Genomic patterns of geographic differentiation in *Drosophila simulans*

Alisa Sedghifar [*†], Perot Saelao[*] and David J. Begun [*†]

Running Head: *D. simulans* geographic differentiation

Key Words: Adaptation, Geographic Variation, Population Genomics

Corresponding Author:

Alisa Sedghifar

Department of Evolution and Ecology

One Shields Ave

Davis CA 9516

Tel (530)754-6359

Fax (530)752-1272

sedghifar@ucdavis.edu

[*]Department of Evolution and Ecology, University of California, Davis, Davis, CA, 95616

[†]Center for Population Biology, University of California, Davis, Davis, CA 95616

Abstract

Geographic patterns of genetic differentiation have long been used to understand population history, and learn about the biological mechanisms of adaptation. Here, we present an examination of genomic patterns of differentiation between northern and southern populations of Australian and North American *Drosophila simulans*, with an emphasis on characterizing signals of parallel differentiation. We report on the genomic scale of differentiation, and functional enrichment of outlier SNPs, consistent with differential selection acting on coding sequence variation. While overall, signals of shared differentiation are modest, we find the strongest support for parallel differentiation in genomic regions that are associated with regulation. Between-species comparisons to *D. melanogaster* yield potential candidate genes involved in local adaptation in both species, providing insight into common selective pressures and responses. In contrast to *D. melanogaster*, in *D. simulans* we observe patterns of variation that are inconsistent with a model of temperate adaptation out of a tropical ancestral range, highlighting potential differences in demographic and colonization histories of this cosmopolitan species pair.

Introduction

The geographic distribution of genetic or phenotypic variation can provide valuable insight into the process of adaptation. For example, consistent patterns of genetic variation across space have long been interpreted as evidence for local adaptation due to spatially varying selection (Endler 1977). This is well illustrated in populations of *Drosophila melanogaster*, a model system showing consistent phenotypic and molecular clines across environmental gradients (Hoffmann and Weeks 2007). Among these, the association of latitude with variation in ecologically relevant traits such as heat knockdown resistance, chill coma recovery and diapause incidence (Hoffmann *et al.* 2002; Schmidt and Paaby 2008), provide strong support for local adaptation to climate.

Despite efforts to understand the potential adaptive nature of molecular variation in populations of *Drosophila*, there remains a disconnect between our understanding of allele frequency clines and phenotypic clines, of which the latter is more easily and intuitively interpretable. For the vast majority of clinal molecular polymorphisms in *Drosophila*, the mechanisms underlying their maintenance is poorly understood. Nevertheless, because gene flow in *Drosophila* is thought to be high, strong differentiation can be often argued as evidence of local adaptation. Moreover, because the physical scale of linkage disequilibrium (LD) in *Drosophlia* is often smaller than the size of genes (Langley *et al.* 2012), differentiation between populations are generally associated with hypotheses regarding individual genes as targets of selection.

Observations of parallel patterns of differentiation furthers the argument for an adaptive basis to differentiation and, in general, comparisons of patterns of variation across independent, replicate geographic transects may contribute to an understanding of the contribution of a variant to fitness under differing environmental conditions. This approach has been utilized often, in *Drosophila* and other systems (Turner *et al.* 2010; Paaby *et al.* 2010;

Anderson and Oakeshott 1984; Colosimo *et al.* 2005; Hohenlohe *et al.* 2010). Parallel patterns not only provide compelling evidence that a particular trait or genetic variant plays a role in adaptation, but also provides insight into the repeatability of adaptation. Even the absence of parallel differentiation contributes to our understanding of the repeatability of adaptive differentiation and the different mechanisms and constraints that influence both phenotypic and molecular evolution.

While there has been a great focus on geographic variation in *D. melanogaster*, investigations of other *Drosophila* have demonstrated the presence of geographic patterns in a number of other species in the genus (e.g. Sturtevant and Dobzhansky 1936; Dobzhansky 1948, 1947; Price *et al.* 2014; Huey 2000; Hallas *et al.* 2002; Arthur *et al.* 2008; Tyukmaeva *et al.* 2011), revealing cross-species convergence in clines for traits such as wing size and cold tolerance. Among these species, *Drosophila simulans* presents an especially attractive system for further study of geographic genetic variation, as it is very recently diverged (Tamura *et al.* 2004, 5mya) from the well studied *D. melaongaster*. In addition to this shared evolutionary history, similarities in recent colonization histories and shared cosmopolitan distributions mean that it may be reasonable to expect that the two species have experienced recent parallel evolution and indeed, *D. melanogaster*/*D. simulans* pair has for a while been a popular focus for comparative population genetics (e.g. Zhao *et al.* 2015; Capy and Gibert 2004b; Parsons 1975a; Singh *et al.* 1987).

Although the two species share recent common ancestry and have broadly similar ecologies, there are several important differences between these species. For example, *D. melanogaster* appears to be more tolerant of high ethanol concentrations, and the two species differ in their seasonal abundances and thermal tolerances (Parsons 1975b, 1977). Moreover, it is known that the geographic centers of diversity vary, with *D. melanogaster* being most diverse in southern-central Africa (Pool *et al.* 2012) and *D. simulans* in Madagascar (Dean and Ballard 2004). Such contrasts emphasize the possibility that the two species are historically

adapted to different environments and have experienced vastly different colonization histories. Potential differences in population histories are further reflected in the contrasting patterns of genetic variation outside of Africa (Begun and Whitley 2000; Andolfatto 2001; Capy and Gibert 2004a, e.g.). Broadly speaking, outside of Africa, *D. simulans* exhibits higher within-population diversity and *D. melanogaster* higher levels of between-population diversity (Singh 1989). Notably, while strong clines are abundant in *D. melanogaster* and have been the focus of extensive investigation, there seems to be less clinal variation in *D. simulans*. For example, Arthur *et al.* (2008) showed that there are no apparent clines for cold tolerance or heat shock in Australian populations of *D. simulans*, despite these traits being strongly clinal in Australian *D. melanogaster* (Hoffmann *et al.* 2002), and (Gibert *et al.* 2004) reported that even when present, clines in *D. simulans* were weak. This could potentially be interpreted as a relative lack of local adaptation in *D. simulans*. More recently (Machado *et al.* 2015) found genomic evidence for clinal variation in *D. simulans* and verified that it is less pronounced than in *D. melanogaster*.

While differences in the strength of clinal variation in *D. simulans* compared to *D. melanogaster* may suggest that the two species are responding to their local environments in different ways, the findings of Machado *et al.* (2015), differentiation in patterns of gene expression (Zhao *et al.* 2015), and phenotypic clines in traits such as body size, indicate that *D. simulans* is likely evolving, in at least some capacity, to spatially varying selection. This is further supported by the observation that there is significant overlap in differentiated genes between *D. simulans* and *D. melanogaster* (Machado *et al.* 2015; Zhao *et al.* 2015). This between-species parallel differentiation in both gene expression (Zhao *et al.* 2015) and allele frequency (Machado *et al.* 2015) raise the additional possibility that weaker phenotypic clines generally reported for *D. simulans* may not accurately reflect the influence of spatially varying selection on this species.

To further investigate patterns of geographic differentiation in *D. simulans* and similarities and differences with respect to *D. melanogaster*, we re-sequenced four *D. simulans*

populations – one northern and one southern – in both North America and Australia. We then employed an F_{ST} outlier approach to identify putative targets of spatially varying selection. The advantages of this two-continent design are twofold: First, we are able to address our direct objective of assessing the degree of parallelism in local adaptation between the two continents and compare them to analogous patterns in *D. melanogaster*. Second, focusing on SNPs that are strongly differentiated on two continents will, to some degree, mitigate potential false discoveries that may arise as a consequence of sampling error, fine-scale local environmental adaptation, or demography. Such comparative population genomic approaches may inform our understanding of parallelism at various levels, from the nucleotide level, to gene annotations, or pathways and may also provide useful information regarding constraints, repeatability and diversity of mechanisms of adaptation in these two species.

Similar genome-wide studies of differentiation in comparable populations of *D. melanogaster* (Turner *et al.* 2008; Kolaczkowski *et al.* 2011; Fabian *et al.* 2012; Reinhardt *et al.* 2014) have detected signals of parallel differentiation, and in particular a strong association of large inversion with differentiation. The prevalence of inversion frequency clines in *D. melanogaster* is thought to reflect some response to spatially varying selection, but their adaptive significance remains unclear. This is noteworthy because inversion polymorphisms are virtually absent in *D. simulans* (Ashburner and Lemeunier 1976) and it remains unknown what the implications are for adaptive differentiation in *D. simulans*. In addition to learning more about general patters of variation and potential mechanisms of adaptation, an assessment here of genomic patterns of geographic variation in *D. simulans* presents the opportunity to begin to gain further insight into general patterns of geographic variation in the two species, as well as common responses to challenges posed by novel environments.

