Genetics: Published Articles Ahead of Print, published on November 8, 2010 as 10.1534/genetics.110.123059

Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*

An Investigation Submitted to the Population and Evolutionary Genetics Section of *Genetics*

Bryan Kolaczkowski*[†], Andrew D. Kern*[†], Alisha K. Holloway[‡], and David J. Begun[‡]

November 3, 2010

*these authors contributed equally

[†]Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA

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

Running Head: *D. melanogaster* clinal variation

Key Words: Population Genomics, Clinal Variation, *Drosophila melanogaster*

Corresponding Author:

David J. Begun

Evolution and Ecology
College of Biological Sciences
University of California, Davis
3350A Storer
Davis, CA 95616
djbegun@ucdavis.edu

Abstract

Determining the genetic basis of environmental adaptation is a central problem of evolutionary biology. This issue has been fruitfully addressed by examining genetic di?erentiation between populations that are recently separated and/or experience high rates of gene ow. A good example of this approach is the decades-long investigation of selection acting along latitudinal clines in *D. melanogaster*. Here we use next-generation genome sequencing to re-examine the well-studied Australian *D. melanogaster* cline. We find evidence for extensive differentiation between temperate and tropical populations, with regulatory regions and unannotated regions showing particularly high levels of differentiation. Although the physical genomic scale of geographic differentiation is small—on the order of gene-sized—we observed several larger highly-differentiated regions. The region spanned by the cosmopolitan inversion polymorphism *In(3R)P* shows higher levels of differentiation, consistent with the major difference in allele frequencies of Standard and *In(3R)P* karyotypes in temperate vs. tropical Australian populations. Our analysis reveals evidence for spatially-varying selection on a number of key biological processes, suggesting fundamental biological differences between flies from these two geographic regions.

INTRODUCTION

Determining the processes maintaining genetic variation within species is a basic goal of biological research and a central problem of evolutionary genetics. Indeed, the relative contributions to segregating variation of 1) low-frequency, unconditionally deleterious mutations, 2) intermediate-frequency small-effect variants maintained by mutation and genetic drift, and 3) adaptive mutations maintained by positive selection—e.g. spatially-varying or negative frequency-dependent selection—remain unknown in any species. Thus, it is also unclear whether different processes predominate in different species, perhaps resulting from differences in population size, ecology or genetics.

One approach for identifying adaptive variants segregating within species is to investigate systems in which there are major phenotypic variants likely influenced by natural selection and which have relatively simple genetics. This is what has traditionally been thought of as ecological genetics. For example, pigmentation variation in vertebrates (e.g. Nachman *et al.* (2003)) is a good example of a trait for which the relatively small number of candidate genes allows the phenotypic effects of natural variants to be directly tested. For major phenotypic variants having a simple genetic basis but no candidate genes, genetic analysis can be used to isolate alternative alleles underlying the phenotypic difference. Examples include diapause variation and foraging behavior in *Drosophila melanogaster* (Schmidt *et al.* 2008; Osborne *et al.* 1997), traits relating to social behavior and copulatory plug formation in *Caenorhabditis elegans* (de Bono and Bargmann 1998; Palopoli *et al.* 2008) and several phenotypes in sticklebacks (Colosimo *et al.* 2004; Miller *et al.* 2007; Chan *et al.* 2010). Besides their simple genetics, such biological examples have the advantage that the targeted traits may have plausible connections to fitness variation in nature (though this is not always the case). In spite of the practical advantages associated with phenotypic variation resulting from simple genetics and alleles of large effect, such variation may not speak very strongly to the general properties of adaptive polymorphisms in natural populations, which may often be characterized by complex genetics or small-effect alleles.

A complementary approach uses population-genetic analysis to identify individual polymorphic variants/genes that may have been influenced by positive selection. This approach offers at least two advantages. First, it can be made genomic in scope, and therefore may provide a less-biased view of the genes and phenotypes influenced by positive selection. There is no comparably comprehensive "omic" concept for phenotypic analysis, because the universe of phenotype space is difficult to define, difficult to measure and highly dimensional (Lewontin 1974). Second, alleles having relatively small effects or effects not associated with easily-defined phenotypes can be identified. A population-genetic approach is a particularly powerful discovery tool when joined with high-quality genome annotation, generating many new hypotheses about the genetic and phenotypic variation influenced by positive selection within species and providing vast opportunities for the downstream functional investigation of such variation.

One population-genetic approach for identifying positively selected polymorphisms is to search the genome for sites exhibiting large allele-frequency differences between recently separated populations or those experiencing high rates of gene flow (Lewontin and Krakauer 1975). Because even low levels of gene flow effectively homogenize neutral allele frequencies (Wright 1931; Maruyama 1970; Slatkin 1981), alleles under spatially-varying selection are expected to appear as outliers with respect to allele-frequency differences across populations. This strategy may be particularly effective when allele frequencies change gradually along a cline, such as with latitude or altitude.

Some of the best-studied cases of latitudinal clines maintained by spatially-varying selection are those of *Drosophila melanogaster*. The majority of work on these clines has investigated various phenotypic traits, chromosome inversion polymorphisms and enzyme-coding genes (Sezgin *et al.* 2004), as well as several other genes harboring clinal variants (Costa *et al.* 1992; McColl and McKechnie 1999; Duvernell *et al.* 2003; Schmidt *et al.* 2000). The cline along the east coast of Australia has received considerable recent attention due to the efforts of Ary Hoffmann and collaborators (e.g. Hoffmann and Weeks

(2007)). The fact that similar clines are often observed on different continents strongly implicates natural selection rather than demography as the cause of clinal variation (Oakeshott *et al.* 1981; Oakeshott *et al.* 1983; Singh and Rhomberg 1987; Singh 1989; SINGH and LONG 1992; Gockel *et al.* 2001; Kennington *et al.* 2003; Hoffmann and Weeks 2007). Importantly, though cosmopolitan chromosome inversion polymorphisms exhibit latitudinal clines (with inversion frequency increasing in more tropical populations), many observations convincingly show that inversions explain only a fraction of clinal variation, even for genes located in inverted regions (Voelker *et al.* 1978; Knibb 1982; Singh and Rhomberg 1987; Frydenberg *et al.* 2003; Umina *et al.* 2006). Indeed, many clinally-varying genes are not physically near inversions (Voelker *et al.* 1978; Singh and Rhomberg 1987; Sezgin *et al.* 2004; Turner *et al.* 2008).

We recently extended the genetic characterization of population differentiation from *D. melanogaster* clines by comparative-genomic hybridization analysis of population samples from opposite ends of well-described clines in Australia and North America (Turner *et al.* 2008). That study generated new information on genomic differentiation, but the crude nature of the data limited the scope of the analysis and the strength of the conclusions that could be drawn. Here we revisit the issue of geographic differentiation between opposite ends of a known *D. melanogaster* cline using next-generation sequencing to characterize genomic variation in flies from Queensland and Tasmania, Australia. These data are used to generate hypotheses regarding the biological differences between flies from these regions and to assess the population-genetic properties of sequence differentiation between these geographic regions.

MATERIALS AND METHODS

Sequencing, assembly and data filtering: Population samples from the east coast of Australia were collected in 2004 (Anderson *et al.* 2005). Twenty isofemale lines from Queensland (Cairns, lat. 16.907 and Cooktown, lat. 15.476) and 19 isofemale lines from Tasmania (Hillwood, lat. 41.237 and Sorell, lat. 42.769) were used. Two females were collected from each Queensland line (n=40 flies). These flies were pooled in a single tube and made into DNA. Similarly, two females were collected from each Tasmania line (n=38 flies), pooled in a single tube and made into DNA. Each of the two DNA samples was then sequenced using Solexa/Illumina technology (BENTLEY *et al.* 2008). Base calls and quality scores were determined using the Solexa GAPipeline v0.3.0. Output files were in fastq format. Reads were mapped against the *D. melanogaster* reference genome R5.8 (ADAMS *et al.* 2000) using Maq v0.6.8 (LI *et al.* 2008). Prior to mapping, we split fastq files into separate files with one million reads/file. The reads are available in the NCBI Sequence Read Archive under Accession SRA012285.16.

Several Maq functions were used for data formatting. Solexa quality scores were converted to Sanger quality scores using Maq function *sol2sanger* and converted from fastq files to binary fastq (bfq) using the Maq function *fastq2bfq*. Bases 1-36 of each read were used; the expected heterozygosity parameter ("-m" flag) was 0.005. Mapped reads were merged using *mapmerge*. The functions *maq assemble* and *maq pileup* were then used to produce pileup files. Finally, pileup files were split by chromosome arm for downstream analysis. Individual base calls with Maq quality scores < 10 were excluded, as were positions with only a singleton variant in the entire Australian sample. We explored the value of increasing the Maq quality threshold to 20, but the reduction in coverage was too costly, given the amount of data. Because we excluded singletons and focused on genomic outliers, errors should not be an important factor with respect to our biological conclusions. We excluded genomic positions with < 6 or > 20 sequence reads in either population, because these sites are associated either with very low power to reject the null hypothesis or with the confounding phenomenon of differentiated copy-number variation.

Because a primary goal of our study was to generate biological, gene-centric hypotheses regarding the nature of selection, most analyses excluded regions of the genome adjacent to centromeres and telomeres associated with low heterozygosity, as determined from genome sequences of a Raleigh sample of inbred lines sequenced as part of the Drosophila Population Genomics Project (DPGP.org). These regions of reduced heterozygosity are expected to be associated with lower power to detect differentiation, and because they experience reduced rates of crossing-over, the physical scale of differentiation may be quite large, limiting opportunities for identifying potential targets of selection. The coordinates corresponding to regions of normal recombination used in our analyses are (*2L*:844225-19946732; *2R*:6063980-20322335; *3L*:447386-18392988; *3R*:7940899-27237549; *X*:1036552-20902578). The regions excluded are roughly consistent with the non- or low-recombining portions of the genome identified in prior studies (e.g. Singh *et al.* (2005)).

Ancestral Sequence Reconstruction: For the purposes of unfolding the site frequency spectrum in our samples, ancestral states were inferred using maximum likelihood (ML) (Yang *et al.* 1995) (provided by PAML v4.3 (Yang 2007)), assuming the reference phylogeny (Consortium *et al.* 2007), the HKY nucleotide substitution model (Hasegawa *et al.* 1985) and gamma-distributed among-site rate variation (Yang 1996). ML reconstruction posterior probabilities were calculated using the empirical Bayesian approach described in Yang *et al.* (1995); the posterior probability of ancestral base b_i, given data x_j at alignment position j is given by: $P(b_i|x_j) = P(x_j|b_i)P(b_i)/\sum_{k=1}^{4} P(x_j|b_k)P(b_k)$, where $P(x_j|b_i)$ is the probability of observing data x_j given base b_i in the ancestral sequence, and $P(b_i)$ is the frequency of base b_i in the dataset. Positions with a ML reconstruction posterior probability < 0.9 were considered potentially unreliable and excluded from the analysis. The data for our ancestral sequence reconstruction were obtained from the MULTIZ 15-way insect alignment available for download from the UCSC genome browser (Blanchette *et al.* 2004; Hinrichs *et al.* 2006).

Population Genetic Estimation of Pooled Sample Reads: The pooled sampling strategy that we have taken is an economical way to get a picture of variation in a population, however it requires a bit of work to correct for biases associated with the sampling alone. In particular, there is an added layer of sampling above the normal population genetic assumptions. Here we provide some salient results which we have derived for bias corrected estimation of heterozygosity and other canonical population genetic summary statistics.

Sequencing pooled DNA leads to an additional round of sampling with replacement, beyond the initial sampling of chromosomes from nature. Let p be the population frequency of an allele A_1. Also consider the case where n chromosomes are sampled from nature and are sequenced to a depth m. We do not treat m as a random variable, although other authors have (Futschik and Schlotterer 2010). The probability of sequencing $X = k$ out of m reads of the A_1 allele, conditional upon the population frequency p and our pooled sample size n, is

$$Prob(X = k|m, n, p) = \sum_{i=0}^{n} \binom{m}{k} (i/n)^k (1 - i/n)^{m-k} \binom{n}{i} p^i (1 - p)^{n-i} \tag{1}$$

Its clear enough to see that the expected value of the sample frequency, $E(k/m)$, should be unbiased with respect to the frequency in the population, as $E(k/m) = E(E(k/m|i/n)) = \sum_i E(k/m|i/n) * Prob(i) = p$. Deriving the second moment of the sample frequency is more involved and can be found in the supplement. The result is $E((k/m)^2) = \frac{p(1-p)(n-1+m)}{nm} + p^2$. This allows us to write down an unbiased estimator of heterozygosity $H = 2p(1-p)$. Under standard binomial sampling, the estimator H is biased and needs to be corrected by a factor of $n/(n-1)$ (Nei 1987). In the case of sequencing into pooled samples, the expectation of H is

$$\begin{aligned} E(H) = E(2p(1-p)) &= 2(E(p) - E(p^2)) \\ &= 2p(1-p)((n-1)/n)((m-1)/m) \end{aligned} \tag{2}$$

So in the correction for our second round of sampling, there is the addition of exactly one term in our estimate of heterozygosity. The correction leads to our estimate of allele-frequency differentiation between Queensland and Tasmania, F_{ST}, which was calculated as:

$$F_{ST} = \frac{\Pi_{total} - \Pi_{within}}{\Pi_{total}},$$

where

$$\begin{aligned}
\Pi_{total} &= H(P_{total}) \\
\Pi_{within} &= \frac{(N_Q \times H(P_Q)) + (N_{TAS} \times H(P_{TAS})))}{N_Q + N_{TAS}} \\
H(P) &= 2p(1-p)\frac{n}{n-1}\frac{m}{m-1}
\end{aligned}$$

Here N_Q and N_{TAS} are the sample sizes from Queensland and Tasmania populations, respectively, and P_Q and P_{TAS} are the corresponding allele frequencies. P_{total} is the allele frequency in the combined (i.e., Queensland + Tasmania) population sample. $H(P)$ is our corrected estimate of heterozygosity from equation 2. In the supplement we provide simulation results that demonstrate our corrected version of Fst is unbiased to coverage.

