Comparative population genomics of latitudinal variation in *D. simulans* and *D. melanogaster*

Heather E. Machado\(^1\)*, Alan O. Bergland\(^1\), Katherine R. O’Brien\(^2,3\), Emily L. Behrman\(^3\), Paul S. Schmidt\(^3\), and Dmitri A. Petrov\(^1\)

\(^1\)Department of Biology, Stanford University, 371 Serra Mall, Stanford, CA 94305-5020
\(^2\)School of Biological Sciences, University of Nebraska-Lincoln, 348 Manter Hall, Lincoln, NE 68588
\(^3\)Department of Biology, University of Pennsylvania, 102 Leidy Laboratories, Philadelphia, PA 19104-6313

Keywords: comparative genomics; latitudinal variation; Drosophila; population genomics; latitudinal cline; parallelism

*Corresponding author: Heather E. Machado
Department of Biology, Stanford University, 371 Serra Mall, Stanford, CA 94305-5020
Fax: 650-723-6132
E-mail: machadoheather@gmail.com

Running title: *Drosophila* latitudinal comparative genomics

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.13446

This article is protected by copyright. All rights reserved.
Abstract

Examples of clinal variation in phenotypes and genotypes across latitudinal transects have served as important models for understanding how spatially varying selection and demographic forces shape variation within species. Here we examine the selective and demographic contributions to latitudinal variation through the largest comparative genomic study to date of *Drosophila simulans* and *D. melanogaster*, with genomic sequence data from 382 individual fruit flies, collected across a spatial transect of 19 degrees latitude and at multiple timepoints over two years. Consistent with phenotypic studies, we find less clinal variation in *D. simulans* than *D. melanogaster*, particularly for the autosomes. Moreover, we find that clinally varying loci in *D. simulans* are less stable over multiple years than comparable clines in *D. melanogaster*. *D. simulans* shows a significantly weaker pattern of isolation by distance than *D. melanogaster* and we find evidence for a strong contribution of annual re-migration to *D. simulans* population genetic structure. While population bottlenecks and migration can plausibly explain the differences in amount and stability of clinal variation between the two species, we also observe a significant enrichment of shared clinal genes, suggesting that the selective forces associated with climate are acting on the same genes and phenotypes in *D. simulans* and *D. melanogaster*.

Introduction

Latitudinal transects have been studied across the tree of life, in a large number of bacteria, plant, and animal species, revealing phenotypic and genetic clines (Baumann & Conover 2011; Feder & Bush 1989; Fuhrman *et al.* 2008; Salgado & Pennings 2005; Weber & Schmid 1998). A correlation between phenotypic variation and latitude is suggestive of local adaptation. For example, local adaptation to temperature is implicated in the correlation between decreased lifespan and latitude in ectotherms (Munch & Salinas 2009) and local adaptation to photoperiod is implicated in the correlation between flowering time and latitude in plants (Keller *et al.* 2011). However, neutral demographic processes also generate clinal variation. For example, “isolation by distance”, where gene flow is decreased between geographically distant populations, can produce patterns of variation...
similar to those resulting from local adaptation (Endler 1977). Strong patterns of clinal variation can also be generated by introgression between separate invading populations (Cruzan 2005) or range expansion of a single founding population (Excoffier et al. 2009). These demographic processes can be coincident with selective processes. Although disentangling selective and demographic scenarios is challenging, genomic datasets have the power to identify patterns associated either with selection or with demography. We perform a genomic study across two closely related Drosophila species, allowing us to elucidate general patterns that are shared between the species as well as refine our understanding of how the processes underlying clinal variation differ between these species.

The genus Drosophila represents a powerful system for the study of selection and demography. This group is composed of several species with broad distribution, and represents old and more recent New World colonizations. D. melanogaster has been studied extensively in a latitudinal context. Several phenotypic traits and genetic loci vary with latitude in D. melanogaster (Coyne & Beecham 1987; David et al. 1985; Emerson et al. 2009; Gockel et al. 2001; Hoffmann et al. 2002; James et al. 1995; Karan et al. 1998; Knibb et al. 1981; Mettler et al. 1977; Mitrovski & Hoffmann 2001; Munjal et al. 1997; Oakeshott et al. 1982; Paaby et al. 2010; Pool & Aquadro 2007; Schmidt et al. 2000; 2005; 2008; Sezgin et al. 2004; Singh et al. 1982; Vigue & Johnson 1973; Voelker et al. 1977). D. melanogaster latitudinal variation has also been studied in a genomic context and on multiple continents (North America, Australia, Europe, Asia, and Africa) (Bergland et al. 2014b; Fabian et al. 2012; Kolaczkowski et al. 2011; Reinhardt et al. 2014; Turner et al. 2008). D. melanogaster is a relatively recent colonizer of temperate climates (10,000 – 20,000 years since expansion out of central Africa; Lachaise et al. 1988; Li & Stephan 2006). Temperate-adapted characters such as cold tolerance and starvation resistance are more pronounced at higher latitudes in D. melanogaster, suggesting that some clinal variation in D. melanogaster is a result of local adaptation to temperate climates (Hoffmann et al. 2002; Karan et al. 1998; Schmidt et al. 2008). Additionally, there is some parallelism in clinal allele frequency patterns along the North American and Australian latitudinal clines, suggesting that there has been convergent adaptation to latitude (Fabian et al. 2012; Reinhardt et al. 2014; Turner et al. 2008). The D. melanogaster latitudinal clines are also subject to confounding demographic effects. Both North American and Australian
populations seem to be a result of admixture (either pre- or post-colonization) between European and African populations (Bergland et al. 2014b; Duchen et al. 2013; Kao et al. 2015). Although the D. melanogaster latitudinal clines are robust and some do result from local adaptation, demography complicates the inference of selection.

Comparative studies can help us understand general patterns of latitudinal variation. The sister species D. simulans and D. melanogaster (∼3 × 10^6 years diverged; Hey & Kliman 1993) represent a powerful system for comparative study. These species are similar in their range, ecology, and evolutionary history (Cariou 1987; Hey & Kliman 1993). They have experienced parallel expansions out of Africa, adaptation to temperate climates and development of human commensalism (David & Capy 1988; Lachaise et al. 1988; Lachaise & Silvain 2004). Unfortunately, the limited amount of research on clinal variation in D. simulans has made a large comparative study of latitudinal variation impossible.

While D. simulans exhibits clinal variation in some of the same traits as D. melanogaster (pigmentation: David et al. 1985; body size: Arthur et al. 2008), D. simulans also seems less temperate-adapted (Arthur et al. 2008; Gibert et al. 2004; McKenzie & Parsons 1974). For example, D. simulans has less physiological tolerance to cold and starvation (reviewed in Hoffmann & Harshman 1999). Another key clinal trait in D. melanogaster is a reproductive diapause, which is hypothesized to be important for survival through the high-latitude winters (Saunders et al. 1989; Schmidt & Conde 2006). Reproductive diapause has not been observed in D. simulans. Certain phenotypes that are clinal in both species vary less with latitude in D. simulans than in D. melanogaster (starvation: Arthur et al. 2008; desiccation: McKenzie & Parsons 1974), supporting the hypothesis of a more shallow cline in D. simulans (reviewed in Gibert et al. 2004).

While local adaptation could explain the above patterns, a shallow cline in D. simulans could also result from demographic patterns. Contemporary demographic patterns such as seasonal bottlenecks and migration may contribute to clinal variation. Although the true demographic patterns in D. simulans are not known, D. simulans has been hypothesized to experience strong bottlenecks and/or employ migratory behavior in response to seasonal fluctuations. This is supported by the temporal abundance patterns found along latitudinal clines in Europe and North America.
Specifically, *D. simulans* tends to be in greater relative abundance in the more equatorial populations, and does not appear at the higher latitudes until later in the year than *D. melanogaster*. Additionally, in temperate North America there are distinct differences between *D. melanogaster* and *D. simulans* in the population age structure across seasonal time that are also indicative of different overwintering strategies (Behrman et al. 2015). In *D. melanogaster* the earliest observed spring populations have a uniformly young age distribution, shifting to a heterogeneous age distribution over time. This pattern is consistent with populations that overwinter locally. In *D. simulans* the earliest observed post-winter populations are already age heterogeneous, which is more consistent with annual re-colonization from a refugia (either local or more distant) than with in-situ overwintering. The *D. simulans* relative abundance and age distribution patterns can be explained by either 1) annual extirpation and re-colonization of high-latitude populations, 2) in-situ overwintering and maintenance of a small resident population, or 3) both a strong annual bottleneck and subsequent input of migrants with the maintenance of a small resident population. Each of these scenarios would also contribute to a shallow cline.