Materials and Methods

Sampling and sequencing

Four populations are represented in this study: Northern Australia, Southern Australia, Northern United States (US) and Southern US (Table 1). The two US subpopulation libraries were generated from pools of single daughters of females sampled directly from the field in 2011. The two Australian subpopulations were generated by pooling a single female from isofemale lines established in 2004. Libraries were prepared according to the NEBNext DNA Library Prep Master Mix Set for Illumina protocol and were sequenced on the Illumina GAII platform at two or three libraries per lane. Reads were trimmed using SolexaQA (Cox *et al.* 2010) with a quality score threshold of 28 and any resulting reads shorter than 36bp were discarded. Both subpopulations from a given continent were sequenced in the same lane, which eliminates concerns of lane effects on within-continent differentiation.

Alignment

Reads were aligned to the *D. simulans w501* assembly from Hu *et al.* (2013) and *Wolbachia pipientis wRi* strain using BWA (Li and Durbin 2009). Reads with mapping quality under 30 were discarded and optical duplicates and reads mapping to multiple regions were removed. Initially, sites with coverage less than 15 and greater than 2 standard deviations from the mean were removed from the analysis as these sites are respectively prone to inflated F_{ST} from sampling error and potential duplications or paralogy. Because of the substantially smaller sample size from the Australian populations, estimated allele frequencies have greater variance compared to the North American population ($Var(\hat{p}) = \frac{m+n-1}{mn}p(1-p)$), where n, m and p are sample size, coverage and allele frequency, respectively (Futschik and Schlötterer 2010)). To reduce noise and potential biases from smaller sample sizes, minimum cutoffs in Queensland and Tasmania were increased to 20 and 29 respectively for outlier-based analyses at the SNP level.

Repetitive regions, defined by Hu *et al.* (2013) were removed after alignment. In order to minimize the effect of sequencing error on population genetic analyses, on either continent only variants supported by two or more reads, and segregating at a frequency greater than 0.05 were considered. Since they are likely to be regions of low recombination, regions of low heterozygosity on the proximal and distal ends of each chromosome arm were removed. These regions were determined by defining uninterrupted sequences of 100kb windows (sliding by 50kb) on the ends of the chromosome arms that were below half the chromosomal average for either mean π or number of segregating sites.

Outlier SNPs and regions

$\hat{\pi}$ and $\hat{F_{ST}}$ were calculated at each position in the genome using estimators described in Kolaczkowski *et al.* (2011). Within each each continent, SNPs in the upper tail of F_{ST} were considered to be highly differentiated. Where indicated, further refinement of candidate loci took place by only considering SNPs that were outliers on both continents, and were differentiated in the same direction with respect to latitude, since these are more likely to be under parallel differential selection. Because sites with lower coverage will have greater variance in F_{ST} due to sampling error, F_{ST} outliers may be enriched for sites with lower coverage. To address this effect of coverage, polymorphic sites were binned by the minimum coverage of the two populations in a continent. These sites were then ranked by F_{ST} within each bin and SNPs were required to be outliers with respect to both genome-wide F_{ST} and coverage-based rank to be classified as a strongly differentiated site. F_{ST} and rank were highly correlated by Spearman's rank-order correlation ($\rho = 0.98$ on both continents). Statistical significance for enrichment was calculated using Fisher's exact test (FET).

Derived alleles:

Gene alignments for *D. melanogaster*, *D. simulans* and *D. yakuba* from Hu *et al.* (2013) were used to determine the ancestral allele at a polymorphic site. If either allele at a bi-allelic site matched the *D. melanogaster* and *D. yakuba* sequence, it was considered to be the ancestral state, and the other allele was considered to be derived. For a given SNP within a continent, the allele present at higher frequency in the higher-latitude population was considered the 'temperate-adapted' allele. As F_{ST} threshold was increased, we compared the number of temperate-adapted alleles that were ancestral to the number that were derived. We expect, as a null, to observe an unchanging proportion of temperate-adapted derived alleles across F_{ST} thresholds, and statistical tests for over/under-representation of temperate-adapted alleles for a given F_{ST} threshold were based on a binomial expectation, with rate given by the proportion of temperate adapted alleles across the whole genome.

All analyses of functional regions used annotations accompanying Hu *et al.* (2013). These annotations were augmented using the assembled transcriptome from Rogers *et al.* (2014). Transcripts from Rogers *et al.* (2014) were matched to *D. melanogaster* annotations by aligning predicted translations to *D. melanogaster* translations in FlyBase release 5.9, using BLAST under default parameters. The top BLAST hits were retained only if protein sequences aligned at the first residue, and the final residue of the *D. simulans* protein aligned to within 5 residues of the *D. melanogaster* stop codon. Some analyses focus on different "annotation classes"; upstream (region 500bp upstream of transcription start site), exon, 3'UTR, CDS, intron, 5'UTR, intergenic (unannotated).

Circular Permutation

Because the positions of F_{ST} outliers appear to be autocorrelated throughout the genome, generating a null expectation for non-SNP-based analyses (such as the expected number of shared outlier genes based on random sampling of SNPs or genes) can be a challenge.

To address this issue, for analyses involving annotations, we generated a null distribution of enrichments by iteratively shifting the relative position of each SNP along a concatenation of all chromosomes by one randomly selected number; the positions at which SNPs occur remain unchanged for all permutations. For each iteration, a new random number is selected and a list of outlier annotations are generated. This approach provides an alternative to explicitly defining independent differentiated regions as the local autocorrelation of F_{ST} is conserved in each iteration and is similar to the strategy used in Nordborg *et al.* (2005).

Gene Ontologies

In order to account for any bias in overrepresented Gene Ontology (GO) categories due to gene size, we permute in a circular fashion the F_{ST} value of each genic site by shifting the relative position of each base by a randomly chosen number and re-calculating the GO enrichment p-value under a hypergeometric model. By iterating this process, we obtained a distribution of enrichment p-values for each GO category, which was then used to obtain a quantile value for the p-value that was observed in the non-permuted data. This preserves the autocorrelation in distribution of F_{ST}. The R Bioconductor (Gentleman *et al.* 2004) and org.Dm.eg.db (Carlson *et al.* 20) packages were used.

Data availability

Genomic data is available as raw sequence reads from the NCBI SRA.

Results

The four study populations were sequenced to mean coverages ranging from 43 to 70 (see Table 1). Mean π_s and F_{ST} for each chromosome are reported in Table S1, and patterns along chromosomes are shown in Figures S1 and S2. Heterozygosity does not differ significantly across autosomal arms in the North American populations (Kruskal-Wallis using

100kb windows; $p = 0.11$ (FL),0.21(RI)). In Australia, however, there is significant heterogeneity among the autosomes in both populations ($p = 3.3e{-}6$ (QLD),$4.5e{-}3$(TAS)). Genome wide, higher latitude populations have higher mean heterozygosity than lower latitude populations, based both on π_s and mean $\hat{\pi}$ in 100kb windows and this pattern is consistent across the genome, (Wilcoxon signed rank on mean π in 100kb windows, $p < 10e{-}16$ for both continents). This contrasts with observations in *D. melanogaster* of higher heterozygosity at lower latitudes (Kolaczkowski *et al.* 2011; Fabian *et al.* 2012; Reinhardt *et al.* 2014). We note that of the four populations, FL has the lowest heterozygosity genome wide, reflecting perhaps more severe drift than the other three populations. This is in contrast to the findings of Machado *et al.* (2015), who did not observe substantially lower levels of heterozygosity in populations from FL compared to other populations sampled in North America.

Mean F_{ST} is heterogeneous across autosomal arms for both continents ($p = 0.006$ and $p = 0.002$ respectively), although the rank order of chromosome arms differs for the two continents. In particular, mean F_{ST} is highest on the X-chromosome in North America (as found by (Machado *et al.* 2015)), but highest on 3R in Australia. Furthermore, there appears to be little shared differentiation on a broad physical scale of 100kb windows; Figure S2. Compared to populations of *D. melanogaster* (Fabian *et al.* 2012; Reinhardt *et al.* 2014) sampled over a similar spatial scale, *D. simulans* appear to exhibit less genome-wide differentiation.

Scale of differentiation

Estimated F_{ST} is expected to be correlated across closely linked SNPs, at least in part because of the effects of linked selection. To summarize the physical scale of differentiation that arises from this non-independence, we measured the mean F_{ST} as a function of distance from SNPs in the top 1% tail of F_{ST}. On a scale measured by 1kb non-overlapping windows,

mean F_{ST} decays to background genomic levels within 60kb in North America (Fig 1a) and on a scale of 100kb in Australia (Fig S3a). Although lack of detailed information on recombination variation in *D. simulans* precludes a formal comparison of recombination rate and differentiation, the observed physical scale scales of genomic variation in recombination rate in *D. melanogaster* (Comeron *et al.* 2012) suggest that the relatively large scale of correlated F_{ST} could be influenced by genome-wide heterogeneity in recombination rate.

On a scale measured in 10bp non-overlapping windows, F_{ST} decays rapidly within 100bp (Fig 1a). This smaller scale of decay is reminiscent of the scale of linkage disequilibrium observed in *D. melanogester* (Langley *et al.* 2012), and is consistent with adaptation from standing variation. Whatever the driving factors behind these heterogenous scales of decay, it is clear that strongly differentiated SNPs do not occur independently throughout the genome.