Estimators of θ: As above in our treatment of heterozygosity, we need to correct estimations of the neutral mutation parameter $\theta = 4Nu$ for our pooled sampling strategy. Some recent work on this problem has been done by Futschik and Schlotterer (2010), who consider the case of pooled samples when the pool is large in comparison to the sequence coverage obtained. Here and in the supplemental materials, we derive results for corrected estimators which are accurate in the case where coverage is of similar size to the pooled sampled. Importantly, we are able to derive the expected site frequency spectrum of our pooled sequencing experiment.

The first result of interest is the probability of observing an allele segregating at frequency k out of m in our sequenced sample, given a pooled sample size of n. This will differ from the quantity in equation 1, because we will sum over possible allele frequencies of the A_1

allele in the sample, i, in accordance with their expected probabilities under the standard neutral model. Thus the unconditional probability is

$$\begin{aligned} Prob(k|m,n) &= \sum_{i=1}^{n-1} Prob(K|m,n,i)Prob(i) \\ &= \sum_{i=1}^{n-1} \binom{m}{k} (i/n)^k (1-i/n)^{m-k} (1/ia_n) \qquad (3) \end{aligned}$$

where $a_n = \sum_{j=1}^{n-1} 1/j$. The last term in equation 3 represents the probability of observing an allele segregating at frequency i out of n chromosomes under the neutral model (EWENS 2004). FU (1995) was able to derive the expected number of sites, X_i segregating at frequency i out of n as $E\{X_i\} = \theta/i$. While Fu derived his result from modeling the genealogical process as a form of the Polya urn model, a simpler derivation comes by conditioning on the total number of segregating sites in a sample, S. Conditional on S, the X_is can be assumed to follow a multinomial distribution where the individual parameters reflect the expected frequencies of sites in the sample. Using this logic then, $E\{X_i\} = E\{S\} \times Prob(i) = \theta a_n \times 1/ia_n = \theta/i$. Similarly we can write down the expected counts of each frequency class in our sequenced sample Y_i,

$$\begin{aligned} E\{Y_k\} &= E\{S\} \times Prob(k|m,n) \\ &= \theta a_n \sum_{i=1}^{n-1} \binom{m}{k} (i/n)^k (1-i/n)^{m-k} (1/ia_n) \qquad (4) \end{aligned}$$

We point the reader to the supplement for simulation results confirming the accuracy of this expression. With the expected site frequency spectrum in hand, we can use the weighted linear combination of ACHAZ (2009) to write down estimators of θ given our sampling regime. In particular given the high sequencing error rates inherent in these data, we want modified estimators of θ, which excluded singletons.

Modified versions of Tajima's nucleotide diversity ($\hat{\theta}_\pi$) and Fay and Wu's ($\hat{\theta}_H$) (TAJIMA 1983; FAY and WU 2000) were computed as follows. Let Y_k represent the number of sites segregating in a region at derived frequency k out of m reads, given a pool of n chromosomes. One can write down an unbiased estimator of θ using arbitrary weights for each frequency class ω_i, such that

$$\hat{\theta}_\omega = \frac{1}{a_n \sum_k \omega_k} \sum_{k=1}^{m-1} \omega_k Y_k \frac{1}{Prob(k|m,n)}. \tag{5}$$

This result allows for generalized weighted estimators of θ given pooled sampling. We present simulation results in the supplement that demonstrate our new estimators are accurate and unbiased with respect to coverage. In the present case, we are interested in two weighting schemes, one to create a modified $\hat{\theta}_\pi$ and the other for a modified $\hat{\theta}_H$ estimator. Let the associated weights be $\omega_{\pi,k}$ and $\omega_{H,k}$, respectively. Then

$$\omega_{\pi,k} = \begin{cases} 0 & k = 1 \\ m-k & 1 < k \leq m-1 \end{cases}$$

and

$$\omega_{H,k} = \begin{cases} 0 & k = 1 \\ k & 1 < k \leq m-1. \end{cases}$$

The modified Fay and Wu's H that excludes singleton sites is the difference between our two estimators. As our estimators are unbiased with respect to coverage, $\hat{\theta}$ over a region where m (coverage) varies is simply the sum of $\hat{\theta}$ at each m.

Outlier approach The relative merit of a model-based inference from theory or simulations vs. an empirical genomic-based outlier approach for detecting targets of positive selection is an ongoing discussion in the literature (Beaumont and Nichols 1996; Akey *et al.* 2002; Beaumont and Balding 2004; Teshima *et al.* 2006; Voight *et al.* 2006; Pickrell *et al.* 2009). For the following reasons, we chose to use an empirically-based outlier approach for identifying candidate targets of selection: 1) the challenges associated with generating a realistic null model for our *D. melanogaster* cline are substantial; 2) we have relatively few data from which to estimate model parameters; 3) there is little doubt that many of the highly-differentiated genomic regions from the east Australian cline result from selection, and 4) the empirical approach represents a simple, transparent treatment of the data. The many consistent biological signals we report here support the value of this approach, though they do not speak to its optimality.

Because the true length distribution of differentiated regions is unknown, two main approaches were used to identify such regions. Mean F_{ST} values were calculated for 1 kb non-overlapping windows across the normally-recombining regions of the genome. The top 1% or top 2.5% of these windows were considered "differentiated" for most analyses. For some analyses, the 5% tail was used (see Fig. S1a and results section below). To identify differentiation on a scale greater than 1 kb, we aggregated 1 kb windows in our top 1% tail. We considered any region of at least 5 consecutive windows that were not in the top 10% of mean 1 kb F_{ST} as "undifferentiated" between Queensland and Tasmania. Any region between two undifferentiated regions that had at least one 1 kb window in the top 1% F_{ST} was considered an independent differentiated region. We additionally investigated very small-scale differentiation by considering the top 0.1% of individual-position F_{ST} values not occurring in the top 10% 1 kb windows as potential outlier variants. Unless otherwise noted, all analyses were restricted to outliers occurring in normally recombining regions.

Genome Annotations were taken from FlyBase R5.24 (Tweedie *et al.* 2009). Genome positions were annotated as coding sequence (CDS), 3'- and 5'-UTR, intron, regulatory and "other." Because regulatory regions are under-represented in the FlyBase annotation, additional regulatory annotations were retrieved from the OregAnno database (Griffith *et al.* 2008) and a recent genome-wide scan for transcription-factor binding sites (MacArthur *et al.* 2009). Polymorphisms within coding sequence were additionally annotated as either nonsynonymous or synonymous.

Gene Ontology (GO) annotations (Ashburner *et al.* 2000) were obtained from FlyBase R5.24 (Tweedie *et al.* 2009). For each GO annotation, the number of genes within all 1 kb normally-recombining windows with that annotation were identified. GO-category enrichment was determined using a hypergeometric test that compared the proportion of genes with a given GO annotation to the proportion of genes in the 2.5% most-differentiated 1 kb windows with that GO annotation. All GO categories with < 4 genes were excluded, as 4 genes is the minimum number for which a significant hypergeometric result is possible at α=0.05. After controlling the false discovery rate using the method of Storey (2002), enriched GO categories with FDR-corrected P-values < 0.05 were determined. Similar GO-category enrichment analyses were performed using individual outlier genomic positions. Of course, differentiation at specific genes could have profound phenotypic consequences without leaving a statistically-significant signature of GO enrichment.

Copy-Number Variation was evaluated by calculating the mean coverage for nonoverlapping 1 kb windows across Queensland and Tasmania genomes. For each window, we calculated the ratio of Queensland/Tasmania coverage and normalized these ratios by the mean coverage ratio across each chromosome arm. The top 1%, 2.5% and 5% most-extreme windows were considered highly-differentiated in copy number (see Fig. S1b). Gene Ontology enrichment analyses were conducted as described above.

Structure Prediction: RNA secondary structures were inferred using the Vienna RNA package v1.8.2 (Hofacker 2003) with default parameters. Protein domain architecture was inferred using a sequence search of the PFam database (Coggill *et al.* 2008; Finn *et al.* 2010). Homology-based 3D structural modeling was performed using MODELLER 9v7 (Eswar *et al.* 2008). Structures were inferred for predicted proteins from a consensus sequence for Queensland and Tasmania genes *Irc* and *NtR*. Searching the Protein Data Bank (Berman *et al.* 2000) using *melanogaster* protein sequences returned structures 3ERH

(Sheikh *et al.* 2009) and 2QC1 (Dellisanti *et al.* 2007) as the best matches to the predicted proteins of *Irc* and *NtR*, respectively. Queensland and Tasmania consensus protein sequences were aligned to each structural template using MAFFT v6.611 with the E-INS-i option (Katoh *et al.* 2002; Katoh and Toh 2008). Five structural models of each sequence were constructed and evaluated using the MODELLER objective function as well as DOPE and GA341 assessment scores (Eramian *et al.* 2008). Results are shown for the best overall models. Sequence not alignable to the structural template was excluded.

RESULTS

After filtering, the average genome coverage was 11.6× in Queensland and 8.2× in Tasmania. Coverage varied little across chromosome arms (Fig. 1). The Queensland/Tasmania coverage ratio was highly consistent, varying from 1.20 to 1.45 across all regions examined. In addition, coverage in normally-recombining regions was nearly equivalent across chromosome arms: the *X* chromosome had the greatest coverage (11.3 and 8.0 in Queensland and Tasmania, respectively), while chromosome *2L* had the lowest (10.4 and 7.3). After filtering, the mean coverage and mean number of SNPs per 1 kb window were 604.7 bp and 9.4, respectively.

[Figure 1 about here.]

Genomic Patterns: Mean F_{ST} across the entire genome was $0.112 \pm 8.23 \times 10^{-5}$. The distribution of 1 kb window F_{ST} estimates has a long right tail (see Fig. S1a); the 5%, 2.5% and 1% thresholds for this tail are F_{ST}=0.23, F_{ST}=0.27 and F_{ST}=0.32, respectively. Among-arm variation in F_{ST} was significantly heterogeneous (Kruskal-Wallis rank sum test: $p < 2.2 \times 10^{-16}$, see also Table S1); the rank order of mean F_{ST} across chromosome arms was *3R* (0.124) > *2L* (0.116) > *3L* (0.111) > *2R* (0.107) > *X* (0.097). Previous studies have demonstrated that *In(3R)P* vs. Standard represents a nearly fixed difference between Queensland and Tasmania (corresponding to F_{ST} close to 1.0), which is considerably greater differentiation than that observed for other cosmopolitan inversions in these populations

(Knibb *et al.* 1981). This suggests that the *In(3R)P* cline is a main cause of the elevated F_{ST} for *3R*. Two aspects of the data support this proposition. First, the region spanned by *In(3R)P* was significantly more differentiated than the rest of *3R* (0.129 vs. 0.113, Wilcoxon rank sum test: $p < 2.2 \times 10^{-16}$, see Figs. 2c,S2). Second, the physical scale of differentiation was significantly greater on chromosome arm *3R*, which exhibited slightly fewer very-small differentiated regions (<2 kb) and significantly more large regions of high-F_{ST} (>10 kb) compared to the other arms (Fisher's Exact Test, $p = 0.000378$, Fig. 2b). Note that F_{ST} of nucleotide variation in the region spanned by *In(3R)P* was dramatically lower than estimates of F_{ST} of the inversion itself, based on previous studies of these populations (Knibb *et al.* 1981; Knibb 1982; Umina *et al.* 2005), suggesting extensive recombination in the history of this arrangement.

In(2L)t also shows clinal variation, though not as steep as that of *In(3R)P* (Knibb *et al.* 1981). There was also a significant difference in F_{ST} for the region spanned by *In(2L)t* (0.116) versus the rest of the arm (0.109) (Wilcoxon rank sum test: $p < 2.2 \times 10^{-16}$), however it appears that most of the difference is explained by the region of low-differentiation in the uninverted region adjacent to the centromere (see Fig. S2). The other two autosomal arms similarly showed only very slightly higher F_{ST} (*3L*) or no difference in F_{ST} (*2R*) for regions spanned by cosmopolitan inversions (there is no such inversion on the *X* chromosome). Much of the difference between Standard and inverted regions for arms other than *3R* is explained by reduced heterozygosity and differentiation in centromere proximal regions that are not included in the inversions (see Fig. S2).

Despite the filtering of regions corresponding to reduced heterozygosity as defined by DPGP, we observed that regions near centromeres (and some telomeres) showed low levels of differentiation, which corresponds to regions of reduced heterozygosity (see Fig. S2). This suggests that some centromere- and telomere-proximal euchromatic sequence experiencing reduced crossing-over may remain in our filtered data. However, the physical scale of differentiated regions was similar in normally- vs. low-recombining regions of the genome (Fig.

2a).

We detected significant heterogeneity in levels of nucleotide diversity ($\hat{\theta}_\pi$) among chromosome arms (Kruskal-Wallis rank sum test: $p < 2.2 \times 10^{-16}$, see also Table S1), with the X chromosome showing the lowest diversity. We also detected systematic differences in nucleotide diversity between population samples, with the Tasmanian population showing consistently lower heterozygosity than the Queensland sample (see Table S1). Additionally, Fay and Wu's H statistic was significantly more negative for Tasmania than for Queensland in both the genome as a whole (Wilcoxon rank sum test: $p < 2.2 \times 10^{-16}$; see Fig. S3) and for the normally-recombining portion of the genome (Wilcoxon rank sum test: $p < 2.2 \times 10^{-16}$). One explanation for the more negative Fay and Wu's H statistic in Tasmania is recent strong selection in this temperate population (Fay and Wu 2000). Consistent with this explanation, we found that the 1 kb regions that were very-highly differentiated also exhibited considerably more negative values of H in Tasmania compared to Queensland, relative to the rest of the genome (Wilcoxon rank sum tests: 5% tail, $p < 2.2 \times 10^{-16}$; 2.5% tail, $p < 2.2 \times 10^{-16}$; 1% tail, $p < 2.2 \times 10^{-16}$).

[Figure 2 about here.]