Genomic analyses of latitudinal variation have been performed in *D. melanogaster*; however, no such studies exist for *D. simulans*. Genomic datasets are critical to understanding general patterns of clinal variation. With genomic data we can statistically differentiate subtle patterns, such as the enrichment of functional genic classes and parallelism in clinal variants between species. Here we present a multi-year, multi-season, genome-wide analysis of population differentiation and latitudinal variation in *D. simulans* and *D. melanogaster*. We directly compare the amount of clinal variation in *D. simulans* and *D. melanogaster* using these genomic data and confirm that, in line with phenotypic observations, *D. simulans* has less clinal variation than *D. melanogaster*. We find evidence for a strong contribution of annual variation to *D. simulans* population genetic structure, which is not found in *D. melanogaster*. The strong, stable cline in *D. melanogaster* is a stark contrast to the weak cline seen across *D. simulans* populations, where we see greater evidence of processes that increase differentiation from year to year, such as migration and bottlenecks. We also observe signatures of spatially varying selection in *D. melanogaster* and to a lesser extent in

This article is protected by copyright. All rights reserved.
D. simulans, and evidence for convergent evolution of clinal variation across genes.

Materials and Methods

Sequence data

D. simulans We sampled individuals from four D. simulans populations along the East Coast of North America, spanning 19 degrees latitude (S1 Table). From north to south, the population and year of collection are as follows: Maine 2011 (ME), Pennsylvania 2011 (PA), Virginia 2010 (VA), and Florida 2011 (FL). Three separate samples of the PA population were taken, one each in August, September, and November (named PA8.2011, PA9.2011, and PA11.2011, respectively). Populations were sampled by direct aspiration of flies from substrates and by collection with banana and yeast baited traps. We extracted DNA from a total of 267 female flies (an average of ∼50 files per sample) using Favorgen 96-well genomic DNA extraction kits, and quantified the DNA with a Picogreen fluorescence assay. Moleculo (now Illumina TruSeq) performed the per-individual library preparation and sequenced paired-end 100bp reads on an Illumina HiSeq 2000. Depth of sequencing coverage per individual varied from 0.01x to 5x. We aligned reads to the D. simulans v2 reference genome (Hu et al. 2013) with BWA version 0.6.2 aln and sampe functions (default parameters; Li et al. 2009). We performed PCR duplicate removal with SAMtools version 0.1.19 (Li et al. 2009) and indel realignment with GATK version 2.4 (McKenna et al. 2010). A total of ∼3.2M single nucleotide polymorphisms (SNPs) were called with GATK Unified Genotyper version 2.4 using all reads combined (DePristo et al. 2011). Genotype calls were made per individual for each SNP (ploidy 2). For each individual and SNP, we randomly chose a single chromosome of the diploid genotype for use in the final analysis, in order to avoid bias from individuals with higher coverage. SNPs were then filtered for ∼2.5M nucleotides of repetitive DNA identified using RepeatMasker (Smit et al. 2013). We also filtered out high coverage sites (upper 95th quantile) and low coverage sites (less than 5x per collection site). For consistency with the D. melanogaster data (see below), we filtered out sites within 5bp of an indel, with low minor allele frequency (MAF) sites (mean MAF < 10% across the four collection sites), and that were non-bi-allelic. Allele frequencies
for each population were calculated relative to the reference genome. The filtered dataset had a mean coverage of 20x per population, with 2.2M SNPs on autosomes and 0.3M SNPs on the X chromosome. For functional analysis, we used a D. simulans cDNA-guided genome annotation (Rogers et al. 2014b) and the SNP functional annotator snpEff v4.0 (Cingolani et al. 2012).

We also sequenced two 2010 temporal samples (July and September) from the Pennsylvania population (named PA7.2010 and PA9.2010, respectively). These population samples consisted of male files sequenced using pooled population sequencing (pool-seq), where individuals from each sample were pooled prior to DNA extraction and sequencing. We extracted DNA from these two samples using a lithium chloride precipitation. Sequencing library construction followed the protocol described in Bergland et al. 2014a. The samples were sequenced with paired-end 100 bp reads on an Illumina HiSeq 2000. The effective number of chromosomes ($N_C$) represented in the pooled samples was calculated as

$$N_C(N, R) = \left( \frac{1}{N} + \frac{1}{R} \right)^{-1}$$

where $N$ is the number of chromosomes in the pool and $R$ is the read depth at that site (Bergland et al. 2014a; Feder et al. 2012; Kolaczkowski et al. 2011). This adjusts for the additional error introduced by sampling of the pool at the time of sequencing. Sequencing reads mapped to autosomes were down-sampled to match the $N_C$ of the X chromosome.

**D. melanogaster** We compared D. simulans data to published D. melanogaster data from a study conducted by Bergland and collaborators (2014a). Three of the four D. melanogaster collection sites (FL, PA, ME) were the same as the D. simulans collection sites. The fourth D. simulans site (VA) was imperfectly matched to a D. melanogaster Georgia (GA) collection site (S1 Table). We used two D. melanogaster temporal samples of the PA site (November 2009 and November 2010). The other sites were sampled once each; FL in 2010, GA in 2008, and ME in 2009. Bergland et al 2014a produced sequence data by pooling males files within each population and sequencing on an Illumina HiSeq 2000. We mapped the raw reads to the D. melanogaster genome version 5.5

This article is protected by copyright. All rights reserved.
using BWA version 0.7.9 aln and sampe algorithms, with default parameters (Li & Durbin 2009). Reads mapping to autosomes were down-sampled to match the $N_C$ of the X chromosome for each population. Allele frequency was calculated relative to the reference allele for each SNP used in Bergland et al. 2014a (~600K SNPs). SNP calling in Bergland et al. 2014a differed from the $D.\ simulans$ SNP calling. The data in Bergland et al. 2014a were exclusively pool-seq data, for which SNPs were called using the program CRISP (Bansal 2010). Additional filtering also took place, notably, the exclusion of SNPs not also identified in the $Drosophila$ Genetic Reference Panel (DGRP). The differences in SNP calling and filtering, along with real differences in genetic diversity between the two species, account for the smaller number of SNPs in the $D.\ melanogaster$ dataset.

Pool-seq error model

In this study we compare pooled $D.\ melanogaster$ population samples with (primarily) non-pooled $D.\ simulans$ population samples. Pool-seq is known to have inherent errors in allele frequency estimation, therefore we must take care to model this variance appropriately (Kofler et al. 2011; Lynch et al. 2014; Zhu et al. 2012). This is particularly important for our analysis of the relative proportion of clinal variation in $D.\ simulans$ and $D.\ melanogaster$. Since all of the $D.\ melanogaster$ samples are pooled, these samples inherently have an additional source of error that is not accounted for, resulting in an overestimate of the sample size. Since clinal patterns are expected to be more pronounced in $D.\ melanogaster$, a perceived increase in clinal variation in $D.\ melanogaster$ could be attributed to the pool-seq variance. In order to arrive at a conservative estimate of clinal variation in $D.\ melanogaster$ a liberal estimate of pool-seq error should be used.