Relative frequencies of derived alleles

Under a model of a tropical *D. simulans* ancestral range, adaptation to temperate climates in recently established populations should generate a pattern, on average, of derived variants segregating at higher frequency at lower latitudes at strongly differentiated SNPs (e.g. Sezgin *et al.* 2004). We therefore identified the derived allele for each SNP, using available alignments to *D. melanogaster* and *D. yakuba* (see Methods). On average, at the most differentiated SNPs the derived allele was found to be segregating at a higher frequency in the tropical population compared to the temperate population. This pattern is present, and significant (with $p < 10^{-12}$ at the 99% cutoff, see Methods) on both continents, but is more pronounced in North America (Fig. 3, Fig. S9). Differences in the derived allele frequency for FL and RI populations for F_{ST} outlier SNPs reflect this observation, with a larger skew towards high-frequency derived alleles in FL for this subset of the genome.

To further investigate this pattern, we compared heterozygosities surrounding outlier SNPs between high and low latitude populations. Since selection reduces local variation within the genome (Maynard Smith and Haigh 1974; Charlesworth *et al.* 1993), we expect small regions that have large differences in heterozygosity between the two populations and also contain an outlier SNP to be the most likely to have experienced recent differential selection. We therefore measured the differences in heterozygosity in 100bp non-overlapping windows between the two populations in a given continent and identified windows that fell in the $\pm 2.5\%$ tail on either side of the distribution of population differences in π (Fig. S10). We then identified windows containing an outlier SNP (1%) and compared the number of such windows with smaller $\hat{\pi}$ in the more tropical population to the number with smaller $\hat{\pi}$ in the temperate population. As shown in Fig. S10, there are more windows that support recent adaptation in the tropical population than in the temperate population (302 compared to 163 in North America, chi-squared test: $p = 1.2e^{-7}$, given an expectation scaled by the relative portion of low heterozygosity windows).

Shared differentiation at SNPs

In the absence of shared differentiation between Australia and the US, the proportion of shared outlier SNPs is expected to be roughly equal to the product of the proportions of SNPs defined as outliers on each continent (for example, within the set of all shared SNPs, we expect 1% of SNPs to be found in the top 10% of F_{ST} on both continents). Because it is unknown *a priori* what an appropriate F_{ST} cutoff is, we evaluated this enrichment for a range of F_{ST} cutoffs (Fig 2). These results were then used to inform suitable F_{ST} outlier cutoffs of 5% in North America and 15% in Australia for downstream analyses.

Enrichment for shared outlier SNPs increases as cutoffs for F_{ST} become more extreme, providing evidence for shared differentiation on a genome-wide scale (Fig 2). The statistical significance of this enrichment under the FET, however, is modest (Figure S5). The pattern

of increased enrichment with F_{ST} cutoff persists at the scale of 100bp and 1kb windows (Fig S4), consistent with the scales of differentiation reported above.

In addition to an enriched sharing of outliers, if a variant is subject to latitudinally varying selection, then we expect the difference in allele frequency between low and high-latitude populations to have the same sign on the two continents (referred to here as same-direction SNPs). In the absence of such parallel differentiation, then the expectation is to observe approximately half of the SNPs to be same-direction, independent of F_{ST}. We tested for parallelism among shared outliers under a binomial model with probability 0.5 of a shared outlier SNP being same-direction and did not observe a significant signal of same-directionality , and instead observed an enrichment of opposite direction SNPs across many outlier cutoffs (Figures S6 and S7).

To further investigate these patterns of enrichment and parallelism, we focused our analysis on different annotation categories (see Methods for details). Within each subset of the genome corresponding to a category, we conducted the same enrichment analyses. The results, shown in Figure 2, indicate a potential signal of enriched parallel differentiation within the 3'UTR regions of the genome, and are consistent with the observed expression-level differentiation found by Zhao *et al.* (2015). Consistent patterns of enriched parallel differentiation were not observed for other regions of the genome, but this could be partly due to limited power.

We now focus on same-direction SNPs that are found in both the top 5% tail in North America and the top 15% tail in Australia, referring to this subset of SNPs as same-direction shared outliers. This subset of SNPs was tested for associations with different annotations classes. As above, null distributions of enrichment values were generated by circular permutation to assess the significance of observed enrichments. We find a significant enrichment of

same-direction shared outliers SNPs in the 3'UTR regions, consistent with the results of the parallel enrichment above (Figure S8), and similar to patterns reported by Kolaczkowski *et al.* (2011) and Reinhardt *et al.* (2014). Although significant enrichment is not observed for any other class, it is again possible that this can be attributed to insufficient power, especially considering that the number of shared outlier SNPs in some annotation classes can be small.

Genes containing same-direction shared outlier SNPs in the 3'UTR, 5'UTR, upstream region, or as a non-synonymous variant are listed in File S1. While many of these genes only have one or two shared same-direction outliers, there are several genes that stand out for having 4 or more differentiated SNPs within relatively short genomic regions. Genes containing 4 or more shared SNPs in the UTR and upstream regions, which may be associated with regulation of expression, are *Dopamine transporter* (*DAT*) and *CG1527*. Dopamine is a neurotransmitter that has many biological roles, of which one is the circadian rhythm (Hirsh *et al.* 2010), a clinally varying trait in *D. melanogaster* (Svetec *et al.* 2015). A mutation in *D. melanogaster DAT* is associated with decreased sleep duration and arousal threshold (Kume *et al.* 2005; Kume 2006), as well as metabolic rate, and thermal preference and tolerance Ueno *et al.* (2012). Little is known about the function of *CG1527*. Another gene, *Lava lamp* (*lva*), containing 4 same-direction non-synonymous SNPs, is a golgin protein involved in transmembrane secretion during development Sisson *et al.* (2000); Papoulas *et al.* (2005).

Differentiated Genes

Convergence in adaptation is also possible through selection on different variants within the same gene (Rosenblum *et al.* 2010). Given the autocorrelation of the position of outlier SNPs, and because larger genes are by chance likely to contain an outlier SNP on both continents, we permuted (10000 iterations) the positions among genic SNPs to assess the extent of enrichment of shared genes (see Methods). In this instance, new outlier gene sets were

generated for North America, and the proportion of outlier genes shared with Australia was used as a measure of sharing. As before, the significance of the enrichment was evaluated by comparing the proportion of shared genes to the distribution generated by the permutations. We tested a range of cutoffs, but found no significant enrichment of shared genes (when requiring $p < 0.01$). The strongest signal of enrichment is present at the 1% F_{ST} cutoffs on both continents, with $p = 0.05$. We compared this subset of genes to genes identified by Reinhardt *et al.* (2014) as differentiated on both continents in *D. melanogaster* and by Zhao *et al.* (2015) as differentially expressed between Maine and Panama population of *D. simulans* (File S2). These genes, which are strongly differentiated on two continents in two species, may be among the most promising candidates for further study on potential targets of spatially varying selection in cosmopolitan *Drosophlia*.

Between-species parallelism in genes associated with insecticide resistance

Two genes, *Cyp6g1* and *Ace*, known to be involved in resistance to insecticides in *D. melanogaster*, appear to have undergone a selective sweep in one or more *D. simulans* populations (Fig 4). The sweep in *Cyp6g1* recapitulates the result of Schlenke and Begun (2004) and appears to be global, although the Tasmanian population appears to retain some diversity in this region. In contrast, the sweep surrounding *Ace* is only present in the QLD population. Kolaczkowski *et al.* (2011) found an overlapping region containing a putative copy number variant segregating at higher frequency in QLD populations of *D. melanogaster* pointing to the possibility that the two species are responding in a parallel manner to insecticides through different molecular mechanisms. This gene was also identified by Fabian *et al.* (2012) as a highly differentiated gene in North America.

In *D. melanogaster* six *Ace* amino-acid polymorphisms have been implicated in insecticide resistance(Fournier *et al.* 1992; Mutero *et al.* 1994). We looked for amino acid poly-

morphisms in *D. simulans-Ace* that were fixed in QLD but at intermediate frequency in TAS and found three candidate residues that are identical to those found in *D. melanogaster* (Table S2). Moreover, these are due to the same DNA polymorphisms, presumably because there is only one substitution in the respective ancestral codons that can produce these specific amino acid polymorphisms. Such specificity in convergence has been observed in the *Resistance to dieldrin (Rdl)* gene where a replacement of *Ala*302 associated with cyclodiene resistance in *D. melanogaster* has been identified in multiple insect species (FFRENCH-CONSTANT *et al.* 2000). There is also evidence to suggest differentiation in expression of *Ace* between high and low latitude populations of *D. simulans*; the 3'UTR region of the gene contains a same-direction shared SNP (File S1) and has been identified by (ZHAO *et al.* 2015) as a differentially expressed gene between Maine and Panama populations of *D. simulans* and *D. melanogaster*.

REINHARDT *et al.* (2014) reported a large continent-specific differentiated region surrounding *Cyp6g1* in Australian populations of *D. melanogaster*. A nearby region in *D. simulans* shows continent specific differentiation in Australia (Fig. S2), but is located adjacent downstream of *Cyp6g1*. Because of this, it is unclear whether or not this is an example of parallel differentiation between species, and if it is, it raises the possibility that the common target is not *Cyp6g1*.