The largest differentiated euchromatic region spanned 854 kb at the tip of the X chromosome (Fig. 3a), a region of low heterozygosity documented in several studies (Aguade *et al.* 1989; Begun and Aquadro 1995; Langley *et al.* 2000). Interestingly, previous studies suggested that the scale of linkage disequilibrium in this region of the genome is not dramatically reduced, in spite of reduced levels of crossing-over (Begun and Aquadro 1995; Langley *et al.* 2000). This suggests that differentiation at the tip-of-the X region corresponds to a mosaic linkage-disequilibrium structure of relatively low small-scale linkage disequilibrium interspersed with scattered large-scale linkage disequilibrium. The largest differentiated segment in the middle of a chromosome arm was a 752 kb region of chromosome *2R* (Fig. 3b). Interestingly, *Cyp6g1*, an insecticide resistance gene (Daborn *et al.* 2002;

Schmidt *et al.* 2010) known to be under recent strong selection, is located in this region and is an excellent candidate for the observed differentiation. Other areas of extended differentiation were observed in the euchromatic portion of the *X* chromosome (a 245 kb region from 18,055kb to 18,300kb) and toward the proximal end of chromosome *2L* (a 131 kb region from 20,172kb to 20,303kb).

[Figure 3 about here.]

The majority of differentiation between the Queensland and Tasmania populations occurs on a small physical scale (see Fig. 2a-b, Table S1). In fact, F_{ST}-outlier regions (see Materials and Methods) were defined by single 1 kb windows in most cases, and most such windows localize to single genes. This small-scale differentiation facilitates effective identification of candidate genes influenced by spatially-varying selection. Figure 4 shows one example in which a 1 kb windows in the top 2.5% F_{ST} tail localizes to *Sfmbt*, a chromatin-binding protein involved in gene regulation (Grimm *et al.* 2009). Differentiation in this gene is primarily attributable to two fixed substitutions in the middle of the gene. Interestingly, *Sfmbt* has been shown through yeast two-hybrid studies to physically interact with 7 other genes (Yu *et al.* 2008), two of which—*CG33275* and *CG17018*—are also highly differentiated between Queensland and Tasmania (1 kb F_{ST}=0.26 and 0.45, respectively). Two additional genes predicted to interact with *Sfmbt* based on known interactions between human homologs—*Hdac3* and *Stam*—are also highly differentiated (1 kb F_{ST}=0.28 and 0.33, respectively).

A genome browser displaying 1 kb windows and their associated F_{ST} estimates is available at http://altair.dartmouth.edu/ucsc/index.html. Significantly differentiated regions showed substantial overlap with outlier regions previously identified in similar Australian samples using comparative genomic hybridization (Turner *et al.* 2008). For example, the proportion of Turner *et al.*'s outlier regions at FDR=0.001 that overlap at least one 1 kb window in our 2.5% or 5% F_{ST} tail was 34% and 58%, respectively.

[Figure 4 about here.]

Differentiation Across Genome Annotations: Among CDS, intron, 5'UTR, 3'UTR, regulatory, and unannotated parts of the genome, mean F_{ST} was highest for 3'UTR (Fisher's Exact Test, $p = 0.0007346$), in spite of the lower power associated with the small size of UTR sequence. Moreover, 3'UTRs were consistently over-represented in the tail of highly-differentiated 1 kb windows (Fig. 5). In contrast, coding sequence and introns were consistently under-represented in the most-differentiated genomic regions. Regions not annotated as either genic or regulatory were also highly enriched in the most-differentiated regions, though less so than 3'UTRs. Interestingly, regulatory regions and 5'UTRs were moderately over-represented in highly-differentiated autosomal regions but under-represented on the X chromosome.

[Figure 5 about here.]

To investigate general biological patterns associated with the observed 3'UTR differentiation, F_{ST} was calculated for each 3'UTR, which was followed by a Gene Ontology enrichment analysis for the genes associated with the top 1% most-differentiated 3'UTRs. This analysis revealed no significant enrichments, which was not unexpected given the limited functional annotations associated with most of the genes. However, a number of highly-differentiated 3'UTRs were associated with either transcriptional regulators or genes involved in protein phosphorylation, supporting an important role for regulatory evolution in Queensland vs. Tasmania differentiation. Other genes with highly-differentiated 3'UTRs code for proteins involved in energy metabolism, development, or seminal fluid (see Table S2).

An example of a gene exhibiting highly-localized 3'UTR-differentiation is *Hex-t2*, a testis-specific hexokinase (Duvernell and Eanes 2000). Figure 6 shows that there is a small region of elevated differentiation toward the 3' end of *Hex-t2*, with peak differentiation occurring in the 3'UTR. Within this differentiated region are two polymorphic sites in the Queensland population (a U/A polymorphism at position 75 in the UTR and an A/G polymorphism at position 55) that are fixed for the minor allele in Tasmania. Computational prediction of the RNA secondary structure of this 3'UTR suggests that the Tasmania fixations induce a marked change in RNA secondary structure, consistent with potential functional importance.

[Figure 6 about here.]

Protein-Coding Differentiation: Despite the fact that many outlier F_{ST} windows fall within exons, coding sequence was not overrepresented in the 1 kb window F_{ST} tail. However, because the windowing analysis does not account for the possibility of different physical scales of selection in DNA sequence space and protein space, alternative methods of characterizing protein differentiation were explored. First, mean F_{ST} for nonsynonymous variants in each gene in the normally recombining portion of the genome was calculated, with the top 1% of individual-gene nonsynonymous F_{ST} considered as coding for highly-differentiated proteins. This analysis favors smaller genes/proteins, for which differentiation is likely to be gene/protein-wide. Alternatively, large multidomain proteins might show significant differentiation only in specific functional domains. To investigate this possibility, the PFam database (Finn *et al.* 2010) was used to annotate known functional domains for all *D. melanogaster* genes. Mean nonsynonymous-F_{ST} was calculated separately for each domain in a gene, with the maximum domain-F_{ST} being recorded for each gene.

Tables S3 and S4 list the top candidate genes from these analyses, which suggest a number of interesting protein-coding genes for further study. For example, figure 7a shows elevated differentiation around a fixed amino-acid difference at position 47 in the disulfide oxidoreductase gene *Txl*. A threonine residue in Tasmania that is conserved throughout *Drosophila* has changed to alanine in Queensland, leading to elevated F_{ST} throughout the first exon. The alanine allele has also been observed in African *melanogaster* populations (DPGP.org). This may represent a more unusual case of recent selection in tropical populations (Queensland and Africa) rather than temperate adaptation.

We also observed elevated F_{ST} around a nonsynonymous fixed substitution in *Irc* (Fig.

7b), an immune-related catalase required to protect flies from microbial infection (Ha *et al.* 2005; Ha *et al.* 2005). Although the observed V317I substitution in Tasmania is conservative and occurs in a disordered loop region, this position is in direct ligand contact in the protein structure, suggesting a potential functional role in modulating molecular interactions (Fig. 7c). Alternatively, these changes could be affecting pre-mRNA processing. The two fixed substitutions in Tasmanian *Irc* are the nonsynonymous V317I change at the 5' end of exon 6 and a synonymous G→A substitution 11 bases downstream. These changes could be involved in splicing regulation, as RNA secondary structure prediction suggests that they could produce a radical reorganization of pre-mRNA structure (see Fig. S4).

[Figure 7 about here.]

One of the most differentiated protein domains in the genome is the ligand binding domain of the *NtR* gene, an extracellular ligand-gated ion channel. Figure 8a shows a large number of polymorphisms across *NtR*, along with a cluster of three amino acid variants in the ligand binding domain. The most differentiated of these variants is an I/V polymorphism for which the major allele in Queensland (I, frequency=0.73) is the minor allele in Tasmania (frequency 0.1); F_{ST} for this site is 0.51. The remaining amino acid polymorphisms in this domain are an L/F polymorphism (F_{ST}=0.14) and an E/D polymorphism (F_{ST}=0.19). While L is the major allele in both populations at the first position, the E/D Queensland polymorphism is fixed for D in Tasmania. Structural homology modeling suggests that this E/D polymorphism occurs in the main immunogenic region (MIR) of the protein (Fig. 8b). This region constitutes a loop sandwiched between $\beta2$ and $\beta3$ that binds autoimmune antibodies in myasthenia gravis patients in the homologous human muscle acetylcholine receptor (Tsouloufis *et al.* 2000; Dellisanti *et al.* 2007). The fact that the I/V polymorphism is found in close proximity to this region suggests the possibility that differentiation at *NtR* could affect interactions with other molecules, possibly those relating to the immune system.

[Figure 8 about here.]

Biological Patterns Underlying Genic Differentiation: The extensive genetic interactions and pleiotropic effects of laboratory mutations in *Drosophila* genes make it challenging to reliably infer from differentiated genes the phenotypes that may be targets of selection. Nevertheless, the small physical scale of differentiation makes it worthwhile to explore general patterns in the data as a means of generating hypotheses regarding pathways and phenotypes that might experience spatially-varying selection in Australian *melanogaster* populations. Our approach was to test for enrichment of GO terms among the genes that overlapped a 1 kb window in the upper 2.5% tail of the distribution, which corresponds to $F_{ST} > 0.27$. These analyses were supplemented by inspection of genetic interactions annotated in FlyBase. We also point to plausible candidates in the 5% tail where appropriate.

Several high-F_{ST} windows overlapped genes functioning in central *Drosophila* signaling pathways, including the *JAK-STAT* pathway, the *torso* pathway, the *EGFR* pathway and the *TGF-B* pathway. In the *JAK-STAT* pathway the ligand *upd2* (1 kb F_{ST}=0.70) and *STAT* (*Stat92E*, 1 kb F_{ST}=0.32) both showed elevated F_{ST}, as did *CycE* (1 kb F_{ST}=0.25) and *Ptp61F* (1 kb F_{ST}=0.28), which regulate that pathway. Other modifiers of *JAK-STAT* signaling that overlapped high-F_{ST} windows included *crb* (1 kb F_{ST}=0.35), *tkv* (1 kb F_{ST}=0.39), *Mad* (1 kb F_{ST}=0.35), and *Stam* (1 kb F_{ST}=0.33). Highly-differentiated genes in the *torso* signaling pathway (which regulates several processes, including metamorphosis and body size) included *tup* (1 kb F_{ST}=0.41), *Gap1* (1 kb F_{ST}=0.26), *pnt* (1 kb F_{ST}=0.60), *tld* (1 kb F_{ST}=0.25) and *csw* (1 kb F_{ST}=0.26). Differentiated genes in the *EGFR* signaling pathway included *vn* (1 kb F_{ST}=0.27), *argos* (1 kb F_{ST}=0.23), *sprouty* (1 kb F_{ST}=0.29), *Star* (1 kb F_{ST}=0.29) and *ed* (1 kb F_{ST}=0.30). Genes in the *TGF-B* pathway were also overrepresented among high-F_{ST} windows and included *dally* (1 kb F_{ST}=0.39), *Mad* and *tkv* (1 kb F_{ST}=0.39). The gene *dpp*, which is centrally located in this pathway, also contained a region of high differentiation (1 kb F_{ST}=0.24). Finally, the hypothesis that ecdysone signaling experiences spatially-varying selection is supported by highly-differentiated windows overlapping the ecdysone receptor, *EcR* (1 kb F_{ST}=0.25), the eclosion hormone gene *Eh* (1 kb F_{ST}=0.33), *Moses* (1 kb F_{ST}=0.41), *taiman* (1 kb F_{ST}=0.37) and the ecdysone-induced protein-coding genes *Eip63E* (1 kb F_{ST}=0.33), *Eip74EF* (1 kb F_{ST}=0.31), *Eip75B* (1 kb F_{ST}=0.30) and *Eip93F* (1kb F_{ST}=0.44). It is worth noting that substantial crosstalk exists between some of these pathways, and that other genes associated with key pathways such as *Notch* show evidence of differentiation in our data.

These results support the existence of pervasive spatially-varying selection acting at key genes throughout multiple *Drosophila* signaling pathways. It is highly plausible that several candidates influence clinal variation in body size, metabolism, and additional important life history traits (see Supplementary Table S5 for a complete list of enriched GO terms). Many genes implicated in body-size variation were highly differentiated, including *InR* (1 kb F_{ST}=0.26, (Paaby *et al.* 2010)), *dally* (1 kb F_{ST}=0.39), *Orct2* and *Pi3K21B* at the tip of *2L*, which contains a highly-differentiated 1 kb window (F_{ST}=0.28) but was not included in most of our analyses because of its location at the distal end of the chromosome arm. Interestingly, many body-size candidate genes revealed by our analysis are located on chromosome arm *3R*, which is consistent with previous genetic analyses showing that most of the body-size variation associated with the Australian cline is inseparable from *In(3R)P* in mapping crosses (Rako *et al.* 2006; Rako *et al.* 2007). Our data—including evidence of extensive recombination between standard and *In(3R)P* arrangements—suggests that the differentiated genes that are located on *3R* are particularly promising targets for investigating the genetic basis of body-size variation in *D. melanogaster*.

A large number of GO terms related to developmental processes are enriched for F_{ST} outliers. The associated genes contribute to many phenotypes, including external morphology (e.g., wing, eye), nervous system development, ovarian follicle development, larval development and embryonic development. The *Toll* signaling pathway, which contains a number of immune-system genes, is enriched. The immunity gene *sick* is also in the 5% tail of F_{ST} windows. Olfactory behavior and olfactory learning are enriched in 1 kb outlier tails. In addition, a number of F_{ST}-outlier nonsynonymous SNPs not located in outlier windows are found in olfactory or gustatory receptors, or odorant binding proteins. Several ionotropic receptors, a new class of odorant receptors, appear in the 5% F_{ST} tail of 1 kb windows. It is interesting to note the evidence that thermal stress disrupts odor learning in flies (Wang *et al.* 2007) via developmental effects on the mushroom body, in light of the observation that "mushroom body development" is among the enriched GO terms in our analysis. A number of ion channel-related genes appear among the outlier 1 kb windows, leading to enrichment of GO categories: calcium-, potassium- and sodium-ion transport. "Calcium ion binding" is the second most significantly enriched molecular function and includes several Cadherins as well as Calmodulin. Selection associated with variation in the visual environment between Queensland and Tasmania is suggested by the enrichment of GO terms such as "phototransduction."