Two methods for accounting for extra variance in pool-seq data are 1) modifying the statistical tests used (e.g., modification of the null expectation, as in Bastide et al. 2013) and 2) translating the additional variance into an effective sample size. We chose the latter, using our comparable barcoded dataset to assess the additional pool-seq error. To model the pool-seq error we compared the level of genetic differentiation among barcoded temporal samples ($D.\ simulans$ PA8.2011, PA9.2011, PA11.2011) with differentiation between pooled temporal samples ($D.\ simulans$ PA7.2010, PA9.2010), with the assumption that within-population samples should have similar
amounts of month-to-month variation from one year to the next. This assumption is reasonable, as we observe this to be the case (see below) for populations within a few months of each other. Note that this may not be the case for certain months, particularly for those during or directly following a winter bottleneck. For this analysis we use the proportion of SNPs found to be at significantly different allele frequencies as a measure of genetic differentiation (Fisher’s exact test for each SNP). Only SNPs with a total of 40 chromosomes between the two samples being tested were used in order to ensure equal power between datasets. The range of chromosomes per population varied (PA8.2011: 8-35; PA9.2011: 8-36; PA11.2011: 4-32; PA7.2010: 18-22; PA9.2010: 22-18). The Fisher’s exact test provides a test of the deviation from panmixia (i.e., variation above binomial sampling error) with a standard expectation of a uniform $P$-value distribution, and is robust to small and unequally distributed sample sizes. Panmixia is rejected if there is enrichment of differentiated SNPs above this expectation. For pool-seq data, we do not expect a uniform $P$-value distribution under panmixia for two reasons: 1) to account for the two levels of sampling (chromosomes and reads), we use a single effective sample size ($N_C$), which is close to but not exactly the same as correctly using the convolution of two binomials, and 2) the average error in allele frequency estimation for pool-seq data may be greater than binomial, even with the effective $N_C$ calculation. We use the Fisher’s exact test on pooled and non-pooled data to estimate this second error component.

We first tested the differentiation among the three barcoded PA 2011 samples. We found that each of the three comparisons had similar levels of differentiation (between 0.99% and 1.05% of SNPs differentiated at $P < 0.01$, an average of 0.01% over expected; S1 Fig.), representing near-uniform $P$-value distributions (the null distribution). In contrast, the pooled temporal samples (PA7.2010 and PA9.2010) showed an enrichment of differentiated SNPs (1.26% of SNPs differentiated at $P < 0.01$). This is consistent with additional sampling error being introduced in the process of pooled DNA extraction, amplification, sequencing, and mapping, resulting in an overestimate of the effective number of chromosomes sampled.

To determine how much additional variance is introduced by pool-seq, above what is accounted for by the $N_C(N,R)$ correction already implemented, we tested two models of pooled error. We used the data from all three barcoded PA temporal comparisons to perform a linear regression of
differentiation with increasing $N_C$, providing the barcoded null model. We then found the additional variance component, which we call $\epsilon$, that results in the best fit of the pooled PA comparison to the barcoded null (lowest sum of square deviations from the null). The first model tested fits an $\epsilon$ that is independent of $R$:

$$N_C(N, R, \epsilon) = \left(\frac{1}{N} + \frac{1}{R} + \epsilon\right)^{-1}$$ (2)

The second model tested fits an $\epsilon$ that is inversely proportional to $R$ (greater error at lower read depth):

$$N_C(N, R, \epsilon) = \left(\frac{1}{N} + \frac{1}{R} + \frac{\epsilon}{R}\right)^{-1}$$ (3)

We found Model 2 (with the $R$ dependence) to be a better fit to the data than Model 1, with a best-fit value of $\epsilon$ to be 0.1 (S5 Fig.). Using this error model we can calculate a more conservative $N_C$, which we use for the calculation of $N_C$ for all pooled samples ($D. \ simulans$ PA7.2010 and PA9.2010 samples; all $D. \ melanogaster$ samples). Applying this $N_C(N, R, \epsilon)$ correction to the pooled $D. \ simulans$ PA7.2010 and PA9.2010 samples, we find a slight depletion of significantly differentiated SNPs compared to the barcoded samples (0.84% at $P < 0.01$, compared with 1.01%). This indicates that our correction for pooled error results in a conservative estimate of the effective number of chromosomes in a pooled sample.

Use of this correction for pooled error also decreases the average coverage per population. However, even with the use our pool-seq error correction, our pool-seq libraries are still more efficient in estimating population allele frequency than our barcoded libraries (per raw sequencing read). For example, from 898M raw barcoded reads, we retrieved a total of 115x coverage across all populations, which is an average of 8.1M reads per 1x coverage. This is compared to 2.7M and 4.3M reads per 1x coverage for the $D. \ simulans$ PA7.2010 and PA9.2010 pool-seq libraries, respectively. In summary, our two $D. \ simulans$ pool-seq libraries were 39–53% more efficient in population allele frequency
estimation per raw sequence read than our barcoded libraries. This increased pool-seq efficiency may be particularly pronounced in our study, as our barcoded libraries had high heterogeneity in coverage across individuals.

Measures of genetic variation, genetic differentiation, and isolation by distance

We calculated two measures of within-population genetic variation—mean expected heterozygosity \((H)\) and Watterson’s theta \((\theta_S)\). For these analyses we considered only sites covered by exactly 20 chromosomes in a given population, in order to avoid any biases resulting from differences in coverage among populations. Mean heterozygosity was calculated as

\[
H = \frac{1}{N} \sum_{i=1}^{N} 2p_i(1 - p_i) \tag{4}
\]

where \(N\) is the number of sites (polymorphic and monomorphic) and \(p\) is the allele frequency of each site. \(\theta_S\) was measured as the proportion of polymorphic SNPs, divided by the sample size correction:

\[
\theta_S = \frac{S}{\sum_{i=1}^{n-1} \frac{1}{n}} \tag{5}
\]

where \(S\) is the proportion of SNPs in the genome and \(n\) is the number of chromosomes (i.e. 20).

We measured between-population genetic differentiation with the \(F_{st}\) statistic (Weir & Crawfordham 1984, equations 1:4). \(F_{st}\) calculations were performed for each pairwise population comparison, for each SNP. Since sample size affects the results of the \(F_{st}\) statistic, we consider only SNPs with a total depth of coverage of 40 – 44 chromosomes between the two populations, with a minimum of 5 per population. In \(D.\ simulans\), the maximum number of chromosomes per population ranged from 36 – 39 (FL: 39; VA: 39; PA8.2011: 37; PA9.2011: 39; PA11.2011: 36; ME: 37; PA7.2010: 39; PA9.2010: 39). In \(D.\ melanogaster\), the maximum ranged from 31 – 39 (FL: 31; GA: 39; PA8.2011: 39; PA.2009: 39; PA.2010: 31; ME: 39). Since the variance in the pool-seq allele frequency estimates are accounted for by the measure of effective number of chromosomes, \(N_C(N, R, \epsilon)\), no additional
pool-seq correction is necessary for $F_{st}$ or genetic variation calculations.

We assessed isolation by distance with a linear regression of $F_{st}$ with geographic distance between populations (degrees latitude). We incorporated into a multiple linear regression model the effect of comparison with the Maine population (versus comparison between two non-Maine populations) and within-year (versus between-year) comparison. The final regression model is of the form:

$$y_i = d + m + y + d \times m + d \times y + m \times y + \epsilon_i$$

(6)

where $y_i$ is the pairwise $F_{st}$, $d$ is the distance (degrees latitude) between two populations, $m$ is whether or not one of the two populations of the comparison is Maine, $y$ is between versus within year comparison, and $\epsilon_i$ is the gaussian error at the $i^{th}$ SNP.

**Measures of clinal variation**

To identify clinal SNPs we used a generalized linear model (conducted in R version 3.1.0; R Core Team 2014) of allele frequency and population latitude, using a binomial error model and weights proportional to the effective number of chromosomes at each site ($N_C$):

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

(7)

where $y_i$ is the allele frequencies at the $i^{th}$ SNP, and $\epsilon_i$ is the binomial error given the $N_C$ at the $i^{th}$ 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. For each species, we used five population measurements sampled from the four populations- one sample from each population, with an additional year’s sample for PA (for *D. simulans*, we used PA7.2010 and PA8.2011). Each year of Pennsylvania samples was treated as a separate datapoint in the regression analysis, with a single timepoint for each year.