Gene Ontology

For each continent, we performed independent Gene Ontology (GO) enrichment analyses on the subsets of genes containing a SNP in the 1% tail and 4% tails in Australia and North America, respectively. To account for any bias in enrichment of GOs introduced by gene size, we permuted the F_{ST} values of SNPs present in annotated genes (see Methods). GOs that had a p-value (hypergeometric test) of less than 0.005 and a quantile value, based on circular permutation, of less than 0.01 are listed in File S3.

Discussion

General patterns of differentiation

Here we have presented a genome-wide analysis of geographic variation in *D. simulans*, specifically aiming to compare populations from high and low latitudes. Our results, consistent with previous studies of spatial variation in this species (Choudhary and Singh 1987; Singh 1989; Long and Singh 1992; Machado *et al.* 2015), indicate that on a genomic scale F_{ST} is lower in *D. simulans* than in *D. melanogaster*, even when the effects of inversions are removed (Fabian *et al.* 2012; Kolaczkowski *et al.* 2011; Reinhardt *et al.* 2014). The X-chromosome is an exception to this, with comparable mean F_{ST} to North American populations of *D. melanogaster* (Reinhardt *et al.* 2014). It should be noted, however, that differences in sampling, sequence quality and criteria for retaining sites for analysis differ between studies, casting some uncertainty on the interpretation of these comparisons.

Average F_{ST} was approximately two-fold higher between the North American *D. simulans* populations compared to the Australian populations. This genome-wide difference could be explained by a more recent colonization of Australia, lower rates of gene flow in North America, or demographic processes (such as a bottleneck) in North America. Pairwise comparisons of F_{ST} indicate that the FL population is the most differentiated compared to the others (Fig. S11). Given that the FL population also has the lowest heterozygosity of the four populations, it is possible that some recent demographic history of the FL population may be contributing to the overall higher levels of differentiation observed in the North American samples, but technical effects related to library construction or sequencing cannot be ruled out. Our findings here are in contrast to those made by Machado *et al.* (2015), who observed that their northernmost population sampled in Maine seemed to be an outlier relative to the other populations. Combined, these results support the role of local perturbation in shaping geographic patterns of variation in *D. simulans*.

Curiously, between North American populations of *D. simulans*, the X-chromosome is most differentiated, but the converse is true in *D. melanogaster*, where the X is the least differentiated arm (Reinhardt *et al.* 2014; Kolaczkowski *et al.* 2011; Machado *et al.* 2015). This is consistent with the results of Machado *et al.* (2015), and is perhaps a continent-specific effect, as the X-chromosome is not the most differentiated arm in Australia. While it seems likely that such a chromosome-wide effect could be due to demography, such as sex-biased dispersal, or extreme bottlenecks, these hypotheses cannot be addressed with the currently available data.

Within chromosomes, sites with high F_{ST} are not uniformly distributed throughout the genome. Mean F_{ST} around outliers decays to approximately 5% more than background levels within 200bp (Fig: 1b), which is roughly consistent with the scale of LD in *D. melanogater* (and the therefore assumed scale of LD in *D. simulans*). However, mean F_{ST} decays completely to background levels on a much larger scale of 100kb. This is consistent with a large scale heterogeneity of F_{ST} across the genome, perhaps associated with recombination rate variation (Begun *et al.* 2007; Comeron *et al.* 2012). This pattern could also reflect reduced heterozygosity as a result of recent adaptation in one population, as this would reduce heterozygosity in (potentially large) genomic regions influenced by selection, resulting in elevated local F_{ST}. It is therefore possible that recent population-specific sweeps are contributing to larger-scale patterns of differentiation. This is distinct from differentiation due to selection against migrants although, if migration is sufficiently high in *D. simulans*, an argument could be made for attributing most strong differentiation to differential selection. We did not observe Megabase-scale regions of elevated F_{ST} such as those present in *D. melanogaster*, perhaps due to the lack of large-inversion polymorphisms in *D. simulans*.

Patterns of parallel differentiation

Parallelism and convergence can occur at many functional levels ranging from phenotype to nucleotide (Rosenblum *et al.* 2010; Manceau 2010). Here, we examined potential patterns of parallelism in SNPs, genes and gene ontologies.

The enrichment of shared SNPs at extreme F_{ST} cutoffs is consistent with the two continents sharing some mechanisms of local adaptation to latitudinally varying selective pressures. While it is difficult to assess how much of the excess sharing of outliers is driven by high F_{ST} caused by linkage to true targets of selection, which is likely to be driving some autocorrelation in outlier SNP positions, the decay we see on a relatively small scale ($\sim$100bp) provides support for some local adaptation from standing variation. The significant number of shared SNPs along similar outlier classes (along the diagonal of S5) indicates that, beyond the most differentiated sites, there is substantial correlation in the patterns of differentiation across the genome.

Among different genomic regions the 3'UTR regions have the strongest patterns of shared differentiation, consistent with differential selection acting on regulatory variation. This result is consistent with evidence from (Zhao *et al.* 2015) of adaptive gene expression differentiation between Maine and Panama populations of *D. simulans*, and with enrichment for clinal SNPs found by Machado *et al.* (2015). Unlike Machado *et al.* (2015), we detected relatively relatively little evidence for patterns of parallel differentiation within other regions of the genome, but this does not necessarily indicate that structural variation, or variation in other genomic regions, does not play an important role in adaptation, as these analyses reflect genome-wide patterns and could also be affected by a lack of power. Additionally, we note that the set of same-direction shared outliers found in 3'UTR comprises a very small subset of the genome, and as such these enrichment results should be treated with caution.

Our investigation of overlap between sets of outliers genes on the two continents detected an enrichment for the number of shared genes but, it is difficult to compare the strengths of sharing at the gene and SNP levels, in part because the physical scale of differentiation makes it challenging to disentangle the effects of selection on the same variant from that on different variants within the same gene. Enrichment of shared genes did not translate into a strong pattern of sharing at higher GO levels - between the two continents, only one GO term – GO:0016021 (integral component of membrane) – is shared. While acknowledging the dangers of post-hoc interpretation of GO analysis (Pavlidis *et al.* 2012), we note that within continents, terms such as GO:009416 (response to light) and GO:0045792 (negative regulation of cell size) are enriched, as these could relate to phenotypic clines observed in traits that are influenced by circadian rhythm (Hut *et al.* 2013; Svetec *et al.* 2015) and body size in *Drosophila* (Zwaan *et al.* 2000).

While we observe signals of parallel differentiation, we note that the enrichment across all levels seem at best modest, especially in comparison to patterns of differentiation in *D. melanogaster* (Reinhardt *et al.* 2014; Fabian *et al.* 2012; Bergland *et al.* 2014). This would suggest that, while there may be some shared differentiation resulting from parallel adaptation, populations on the two continents are largely responding differently on the molecular level to their local environments. This is consistent with our current understanding of adaptation in *Drosophila*; given that many ecologically relevant phenotypes (e.g. body size) are likely to be polygenic, we should not expect to detect strong differentiation at loci of relatively small effect. This is especially true for the present study, which only investigates differentiation between population pairs. On the other hand, in *D. melanogaster*, large inversions such as In(3R)P are likely to have a substantial effect on fitness, and accordingly show strong patterns of parallel differentiation. Our results, in light of earlier findings of similar studies in *D. melanogaster*, reiterate the important contribution of inversion fre-

quency clines in shaping patterns of shared differentiation, and suggest that in *D. simulans* local adaptation from selection on large-effect loci may be relatively uncommon. Lastly, we note that incomplete annotation of the *D. simulans* genome, especially in comparison to *D. melanogaster*, may have influenced the results of all annotation-based analyses, mostly by reducing power to detect shared differentiation.

Recent adaptation in *D. simulans*

Although one objective of this study was to gain insight into potential mechanisms of local adaptation in *D. simulans*, as mentioned above, it is possible that high F_{ST} between two populations is driven by reduced diversity in one population rather than selection against migrants as described in classical cline models HALDANE (1948). Because *D. simulans* has a large population size, however, and because it is thought that rates of gene flow are high, we have assumed that the most differentiated sites are, or are closely linked to, targets of spatially varying selection. Even if we are detecting strong differentiation due to selection in a single population, these differentiated sites can provide valuable information about recent adaptation.

The population-specific sweep in a region surrounding *Ace* – a gene associated with insecticide resistance – is a case in which strong selection in one population (QLD) has driven high levels of differentiation. Given that there is no reduced diversity around *Ace* in TAS populations, the differentiation of SNPs within the region is likely to have been driven by differences in pesticide application in the Australian populations. While there is evidence that insecticide resistance can have a negative pleiotropic effect on fitness in the absence of insecticide (e.g. LENORMAND *et al.* 1999), whether or not differentiation at this locus is maintained by selection against migrants, or that migration-selection is in a non-equilibrium state, remains unknown. Nevertheless, this speaks to the point that some portion of the differentiation we have observed may not be driven by latitudinally varying climatic factors,

but may also be influenced by much more localized variation in environment, such as agriculture. These local factors are unlikely to drive a signal of parallel signal of parallel latitudinal differentiation between continents and may perhaps account for some of the differences in differentiated loci between the two continents.

