Parallel Geographic Variation in *Drosophila melanogaster*

Josie A. Reinhardt,* Bryan Kolaczkowski,† Corbin D. Jones,*,‡ David J. Begun,§ and Andrew D. Kern,[1]**
*Department of Biology and ‡Carolina Center for Genome Science, University of North Carolina, Chapel Hill, North Carolina 27599, †Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611, §Department of Evolution and Ecology, University of California, Davis, California 95616, and **Department of Genetics, Rutgers University, New Brunswick, New Jersey 08854

ABSTRACT *Drosophila melanogaster*, an ancestrally African species, has recently spread throughout the world, associated with human activity. The species has served as the focus of many studies investigating local adaptation relating to latitudinal variation in non-African populations, especially those from the United States and Australia. These studies have documented the existence of shared, genetically determined phenotypic clines for several life history and morphological traits. However, there are no studies designed to formally address the degree of shared latitudinal differentiation at the genomic level. Here we present our comparative analysis of such differentiation. Not surprisingly, we find evidence of substantial, shared selection responses on the two continents, probably resulting from selection on standing ancestral variation. The polymorphic inversion In(3R)P has an important effect on this pattern, but considerable parallelism is also observed across the genome in regions not associated with inversion polymorphism. Interestingly, parallel latitudinal differentiation is observed even for variants that are not particularly strongly differentiated, which suggests that very large numbers of polymorphisms are targets of spatially varying selection in this species.

HOW organisms adapt to the ecological challenges of a new environment remains poorly understood. Indeed, while observations from comparative biology show that organisms often evolve convergent phenotypes when faced with similar selection pressures, we have little insight into how underlying historical and population genetic processes determine the degree of shared or divergent selection responses across populations or species. The latitudinal clines of *Drosophila melanogaster* provide a rich system for exploring these questions.

While there is general agreement that *D. melanogaster* evolved in Africa, spread through Eurasia several thousand years ago, and only recently colonized the Americas and Australia (David and Capy 1988; Lachaise *et al.* 1988; Begun and Aquadro 1993; Keller 2007; Stephan and Li 2007; Duchen *et al.* 2013), our understanding of the species' historical biogeography is incomplete. There are at least two potential clines that have received much attention—one in Australia and one in North America—that likely represent independent samplings of shared ancestral variation (Knibb 1982; Hoffmann and Weeks 2007).

Decades of research on *D. melanogaster* clines have revealed broad shared patterns of adaptive phenotypic divergence along latitudinal transects in the Americas and Australia. For example, several phenotypes including body size, multiple physiological traits, and multiple allozyme variants show parallel clines on these continents (Singh and Long 1992). Paracentric chromosome inversion polymorphism is also well documented as showing similar patterns of clinal variation on the two continents (Voelker *et al.* 1978; Knibb 1982). Genomic description of the two clines is minimal. The most comprehensive data bearing on the issue of shared clinality on these continents are from tiling array data from Queensland, Tasmania, Maine, and Florida (Turner *et al.* 2008). These data revealed, perhaps not surprisingly given previous data from phenotypes, inversions, and allozymes, that a considerable portion of clinal variation appears to be shared at the genomic level. Interestingly, however, there were also major differences between the

doi: 10.1534/genetics.114.161463
Manuscript received January 22, 2014; accepted for publication February 28, 2014; published Early Online March 7, 2014.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161463/-/DC1.
[1]Corresponding author: Department of Genetics, Rutgers University, Nelson Biolabs, 604 Alison Rd., Piscataway, NJ 08854. E-mail: kern@biology.rutgers.edu

Figure 1 Sampling locations in North America and Australia. In North America more tropical samples come from Florida and temperate samples are from Maine. In Australia tropical flies are from Queensland and temperate collections are from Tasmania.

continents. For example, some of the most strongly differentiated regions between northern and southern Australian populations showed no evidence of differentiation between northern and southern populations from North America. However, the technology at that time prohibited a base-level examination. A recent population genomic analysis of the North American cline (Fabian *et al.* 2012) supported the conclusion from the tiling array analysis (Turner *et al.* 2008) that there is substantial parallel differentiation on the two continents, but there was no formal comparison of comparable data from the two continents analyzed in the same way. Here we use whole-genome sequencing to elucidate patterns of genomic differentiation in North America and Australia based on genomic sequencing to tease out the degree to which phenotypic convergence in these parallel clines has resulted from convergent evolution at the genetic level. In so doing, we aim to understand the underlying historical and population genetic processes that explain both the degree of shared selection response to latitudinal gradients in these populations and the differences between them.

Materials and Methods

Population samples

The populations investigated here are from Queensland (QUE), Tasmania (TAS), Maine (MAI), and Florida (FLA) and were described previously (Turner *et al.* 2008). Figure 1 shows the location of each of the populations sampled from the four different locations. Two samples were taken at each location (see Turner *et al.* 2008 for details). Sample locations in Figure 1 are labeled red and blue to indicate their low-latitude (red) *vs.* high-latitude (blue) environment. Two samples from each of the four geographic locations were taken, and these two samples were then pooled for sequencing. For DNA isolation and pooling we used females collected from 16 isofemale lines from MAI, 16 lines from FLA, 17 lines from QUE, and 15 lines from TAS.