The average $N_C$ across the populations used in the clinal regression varied little from chromosome

This article is protected by copyright. All rights reserved.
to chromosome, ranging from 21.2 − 21.4 in *D. melanogaster* and 20.9 − 21.8 in *D. simulans*. There was no significant difference between the two species in mean $N_C$ (t-test $P = 0.19$) or total $N$ summed over all five populations (t-test $P = 0.13$) (S2 Fig.). This equality of sample sizes is important because it allows us to compare the two datasets without confounding differences in power.

We identified two sets of clinal SNPs based on the results of the clinal regressions- SNPs that were statistically significant at $P < 0.01$, and SNPs that were statistically significant at false discovery rate (FDR) of $Q < 0.2$. FDR $Q$-values represent the proportion of false positives in a set of tests, and were calculated with the R package *qvalue* (Storey 2015). We use the $P < 0.01$ set to estimate the relative proportion of clinal loci in *D. melanogaster* and *D. simulans*, allowing us account for the number of false positives due to multiple testing (using the null expectation) in a way that does not skew the false negative rates. The proportion of clinal loci (SNPs) is calculated as:

$$\frac{\text{obs}_{P<0.01} - \text{exp}_{P<0.01}}{L}$$

(8) where $L$ is the number of SNPs tested, $\text{obs}_{P<0.01}$ is the observed number of tests with $P < 0.01$, and $\text{exp}_{P<0.01}$ is the expected number of tests with $P < 0.01$ under the null expectation ($L \cdot 0.01$). For the remainder of the analyses (i.e., clinal consistency, functional genic classes, shared clinal genes) we use FDR $Q$-values, ensuring equal proportions of false positives in the *D. melanogaster* and *D. simulans* datasets.

To test the consistency of clinal patterns of allele frequency across years, we measured how well the regression coefficient from one year predicts the directionality in a second year. Allele frequency measures from three *D. simulans* sites from 2011 and two from 2010 were available. We performed a logistic regression across the three 2011 sites (FL, PA2011, ME), and asked if the same trend of either increasing or decreasing frequency with latitude was observed in 2010. Specifically, we asked if the sign of the regression coefficient agreed with the sign of the difference between the 2010 populations (VA, PA2010). If there was agreement, these SNPs were deemed to be “consistently clinal”. If SNPs truly are clinal from year to year, it is expected that the proportion of SNPs found to be consistently clinal to increase with the stringency of the regression test (lower $Q$-value). We then
performed a similar analysis in *D. melanogaster*, comparing the regression of the three 2008/2009 sites (GA2008, PA2009, ME2009) with two 2010 sites (FL2010, PA2010). Since the inclusion of sites from two different years in the regression might bias towards identifying sites that truly are persistently clinal, thereby increasing the amount of clinal consistency detected, we compared this analysis with a mixed-year analysis of *D. simulans*. For this analysis, we performed a regression of *D. simulans* VA2010, PA2010, and ME2011, compared with the difference between the FL2011 and PA2011 sites. This provided a comparison that was liberal to finding clinal consistency in *D. simulans*. Results from the *D. simulans* mixed-year analysis were not significantly different from the single-year analysis (within two standard deviation), with the exception of chromosome 2L, for which the mixed-year analysis shows a decrease in clinal consistency (S3 Fig.).

**Enrichment tests**

To test for enrichment of genic categories and of polymorphisms shared between *D. melanogaster* and *D. simulans* in sets of clinal SNPs ($Q < 0.2$) we compared our datasets with 100 bootstrap control datasets matched for mean allele frequency across the populations (by 20th quantile bin), inversion status (within the same inversion or outside inversions, applicable to *D. melanogaster* only; by 7th quantile bin), chromosome, and effective sample size $N_C$ (by 10th quantile bin). The sizes of matching bins were chosen to result in the most well-matched controls that were also independent of one another. Genic categories for each species were identified with SNPeff (Cingolani *et al.* 2012), except for short introns. We used the set of *D. melanogaster* short introns identified in (Lawrie *et al.* 2013) and identified short introns in *D. simulans* as those less than 68bp in the annotation by Rogers *et al.* (2014b). We used the same *D. melanogaster* inversion breakpoints as in (Corbett-Detig & Hartl 2012).

We tested for an enrichment of genes identified as clinal in both *D. melanogaster* and *D. simulans*. We identified a gene as clinal if it had at least one clinal genic SNP (i.e., in the CDS, UTR, or intronic regions). We measured the percent of shared clinal genes as the overlap of *D. simulans* clinal genes with *D. melanogaster* clinal genes (contains at least one SNP with $Q < 0.2$). This was performed for five sets of *D. simulans* clinal genes, ranging in stringency from $Q < 0.5$ to $Q < 0.1$. 

This article is protected by copyright. All rights reserved.
For each set of *D. simulans* clinal genes, we produced 100 control sets of *D. simulans* genes matched for gene length (by 10th quantile bin) and SNP density (by 10th quantile bin), and measured the proportion of control genes shared with *D. melanogaster* clinal genes. Genes were omitted if less than 85 unique control genes could be identified. The distributions of gene length and SNP density for the clinal compared with the control gene sets overlapped well, and the majority (87%) of control genes were unique across permutations (S4 Fig.).

**Results**

**D. simulans** SNPs across space and time

Here we study *D. simulans* population genetic variation using genomic sequence data from 382 individual fruit flies (267 individually barcoded and 115 in pooled samples). Samples represented a spatial transect of 4 populations over 19 degrees latitude and a temporal transect of multiple timepoints over the course of two years (S1 Table). We identified $2.5 \times 10^6$ bi-allelic *D. simulans* single nucleotide polymorphisms (SNPs) across the four major autosomal chromosome arms and the X chromosome (see Methods for filtering parameters). We utilized a matched *D. melanogaster* dataset of pooled population sequence data ($\sim 6 \times 10^5$ SNPs; Bergland *et al.* 2014a) to compare patterns of within, between, inter-annual, and latitudinal population genetic variation. For all pool-seq samples, we applied a stringent pool-seq error correction that accounted for finite sampling and additional pool-seq variance (see Methods), allowing us to confidently compare the *D. melanogaster* dataset with the *D. simulans* dataset.

Larger proportion of clinal variants in *D. melanogaster* than *D. simulans*

We found a larger proportion of latitudinally clinal variants in *D. melanogaster* (3.7%) than in *D. simulans* (2.5%) ($P < 0.01$; Fig. 1D). The difference in the proportion of clinal variants was even greater when we considered only autosomal SNPs (4.3% in *D. melanogaster* compared with 2.1% in *D. simulans*; Fig. 2). As major chromosomal inversions in *D. melanogaster* show clinal...
patterns in frequency (Mettler et al. 1977), we asked if inversions account for the difference between species. We found an elevated proportion of clinal SNPs in D. melanogaster inversions; however, D. melanogaster had a higher proportion of clinal SNPs than D. simulans in non-inverted regions as well (S2 Table). Similarly, although we did see an enrichment of clinal SNPs in low-recombination regions for D. melanogaster, the proportion of clinal SNPs outside low-recombination regions was still greater for D. melanogaster than D. simulans (S2 Table).

We found substantial variation in clinality among chromosomes. The most striking pattern in D. melanogaster was the strong enrichment of clinal variants on chromosome 3R (9% clinal; Fig. 2). In D. melanogaster much of the 3R chromosome is covered by three large cosmopolitan inversions. These inversions, particularly In(3R)P, have previously been found to be strongly clinal (Kapun et al. 2014; Mettler et al. 1977). On the X chromosome D. melanogaster and D. simulans had the opposite patterns of clinal variation. D. melanogaster had less clinal variation on the X chromosome (1% clinal) than any of the autosomes, whereas D. simulans had more clinal variation on the X chromosome (4% clinal) than any of the autosomes. Lower levels of clinal variation on the D. melanogaster X chromosome have been observed in previous studies (David & Capy 1988; Fabian et al. 2012; Kolaczkowski et al. 2011).