Very highly differentiated and non-synonymous SNPs identified in *Ace* are associated with insecticide resistance in *D. melanogaster*, presenting a compelling example of convergent evolution between species at the nucleotide level. However, additional patterns of differentiation surrounding this gene indicate that there may be more to its role in adaptation: Zhao *et al.* (2015) report that *Ace* in *D. simulans* and *D. melanogaster* is differentially expressed between Maine and Panama populations, and in our study we identify a same-direction SNP in the 3'UTR of the gene. Furthermore, Kolaczkowski *et al.* (2011) identified a putative duplication spanning *Ace* in *D. melanogaster* to be segregating at a higher frequency in QLD compared to TAS. This suggest that both structural and regulatory variation in *Ace* may be responding to selection in *Drosophila*.

Lastly, the signals of selection surrounding *Ace* provide evidence that patterns of variation are influenced by recent human activity. This is consistent with the findings of Wurmser *et al.* (2013) that some of the most pronounced differences in expression profiles of African and non-African *D. simulans* are potentially attributable to adaptation to insecticides outside of Africa, and our observation that there is reduced variation around *Cyp6g1* in all of our sampled populations. It should be noted that the strong signals indicating responses to selection from insecticides reflect the effect sizes and initial frequencies of the loci contributing to resistance and the fact that they are easily detected should not downplay the relative importance of adaptation to other environmental variables. Our understanding of patterns of genetic variation pertaining to other ecologically relevant traits will improve with a better understanding of the underlying mechanisms of ecologically important phenotypes.

Adaptation out of ancestral range Given its ancestral range of East Africa/Madagascar (Lachaise *et al.* 1988; Dean and Ballard 2004; Kopp *et al.* 2006), we looked for evidence that *D. simulans* has been experiencing adaptation to temperate environments by comparing the frequencies of derived alleles in tropical and temperate populations. We found that derived variants are at higher frequency in the subtropical populations (Fig: 3), and the diversity around high-F_{ST} SNPs is lower in subtropical populations than in temperate populations (Fig S10). Furthermore, on both continents the genome-wide mean heterozygosity is lower in tropical populations than in temperate ones. This is in contrast to observations in populations of *D. melanogaster*, which show reduced genomic diversity (Kolaczkowski *et al.* 2011; Fabian *et al.* 2012) and higher frequency of derived alleles in temperate populations for some strongly differentiated loci (Sezgin *et al.* 2004; Turner *et al.* 2008; Kolaczkowski *et al.* 2011; Reinhardt *et al.* 2014).

Combined, our results indicate that, unlike *D. melanogaster*, there is little support that *D. simulans* is ancestrally tropical-adapted with recent adaptation to temperate climates outside of Africa. Rather, the smaller population size and increased frequency of derived alleles in lower-latitude populations are consistent with these populations experiencing greater environmental stresses than their more temperate counterparts. This is supported by studies suggesting that *D. simulans* may be better adapted to cold temperatures (Chakir *et al.* 2002; Petavy *et al.* 2001) and less adapted to hot temperatures (Jenkins and Hoffmann 1994; Kellermann *et al.* 2012), although results across studies are somewhat equivocal in their conclusions (Parsons 1977; Boulétreau-Merle *et al.* 2003; David *et al.* 2004). Our own results, in contrast to the genomic results of Machado *et al.* (2015), would require confirmation from comparing patterns of diversity among additional populations along latitudinal clines.

While the mechanism may remain unclear, contrasting patterns between *D. melanogaster* and *D. simulans* emphasizes potential differences in the biogeographic histories of the two species. While both are considered to be African in origin, the ancestral ranges may have differed substantially (Lachaise *et al.* 1988). Specifically, the *D. simulans* ancestral range is believed to be in Madacascar/East Africa (Dean and Ballard 2004; Rogers *et al.* 2014), while *D. melanogaster* is thought to have an ancestral range further to the west (Pool *et al.* 2012). It is conceivable that these two regions have historically experienced substantially different climates leading to phenotypic differences between ancestral populations of the two species. It has also been proposed that there are substantial differences in adaptive strategies between the two species (Choudhary and Singh 1987), for example the role of phenotypic plasticity in the ability of *D. simulans* to persist in novel environments (van Heerwaarden *et al.* 2012; Austin and Moehring 2013).

Continent specific adaptation and clinal variation

The analysis presented here highlights the differences between two cosmopolitan species, and suggests that within *D. simulans*, Australian and North American populations are adapting to their local environments via both shared and different mechanisms. These results point to several aspects of biology that are potentially important for local adaptation in this species, including regulation, light response and insecticide resistance.

With the current dataset, we are likely to detect either genome-wide patterns, or signatures of selection at specific loci of large effect. While this has provided us with some additional insight into recent adaptation in *D. simulans*, a substantially larger dataset would be required to gain a deeper and more detailed understanding of the demographic and adaptive processes influencing this species. In light of this, and our observation that much of the differentiation within continents is not shared between continents, it seems that a dense

sampling of a single clinal transect would perhaps be the best strategy for understanding the genetics of local adaptation. This would also address whether differentiation reflects continuously varying environment, or is influenced by local, discontinuous environmental heterogeneity. Lastly, we have assumed, like many others before us, that gene flow is high in *D. simulans*, and that clines are stable (i.e. allele frequencies do not change substantially in time). Based on temporal sampling of a single population, MACHADO *et al.* (2015) find that while somewhat stable, allele frequency clines in *D. simulans* are less stable than clines in *D. melanogaster*. Although evidence for temporally stable clines indicate that this assumption is appropriate for some variants, the extent of how true this is on a genome-wide scale will be addressed as datasets with denser temporal and spatial sampling become available.

Acknowledgments

We thank A. Hoffmann for providing us with *D. simulans* stock lines and thank Y. Brandvain, G. Coop, J. Cridland, N. Svetec, L. Zhao and T. Seher for valuable insights and comments. This study was funded by a National Institute of Health grant to D.J. Begun (GM084056) and a National Science Foundation Graduate Research Fellowship to A. Sedghifar (1148897)

LITERATURE CITED

ANDERSON, P. R., and J. G. OAKESHOTT, 1984 Parallel geographical patterns of allozyme variation in two sibling Drosophila species. Nature **308**: 729–731.

ANDOLFATTO, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in Drosophila melanogaster and Drosophila simulans. Molecular Biology and Evolution : 279–290.

ARTHUR, A. L., A. R. WEEKS, and C. M. SGRÒ, 2008 Investigating latitudinal clines for life history and stress resistance traits in Drosophila simulans from eastern Australia. Journal of evolutionary biology **21**: 1470–9.

Ashburner, M., and F. Lemeunier, 1976 Relationships within the melanogaster Species Subgroup of the Genus Drosophila (Sophophora). I. Inversion Polymorphisms in Drosophila melanogaster and Drosophila simulans. Proceedings of the Royal Society B: Biological Sciences **193**: 137–157.

Austin, C. J., and A. J. Moehring, 2013 Optimal temperature range of a plastic species, Drosophila simulans. The Journal of animal ecology : 663–672.

Begun, D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y.-P. Poh, M. W. Hahn, P. M. Nista, C. D. Jones, A. D. Kern, C. N. Dewey, L. Pachter, E. Myers, and C. H. Langley, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. PLoS biology **5**: e310.

Begun, D. J., and P. Whitley, 2000 Reduced X-linked nucleotide polymorphism in Drosophila simulans. Proceedings of the National Academy of Sciences **97**: 5960–5965.

Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. a. Petrov, 2014 Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila. PLoS genetics **10**: e1004775.

Boulétreau-Merle, J., P. Fouillet, and J. Varaldi, 2003 Divergent strategies in low temperature environment for the sibling species Drosophila melanogaster and D . simulans: overwintering in extension border areas of France and comparison with African populations. Evolutionary Ecology **17**: 523–548.

Capy, P., and P. Gibert, 2004a Drosophila melanogaster, Drosophila simulans: so similar yet so different. Genetica **120**: 5–16.

Capy, P., and P. E. Gibert, 2004b Drosophila melanogaster, Drosophila simulans: so similar, yet so differnt. [Special issue] Genetica **120**.

642 Carlson, M., S. Falcon, H. Pages, and N. Li, 20 org.Dm.eg.db: Genome wide annota-
643 tion for Fly. R package version 2.6.4.

644 Chakir, M., A. Chafik, B. Moreteau, P. Gibert, and J. R. David, 2002 Male
645 sterility thermal thresholds in Drosophila: D. simulans appears more cold-adapted than
646 its sibling D. melanogaster. Genetica **114**: 195–205.

647 Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993 The effect of dele-
648 terious mutations on neutral molecular variation. Genetics **134**: 1289–303.

649 Choudhary, M., and R. Singh, 1987 A Comprehensive Study of Genetic Variation in
650 Natural Populations of Drosophila melanogaster. III. Variations in Genetic Structure and
651 Their Causes Between Drosophila melanogaster and Its Sibling Species Drosophila simu-
652 lans. Genetics : 697–710.

653 Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. V. Jr, M. Dickson, J. Grim-
654 wood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley, 2005
655 Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin
656 Alleles. Science : 1928–1934.

