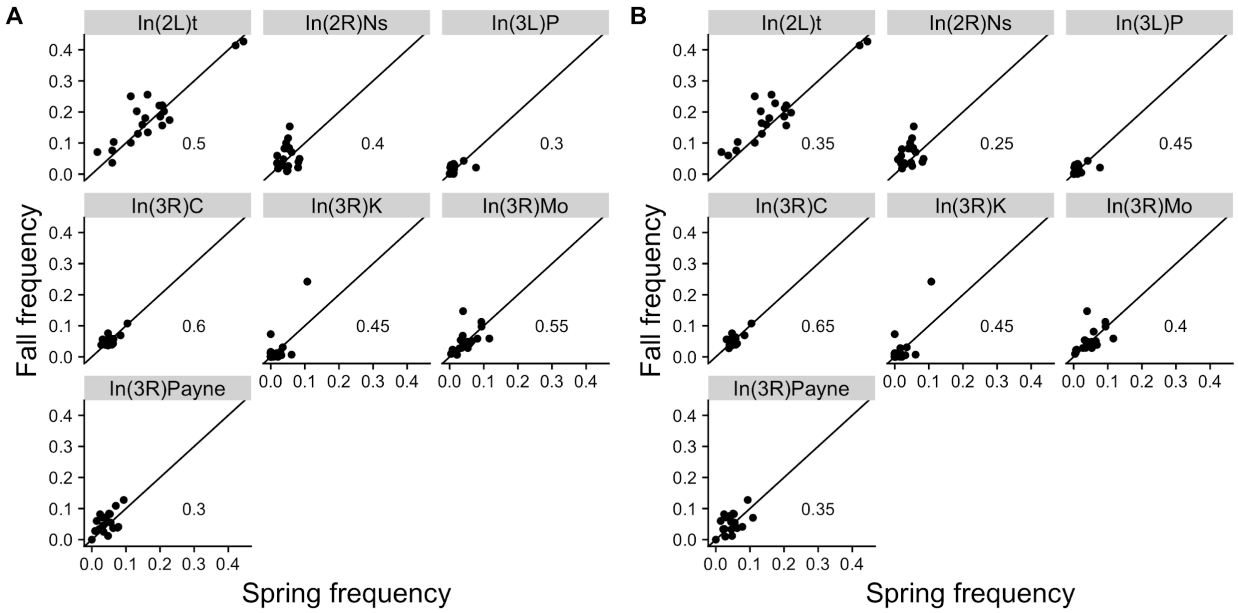
Supplemental Figure 7. Genomic characteristic of seasonal variants for the flipped dataset. A) P-value distribution of per-SNP seasonal regressions (blue and red lines), permutations (grey lines). P-value bins are of size 0.001, with a corresponding expected uniform distribution at 0.001 (dashed line). B) Average spring/fall allele frequency change for each of the top 1% of seasonally varying SNPs (blue and red) and non-seasonal, matched control SNPs (black), as a function of spring allele frequency. Lines represent a moving average across SNPs with a given spring allele frequency. C) Spring/fall genetic differentiation (F_{st}) at highly seasonal SNPs (distance = 0) and with distance (bp). Also plotted is the mean (black) and 96% confidence interval (grey area) of 100 matched controls.



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 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. Fisher's method ABC with subsampling. A,D) No subsampling. B,E) Sampling one SNP per 1Kb. C,F) Sampling one SNP per 5Kb. Top row: original dataset. Bottom row: flipped dataset.



Supplemental Figure 10. 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.