We asked if the increased amount of clinal variation observed in D. melanogaster could be explained by greater D. melanogaster population structure. We looked at the effect of population structure by comparing genome-wide mean pairwise $F_{st}$. First we noticed that on average (across all SNPs) D. simulans had a greater mean $F_{st}$ than D. melanogaster, indicating that a net increase in population structure was not driving the increased proportion of clinal variants in D. melanogaster. To look at the effect population structure had on the magnitude of clinal variation, we asked how mean $F_{st}$ scaled with the clinal effect size $\beta$ (regression coefficient). We found that D. melanogaster had a stronger relationship between $F_{st}$ and $\beta$ than D. simulans (S5 Fig.), indicating that in D. melanogaster more of the observed population structure was due to clinal genetic differentiation.

This article is protected by copyright. All rights reserved.
Consistency of clinal variants from year to year

To assess the stability of clinal variation over time, we measured how well the clinal regression coefficient in one year predicted the allele frequency directionality in a second year. In order to ensure equal power and noise for the *D. melanogaster* and *D. simulans* analyses, we used false discovery rate (FDR) corrected *Q*-value significance thresholds for the clinal regressions and down-sampled the number of SNPs to the same number in each species and chromosome. We found evidence for clinal consistency from year to year in both species, with the proportion of clinal consistency increasing with *Q*-value stringency to 67% and 54% for *D. melanogaster* and *D. simulans*, respectively (at clinal *Q* < 0.3; Fig. 3; all chromosomes). Note that the *Q*-values are generally higher in this analysis than in the full clinal regression, as we use three populations instead of five. We found that *D. melanogaster* had significantly greater clinal consistency from year to year than *D. simulans* for each chromosome (Fisher’s exact test *P* < 10^{-14}) except the X chromosome (*P* = 0.3).

Selection and parallelism in clinal variants

If clinal SNPs have phenotypic effects that are under spatially varying selection, we expect functional sites to be over-represented in the sets of clinal SNPs. Our expectation is that intergenic regions, short introns, and synonymous sites are less likely to be functional than UTR’s, non-synonymous sites, and long introns. We used a constant FDR (*Q* < 0.2) and number of SNPs per species (25,134 autosomal and 805 X chromosome SNPs) in order to ensure equal noise and power for the *D. melanogaster* and *D. simulans* analyses. For the set of SNPs clinal in *D. melanogaster* autosomes we found a significant enrichment of all genic classes (UTR’s, long intron, synonymous coding, and non-synonymous coding) except short introns, and found a depletion of intergenic regions, compared with 100 bootstrap control datasets matched for chromosome, mean minor allele frequency, sample size, recombination rate, and inversion status (see Methods; Fig. 4A). Additionally, we found a marginal increase in the proportion of non-synonymous SNPs compared with synonymous SNPs (*P* = 0.1). Conversely, the *D. melanogaster* X chromosome was enriched for intergenic SNPs and depleted for long introns and non-synonymous SNPs (S6 Fig.). The set of *D. simulans* clinal SNPs showed a marginal enrichment (*P* < 0.1) of 5’UTR SNPs and a marginal depletion of intergenic
SNPs (autosomes; Fig. 4B).

If selection is acting similarly on both species, we might find evidence of convergent evolution of clinal variants. We asked if there was an enrichment for SNPs or genes that are clinal in both *D. simulans* and *D. melanogaster*. We found no significant enrichment for shared clinal SNPs (61 shared clinal polymorphisms out of 32,136 shared polymorphisms total). However, we did observe an enrichment of shared clinal genes (Fig. 5). We compared the proportion of shared clinal genes with the proportion for 100 bootstrap control sets of genes, matched for *D. simulans* gene length and SNP density (see Methods). Of the genes with at least one clinal SNP ($Q < 0.2$; 5559 *D. simulans* genes and 5556 *D. melanogaster* genes), 56% were clinal in both species, compared to a mean of 45% across the bootstrap replicates ($P = 0.01$). This enrichment became even more pronounced at more stringent *D. simulans* clinal regression thresholds (for *D. simulans* clinal regression $Q < 0.1$, observed: 65%, control: 46%; Fig. 5). We did not find the shared clinal genes to be enriched in SNPs that were also clinally consistent (Fisher’s exact test; $P > 0.3$ for both species).

We next queried the list of 3342 shared clinal genes for its overlap with a set of 13 genes previously found to be clinal in *D. melanogaster*. In order to arrive at a set of putatively clinal genes, we gathered genes from targeted studies of clinal variation (rather than genomic scans). The result was 13 genes with strong support in the literature, and was comprised of the seven metabolism genes *Pgm* (Sezgin et al. 2004; Verrelli & Eanes 2001), *G6pd* (Oakeshott et al. 1983), *Gpdh* (Oakeshott et al. 1982), *UGP* (Sezgin et al. 2004), *Treh* (Sezgin et al. 2004), *Pgd* (Oakeshott et al. 1983), and *Hex-C* (Duvernell & Eanes 2000) and the six non-metabolism genes *sgg* (Rand et al. 2010), *mth* (Duvernell et al. 2003; Schmidt et al. 2000), *cpo* (Schmidt et al. 2008), *per* (Costa et al. 1992), *Adh* (Berry & Kreitman 1993; Vigue & Johnson 1973), and *InR* (Paaby et al. 2010). All except one of these genes (mth) were analyzed in both species, leaving a final set of 12 genes. Of these 12 genes, 10 were clinal in both species. The two genes that were not found to be clinal in both species were *Pgd* and *Hex-C*.

We also compared our results to a recent study of gene expression in *D. melanogaster* and *D. simulans* low- (Panama) and high- (Maine) latitude populations (Zhao et al. 2015). For each
population, gene expression was measured at 21°C and 29°C. Zhao and colleagues identified sets of 76 and 106 genes with latitude-specific expression in both species, at 21°C and 29°C, respectively (Zhao et al. 2015, S8 Table). We compared the intersection of these datasets and our shared clinal genes dataset with the intersection for 100 bootstrap control datasets matched for D. simulans gene length and SNP density (see Methods). We found only a marginal ($P = 0.1$) enrichment of latitude-specific genes at 29°C, and no enrichment of latitude-specific genes at 21°C, in our set of shared clinal genes. Zhao and colleagues also identified sets of genes with differential expression between temperatures (21°C and 29°C) in both species- 375 genes in the Maine populations and 861 in the Panama populations (S10 Table, Zhao et al.). Also controlling for gene length and SNP density, we did find an enrichment of temperature-responsive genes in our set of shared clinal genes; however, this was only true for the Panama populations ($P = 0.02$) and not the Maine populations ($P = 0.18$).

Population genetic patterns in space

Visual inspection of frequency trajectories along the cline showed a more monotonic increase in allele frequency with latitude in D. melanogaster than D. simulans (Fig. 1). To further investigate this, we asked if genetic differentiation between populations increased monotonically with physical distance between populations, a pattern know as “isolation by distance”. We found that D. simulans had a weaker pattern of isolation by distance than D. melanogaster (Fig. 6). While in D. melanogaster the regression of genetic differentiation ($F_{st}$) and physical distance between populations (degrees latitude) was significant ($P < 10^{-5}$, $R^2 = 0.94$), in D. simulans this was only significant ($P = 0.001$) in a regression model that included Maine (ME) as an explanatory variable (S3 Table). In D. melanogaster there was no effect of ME comparison. The significant effect of ME comparison in D. simulans was due to the disproportionate amount of divergence of ME from the other populations. Interestingly, we also found less genetic diversity in the D. simulans ME population than the other D. simulans populations (S7 Fig.). In addition, the level of differentiation among the three southern D. simulans populations was considerably lower than for the three southern D. melanogaster populations (Fig. 6).