657 Comeron, J. M., R. Ratnappan, and S. Bailin, 2012 The many landscapes of recom-
658 bination in Drosophila melanogaster. PLoS genetics **8**: e1002905.

659 Cox, M. P., D. a. Peterson, and P. J. Biggs, 2010 SolexaQA: At-a-glance quality
660 assessment of Illumina second-generation sequencing data. BMC bioinformatics **11**: 485.

661 David, J. R., R. Allemand, P. Capy, M. Chakir, P. Gibert, G. Pétavy,
662 and B. Moreteau, 2004 Comparative life histories and ecophysiology of Drosophila
663 melanogaster and D. simulans. Genetica **120**: 151–63.

664 Dean, M. D., and J. W. O. Ballard, 2004 Linking phylogenetics with population genet-
665 ics to reconstruct the geographic origin of a species. Molecular phylogenetics and evolution
666 **32**: 998–1009.

667 Dobzhansky, T., 1947 Adaptive changes induced by natural selection in wild populations
668 of Drosophila. Evolution **1**: 1–16.

669 Dobzhansky, T., 1948 Genetics of natural populations. XVI. Altitudinal and seasonal
670 changes produced by natural selection in certain populations of Drosophila pseudoobscura
671 and Drosophila persimilis. Genetics **33**: 158–176.

672 Endler, J. A., 1977 *Geographic Variation, Speciation, and Clines*. Princeton University
673 Press, Princeton, New Jersey.

674 Fabian, D. K., M. Kapun, V. Nolte, R. Kofler, P. S. Schmidt, C. Schlötterer,
675 and T. Flatt, 2012 Genome-wide patterns of latitudinal differentiation among popula-
676 tions of Drosophila melanogaster from North America. Molecular ecology **21**: 4748–69.

677 Ffrench-Constant, R., N. Anthony, K. Aronstein, T. Rocheleau, and G. Stil-
678 well, 2000 Cyclodiene Insecticide Reistance: From Molecular to Population Genetics.
679 Annual Review of Entomology **48**: 449–466.

680 Fournier, D., J.-M. Bride, F. Hoffmann, and F. Karch, 1992 Two types of modifi-
681 cations confer resistance to insecticide. The Journal of biological chemistry **267**: 14270–
682 14274.

683 Futschik, A., and C. Schlötterer, 2010 The next generation of molecular markers from
684 massively parallel sequencing of pooled DNA samples. Genetics **186**: 207–18.

685 Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Du-
686 doit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Hu-
687 ber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Saw-

itzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang, 2004 Bioconductor: open software development for computational biology and bioinformatics. Genome biology **5**.

Gibert, P., P. Capy, and A. Imasheva, 2004 Comparative analysis of morphological traits among Drosophila melanogaster and D. simulans: genetic variability, clines and phenotypic plasticity. Genetica **120**: 165–179.

Haldane, S., 1948 The theory of a cline. Journal of Genetics **48**: 277–284.

Hallas, R., M. Schiffer, and A. a. Hoffmann, 2002 Clinal variation in Drosophila serrata for stress resistance and body size. Genetical research **79**: 141–8.

Hirsh, J., T. Riemensperger, H. Coulom, M. Iché, J. Coupar, and S. Birman, 2010 Roles of dopamine in circadian rhythmicity and extreme light sensitivity of circadian entrainment. Current biology : CB **20**: 209–14.

Hoffmann, A. A., A. Anderson, and R. Hallas, 2002 Opposing clines for high and low temperature resistance in Drosophila melanogaster. Ecology Letters **5**: 614–618.

Hoffmann, A. A., and A. R. Weeks, 2007 Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. Genetica **129**: 133–47.

Hohenlohe, P. a., S. Bassham, P. D. Etter, N. Stiffler, E. a. Johnson, and W. a. Cresko, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS genetics **6**: e1000862.

Hu, T., M. Eisen, K. Thornton, and P. Andolfatto, 2013 A second-generation assembly of the Drosophila simulans genome provides new insights into patterns of lineage-specific divergence. Genome research : 89–98.

711 Huey, R. B., 2000 Rapid Evolution of a Geographic Cline in Size in an Introduced Fly.
712 Science **287**: 308–309.

713 Hut, R. A., S. Paolucci, R. Dor, C. P. Kyriacou, and S. Daan, 2013 Latitudinal
714 clines : an evolutionary view on biological rhythms Latitudinal clines : an evolutionary
715 view on biological rhythms. Proceedings of the Royal Society B: Biological Sciences **280**:
716 20130433.

717 Jenkins, N., and A. Hoffmann, 1994 Genetic and maternal variation for heat resistance
718 in Drosophila from the field. Genetics **137**: 783–789.

719 Kellermann, V., J. Overgaard, A. a. Hoffmann, C. Fløjgaard, J.-C. Svenning,
720 and V. Loeschcke, 2012 Upper thermal limits of Drosophila are linked to species distri-
721 butions and strongly constrained phylogenetically. Proceedings of the National Academy
722 of Sciences of the United States of America **109**: 16228–33.

723 Kolaczkowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun, 2011 Ge-
724 nomic differentiation between temperate and tropical Australian populations of Drosophila
725 melanogaster. Genetics **187**: 245–60.

726 Kopp, A., A. Frank, and J. Fu, 2006 Historical biogeography of Drosophila simulans
727 based on Y-chromosomal sequences. Molecular phylogenetics and evolution **38**: 355–62.

728 Kume, K., 2006 A Drosophila dopamine transporter mutant, fumin (fmn), is defective in
729 arousal regulation. Sleep and Biological Rhythms **4**: 263–273.

730 Kume, K., S. Kume, S. K. Park, J. Hirsh, and F. R. Jackson, 2005 Dopamine is a
731 regulator of arousal in the fruit fly. The Journal of neuroscience : the official journal of
732 the Society for Neuroscience **25**: 7377–84.

733 Lachaise, D., M. Cariou, and J. David, 1988 Historical biogeography of the Drosophila
734 melanogaster species subgroup. Evolutionary Biology **22**: 159–225.

735 Langley, C. H., K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider, J. E.
736 Pool, S. a. Langley, C. Suarez, R. B. Corbett-Detig, B. Kolaczkowski,
737 S. Fang, P. M. Nista, A. K. Holloway, A. D. Kern, C. N. Dewey, Y. S. Song,
738 M. W. Hahn, and D. J. Begun, 2012 Genomic variation in natural populations of
739 Drosophila melanogaster. Genetics **192**: 533–98.

740 Lenormand, T., D. Bourguet, T. Guillemaud, and M. Raymond, 1999 Tracking the
741 evolution of insecticide resistance in the mosquito Culex pipiens. Nature **400**: 861–864.

742 Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler
743 transform. Bioinformatics (Oxford, England) **25**: 1754–60.

744 Long, A. D., and R. S. Singh, 1992 Geographic Variation in Dorophila: From Molecules
745 to Morphology and Back. Trends in Ecology & Evolution **7**: 340–345.

746 Machado, H., A. O. Bergland, K. R. O'Brien, E. L. Behrman, P. Schmidt, and
747 D. A. Petrov, 2015 Comparative population genomics of latitudinal variation in D.
748 simulans and D. melanogaster. Molecular Ecology .

749 Mackay, T. F. C., S. Richards, E. a. Stone, A. Barbadilla, J. F. Ayroles,
750 D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richard-
751 son, R. R. H. Anholt, M. Barrón, C. Bess, K. P. Blankenburg, M. A. Car-
752 bone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C.
753 Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee,
754 P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M.
755 Muzny, L. Nazareth, I. Newsham, L. Perales, L.-L. Pu, C. Qu, M. Ràmia, J. G.
756 Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley, Y.-Q.
757 Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman,
758 and R. a. Gibbs, 2012 The Drosophila melanogaster Genetic Reference Panel. Nature
759 **482**: 173–8.

760 Manceau, M., 2010 Convergence in pigmentation at multiple levels: mutations, genes and
761 function. Philosophical transactions of the Royal Society of London B Biological Sciences
762 **365**: 2439–2450.

763 Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favorable gene.
764 Genetical research **23**: 23–35.

765 Mutero, A., M. Pralavorio, J.-M. Bride, and D. Fournier, 1994 Resistance-
766 associated point mutations in insecticide-insensitive acetylcholinesterase. Proceedings of
767 the National Academy of Sciences **91**: 5922–5926.

768 Nordborg, M., T. T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian, H. Zheng,
769 E. Bakker, P. Calabrese, J. Gladstone, R. Goyal, M. Jakobsson, S. Kim,
770 Y. Morozov, B. Padhukasahasram, V. Plagnol, N. a. Rosenberg, C. Shah,
771 J. D. Wall, J. Wang, K. Zhao, T. Kalbfleisch, V. Schulz, M. Kreitman, and
772 J. Bergelson, 2005 The pattern of polymorphism in Arabidopsis thaliana. PLoS biology
773 **3**: e196.

774 Paaby, A. B., M. J. Blacket, A. a. Hoffmann, and P. S. Schmidt, 2010 Identification
775 of a candidate adaptive polymorphism for Drosophila life history by parallel independent
776 clines on two continents. Molecular ecology **19**: 760–74.