Although circadian rhythm genes are not overrepresented among the F_{ST} outliers, several genes relating to circadian biology are found among the most differentiated 1 kb windows. The cryptochrome gene, which regulates circadian rhythm, is highly differentiated (F_{ST}=0.30), as are *couch potato* (F_{ST}=0.23) and *timeless* (F_{ST}=0.20), which have already been implicated in spatially-varying selection in *D. melanogaster* (Sandrelli *et al.* 2007; Tauber *et al.* 2007; Schmidt *et al.* 2008). Another interesting candidate is *norpA*, a phospholipase C gene required for thermal synchronization of the circadian clock (Glaser and Stanewsky 2005). This gene is in the 2.5% F_{ST} tail and highly differentiated across its entire length (see Fig. 9a). Four of its 7 interacting partners annotated in FlyBase are also in the 2.5% tail (see Table S7). Additionally, *norpA* is known to regulate splicing in the 3'UTR of *per*, a central circadian-clock gene in *Drosophila* (Collins *et al.* 2004; Majercak *et al.* 2004) which shows a highly-localzed 3'UTR-elevation in F_{ST} in our data (Fig. 9b). Together, these results strongly suggest a cluster of correlated differentiation occurring across several genes at the interface between thermal- and light-entrainment of the circadian clock.

[Figure 9 about here.]

Finally, transcription and chromatin regulation appear to be under widespread selection, as seven related biological process GO terms are enriched among the F_{ST} outlier windows. Additionally, "transcription factor" is the second most significantly-enriched GO molecular function term. Particularly interesting differentiated genes include *Trl*, *HDAC4*, *additional sex combs*, *Enhancer of polycomb*, *histoneacetlytransferase Tip60*, *Ino80*, *JIL-1*, *14-3-3ϵ* and *Sfmbt*.

Copy-Number Variation: Differences in copy-number between Queensland and Tasmania were investigated using an outlier approach analogous to that used for F_{ST}. The normalized ratio of Queensland/Tasmania coverage for 1 kb non-overlapping windows was calculated across the genome (see Materials and Methods), with the top 1% most-extreme estimates considered highly-differentiated regions. Note that frequency variation and ploidy-level variation are confounded in this analysis. Relative to the genome-wide average of copy-number differentiation, slightly more than half (55%) of the 1 kb windows had more coverage in the Queensland population. However, significantly more (62.5%) of the highly-differentiated windows showed increased copy number in the Tasmania population (P=2.2×10^{-16}), suggesting that duplication events could be important for local adaptation in Tasmania.

The largest region exhibiting significant copy-number variation (CNV) is a 107 kb region of chromosome *3R* (Fig. 10), which spans a small number of protein-coding genes including the last few exons of *timeout* and the entire *Ace* gene. *Ace* codes for an acetylcholinesterase associated with pesticide resistance (Menozzi *et al.* 2004), which was previously identified as a differentiated CNV between these populations (Turner *et al.* 2008). Interestingly, *Ace* expression has been shown to vary over the circadian cycle (Hooven *et al.* 2009), and acetylcholinesterase levels are highly correlated with pesticide resistance (Charpentier and Fournier 2001).

[Figure 10 about here.]

Gene Ontology enrichment analysis of genes found in highly-differentiated CNV regions revealed categories similar to those observed for our F_{ST} enrichment analysis (see Table S6), including transcription factors and ion-channel genes. Across both GO-enrichment analyses, 185 unique GO terms were enriched, 66 of which (36%) were found in both analyses. Interestingly, despite the large degree of overlap between GO enrichment terms in the F_{ST} and CNV analysis, the specific genes associated with each enriched GO category did not overlap to a large degree. Of the 719 genes in the copy-number 1% outlier set and the 551 genes in the corresponding F_{ST} outlier set, only 72 (6%) were found in both (as expected given the upper bound of coverage included in the F_{ST} analysis). This suggests the possibility that selection may often result in recruitment of alleles resulting from both nucleotide and copy-number differences. Several terms enriched in the CNV GO analysis did not appear in the F_{ST} GO enrichment, including "circadian rhythm," "sex determination," "courtship and mating behavior," "female meiosis chromosome segregation" and "chorion-containing eggshell formation" (which was also detected by TURNER *et al.* (2008)).

DISCUSSION

A large body of evidence supports the idea that much of the phenotypic and genetic differentiation along the Australian *D. melanogaster* latitudinal cline is driven by spatially-varying selection (OAKESHOTT *et al.* 1981; OAKESHOTT *et al.* 1983; SINGH and RHOMBERG 1987; SINGH 1989; SINGH and LONG 1992; GOCKEL *et al.* 2001; KENNINGTON *et al.* 2003; HOFFMANN and WEEKS 2007). Here we have presented the first genome-sequence based analysis of population differentiation associated with this cline. Although our analysis included only populations from each end of the cline, it is likely that the set of highly-differentiated genomic regions between these cline endpoints is considerably enriched for targets of spatially-varying selection. Indeed, the fact that the most highly-differentiated genomic regions show much more negative Fay and Wu's H estimates in Tasmania is consistent with the hypothesis that the observed differentiation is associated with recent strong selection in temperate populations (SEZGIN *et al.* 2004). The dramatic enrichment of several

GO terms among the genes overlapping differentiated regions also supports the notion that selection plays a major role, because it is difficult to envision a neutral demographic process that could result in such enrichment patterns.

Two main lines of evidence support the proposition that gene regulation is an important target of spatially-varying selection in these populations. First, 3'UTRs and unannotated sequence are the most over-represented sequence classes among the outlier 1 kb F_{ST} windows. 3'UTRs, which exhibit the strongest enrichment in our analysis, play an important role in gene regulation (Lai 2002; Kuersten and Goodwin 2003; de Moor *et al.* 2005; Stark *et al.* 2005; Chatterjee and Pal 2009; Mangone *et al.* 2010). Recent studies have found substantial cis-acting effects on regulatory variation in *Drosophila* (Hughes *et al.* 2006; Lawniczak *et al.* 2008; Lemos *et al.* 2008; Graze *et al.* 2009; McManus *et al.* 2010); our results raise the intriguing possibility that variation in 3'UTRs may make a significant contribution to adaptive cis-acting regulatory variation. The over-representation of noncoding DNA among F_{ST} outlier windows is consistent with previous population genetic results supporting the importance of noncoding sequence for adaptive divergence over longer time scales in *Drosophila melanogaster* (Andolfatto 2005). It will be interesting to investigate these currently unannotated regions in the context of ongoing efforts to improve the annotation of the *D. melanogaster* genome (Celniker *et al.* 2009). The second line of evidence supporting the importance of selection on gene regulation along the cline is the finding that transcription and chromatin-related genes are among the most differentiated in the genome, which is consistent with previous analyses of these populations (Levine and Begun 2008; Turner *et al.* 2008) and with genomic inferences on the importance of recurrent directional selection on proteins regulating chromatin and transcription in *D. simulans* (Begun *et al.* 2007).

Although protein-coding sequence was underrepresented among the most extremely-differentiated 1 kb windows, one should not conclude that amino acid variants are unimportant for selection along the cline, as a large number of outlier windows overlap coding sequence. It is interesting to consider possible population-genetic explanations for why CDS is underrepresented. The timescale of differentiation between Queensland and Tasmanian populations is very small, perhaps on the order of 1000 generations (Hoffmann and Weeks 2007). Because the mutation rate per base pair is small, much of the selective response during the initial colonization of Australia was likely the result of frequency changes of alleles already segregating in ancestral populations rather than from invasion into the populations of new mutations that occurred subsequent to colonization. Whole-genome surveys of polymorphism in *Drosophila* suggest that nonsynonymous sites are several-fold less polymorphic than synonymous or non-coding sites (e.g. Begun *et al.* (2007),Sackton *et al.* (2009)). Thus, on a per-site basis compared to noncoding variants, amino acid variants are considerably less available to selection on standing variation following a radical change of the environment. The physical scale of differentiation predicted under the selection-on-standing-variation model depends on the amount of linkage disequilibrium associated with the site destined to experience selection after the environment changes. Surveys of linkage disequilibrium in normally-recombining regions from large samples of cosmopolitan *D. melanogaster* consistently find that sites in strong linkage disequilibrium tend to be within 2 kb of each other (Miyashita and Langley 1988; Palsson *et al.* 2004; Macdonald *et al.* 2005). This is consistent with the scale of geographic differentiation observed in our data and with the hypothesis that much of the observed differentiation between temperate and tropical populations is the result of recent strong selection on standing variants. Genomic data on the frequency distribution of variation and the scale of linkage disequilbrium from populations along the Australian cline and from African and European populations should provide the resources necessary for addressing issues relating to the geographic origins, frequencies and fitnesses of variants experiencing selection in Australia.

One of the general findings from our analysis is that many genes and pathways centrally important to *Drosophila* biology appear to experience spatially-varying selection. The fact that laboratory mutations in these genes and pathways tend to be highly pleiotropic is, in the conventional thinking, associated with reduced mutation rate to beneficial alleles. It is important to realize, however, that it is the individual mutation—rather than the gene—that is more or less pleiotropic. The distribution of pleiotropic effects of natural variants is likely to be quite different and dramatically smaller than those of laboratory mutations. Moreover, the large population sizes of *Drosophila* suggest that drift may be relatively unimportant, and that variants that reach appreciable frequencies may have special genetic and population-genetic properties. Thus, the candidate variants identified here may have very small pleiotropic effects, in spite of the fundamental biological roles of the corresponding genes. Alternatively, natural alleles that were pleiotropic along the axes favored by correlated natural selection would be strongly favored, and these too could constitute a considerable fraction of the variants in fundamental signaling pathways that show differentiation between these populations.

The genomic results regarding the dramatic biological differences between these fly populations raises the obvious question—unanswerable with these data—as to the phenotypic and fitness effects of the selected mutations and how the distribution of such effects may vary across biological functions and positions in genetic pathways. For example, one class of selected mutations may contribute to phenotypic differences between temperate and tropical flies, while a second—potentially larger—class exhibiting genotype x environment interactions may exhibit latitudinal clines, because different genotypes are required to produce a single optimal phenotype in different environments (e.g., (Levine *et al.* 2010)). Larger genomic datasets and functional analyses should produce much sharper inferences regarding the specific polymorphisms, pathways and biological functions that have diverged under selection between temperate and tropical populations and further reveal the genetic and population-genetic principles of adaptation in this model species.

ACKNOWLEDGMENTS

We thank Ary Hoffmann for generously sharing flies and A. Hoffmann and P. Schmidt for their thoughts on clinal variation in *Drosophila*. We thank Michael Nachman, J. Anderson and two anonymous reviewers for comments that improved this manuscript. We also gratefully acknowledge Charis Cardeno, Kristian Stevens, Melissa Eckert, and Thaddeus Seher for technical assistance, and Phil Nista for early contributions to the analysis. This work was funded by NIH grants GM071926 and GM084056 (DJB), by the Drosophila Population Genomics Project (Chuck Langley, PI), and by Dartmouth College and the Neukom Institute (ADK).

LITERATURE CITED

Achaz, G., 2009 Frequency spectrum neutrality tests: one for all and all for one. Genetics *183*(1)**:** 249–58.

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y. H. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, J. F. Abril, A. Agbayani, H. J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson, L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista,

C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M. H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kennison, K. A. Ketchum, B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L. Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacleb, M. Palazzolo, G. S. Pittman, S. Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. Saunders, F. Scheeler, H. Shen, B. C. Shue, I. Sidén-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z. Y. Wang, D. A. Wassarman, G. M. Weinstock, J. Weissenbach, S. M. Williams, WoodageT, K. C. Worley, D. Wu, S. Yang, Q. A. Yao, J. Ye, R. F. Yeh, J. S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X. H. Zheng, F. N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H. O. Smith, R. A. Gibbs, E. W. Myers, G. M. Rubin, and J. C. Venter, 2000, Mar)The genome sequence of Drosophila melanogaster. Science *287*(5461): 2185–95.

Aguade, M., N. Miyashita, and C. H. Langley, 1989, Jul)Reduced Variation in the Yellow-Achaete-Scute Region in Natural Populations of Drosophila Melanogaster. Genetics *122*(3): 607–615.

Akey, J. M., G. Zhang, K. Zhang, L. Jin, and M. D. Shriver, 2002, Dec)Interrogating a high-density SNP map for signatures of natural selection. Genome Res *12*(12): 1805–

1814.

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

Andolfatto, P., 2005, Oct)Adaptive evolution of non-coding DNA in Drosophila. Nature *437*(7062): 1149–1152.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, 2000, May)Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*(1): 25–29.

Beaumont, M. and R. Nichols, 1996 Evaluating loci for use in the genetic anlaysis of population structure. Proceedings of the Royal Society B: Biological Sciences *263*(1377): 1619–1626.

Beaumont, M. A. and D. J. Balding, 2004, Apr)Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol *13*(4): 969–980.

Begun, D., A. Holloway, K. Stevens, L. Hillier, Y. Poh, M. Hahn, P. Nista, C. Jones, A. Kern, C. Dewey, L. Pachter, E. Myers, and C. Langley, 2007, Nov)Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. PLoS Biol. *5*(11): e310.

Begun, D. J. and C. F. Aquadro, 1995, May)Evolution at the tip and base of the X chromosome in an African population of Drosophila melanogaster. Mol Biol Evol *12*(3): 382–390.