This article is protected by copyright. All rights reserved.
Population genetic patterns in time

The analysis of isolation by distance incorporated data from different years. We used this to determine if there was a difference in the amount of inter-annual variation between *D. melanogaster* and *D. simulans*. Since *D. simulans* has low clinal consistency, we might expect to also find a greater amount of inter-annual variation in *D. simulans*. We can test this with the isolation by distance regression model, and ask if there is a significant effect of between- versus within-year comparison. Specifically, between-year comparisons should have greater $F_{st}$ than predicted by a regression of within-year comparisons. In *D. simulans*, we did indeed find that the effect of within-versus between-year comparison was significant in the regression model ($P = 0.002$), with between-year comparisons showing greater genetic differentiation (Fig. 6; S3 Table). The significant effect of between-year sampling implies that there was a detectable level of inter-annual variation in *D. simulans*. In contrast, in *D. melanogaster* there was no effect of between-year comparison.

Although much of the clinal variation in *D. simulans* is not maintained from year to year (low clinal consistency) and there is inter-annual variation, can we still find evidence of genetic continuity in a population from year to year? We assessed the level of genetic continuity across years by comparing the level of differentiation ($F_{st}$) among populations within a year to the level of differentiation within a population across years. We asked if the PA.2010 samples were most similar to the PA.2011 samples (genetic continuity between years) or to the VA.2010 sample (genetic similarity between sites, within a year). We found significantly lower within-population differentiation (PA.2010 / PA.2011) than between-site within-year differentiation (PA.2010 / VA.2010) (chi-squared $P < 0.0001$; S4 Fig.), indicating that a given *D. simulans* population does maintain some degree of genetic similarity from year to year.

Increased X chromosome differentiation and clinal variation in *D. simulans*

The X chromosome in *D. simulans* showed two patterns not observed in *D. melanogaster*- an increased proportion of clinal variants and increased population genetic differentiation compared with the autosomes. The increased level of X chromosome differentiation was particularly pronounced in
any comparisons with ME (Fig. 7). We asked if the increased differentiation on the X chromosome was consistent with its reduced effective population size resulting from hemizygosity in males. We used the formula proposed by Ramachandran et al. 2004 that predicts the relationship between autosomal $F_{st}$ and X chromosome $F_{st}$, given a particular sex-ratio. In order to perform this analysis we calculated pairwise $F_{st}$ for autosomal loci and the corresponding expected X chromosome $F_{st}$ values, assuming equal proportions of breeding males and females. Only in the ME comparisons were the X chromosome $F_{st}$ values significantly greater than expected when accounting for decreased effective population size (Fig. 7). With regard to the proportion of clinal variants, it is impossible to say if the increased level of clinal variation on the X chromosome was due to the general pattern of increased X chromosome differentiation because the two signals are both strongly affected by increased ME differentiation.

**Discussion**

Our study is the first to conduct a comparative genomic analysis of *D. simulans* and *D. melanogaster* latitudinal variation. We expect *D. melanogaster* to have a larger proportion of clinal genetic variants than *D. simulans*, as *D. melanogaster* has been documented to have more strongly clinal phenotypes (Arthur et al. 2008; Gibert et al. 2004). The absence of *D. simulans* at high latitudes early in the year (Behrman et al. 2015; Boulétreau-Merle et al. 2003; Fleury et al. 2004) can be explained by either a stronger *D. simulans* winter bottleneck or population extinction and re-colonization, both of which would result in a less stable cline from year to year. Our experimental design focuses on testing these predictions of less clinal variation and less clinal stability in *D. simulans*, as compared with *D. melanogaster*.

**Less clinal variation in *D. simulans* than *D. melanogaster***

We find strong support for a larger proportion of clinal variants in *D. melanogaster* than in *D. simulans*, particularly for *D. melanogaster* autosomes, which harbor twice as much clinal variation as *D. simulans* autosomes (4.3% and 2.1%, respectively). We have ensured that this result is not
confounded by differences in power or the additional sampling error of pool-seq. With a greater sample size (i.e. additional populations), it is possible that we would find an even greater proportion of clinal variants. For example, using deeper coverage and additional populations, Bergland et al (2014a) identified approximately one-third of common D. melanogaster SNPs as clinal. Our study design of four populations along a latitudinal transect makes our measurements of clinal variation sensitive to outlier allele frequencies at the Florida and Maine populations. In D. simulans, we do find that Maine is a genetic outlier, which could be contributing to the lower proportion of clinal variation identified. However, multiple lines of evidence from this study do support the conclusion of a more robust cline in D. melanogaster than in D. simulans, including increased clinal consistency, stronger isolation by distance, and more clear signatures of selection in D. melanogaster than D. simulans.

The strong pattern of clinal consistency in D. melanogaster, where clinal SNPs tend to show the same allele frequency pattern from year to year, indicates that the D. melanogaster cline is stable, rather than transient and re-established on an annual basis. Not only does D. simulans have a smaller proportion of clinal variants, the variants that are clinal are much less likely to be clinal from year to year than D. melanogaster. This indicates that the D. simulans cline is less stable, with a greater proportion of clinal variants due to processes operating on annual timescales.

The strong pattern of isolation by distance in D. melanogaster is also indicative of a robust cline. The pattern of isolation by distance in D. melanogaster is independent of whether or not the population pair was sampled in the same or different years. In contrast, in the D. simulans isolation by distance regression model there is a significant effect of within- versus between-year sampling of population pairs, indicating that inter-annual variation drives a detectable amount of population genetic variation. One important note is that in D. simulans the genetic continuity at a collection site (i.e. across years) is still greater than the genetic similarity between collection sites (within a year), indicating that there is a balance between the processes resulting in these two patterns. For D. melanogaster, the pattern of isolation by distance is unperturbed by inter-annual variation, possibly indicating low effective migration rate between populations or a balance between selection and migration not seen in D. simulans.
Although a demography-driven pattern of isolation by distance can result in stable clinal variation, stability can also result from local adaptation to variable conditions along a transect. We find that *D. melanogaster* clinal SNPs are significantly enriched for functional genic classes, including UTR’s, coding regions, and long introns, and have a marginally elevated proportion of non-synonymous to synonymous sites. This suggests that *D. melanogaster* clinal variants are under selection. We see weak evidence for selection in *D. simulans*, which shows a marginal enrichment for 5’UTR’s and no enrichment for other genic classes, suggesting that neutral processes play a stronger role.

Our comparisons of clinal variation in these two species reveal robust patterns of allele frequency with latitude in *D. melanogaster*, and weaker patterns in *D. simulans*. *D. melanogaster* not only harbors a larger proportion of clinal SNPs, but allele frequency patterns of clinal variants persist more from year to year, and there is evidence that clinal variants are under increased spatially varying selection. These results are consistent with previous studies that suggest less clinality in *D. simulans*. Specifically, some characters show no clinality in *D. simulans* (weight, wing length: Gibert et al. 2004; hexokinases: Duvernell & Eanes 2000; absence of diapause: Schmidt et al. 2005), while others show a decreased amplitude of clinality (wing length, thorax length, ovariole number: Gibert et al. 2004, cold tolerance, starvation tolerance: Hoffmann & Harshman 1999).

**Shared clinal genes**

A given selection pressure may act on the same genes in closely related species. Since selection pressures along the latitudinal cline are expected to vary in the same manner for *D. melanogaster* as for *D. simulans*, the two species may exhibit similar genetic responses. We find a significant enrichment for genes that are clinal in both species. Fifty-six percent of the 5559 *D. simulans* clinal genes were also clinal in *D. melanogaster*, compared to 45% in the matched controls. The enrichment of shared clinal genes increases with increasing stringency of the clinal regression. This supports the hypothesis of convergent evolution in these species due to the action of similar selection pressures on similar genetic backgrounds. This result is also consistent with the finding of parallel latitudinal gene expression in *D. melanogaster* and *D. simulans* (Zhao et al. 2015).
Although there is a significant enrichment of shared clinal genes (∼ 20% more shared clinal genes than expected), we still cannot say which of the ∼ 3000 shared clinal genes are true positives. However, we can ask if genes previously identified as clinal tend to be shared clinal genes in our dataset. When we look at a set of 12 genes with substantial literature support for latitudinal variation in *D. melanogaster*, 10 are clinal in both *D. melanogaster* and *D. simulans*. These genes include *Pgm* (Sezgin et al. 2004; Verrelli & Eanes 2001), *G6pd* (Oakeshott et al. 1983), *Gpdh* (Oakeshott et al. 1982), *UGP* (Sezgin et al. 2004), *Treh* (Sezgin et al. 2004), *sgg* (Rand et al. 2010), *mth* (Duvernell et al. 2003; Schmidt et al. 2000), *cpo* (Schmidt et al. 2008), *per* (Costa et al. 1992), *Adh* (Berry & Kreitman 1993; Vigue & Johnson 1973), and *InR* (Paaby et al. 2010).