777 Papoulas, O., T. S. Hays, and J. C. Sisson, 2005 The golgin Lava lamp mediates
778 dynein-based Golgi movements during Drosophila cellularization. Nature cell biology **7**:
779 612–8.

780 Parsons, P., 1975a Phototactic responses along a gradient of light intensities for the sibling
781 species Drosophila melanogaster and Drosophila simulans. Behavior genetics **5**: 17–25.

782 Parsons, P., 1975b The comparative evolutionary biology of the sibling species, Drosophila
783 melanogaster and D. simulans. Quarterly Review of Biology **50**: 151–169.

784 Parsons, P., 1977 Resistance to Cold Temperature Stress in Populations of Drosophila
785 melanogaster and D. simulans. Australian Journal of Zoology **25**: 693–98.

786 Pavlidis, P., J. D. Jensen, W. Stephan, and A. Stamatakis, 2012 A critical assessment
787 of storytelling: gene ontology categories and the importance of validating genomic scans.
788 Molecular biology and evolution **29**: 3237–48.

789 Petavy, G., J. David, P. Gibert, and B. Moreteau, 2001 Viability and rate of devel-
790 opment at different temperatures in Drosophila: a comparison of constant and alternating
791 thermal regimes. Journal of Thermal Biology **26**: 29–39.

792 Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. a. Stevens, C. M. Cardeno,
793 M. W. Crepeau, P. Duchen, J. J. Emerson, P. Saelao, D. J. Begun, and C. H.
794 Langley, 2012 Population Genomics of sub-saharan Drosophila melanogaster: African
795 diversity and non-African admixture. PLoS genetics **8**: e1003080.

796 Price, T. a. R., A. Bretman, A. C. Gradilla, J. Reger, M. L. Taylor,
797 P. Giraldo-Perez, A. Campbell, G. D. D. Hurst, and N. Wedell, 2014 Does
798 polyandry control population sex ratio via regulation of a selfish gene? Proceedings.
799 Biological sciences / The Royal Society **281**: 20133259.

800 Reinhardt, J. A., B. Kolaczkowski, C. D. Jones, D. J. Begun, and A. D.
801 Kern, 2014 Parallel Geographic Variation in Drosophila melanogaster. Genetics : ge-
802 netics.114.161.161463.

803 Rogers, R. L., L. Shao, J. S. Sanjak, P. Andolfatto, and K. R. Thornton, 2014
804 Revised Annotations, Sex-Biased Expression, and Lineage-Specific Genes in the Drosophila
805 melanogaster Group. G3 (Bethesda, Md.) **4**: 2345–51.

806 Rosenblum, E. B., H. Römpler, T. Schöneberg, and H. E. Hoekstra, 2010 Molec-
807 ular and functional basis of phenotypic convergence in white lizards at White Sands.

Proceedings of the National Academy of Sciences of the United States of America **107**: 2113–7.

Schlenke, T., and D. Begun, 2004 Strong selective sweep associated with a transposon insertion in Drosophila simulans. Proceedings of the National Academy of Sciences **101**: 1626–1631.

Schmidt, P. S., and A. B. Paaby, 2008 Reproductive diapause and life-history clines in North American populations of Drosophila melanogaster. Evolution; international journal of organic evolution **62**: 1204–15.

Sezgin, E., D. D. Duvernell, L. M. Matzkin, Y. Duan, C.-T. Zhu, B. C. Verrelli, and W. F. Eanes, 2004 Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in Drosophila melanogaster. Genetics **168**: 923–31.

Singh, R., M. Choudhary, and J. David, 1987 Contrasting patterns of geographic variation in the cosmopolitan sibling species Drosophila melanogaster and Drosophila simulans. Biochemical genetics **25**: 27–40.

Singh, R. S., 1989 Population Genetics and Evolution of Species Related to Drosophila melanogaster. Ann. Rev. Genet. **23**: 245–53.

Sisson, J., C. Field, and R. Ventura, 2000 Lava lamp, a novel peripheral golgi protein, is required for Drosophila melanogaster cellularization. The Journal of cell . . . **151**: 905–917.

Sturtevant, A., and T. Dobzhansky, 1936 Geographical distribution and cytology of "sex ratio" in Drosophila pseudoobscura and related species. Genetics **21**: 473–490.

Svetec, N., L. Zhao, P. Saelao, J. C. Chiu, and D. J. Begun, 2015 Evidence that natural selection maintains genetic variation for sleep in Drosophila melanogaster. BMC evolutionary biology **15**: 41.

Tamura, K., S. Subramanian, and S. Kumar, 2004 Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Molecular biology and evolution **21**: 36–44.

Turner, T. L., E. C. Bourne, E. J. Von Wettberg, T. T. Hu, and S. V. Nuzhdin, 2010 Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. Nature genetics **42**: 260–3.

Turner, T. L., M. T. Levine, M. L. Eckert, and D. J. Begun, 2008 Genomic analysis of adaptive differentiation in Drosophila melanogaster. Genetics **179**: 455–73.

Tyukmaeva, V. I., T. S. Salminen, M. Kankare, K. E. Knott, and A. Hoikkala, 2011 Adaptation to a seasonally varying environment: a strong latitudinal cline in reproductive diapause combined with high gene flow in Drosophila montana. Ecology and evolution **1**: 160–8.

Ueno, T., J. Tomita, S. Kume, and K. Kume, 2012 Dopamine modulates metabolic rate and temperature sensitivity in Drosophila melanogaster. PloS one **7**: e31513.

van Heerwaarden, B., R. F. H. Lee, B. Wegener, a. R. Weeks, and C. M. Sgró, 2012 Complex patterns of local adaptation in heat tolerance in Drosophila simulans from eastern Australia. Journal of evolutionary biology **25**: 1765–78.

Wurmser, F., T. Mary-Huard, J.-J. Daudin, D. Joly, and C. Montchamp-Moreau, 2013 Variation of gene expression associated with colonisation of an anthropized environment: comparison between African and European populations of Drosophila simulans. PloS one **8**: e79750.

Zhao, L., J. Wit, N. Svetec, and D. J. Begun, 2015 Parallel Gene Expression Differences between Low and High Latitude Populations of Drosophila melanogaster and D. simulans. PLOS Genetics **11**: e1005184.

856 Zwaan, B. J., R. B. Azevedo, a. C. James, J. Van 't Land, and L. Partridge,
857 2000 Cellular basis of wing size variation in Drosophila melanogaster: a comparison of
858 latitudinal clines on two continents. Heredity **84**: 338–47.

Table 1: Size, collection dates and locations of samples.

key	latitude	chromosomes	year	source
FL (US)	25.47N	66	September 2011	daughters of field-caught females
RI(US)	41.84N	66	September 2011	daughters of field-caught females
QLD(AU)	42.77S	22	Feb-March 2004	isofemale lines
TAS (AU)	25.54S	16	Feb-March 2004	isofemale lines

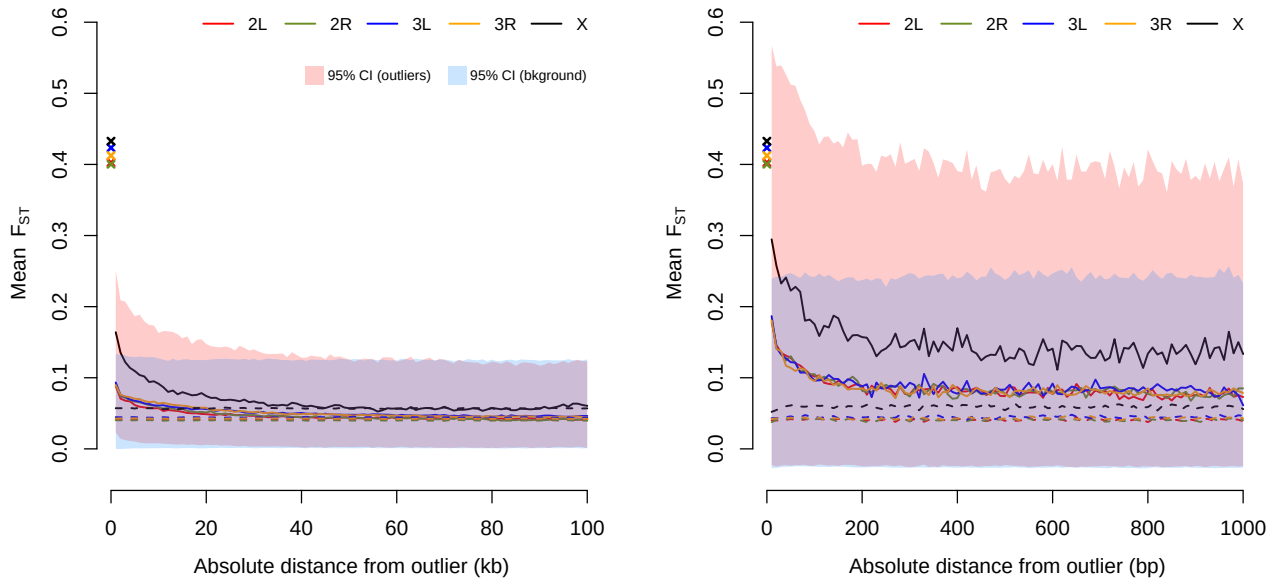

Figure 1: **Left** Mean F_{ST} in increments of non-overlapping 1kb windows as a function of distance from an outlier SNP in the top 1% tail. Crosses denote mean F_{ST} of outlier windows. Background values represent mean F_{ST} as a function of distance from 10000 randomly selected non-outlier SNPs. Confidence Intervals are defined by the upper and lower 2.5% quantiles. **Right** Decay measured in 10bp non-overlapping windows away from outliers in North America.