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, J. M. Boutell, J. Bryant, R. J. Carter, R. Keira Cheetham, A. J. Cox, D. J. Ellis, M. R. Flatbush, N. A. Gormley, S. J. Humphray, L. J. Irving, M. S. Karbelashvili, S. M. Kirk, H. Li, X. Liu, K. S. Maisinger, L. J. Murray, B. Obradovic, T. Ost, M. L. Parkinson, M. R. Pratt, I. M. J. Rasolonjatovo, M. T. Reed, R. Rigatti, C. Rodighiero, M. T. Ross, A. Sabot, S. V. Sankar, A. Scally, G. P. Schroth, M. E. Smith, V. P. Smith, A. Spiridou, P. E. Torrance, S. S. Tzonev, E. H. Vermaas, K. Walter, X. Wu, L. Zhang, M. D. Alam, C. Anastasi, I. C. Aniebo, D. M. D. Bailey, I. R. Bancarz, S. Banerjee, S. G. Barbour, P. A. Baybayan, V. A. Benoit, K. F. Benson, C. Bevis, P. J. Black, A. Boodhun, J. S. Brennan, J. A. Bridgham, R. C. Brown, A. A. Brown, D. H. Buermann, A. A. Bundu, J. C. Burrows, N. P. Carter, N. Castillo, M. Chiara E Catenazzi, S. Chang, R. Neil Cooley, N. R. Crake, O. O. Dada, K. D. Diakoumakos, B. Dominguez-Fernandez, D. J. Earnshaw, U. C. Egbujor, D. W. Elmore, S. S. Etchin, M. R. Ewan, M. Fedurco, L. J. Fraser, K. V. Fuentes Fajardo, W. Scott Furey, D. George, K. J. Gietzen, C. P. Goddard, G. S. Golda, P. A. Granieri, D. E. Green, D. L. Gustafson, N. F. Hansen, K. Harnish, C. D. Haudenschild, N. I. Heyer, M. M. Hims, J. T. Ho, A. M. Horgan, K. Hoschler, S. Hurwitz, D. V. Ivanov, M. Q. Johnson, T. James, T. A. Huw Jones, G.-D. Kang, T. H. Kerelska, A. D. Kersey, I. Khrebtukova, A. P. Kindwall, Z. Kingsbury, P. I. Kokko-Gonzales, A. Kumar, M. A. Laurent, C. T. Lawley, S. E. Lee, X. Lee, A. K. Liao, J. A. Loch, M. Lok, S. Luo, R. M. Mammen, J. W. Martin, P. G. McCauley, P. McNitt, P. Mehta, K. W. Moon, J. W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S. M. Novo, M. J. O'Neill, M. A. Osborne, A. Osnowski, O. Ostadan, L. L. Paraschos, L. Pickering, A. C. Pike, A. C. Pike, D. Chris Pinkard, D. P. Pliskin, J. Podhasky, V. J. Quijano, C. Raczy,

V. H. Rae, S. R. Rawlings, A. Chiva Rodriguez, P. M. Roe, J. Rogers, M. C. Rogert Bacigalupo, N. Romanov, A. Romieu, R. K. Roth, N. J. Rourke, S. T. Ruediger, E. Rusman, R. M. Sanches-Kuiper, M. R. Schenker, J. M. Seoane, R. J. Shaw, M. K. Shiver, S. W. Short, N. L. Sizto, J. P. Sluis, M. A. Smith, J. Ernest Sohna Sohna, E. J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C. L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovsky, S. M. Virk, S. Wakelin, G. C. Walcott, J. Wang, G. J. Worsley, J. Yan, L. Yau, M. Zuerlein, J. Rogers, J. C. Mullikin, M. E. Hurles, N. J. McCooke, J. S. West, F. L. Oaks, P. L. Lundberg, D. Klenerman, R. Durbin, and A. J. Smith, 2008, Nov)Accurate whole human genome sequencing using reversible terminator chemistry. Nature *456*(7218)**:** 53–9.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, 2000, Jan)The Protein Data Bank. Nucleic Acids Res *28*(1)**:** 235–242.

Blanchette, M., W. J. Kent, C. Riemer, L. Elnitski, A. F. A. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, and W. Miller, 2004, Apr)Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res *14*(4)**:** 708–15.

Celniker, S. E., L. A. Dillon, M. B. Gerstein, K. C. Gunsalus, S. Henikoff, G. H. Karpen, M. Kellis, E. C. Lai, J. D. Lieb, D. M. MacAlpine, G. Micklem, F. Piano, M. Snyder, L. Stein, K. P. White, R. H. Waterston, and modENCODE Consortium, 2009, Jun)Unlocking the secrets of the genome. Nature *459*(7249)**:** 927–930.

Chan, Y., M. Marks, F. Jones, G. Villarreal Jr, M. Shapiro, S. Brady, A. Southwick, D. Absher, J. Grimwood, J. Schmutz, and others, 2010 Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1

Enhancer. science *327*(5963)**:** 302.

Charpentier, A. and D. Fournier, 2001 Levels of Total Acetylcholinesterase in Drosophila melanogaster in Relation to Insecticide Resistance. Pesticide Biochemistry and Physiology *70*(2)**:** 100 – 107.

Chatterjee, S. and J. K. Pal, 2009, May)Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. Biol Cell *101*(5)**:** 251–262.

Coggill, P., R. D. Finn, and A. Bateman, 2008, Sep)Identifying protein domains with the Pfam database. Curr Protoc Bioinformatics **Chapter 2**.

Collins, B. H., E. Rosato, and C. P. Kyriacou, 2004, Feb)Seasonal behavior in Drosophila melanogaster requires the photoreceptors, the circadian clock, and phospholipase C. Proc Natl Acad Sci U S A *101*(7)**:** 1945–1950.

Colosimo, P., C. Peichel, K. Nereng, B. Blackman, M. Shapiro, D. Schluter, and D. Kingsley, 2004 The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biology *2*(5)**:** 635–641.

Consortium, D. . G., A. G. Clark, M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, T. A. Markow, T. C. Kaufman, M. Kellis, W. Gelbart, V. N. Iyer, D. A. Pollard, T. B. Sackton, A. M. Larracuente, N. D. Singh, J. P. Abad, D. N. Abt, B. Adryan, M. Aguade, H. Akashi, W. W. Anderson, C. F. Aquadro, D. H. Ardell, R. Arguello, C. G. Artieri, D. A. Barbash, D. Barker, P. Barsanti, P. Batterham, S. Batzoglou, D. Begun, A. Bhutkar, E. Blanco, S. A. Bosak, R. K. Bradley, A. D. Brand, M. R. Brent, A. N. Brooks, R. H. Brown, R. K. Butlin, C. Caggese, B. R. Calvi, A. Bernardo de Carvalho, A. Caspi, S. Castrezana, S. E. Celniker, J. L. Chang, C. Chapple, S. Chatterji, A. Chinwalla, A. Civetta, S. W. Clifton, J. M. Comeron, J. C. Costello, J. A. Coyne, J. Daub, R. G. David, A. L. Delcher, K. Delehaunty, C. B. Do, H. Ebling, K. Edwards, T. Eickbush, J. D. Evans, A. Filipski, S. Findeiss, E. Freyhult, L. Fulton, R. Fulton, A. C. L. Garcia, A. Gardiner, D. A. Garfield, B. E. Garvin, G. Gibson, D. Gilbert, S. Gnerre, J. Godfrey, R. Good, V. Gotea, B. Gravely, A. J. Greenberg, S. Griffiths-Jones, S. Gross, R. Guigo, E. A. Gustafson, W. Haerty, M. W. Hahn, D. L. Halligan, A. L. Halpern, G. M. Halter, M. V. Han, A. Heger, L. Hillier, A. S. Hinrichs, I. Holmes, R. A. Hoskins, M. J. Hubisz, D. Hultmark, M. A. Huntley, D. B. Jaffe, S. Jagadeeshan, W. R. Jeck, J. Johnson, C. D. Jones, W. C. Jordan, G. H. Karpen, E. Kataoka, P. D. Keightley, P. Kheradpour, E. F. Kirkness, L. B. Koerich, K. Kristiansen, D. Kudrna, R. J. Kulathinal, S. Kumar, R. Kwok, E. Lander, C. H. Langley, R. Lapoint, B. P. Lazzaro, S.-J. Lee, L. Levesque, R. Li, C.-F. Lin, M. F. Lin, K. Lindblad-Toh, A. Llopart, M. Long, L. Low, E. Lozovsky, J. Lu, M. Luo, C. A. Machado, W. Makalowski, M. Marzo, M. Matsuda, L. Matzkin, B. McAllister, C. S. McBride, B. McKernan, K. McKernan, M. Mendez-Lago, P. Minx, M. U. Mollenhauer, K. Montooth, S. M. Mount, X. Mu, E. Myers, B. Negre, S. Newfeld, R. Nielsen, M. A. F. Noor, P. O'Grady, L. Pachter, M. Papaceit, M. J. Parisi, M. Parisi, L. Parts, J. S. Pedersen, G. Pesole, A. M. Phillippy, C. P. Ponting, M. Pop, D. Porcelli, J. R. Powell, S. Prohaska, K. Pruitt, M. Puig, H. Quesneville, K. R. Ram, D. Rand, M. D. Rasmussen, L. K. Reed, R. Reenan, A. Reily, K. A. Remington, T. T. Rieger, M. G. Ritchie, C. Robin, Y.-H. Rogers, C. Rohde, J. Rozas, M. J. Rubenfield, A. Ruiz, S. Russo, S. L. Salzberg, A. Sanchez-Gracia, D. J. Saranga, H. Sato, S. W. Schaeffer, M. C. Schatz, T. Schlenke, R. Schwartz, C. Segarra, R. S. Singh, L. Sirot, M. Sirota, N. B. Sisneros, C. D. Smith, T. F. Smith, J. Spieth, D. E. Stage, A. Stark, W. Stephan, R. L. Strausberg, S. Strempel, D. Sturgill, G. Sutton, G. G. Sutton, W. Tao, S. Teichmann, Y. N. Tobari, Y. Tomimura, J. M. Tsolas, V. L. S. Valente,

E. Venter, J. C. Venter, S. Vicario, F. G. Vieira, A. J. Vilella, A. Villasante, B. Walenz, J. Wang, M. Wasserman, T. Watts, D. Wilson, R. K. Wilson, , 2007, Nov)Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*(7167)**:** 203–18.

Costa, R., A. A. Peixoto, G. Barbujani, and C. P. Kyriacou, 1992, Oct)A latitudinal cline in a Drosophila clock gene. Proc Biol Sci *250*(1327)**:** 43–49.

Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil, S. Jeffers, N. Tijet, T. Perry, D. Heckel, P. Batterham, R. Feyereisen, T. G. Wilson, and R. H. ffrench Constant, 2002, Sep)A single p450 allele associated with insecticide resistance in Drosophila. Science *297*(5590)**:** 2253–6.

de Bono, M. and C. I. Bargmann, 1998, Sep)Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in C. elegans. Cell *94*(5)**:** 679–689.

de Moor, C. H., H. Meijer, and S. Lissenden, 2005, Feb)Mechanisms of translational control by the 3' UTR in development and differentiation. Semin Cell Dev Biol *16*(1)**:** 49–58.

Dellisanti, C. D., Y. Yao, J. C. Stroud, Z. Z. Wang, and L. Chen, 2007, Aug)Crystal structure of the extracellular domain of nAChR alpha1 bound to alpha-bungarotoxin at 1.94 A resolution. Nat Neurosci *10*(8)**:** 953–962.

Duvernell, D. D. and W. F. Eanes, 2000, Nov)Contrasting molecular population genetics of four hexokinases in Drosophila melanogaster, D. simulans and D. yakuba. Genetics *156*(3)**:** 1191–1201.

Duvernell, D. D., P. S. Schmidt, and W. F. Eanes, 2003, May)Clines and adaptive evolution in the methuselah gene region in Drosophila melanogaster. Mol Ecol *12*(5)**:** 1277–1285.

Eramian, D., N. Eswar, M. Y. Shen, and A. Sali, 2008, Nov)How well can the accuracy of comparative protein structure models be predicted? Protein Sci *17*(11)**:** 1881–1893.

Eswar, N., D. Eramian, B. Webb, M. Y. Shen, and A. Sali, 2008 Protein structure modeling with MODELLER. Methods Mol Biol **426:** 145–159.

Ewens, W. J., 2004 *Mathematical population genetics* (2nd ed ed.), Volume v. 27. New York: Springer.

Fay, J. C. and C. I. Wu, 2000, Jul)Hitchhiking under positive Darwinian selection. Genetics *155*(3)**:** 1405–13.

Finn, R. D., J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy, and A. Bateman, 2010, Jan)The Pfam protein families database. Nucleic Acids Res *38*(Database issue)**:** 211–222.

Frydenberg, J., A. A. Hoffmann, and V. Loeschcke, 2003, Aug)DNA sequence variation and latitudinal associations in hsp23, hsp26 and hsp27 from natural populations of Drosophila melanogaster. Mol Ecol *12*(8)**:** 2025–32.

Fu, Y. X., 1995 Statistical Properties of Segregating Sites. Theoretical Population Biology **48:** 172–197.

Futschik, A. and C. Schlotterer, 2010 Massively Parallel Sequencing of Pooled DNA Samples–The Next Generation of Molecular Markers. Genetics.

Glaser, F. T. and R. Stanewsky, 2005, Aug)Temperature synchronization of the Drosophila circadian clock. Curr Biol *15*(15)**:** 1352–1363.

Gockel, J., W. J. Kennington, A. Hoffmann, D. B. Goldstein, and L. Partridge, 2001, May)Nonclinality of molecular variation implicates selection in maintaining a morphological cline of Drosophila melanogaster. Genetics *158*(1)**:** 319–323.

Graze, R. M., L. M. McIntyre, B. J. Main, M. L. Wayne, and S. V. Nuzhdin, 2009, Oct)Regulatory divergence in Drosophila melanogaster and D. simulans, a genomewide analysis of allele-specific expression. Genetics *183*(2): 547–561.

Griffith, O. L., S. B. Montgomery, B. Bernier, B. Chu, K. Kasaian, S. Aerts, S. Mahony, M. C. Sleumer, M. Bilenky, M. Haeussler, M. Griffith, S. M. Gallo, B. Giardine, B. Hooghe, P. Van Loo, E. Blanco, A. Ticoll, S. Lithwick, E. Portales-Casamar, I. J. Donaldson, G. Robertson, C. Wadelius, P. De Bleser, D. Vlieghe, M. S. Halfon, W. Wasserman, R. Hardison, C. M. Bergman, S. J. M. Jones, and Open Regulatory Annotation Consortium, 2008, Jan)ORegAnno: an open-access community-driven resource for regulatory annotation. Nucleic Acids Res *36*(Database issue): D107–13.

Grimm, C., R. Matos, N. Ly-Hartig, U. Steuerwald, D. Lindner, V. Rybin, J. Müller, and C. W. Müller, 2009, Jul)Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. EMBO J *28*(13): 1965–1977.

Ha, E. M., C. T. Oh, Y. S. Bae, and W. J. Lee, 2005, Nov)A direct role for dual oxidase in Drosophila gut immunity. Science *310*(5749): 847–850.