We also find that our set of shared clinal genes is enriched for genes recently identified by Zhao and colleagues (2015) to have temperature-dependent expression in both *D. melanogaster* and *D. simulans* (Panama populations). Interestingly, we find only a marginal enrichment (\(P = 0.1\)) for genes with latitude-specific expression (Panama versus Maine) in both species. One explanation for the lack of enrichment is the difference in sampling schemes. We sampled four populations along a continuous transect and identified loci that vary consistently with latitude. Zhao et al. 2015 sampled two populations from separate continents, and identified gene expression differences between these two diverged groups.

**Demographic implications of *Drosophila* clinal patterns**

It is possible that *D. simulans* and *D. melanogaster* differ in both the initial establishment of clinal variation and the potential for that variation to be maintained. There is evidence that some of the latitudinal variation that we see in *D. melanogaster* is due to introgression between founding European and African populations (Bergland et al. 2014b; Duchen et al. 2013; Kao et al. 2015). We have no evidence that this occurred in *D. simulans*. Additionally, the potential for maintenance of clinal variation might be diminished in *D. simulans*. As we discuss below, *D. simulans* population structure may be disproportionately affected by processes such as bottlenecks and migration.

**D. simulans overwintering.** *Drosophila* populations experience a contraction as a result of temperate winters (Ives 1970). The decreased genetic diversity observed in high- relative to low-
latitude populations of both *D. melanogaster* (Reinhardt et al. 2014) and *D. simulans* (S3 Fig.) is consistent with stronger bottlenecks at high-latitudes. *D. simulans* seems to be physiologically less winter-adapted than *D. melanogaster* (Hoffmann & Harshman 1999) and *D. simulans* is not observed at high latitudes until later in the year (Behrman et al. 2015; Boulétreau-Merle et al. 2003; Fleury et al. 2004; Schmidt 2011), suggesting a stronger bottleneck for *D. simulans* high-latitude populations than for *D. melanogaster* high-latitude populations. In addition to the decreased genetic variation we observe in the high latitude *D. simulans* Maine population, we find that this population is much more genetically differentiated from the other three populations, a result that could be explained by strong bottlenecks or by complete extirpation and re-colonization. Alternatively, these genetic patterns could be explained by selective sweeps in the Maine population or by effects due to the Maine population existing at the edge of the *D. simulans* range. Although we find evidence of year-to-year genetic continuity of the lower-latitude Pennsylvania population, indicating that there is not complete annual extirpation at the Pennsylvania site, additional sampling is needed to determine if *D. simulans* is able to overwinter at latitudes as high as Maine (45° latitude).

**Migration.** While *D. melanogaster* has a strong, clear pattern of genetic isolation by distance, this is not true of *D. simulans*. A weak pattern of isolation by distance can be indicative of substantial continued gene flow among populations (Endler 1977). Genetic differentiation is particularly low among the three southern *D. simulans* populations (median $F_{st}$ 0.001-0.006, compared with 0.003-0.012 in *D. melanogaster*). The low level of differentiation indicates that there is a stronger effect of migration among these populations. Such a contribution of migration to *D. simulans* population genetic patterns is consistent with the reduced amount of clinal variation in *D. simulans*, as migration can disrupt clinal patterns resulting from demographic processes or local adaptation. A strong effect of migration in *D. simulans* and not in *D. melanogaster* could also contribute to the increased inter-annual variation observed in *D. simulans*, as evidenced by the significant effect of between-year comparison in the isolation by distance regressions (between-year comparisons show increased differentiation) and by the reduced level of clinal consistency (the same variants are not clinal from year to year). The effect of annual migration would be more acute in *D. simulans* than in *D. melanogaster* if *D. simulans* does indeed experience stronger annual bottlenecks, such that

This article is protected by copyright. All rights reserved.
migrants overwhelm the local population. An additional contributor to weaker population structure in *D. simulans* than *D. melanogaster* could be the lack of large cosmopolitan inversions, which could act as a barrier to gene flow among *D. melanogaster* populations (Hoffmann & Weeks 2007; Knibb *et al.* 1981; Mettler *et al.* 1977; Noor *et al.* 2001).

One caveat to each of the analyses that utilize inter-annual data is the reliance of the conclusions on few between-year comparisons. For example, if the Virginia sample from 2010 was aberrant in its genetic composition, such as might occur with human-mediated migration from a distant population, our conclusions of low clinal consistency and the interaction of sampling year with isolation by distance in *D. simulans* might change. Further temporal sampling could bolster these findings.

**Increased differentiation on the *D. simulans* X chromosome**

We find more population genetic differentiation on the X chromosome than on autosomes in *D. simulans*. This pattern is opposite of what we find in *D. melanogaster*, and is particularly pronounced for any comparisons with Maine. Additionally, only in the Maine comparisons are the X chromosome $F_{st}$ values significantly greater than expected when accounting for decreased effective population size (Fig. 7). In contrast, we see a lack of differentiation on the *D. melanogaster* X chromosome, consistent with previous findings of a drop in X chromosome diversity relative to autosomal diversity in non-African populations (Andolfatto 2001). There are multiple evolutionary processes that can affect the relative rates of divergence of the X and the autosomal chromosomes. Examples of a “faster-X” effect are found across various taxa, including in *D. simulans*, and to a lesser extent in *D. melanogaster* (Begun *et al.* 2007). Certain classes of genes, such as those with greater expression in males than females (Baines *et al.* 2008), have shown faster-X patterns in *Drosophila*, as have certain classes of genomic sites, such as non-synonymous sites, UTR, and long introns (in *D. melanogaster* and *D. simulans*; Hu *et al.* 2013). In addition, gene expression differences have accumulated faster between *Drosophila* species on the X than on autosomes (Meisel *et al.* 2012). Further evidence for the contribution of selection to faster-X evolution in *Drosophila* includes the increased selection on tandem duplication on the X chromosome (in *D. simulans*; Rogers *et al.* 2014a) and faster-X
evolution in non-synonymous sites, UTR, and long introns, but not found in synonymous sites and short introns (Hu et al. 2013). The latter study again finds the effect present in both *D. simulans* and *D. melanogaster* but more marked in *D. simulans*.

The increased divergence of the *D. simulans* Maine X-chromosome could be due to Maine suffering more extreme winter population bottlenecks. This is consistent with our findings of decreased genetic diversity and high levels of divergence on the autosomes as well as the X chromosome. Strong drift and divergence of the Maine population could also be driving clinal variation. A demographic explanation for the observed clinal variation is consistent with the weak evidence for selection on clinal variants in *D. simulans*. Another process that could contribute to X chromosome divergence is that of un-equal sex-ratios. Although we do not have sex-ratio data for our populations, multiple sex-distorter systems have been found in other *D. simulans* populations (Bastide et al. 2013).

Conclusions

We have presented genomic evidence that *D. melanogaster* has a greater proportion of latitudinally varying loci than *D. simulans*. In *D. simulans* we observe a weak pattern of isolation by distance, with a significant effect of between-year differentiation, low consistency of clinal SNPs from year to year and less evidence for selection on clinal variants than in *D. melanogaster*. In *D. melanogaster* we observe the opposite patterns- strong isolation by distance, strong clinal consistency, low inter-annual variation and clear evidence for selection acting on clinal variants. We argue that one contributing factor to these differences is the ability of the two species to overwinter in temperate climates, causing differences in bottlenecks and migration. However, despite differences in demography, we do see an enrichment of shared clinal genes between the two species, suggesting that climate-associated selection might act on similar genes and phenotypes in the two taxa.