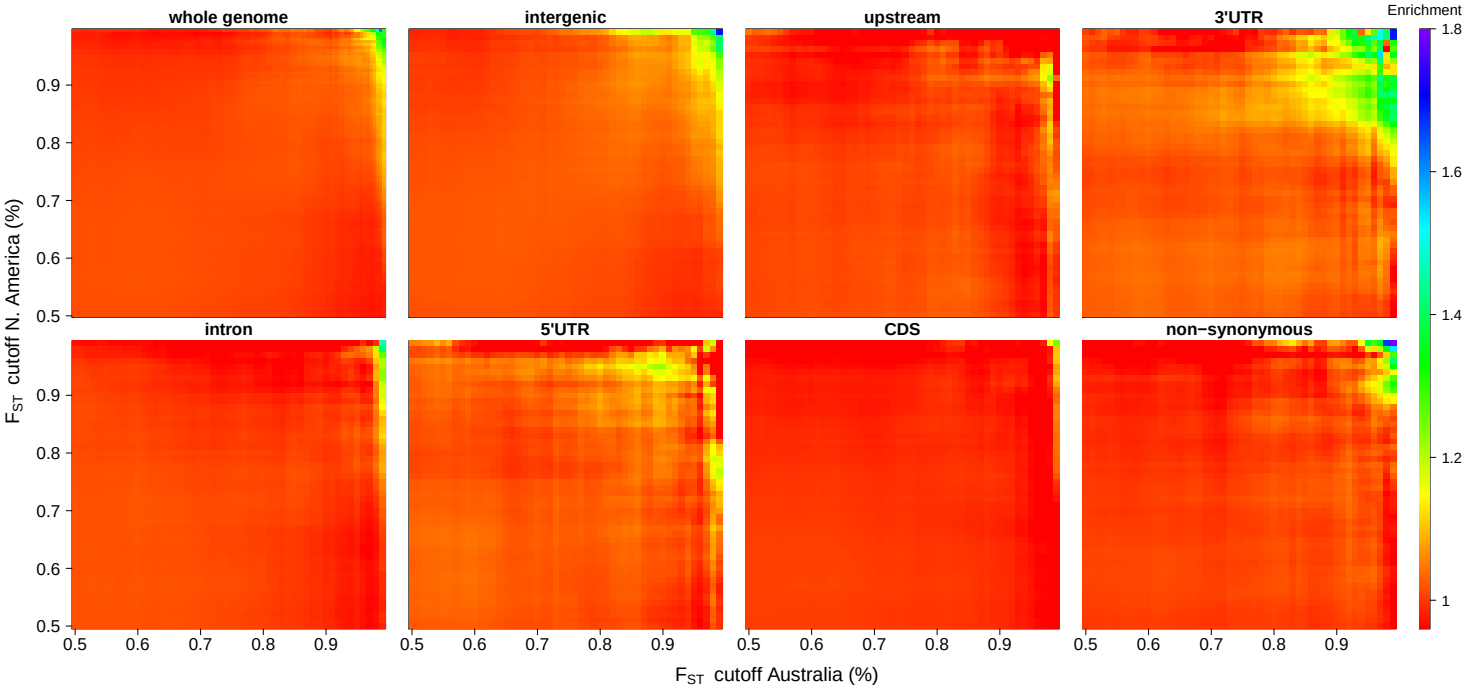

Figure 2: Enrichment for number of shared outlier SNPs for pairwise outlier quantiles increasing in 0.5% increments on both continents, within given subsets of the genome. Each point in the heat maps are cumulative, (i.e. that the 95th percentile is a subset of the 90th percentile.)

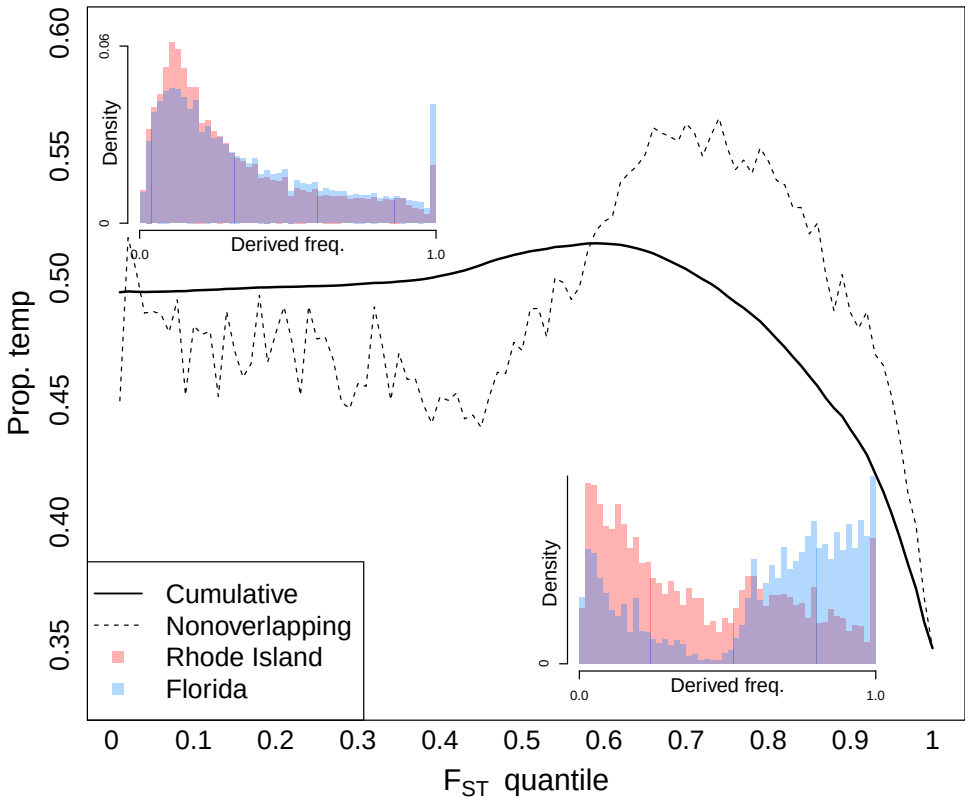

Figure 3: Proportion of derived alleles that are at higher frequency in temperate populations, as a function of F_{ST} in North America. Dotted lines represent the proportion in non-overlapping F_{ST} bins. Solid line represents the cumulative distribution of the dotted line. Left inset is the derived allele frequency spectra for the two North American populations. Right inset is the genome-wide derived allele frequency spectra for SNPs in the 0.99 F_{ST} tail. Only SNPs segregating at a total frequency greater than 0.05 in North America were considered.

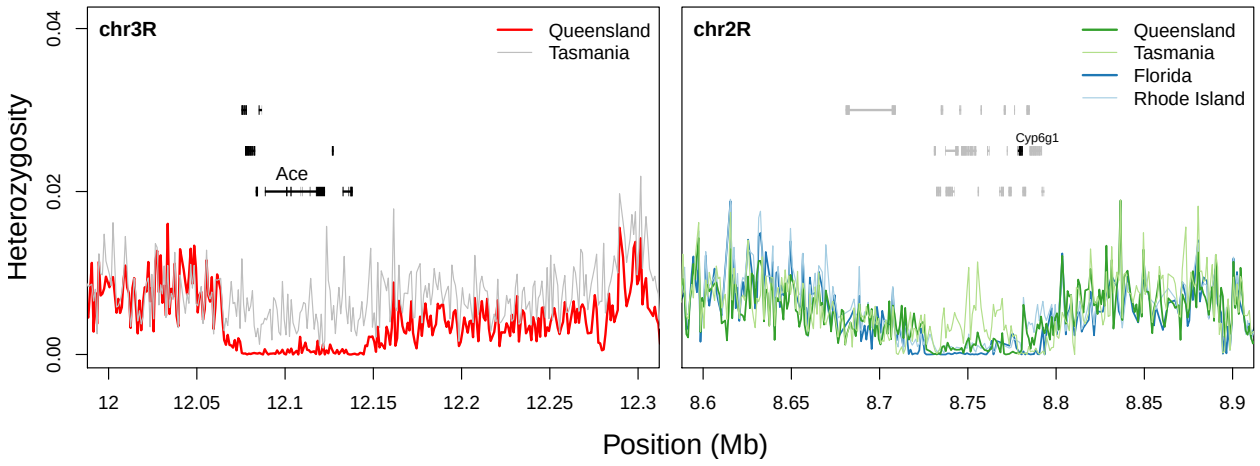

Figure 4: Large regions of reduced diversity around known insecticide resistance loci, shown in non-overlapping 1kb windows. *Left panel:* Region of reduced diversity surrounding *Ace* in Queensland. *Right panel:* Region of reduced diversity around *Cyp6g1*, as identified in Schlenke et al.