Ha, E. M., C. T. Oh, J. H. Ryu, Y. S. Bae, S. W. Kang, I. H. Jang, P. T. Brey, and W. J. Lee, 2005, Jan)An antioxidant system required for host protection against gut infection in Drosophila. Dev Cell *8*(1): 125–132.

Hasegawa, M., H. Kishino, and T. Yano, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol *22*(2): 160–74.

Hinrichs, A. S., D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, and W. J. Kent, 2006, Jan)The UCSC Genome Browser Database: update 2006.

Nucleic Acids Res *34* (Database issue): D590–8.

Hofacker, I. L., 2003, Jul)Vienna RNA secondary structure server. Nucleic Acids Res *31* (13): 3429–3431.

Hoffmann, A. A. and A. R. Weeks, 2007, Feb)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* (2): 133–147.

Hooven, L. A., K. A. Sherman, S. Butcher, and J. M. Giebultowicz, 2009 Does the clock make the poison? Circadian variation in response to pesticides. PLoS One *4* (7).

Hughes, K. A., J. F. Ayroles, M. M. Reedy, J. M. Drnevich, K. C. Rowe, E. A. Ruedi, C. E. Cáceres, and K. N. Paige, 2006, Jul)Segregating variation in the transcriptome: cis regulation and additivity of effects. Genetics *173* (3): 1347–1355.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata, 2002, Jul)MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res *30* (14): 3059–3066.

Katoh, K. and H. Toh, 2008, Jul)Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform *9* (4): 286–298.

Kennington, W. J., J. Gockel, and L. Partridge, 2003, Oct)Testing for asymmetrical gene flow in a Drosophila melanogaster body-size cline. Genetics *165* (2): 667–673.

Knibb, W. R., 1982 Chromosome inversion polymorphisms in Drosophila melanogaster. II. Geographic clines and climatic associations in Australasia, North America and Asia. Genetica *53* (3): 213–221.

Knibb, W. R., J. G. Oakeshott, and J. B. Gibson, 1981, Aug)Chromosome Inversion Polymorphisms in DROSOPHILA MELANOGASTER. I. Latitudinal Clines and Associations between Inversions in Australasian Populations. Genetics *98* (4): 833–847.

Kuersten, S. and E. B. Goodwin, 2003, Aug)The power of the 3' UTR: translational control and development. Nat Rev Genet *4*(8)**:** 626–637.

Lai, E. C., 2002, Apr)Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet *30*(4)**:** 363–364.

Langley, C. H., B. P. Lazzaro, W. Phillips, E. Heikkinen, and J. M. Braverman, 2000, Dec)Linkage disequilibria and the site frequency spectra in the su(s) and su(w(a)) regions of the Drosophila melanogaster X chromosome. Genetics *156*(4)**:** 1837–1852.

Lawniczak, M. K., A. K. Holloway, D. J. Begun, and C. D. Jones, 2008 Genomic analysis of the relationship between gene expression variation and DNA polymorphism in Drosophila simulans. Genome Biol *9*(8).

Lemos, B., L. O. Araripe, P. Fontanillas, and D. L. Hartl, 2008, Sep)Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. Proc Natl Acad Sci U S A *105*(38)**:** 14471–14476.

Levine, M. T. and D. J. Begun, 2008, May)Evidence of spatially varying selection acting on four chromatin-remodeling loci in Drosophila melanogaster. Genetics *179*(1)**:** 475–85.

Levine, M. T., M. Eckert, and D. J. Begun, 2010, Jul)Whole genome expression plasticity across tropical and temperate Drosophila melanogaster populations from eastern Australia. Mol Biol Evol.

Lewontin, R. C., 1974 *The genetic basis of evolutionary change*, Volume no. 25. New York: Columbia University Press.

Lewontin, R. C. and J. Krakauer, 1975, Jun)Letters to the editors: Testing the heterogeneity of F values. Genetics *80*(2)**:** 397–398.

Li, H., J. Ruan, and R. Durbin, 2008, Nov)Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res *18*(11)**:** 1851–8.

MacArthur, S., X.-Y. Li, J. Li, J. B. Brown, H. C. Chu, L. Zeng, B. P. Grondona,

A. Hechmer, L. Simirenko, S. V. E. Keränen, D. W. Knowles, M. Stapleton, P. Bickel, M. D. Biggin, and M. B. Eisen, 2009 Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol *10*(7)**:** R80.

Macdonald, S. J., T. Pastinen, and A. D. Long, 2005, Dec)The effect of polymorphisms in the enhancer of split gene complex on bristle number variation in a large wild-caught cohort of Drosophila melanogaster. Genetics *171*(4)**:** 1741–1756.

Majercak, J., W. F. Chen, and I. Edery, 2004, Apr)Splicing of the period gene 3'-terminal intron is regulated by light, circadian clock factors, and phospholipase C. Mol Cell Biol *24*(8)**:** 3359–3372.

Mangone, M., A. P. Manoharan, D. Thierry-Mieg, J. Thierry-Mieg, T. Han, S. D. Mackowiak, E. Mis, C. Zegar, M. R. Gutwein, V. Khivansara, O. Attie, K. Chen, K. Salehi-Ashtiani, M. Vidal, T. T. Harkins, P. Bouffard, Y. Suzuki, S. Sugano, Y. Kohara, N. Rajewsky, F. Piano, K. C. Gunsalus, and J. K. Kim, 2010, Jul)The landscape of C. elegans 3'UTRs. Science *329*(5990)**:** 432–435.

Maruyama, T., 1970 On the rate of decrease of heterozygosity in circular stepping stone models of populations* 1. Theoretical Population Biology *1*(1)**:** 101–119.

McColl, G. and S. W. McKechnie, 1999, Nov)The Drosophila heat shock hsr-omega gene: an allele frequency cline detected by quantitative PCR. Mol Biol Evol *16*(11)**:** 1568–1574.

McManus, C. J., J. D. Coolon, M. O. Duff, J. Eipper-Mains, B. R. Graveley, and P. J. Wittkopp, 2010, Jun)Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res *20*(6)**:** 816–825.

Menozzi, P., M. A. Shi, A. Lougarre, Z. H. Tang, and D. Fournier, 2004, Feb)Mutations of acetylcholinesterase which confer insecticide resistance in Drosophila melanogaster populations. BMC Evol Biol **4:** 4–4.

Miller, C., S. Beleza, A. Pollen, D. Schluter, R. Kittles, M. Shriver, and D. Kingsley, 2007 cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. Cell *131* (6)**:** 1179–1189.

Miyashita, N. and C. H. Langley, 1988, Sep)Molecular and phenotypic variation of the white locus region in Drosophila melanogaster. Genetics *120* (1)**:** 199–212.

Nachman, M., H. Hoekstra, and S. D'Agostino, 2003 The genetic basis of adaptive melanism in pocket mice. Proceedings of the National Academy of Sciences of the United States of America *100* (9)**:** 5268.

Nei, M., 1987 *Molecular evolutionary genetics*. Columbia Univ Pr.

Oakeshott, J. G., G. K. Chambers, J. B. Gibson, W. F. Eanes, and D. A. Willcocks, 1983, Feb)Geographic variation in G6pd and Pgd allele frequencies in Drosophila melanogaster. Heredity **50 (Pt 1):** 67–72.

Oakeshott, J. G., G. K. Chambers, J. B. Gibson, and D. A. Willcocks, 1981, Dec)Latitudinal relationships of esterase-6 and phosphoglucomutase gene frequencies in Drosophila melanogaster. Heredity *47*(Pt 3)**:** 385–396.

Osborne, K., A. Robichon, E. Burgess, S. Butland, R. Shaw, A. Coulthard, H. Pereira, R. Greenspan, and M. Sokolowski, 1997 Natural behavior polymorphism due to a cGMP-dependent protein kinase of Drosophila. Science *277*(5327)**:** 834.

Paaby, A. B., M. J. Blacket, A. A. Hoffmann, and P. S. Schmidt, 2010, Feb)Identification of a candidate adaptive polymorphism for Drosophila life history by parallel independent clines on two continents. Mol Ecol *19* (4)**:** 760–774.

Palopoli, M., M. Rockman, A. TinMaung, C. Ramsay, S. Curwen, A. Aduna, J. Laurita, and L. Kruglyak, 2008 Molecular basis of the copulatory plug polymorphism in Caenorhabditis elegans. Nature *454* (7207)**:** 1019–1022.

Palsson, A., A. Rouse, R. Riley-Berger, I. Dworkin, and G. Gibson, 2004, Jul)Nucleotide variation in the Egfr locus of Drosophila melanogaster. Genetics *167*(3)**:** 1199–1212.

Pickrell, J. K., G. Coop, J. Novembre, S. Kudaravalli, J. Z. Li, D. Absher, B. S. Srinivasan, G. S. Barsh, R. M. Myers, M. W. Feldman, and J. K. Pritchard, 2009, May)Signals of recent positive selection in a worldwide sample of human populations. Genome Res *19*(5)**:** 826–837.

Rako, L., A. R. Anderson, C. M. Sgrò, A. J. Stocker, and A. A. Hoffmann, 2006 The association between inversion In(3R)Payne and clinally varying traits in Drosophila melanogaster. Genetica *128*(1-3)**:** 373–84.

Rako, L., M. J. Blacket, S. W. McKechnie, and A. A. Hoffmann, 2007, Jul)Candidate genes and thermal phenotypes: identifying ecologically important genetic variation for thermotolerance in the Australian Drosophila melanogaster cline. Mol Ecol *16*(14)**:** 2948–57.

Sackton, T. B., R. J. Kulathinal, C. M. Bergman, A. R. Quinlan, E. B. Dopman, M. Carneiro, G. T. Marth, D. L. Hartl, and A. G. Clark, 2009 Population genomic inferences from sparse high-throughput sequencing of two populations of Drosophila melanogaster. Genome Biol Evol **1:** 449–465.

Sandrelli, F., E. Tauber, M. Pegoraro, G. Mazzotta, P. Cisotto, J. Landskron, R. Stanewsky, A. Piccin, E. Rosato, M. Zordan, R. Costa, and C. P. Kyriacou, 2007, Jun)A molecular basis for natural selection at the timeless locus in Drosophila melanogaster. Science *316*(5833)**:** 1898–1900.

Schmidt, J. M., R. T. Good, B. Appleton, J. Sherrard, G. C. Raymant, M. R. Bogwitz, J. Martin, P. J. Daborn, M. E. Goddard, P. Batterham, and C. Robin, 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. PLoS Genet *6*(6).

Schmidt, P., C. Zhu, J. Das, M. Batavia, L. Yang, and W. Eanes, 2008 An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in Drosophila melanogaster. Proceedings of the National Academy of Sciences *105*(42)**:** 16207.

Schmidt, P. S., D. D. Duvernell, and W. F. Eanes, 2000, Sep)Adaptive evolution of a candidate gene for aging in Drosophila. Proc Natl Acad Sci U S A *97*(20)**:** 10861–10865.

Sezgin, E., D. D. Duvernell, L. M. Matzkin, Y. Duan, C. T. Zhu, B. C. Verrelli, and W. F. Eanes, 2004, Oct)Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in Drosophila melanogaster. Genetics *168*(2)**:** 923–931.

Sheikh, I. A., A. K. Singh, N. Singh, M. Sinha, S. B. Singh, A. Bhushan, P. Kaur, A. Srinivasan, S. Sharma, and T. P. Singh, 2009, May)Structural evidence of substrate specificity in mammalian peroxidases: structure of the thiocyanate complex with lactoperoxidase and its interactions at 2.4 A resolution. J Biol Chem *284*(22)**:** 14849–14856.

Singh, N. D., P. F. Arndt, and D. A. Petrov, 2005, Feb)Genomic heterogeneity of background substitutional patterns in Drosophila melanogaster. Genetics *169*(2)**:** 709–22.

SINGH, R. and A. LONG, 1992, October)GEOGRAPHIC-VARIATION IN DROSOPHILA - FROM MOLECULES TO MORPHOLOGY AND BACK. Trends In Ecology & Evolution *7*(10)**:** 340–345.

Singh, R. S., 1989 Population genetics and evolution of species related to Drosophila melanogaster. Annu Rev Genet **23:** 425–453.

Singh, R. S. and L. R. Rhomberg, 1987, Oct)A Comprehensive Study of Genic Variation in Natural Populations of Drosophila melanogaster. II. Estimates of Heterozygosity and Patterns of Geographic Differentiation. Genetics *117*(2)**:** 255–271.

Slatkin, M., 1981, Oct)Estimating Levels of Gene Flow in Natural Populations. Genetics *99*(2)**:** 323–335.

Stark, A., J. Brennecke, N. Bushati, R. B. Russell, and S. M. Cohen, 2005, Dec)Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell *123*(6)**:** 1133–1146.

Storey, J., 2002 A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B-Statistical Methodology **64:** 479–498.

Tajima, F., 1983, Oct)Evolutionary relationship of DNA sequences in finite populations. Genetics *105*(2)**:** 437–60.

Tauber, E., M. Zordan, F. Sandrelli, M. Pegoraro, N. Osterwalder, C. Breda, A. Daga, A. Selmin, K. Monger, C. Benna, E. Rosato, C. P. Kyriacou, and R. Costa, 2007, Jun)Natural selection favors a newly derived timeless allele in Drosophila melanogaster. Science *316*(5833)**:** 1895–1898.

Teshima, K. M., G. Coop, and M. Przeworski, 2006, Jun)How reliable are empirical genomic scans for selective sweeps? Genome Res *16*(6)**:** 702–712.

Tsouloufis, T., A. Mamalaki, M. Remoundos, and S. J. Tzartos, 2000, Sep)Reconstitution of conformationally dependent epitopes on the N-terminal extracellular domain of the human muscle acetylcholine receptor alpha subunit expressed in Escherichia coli: implications for myasthenia gravis therapeutic approaches. Int Immunol *12*(9)**:** 1255–1265.

Turner, T. L., M. T. Levine, M. L. Eckert, and D. J. Begun, 2008, May)Genomic analysis of adaptive differentiation in Drosophila melanogaster. Genetics *179*(1)**:** 455–473.

Tweedie, S., M. Ashburner, K. Falls, P. Leyland, P. McQuilton, S. Marygold, G. Millburn, D. Osumi-Sutherland, A. Schroeder, R. Seal, H. Zhang, and FlyBase Consortium, 2009, Jan)FlyBase: enhancing Drosophila Gene Ontology annotations. Nucleic Acids Res *37*(Database issue)**:** 555–559.