Acknowledgments

The authors would like to thank Marc Feldman and Jamie Blundell for helpful discussions, and Alison Feder, Nandita Garud, David Enard and Zoe Assaf for comments on the manuscript. This work was supported by the National Institute of Health (www.nih.gov) grants R01 GM097415, R01
GM089926 to DAP, R01 GM100366 to DAP and PS, and F32 GM097837 to AOB, and by the National Science Foundation (www.nsf.gov) grant DEB 0921307 to PS.

References


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


R Core Team (2014) R: A Language and Environment for Statistical Computing.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


**Data Accessibility**

*D. simulans* sequence fastq files and alignment bam files: NCBI SRA: SRP063680
*D. melanogaster* sequence fastq files (Bergland *et al.* 2014a): NCBI SRA: PRJNA256231
Allele frequency data: Dryad doi:10.5061/dryad.3hf2q

*F*<sub>s</sub><sub>st</sub> and latitude matrices for isolation by distance analyses: Dryad doi:10.5061/dryad.3hf2q

GLM results for clinal regressions: Dryad doi:10.5061/dryad.3hf2q

*D. melanogaster* and *D. simulans* shared clinal genes: Dryad doi:10.5061/dryad.3hf2q

**Author Contributions**

Designed research: PS, DAP, AOB and HEM. Contributed samples: PS, KRO and ELB. Performed experiments: HEM. Analyzed the data: HEM. Discussed conclusions: PS, DAP, AOB, KRO, ELB and HEM. Wrote the manuscript: HEM.

**Supporting Information**

**S1 Table.** Populations sampled along the east coast of the North America. *N*: number of individuals sequenced. *C*: average coverage (number of individual chromosomes) across autosomal sites in the final SNP dataset. For pool-seq populations this is the average effective number of chromosomes across sites, calculated as *N*<sub>C</sub>(*N*, *R*, *ε*) = (∑<sup>1</sup><sub>*N*<sup>-1</sub> + <sup>1</sup><sub>*R*<sup>-1</sub> + <sup>0.1</sup><sub>*R*<sup>-1</sub>)<sup>-1</sup> (see Methods). Seq: sequencing type: barcoded individuals (I) or pooled (P).

**S2 Table.** Proportion of clinal SNPs. Proportion of clinal SNPs (logistic regression *P* < 0.01). Values in parentheses are 2 standard error.

**S3 Table.** Isolation by distance linear model. This includes the effect of comparison with ME and between-year comparisons. *P* < 0.001

**S4 Table.** Annotation of *D. simulans* genome and SNPs by genic category.

This article is protected by copyright. All rights reserved.
S1 Figure. Levels of differentiation between barcoded samples compared with differentiation between pooled samples. The horizontal line is the expected proportion of SNPs for each $P$-value bin. Top row: barcoded comparisons. Bottom row: pooled comparisons, 1) without an additional pooled-error variance component ($\epsilon = 0$), 2) pooled error model fitting, with SSE (sum of square residuals from the null barcoded model) versus $\epsilon$ for each model (model 1 $\epsilon$ is scaled by a factor of 50), 3) pooled comparison corrected.

S2 Figure. Sample size ($N_C$) for clinal regressions. A) Average $N_C$ by chromosome. Error bars are 2 standard error (all < 0.03). B) Distribution of total $N_C$ (sum over populations).

S3 Figure. Clinal consistency across years. This analysis compares the pattern of clinal variation seen in samples from one year to the patterns of allele frequency in two samples in a separate year. For *D. simulans* this is a comparison of a regression across FL.2011, PA2011, ME.2011 (Dsim 2011: blue solid line) with the allele frequency change between VA.2010/PA.2010. The *D. melanogaster* regression is across GA.2008, PA.2009, and ME.2009 (Dmel 2009/2010: red line) and compares with allele frequency change in a third year (FL.2010/PA.2010), and is therefore less conservative. We perform a similarly less-conservative analysis with *D. simulans*, with the initial regression using populations from two years (VA.2010, PA.2010, ME.2011; blue dashed line).

S4 Figure. Distribution of *D. simulans* A) coding sequence (CDS) length and B) SNP density (number of SNPs divided by the gene length), for the observed clinal genes ($Q < 0.2$; pink) and the matched control genes (blue). The overlap of these distributions appears purple.

S5 Figure. Relationship between population structure (mean $F_{st}$) and clinal effect size ($\beta$). The shaded regions show 2 standard error of the mean.

S6 Figure. Enrichment of clinal X chromosome SNPs in each functional genic class. A) *D. melanogaster*. B) *D. simulans*. Clinal SNPs are identified as significant at $P < 0.01$ and $\beta > 90$th percentile. The control is 100 bootstrapped sets, matched for chromosome, recombination rate, minor allele frequency, coverage, and inversion status (*D. melanogaster* only). Error bars are one standard deviation. *$P <= 0.05$; **$P <= 0.01$.

S7 Figure. *D. simulans* autosomal diversity across populations. Heterozygosity (solid), $\pi$ (dashed) and $\theta_S$ (dotted) measures are significantly different among all samples, with the exception of two within-population samples (PA7.2011 vs PA9.2011; TukeyHSP $P < 0.00001$).

S8 Figure. $F_{st}$ between the two PA 2010 samples and each of the other population samples in *D. simulans*. We find less differentiation between the PA samples across years than in any of the between-population comparisons.
Figure 1: Clinal genetic variation with latitude. A, B): Allele frequency trajectories for clinal SNPs ($P < 0.01$, sample of 100). Allele frequencies are polarized such that FL < ME. C) Distribution of populations used to assess clinal variation. D) $P$-value distributions from logistic regressions of allele frequency with latitude (bins of 0.01). Error bars are two standard error (not visible).
Figure 2: The distribution of clinal SNPs across the genome. The mean proportion of clinal SNPs ($P < 0.01$) per 1Mb window is plotted across the *D. melanogaster* genome. Shaded areas represent the *D. melanogaster* major inversions. Black along the x-axis represents low recombination rate regions (< 0.5 cM/Mb/female meiosis, 100kb bins). The proportion of SNPs clinal on each chromosome is listed in the legends.
Figure 3: Consistency of clinal variation across years. The proportion of SNPs for which the clinal regression coefficient from one year predicts the directionality in a second year is plotted for sets of clinal SNPs of increasing clinal stringency (decreasing Q-value). Error bars are two standard error.

Figure 4: Enrichment of clinal autosomal SNPs (Q ≤ 0.2, down sampled to 25134 SNPs) in each functional genic class. Plotted is the log of the odds ratio of the proportion of each genic class in the set of clinal SNPs compared with 100 matched controls. Error bars are one standard deviation. Bootstrap P-value *P ≤ 0.05; **P ≤ 0.01.
Figure 5: Percent overlap of *D. simulans* clinal genes with *D. melanogaster* clinal genes ($Q < 0.2$), over increasing stringency of *D. simulans* clinal regression.

Figure 6: Isolation by distance. Between population genetic differentiation (median $F_{st}$) is plotted against geographic distance (degrees latitude). A) *D. melanogaster*. B) *D. simulans*. For *D. simulans*, regression lines are plotted separately for population comparisons without ME within a year, without ME between years, and with ME within a year, reflecting the significant effect of distance, ME vs non-ME comparison, and within- vs between-year comparison in the regression model. ME: $F_{st}$ between one non-ME population and ME; NonME: $F_{st}$ between two non-ME populations; b/t: $F_{st}$ between two samples taken between years; w/i: $F_{st}$ between two samples taken within a year.
Figure 7: Expected versus observed X chromosome $F_{st}$ in *D. simulans*. An expectation of X chromosome median $F_{st}$ is calculated from the autosomal $F_{st}$ values. Within-population $F_{st}$ measures are from the three PA samples taken over the course of 2011. Between-population $F_{st}$ measures are divided up into comparisons that include ME and those that do not include ME. Error bars are 2 standard error.