Umina, P. A., A. A. Hoffmann, A. R. Weeks, and S. W. McKechnie, 2006, Feb)An independent non-linear latitudinal cline for the sn-glycerol-3-phosphate (alpha- Gpdh) polymorphism of Drosophila melanogaster from eastern Australia. Genet Res *87*(1)**:** 13–21.

Umina, P. A., A. R. Weeks, M. R. Kearney, S. W. McKechnie, and A. A. Hoffmann, 2005, Apr)A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. Science *308*(5722)**:** 691–693.

Voelker, R. A., C. C. Cockerham, F. M. Johnson, H. E. Schaffer, T. Mukai, and L. E. Mettler, 1978, Mar)Inversions Fail to Account for Allozyme Clines. Genetics *88*(3)**:** 515–527.

Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard, 2006, Mar)A map of recent positive selection in the human genome. PLoS Biol *4*(3).

Wang, X., D. S. Green, S. P. Roberts, and J. S. de Belle, 2007 Thermal disruption of mushroom body development and odor learning in Drosophila. PLoS One *2*(11)**:** e1125.

Wright, S., 1931, Mar)Evolution in Mendelian Populations. Genetics *16*(2)**:** 97–159.

Yang, Z., 1996, September)Among-site rate variation and its impact on phylogenetic analyses. Trends In Ecology & Evolution *11*(9)**:** 367–372.

Yang, Z., 2007, Aug)PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol *24*(8)**:** 1586–91.

Yang, Z., S. Kumar, and M. Nei, 1995, Dec)A new method of inference of ancestral nucleotide and amino acid sequences. Genetics *141*(4)**:** 1641–50.

Yu, J., S. Pacifico, G. Liu, and R. L. Finley, 2008 DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. BMC Genomics **9:** 461–461.

List of Figures

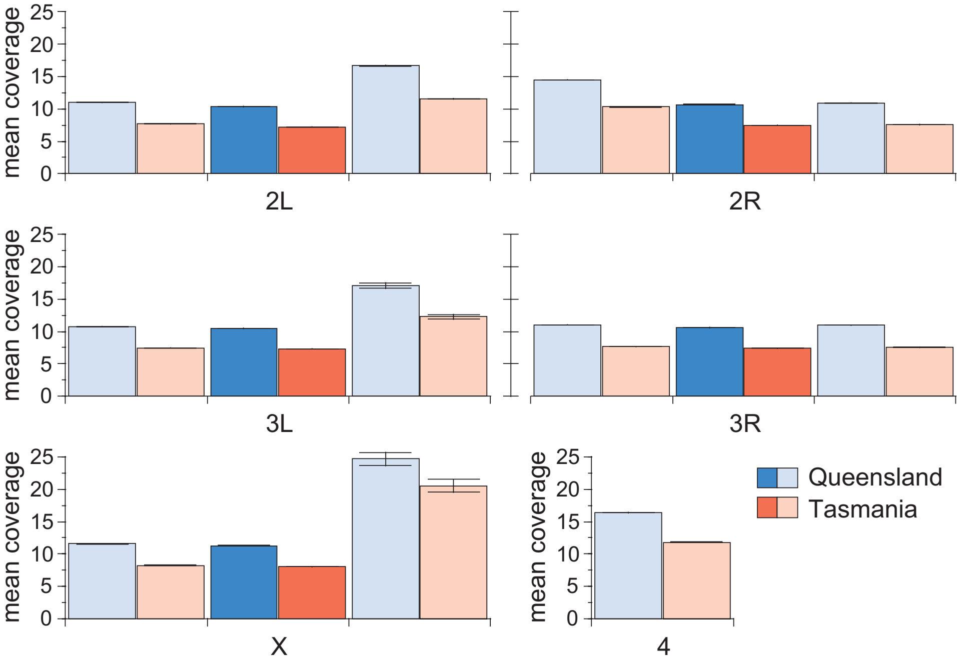

Figure 1: Genome-sequence coverage is equivalent across chromosome arms in normally-recombining regions, more variable in low-recombining regions. Mean sequencing coverage is plotted for Queensland (blue) and Tasmania (red) populations. Dark colors indicate regions of normal recombination; lighter colors indicate low-recombining centromeric and telomeric regions. Bars give standard error.

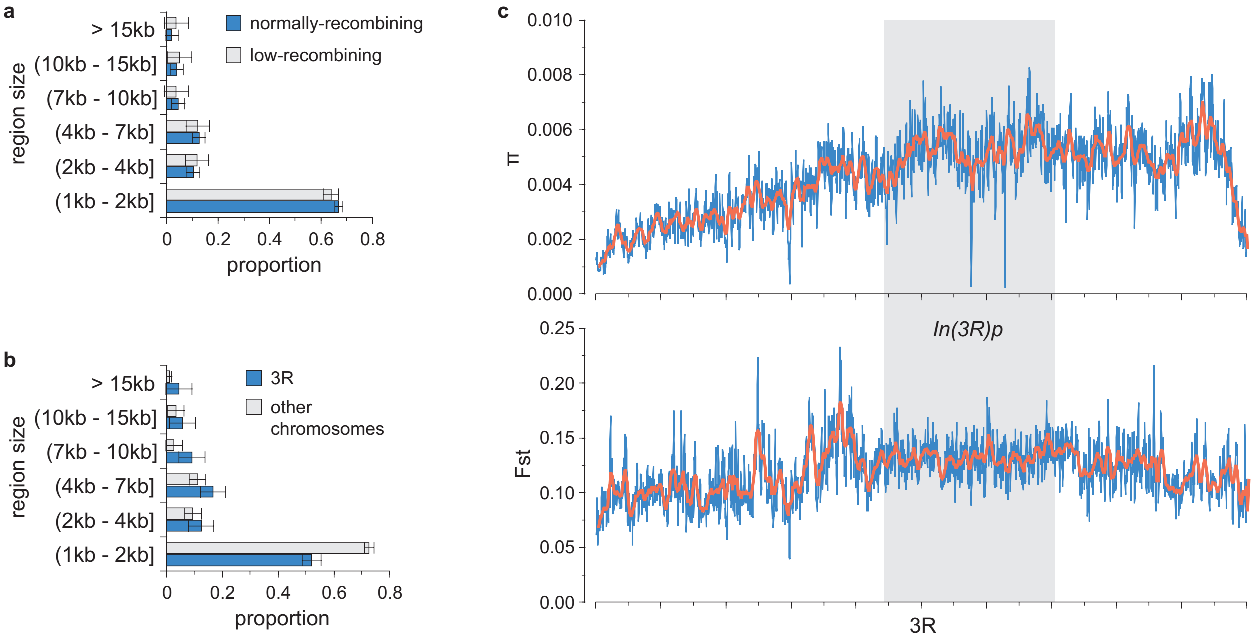

Figure 2: Size of differentiated regions is similar in areas of normal- and low-recombination, larger on chromosome *3R*. We calculated mean F_{ST} in nonoverlapping 1 kb windows across the *D. melanogaster* genome. Groups of windows in the top 1% tail of the F_{ST} distribution were grouped together into larger differentiated regions separated from one another by at least 5 consecutive windows with mean F_{ST} in the bottom 90% tail (see Materials and Methods). **a.** We plot the size distribution of these differentiated regions for normally-recombining (blue) and low-recombining (gray) areas of the genome. Bars indicate standard error. **b.** We plot the size distribution of differentiated regions found in normally-recombining regions of chromosome *3R* (blue) and the size distribution of differentiated regions in normally-recombining regions of other chromosome arms (gray). **c.** We plot mean F_{ST} (bottom) and mean polymorphism (π, top) across chromosome *3R*. Blue lines indicate average values over 25 kb windows slid every 10 kb; red lines show 200 kb windows slid 50 kb at a time. Gray box indicates the location of the cosmopolitan *3R-Payne* inversion.

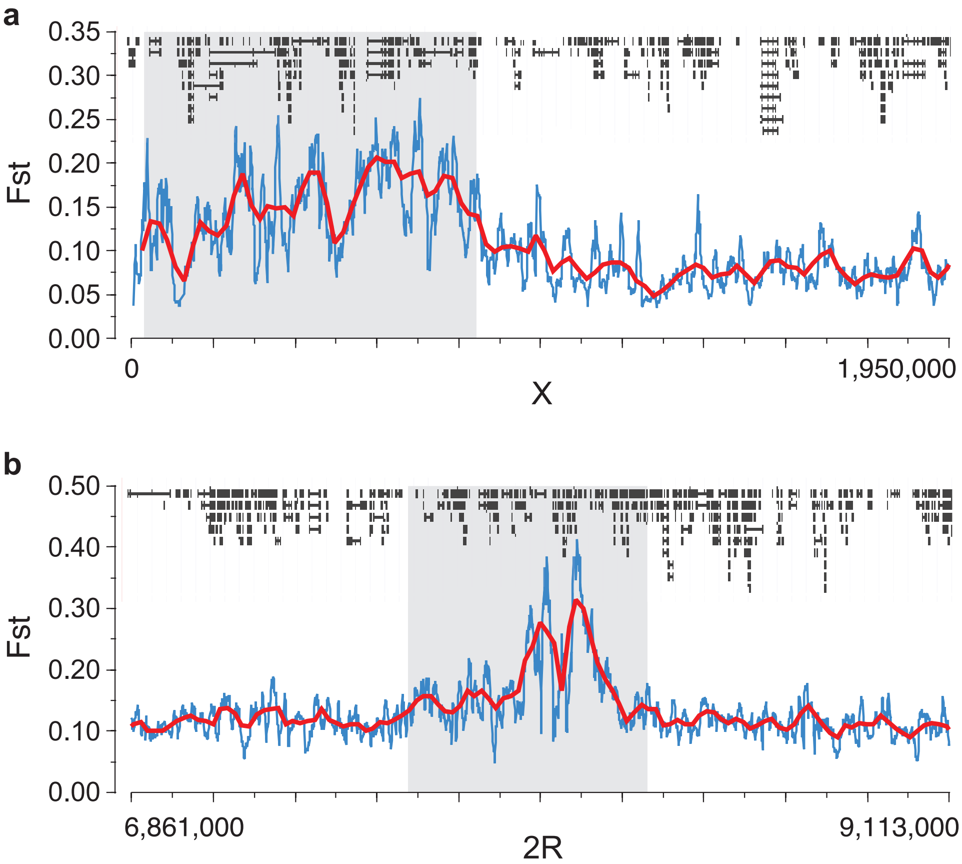

Figure 3: Largest highly-differentiated regions occurred at the tip of the *X* chromosome (**a**) and in the middle of chromosome *2R* (**b**). Highly-differentiated regions are indicated in gray. We plot mean F_{ST} across each chromosomal region, blue lines indicating 10 kb windows with 1 kb slides and red lines indicating 50 kb windows, 20 kb slides. Annotated genes are drawn across the top of each panel.

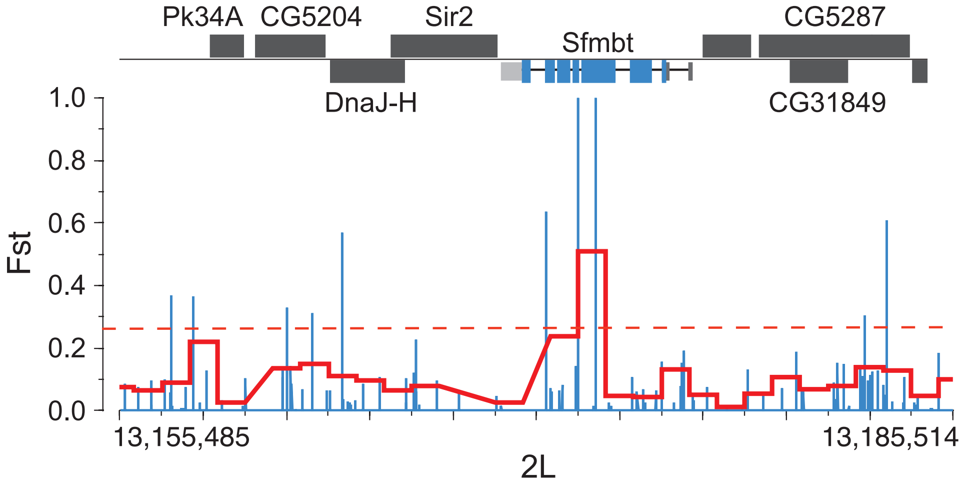

Figure 4: Regions of high population-differentiation localize within the *Sfmbt* gene on chromosome *2L*. We plot individual-position F_{ST} (blue) and mean F_{ST} within 1 kb windows (red) across the chromosome. Red dotted line indicates F_{ST} cutoff for top 2.5% of 1 kb windows. Individual genes are drawn across the top (black); exons are in blue, 3'UTRs in light gray, and 5'UTRs in dark gray.

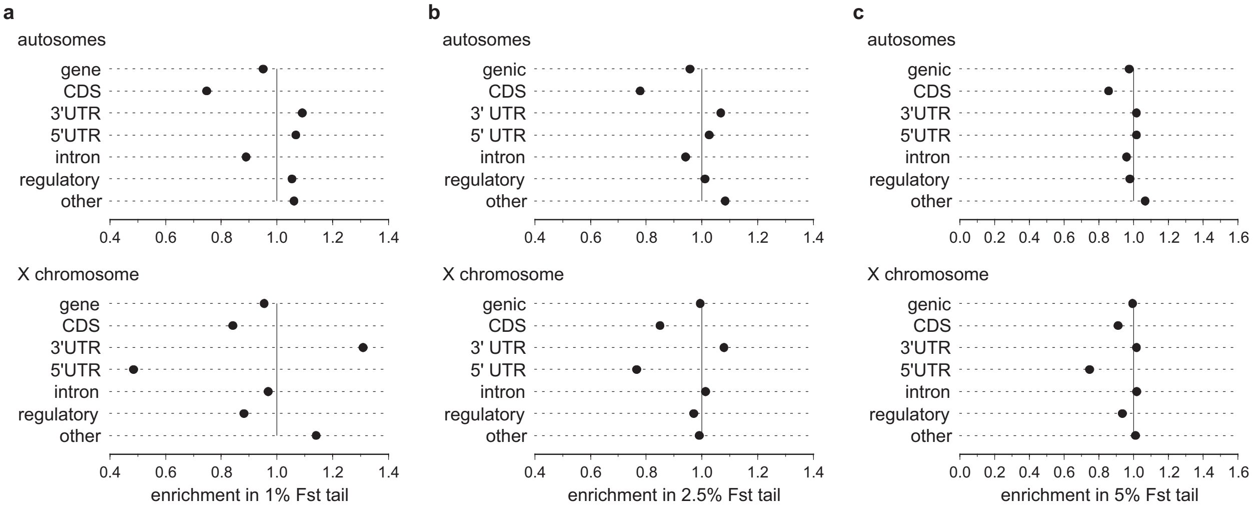

Figure 5: 3'UTRs and unannotated regions are over-represented in the most-differentiated genomic regions. We calculated the enrichment for each annotation type in the 1% (**a**), 2.5% (**b**) and 5% (**c**) tail of 1 kb F_{ST} regions, relative to each type's distribution across all 1 kb windows in the normally-recombining portion of the genome. Results are shown separately for autosomes and the X chromosome. An enrichment score of 1.0 (indicated by solid vertical line) indicates no enrichment or depletion; values > 1 indicate an overabundance of that type in the F_{ST} tail, whereas values < 1 indicate under-abundance.

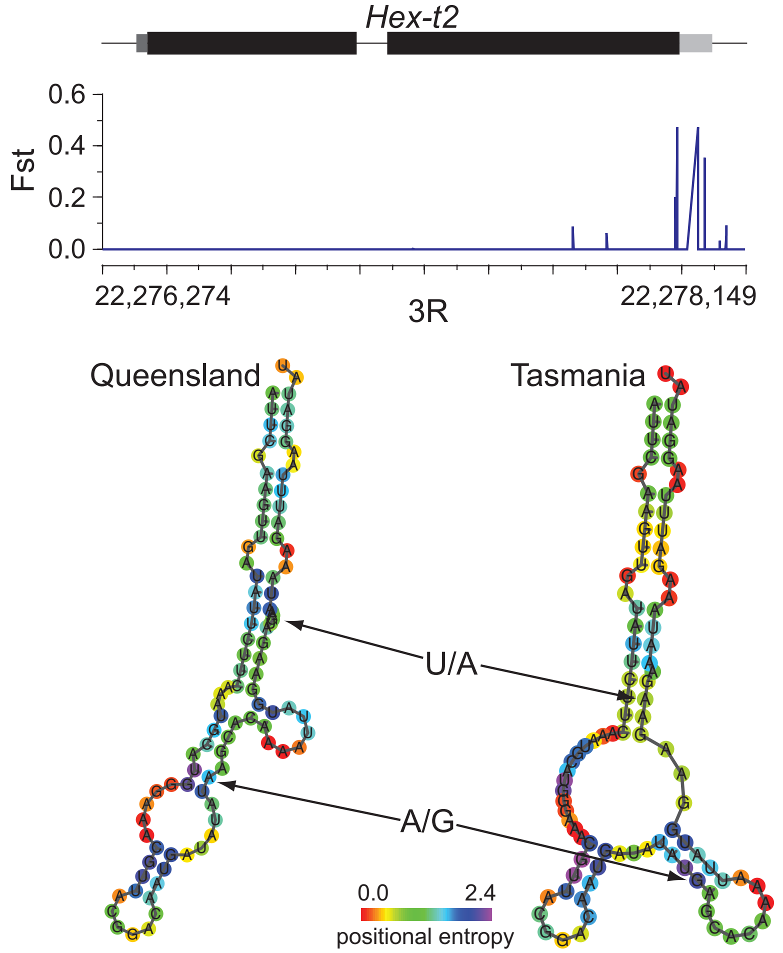

Figure 6: Elevated differentiation between Queensland and Tasmania populations localizes to the 3'UTR of the *Hex-t2* gene. We plot the F_{ST} of individual genomic positions against the structure of the *Hex-t2* gene. Exons are drawn in black, the 5'UTR is dark gray, and the 3'UTR is light gray. Bottom panel shows predicted secondary structures of Queensland and Tasmania 3'UTR regions. Queensland positions indicated by arrows are polymorphic, with the major allele at left; corresponding positions in Tasmania are fixed for what is the minor allele in Queensland.

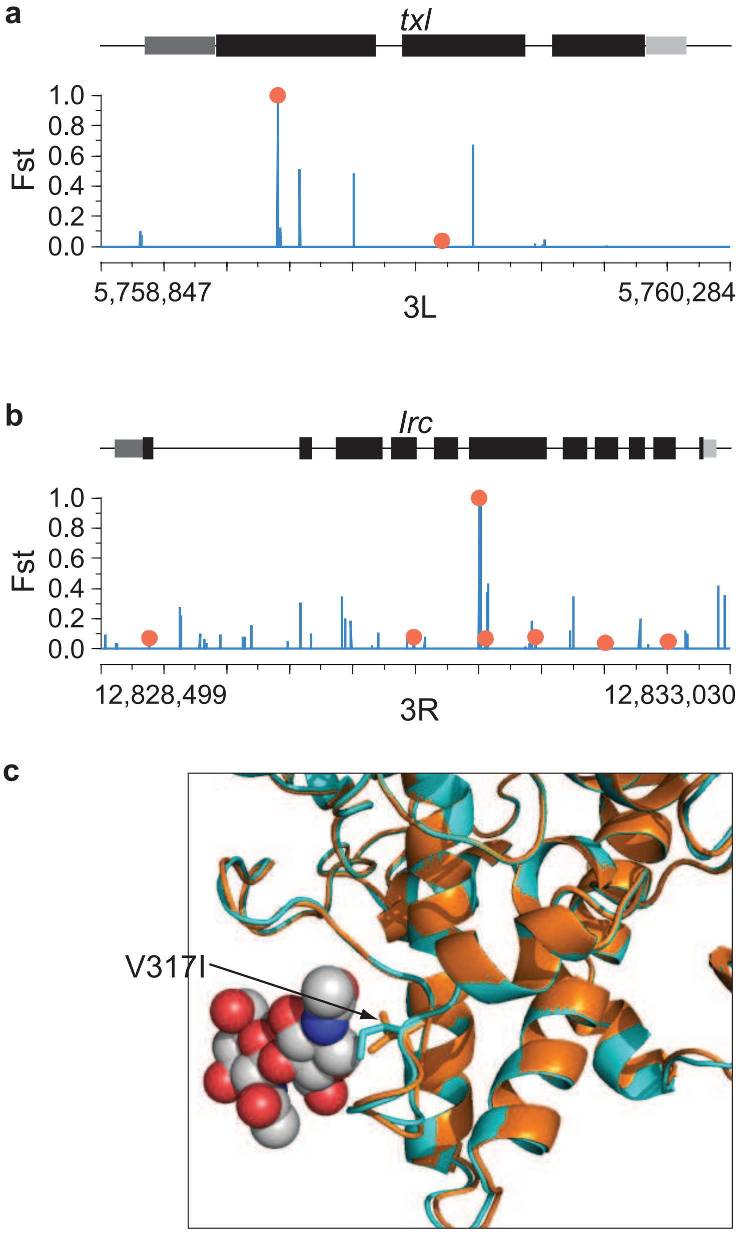

Figure 7: Elevated nonsynonymous F_{ST} in two *melanogaster* protein-coding genes. We plot individual-position F_{ST} along the gene structure. Exons are drawn in black, the 5'UTR is dark gray, and the 3'UTR is light gray. Nonsynonymous polymorphisms are shown in red; synonymous and noncoding polymorphisms are shown in blue. **a.** A nonsynonymous fixed-difference between Queensland and Tasmania is associated with elevated F_{ST} at the *txl* gene. **b.** Elevated F_{ST} at a fixed protein-coding change in *Irc*. **c.** Structural homology models of Queensland (orange) and Tasmania (turquoise) *Irc*; the V317I substitution is potentially involved in direct ligand interaction.

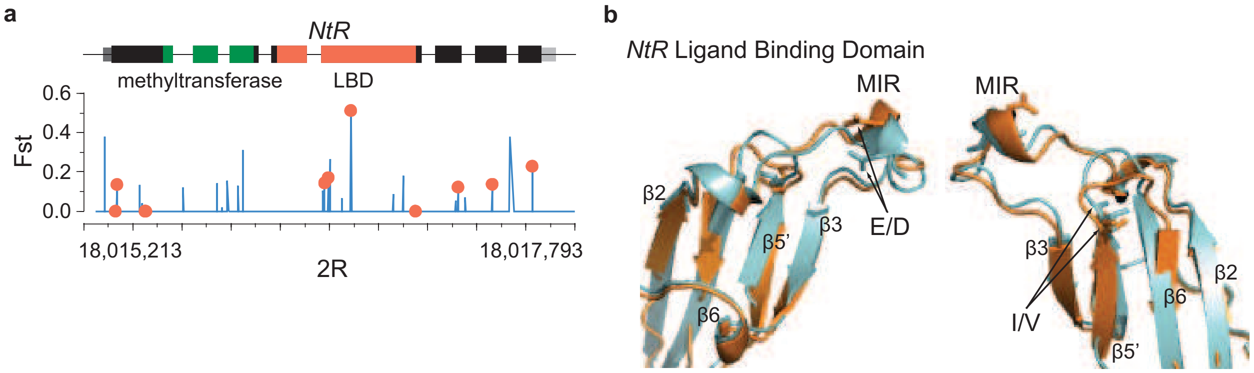

Figure 8: Elevated nonsynonymous differentiation in *NtR* localizes to the major immunogenic region (MIR) of the ligand-binding domain (LBD). **a.** We plot positional F_{ST} across gene structure, with exons drawn in black, 5'UTR in dark gray and 3'UTR in light gray; methyltransferase and ligand-binding domains are indicated by green and red, respectively. Nonsynonymous polymorphisms are shown by red dots. **b.** We plot highly-differentiated E/D and I/V polymorphisms on the predicted 3D structure of the *NtR* LBD. In both cases, the major allele in Queensland (E,I) is shown in orange, and the major allele in Tasmania (D,V) is shown in turquoise.

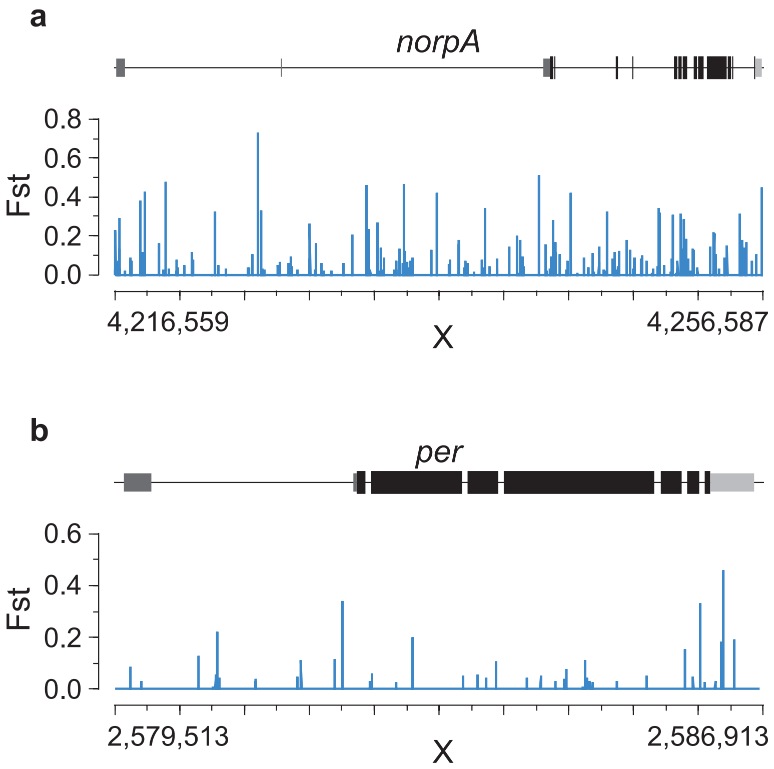

Figure 9: Coordinated differentiation in *norpA* (**a**) and the 3'UTR of *per* (**b**), a known target of *norpA* splicing regulation. We plot individual-position F_{ST} along the gene structure. Exons are drawn in black, the 5'UTR is dark gray, and the 3'UTR is light gray.

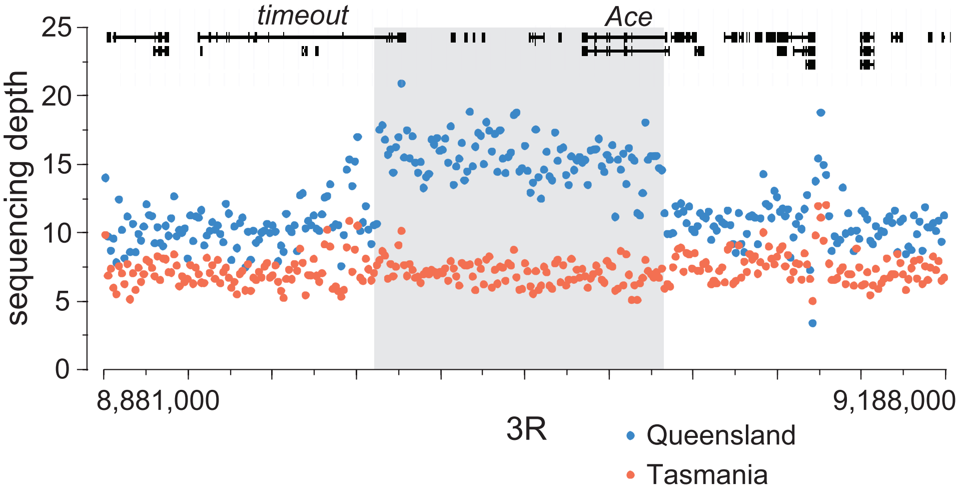

Figure 10: A large region of increased copy-number in Queensland occurs on chromosome *3R*. We plot the average number of sequence reads for each 1 kb window across this region, both for the Queensland (blue) and Tasmania (red) populations. Genes in this region are drawn across the top. Gray box indicates inferred region of increased copy number in Queensland.