Genomic sequencing and mapping

Genomic DNA from pooled samples from each location was run on a single lane of an Illumina GA2 sequencer for 2×75 cycles, using the standard flow cell, yielding $\sim 28\times$ coverage

Table 1 Summary statistics of DNA polymorphism from 1-kb windows throughout the recombining portion of the genome

Population	Chr	Θ_W	Θ_π	Θ_H	H
Low latitude					
Florida	X	3.54	3.12	4.22	−1.089
	2L	5.92	5.47	5.90	−0.421
	2R	5.14	4.67	5.27	−0.586
	3L	5.28	4.80	5.34	−0.533
	3R	5.27	4.82	5.16	−0.333
	GW	5.05	4.60	5.18	−0.579
Queensland	X	3.57	3.12	4.09	−0.957
	2L	5.8	5.36	5.75	−0.377
	2R	4.88	4.38	4.90	−0.507
	3L	5.39	4.87	5.23	−0.353
	3R	5.23	4.80	5.04	−0.233
	GW	5.00	4.54	5.01	−0.469
High latitude					
Maine	X	3.41	3.08	3.99	−0.908
	2L	5.33	4.91	5.69	−0.779
	2R	4.86	4.49	5.05	−0.558
	3L	4.78	4.41	5.08	−0.676
	3R	4.65	4.24	4.88	−0.635
	GW	4.61	4.23	4.94	−0.709
Tasmania	X	3.04	2.74	3.78	−1.038
	2L	4.68	4.34	5.48	−1.140
	2R	4.38	4.09	4.94	−0.848
	3L	4.27	3.96	4.92	−0.968
	3R	4.234	3.87	4.80	−0.938
	GW	4.13	3.80	4.79	−0.986

Mean estimates of 4*Nu* (1e-03/bp) among 1-kb windows are given and are the pooled-sampling, singleton corrected estimators provided in Kolaczkowski *et al.* (2011). Chr, chromosome; GW, genome-wide.

apiece. Raw Illumina GA2 image data were phased and filtered for quality, using default GERALD parameters for unaligned reads. Sequencing reads were mapped back to the *D. melanogaster* reference sequence with BWA v. 0.5.8 (Li and Durbin 2009). As these data are derived from pools of multiple individuals, polymorphism may affect the ability of BWA to align reads harboring SNPs. To control for this effect, we altered the alignment parameters (k, the number of errors allowed in the seed; and n, the number of errors allowed in the whole read) from BWA and compared the number of reads aligned and the accuracy of those alignments (Supporting Information, Table S23). Based on these data, we chose $k = 2$ and $n = 5$ for the alignment of all four population samples. Data have been submitted to the Short Read Archive and can be found under bioproject accession no. PRJNA237820.

Postmapping filtering steps

After alignment we further filtered each two-population data set in a number of ways. First, we filtered any bases that were triallelic or had coverage <6 or >40. We also filtered any bases that had only a single read carrying the minor allele on a continent. Next, we used repeat masker (v. 3.3) to filter positions associated with known repeats or low sequence complexity in the reference sequence. We also removed regions of the genome thought to experience reduced rates of crossing over because their associated reduced heterozygosity could

Table 2 Summary statistics of F_{st} in 1-kb windows from two clines

Cline	Mean	SD	5%	2.5%	1%
Australia	0.0716	0.0392	0.1425	0.1165	0.2002
North America	0.0657	0.0311	0.1232	0.141	0.1658

Shown are the mean; the standard deviation; and the 5%, 2.5%, and 1% cutoffs of the empirical distribution of F_{st}.

reduce the power to detect differentiation and because the larger physical scale of differentiation expected in such regions might compromise one's ability to identify potential targets of selection. The coordinates corresponding to regions of normal recombination used in our analyses were defined by the *Drosophila* Population Genomics Project (dpgp.org) and include 2L:844,225–19,946,732; 2R:6,063,980–20,322,335; 3L:447,386–18,392,988; 3R:7,940,899–27,237,549; and X:1,036,552–20,902,578.

Population genetic analysis

Summary statistics of polymorphism and differentiation were calculated following Kolaczkowski *et al.* (2011). We considered individual populations for single-population summary statistics and pairs of populations for calculations of F_{st}. All downstream analysis uses this pairwise F_{st} information. As the main focus of this article is a comparison of differentiation on two continents, variable coverage across population samples was a potential problem, as it would increase variation in power across sites and continents. To minimize this problem we created a "trimmed" data set. For each position that met the criteria noted in the previous sections the minimum depth among the four population samples was noted and the other three populations were randomly down-sampled to provide equal depth in all four populations. This trimmed data set was the object for the majority of analyses.

Empirical outlier approach

There are benefits and pitfalls of using a genome-wide empirical distribution rather than a model-based approach for the detection of candidates (Beaumont and Nichols 1996; Akey *et al.* 2002; Teshima *et al.* 2006; Voight *et al.* 2006; Pickrell *et al.* 2009). As in Kolaczkowski *et al.* (2011), we have, for a few reasons, opted for an empirical rather than a model-based approach for most analyses. First, model-based demographic inference using pooled population genomic data is an unsolved problem. Second, given the pervasive genomic effect of selection on polymorphism in *D. melanogaster* (Langley *et al.* 2012), demographic model fitting based on the assumption of strict neutrality is likely to be misleading. Third, because we focus on outliers present on two continents, even if our empirical approach were not optimal, we have a strong expectation that it will reveal a substantial component of parallel adaptive allele frequency change.

As the physical distribution of differentiation in the genome is unknown, we used two complementary approaches. First, we calculated F_{st} in nonoverlapping 1-kb windows throughout

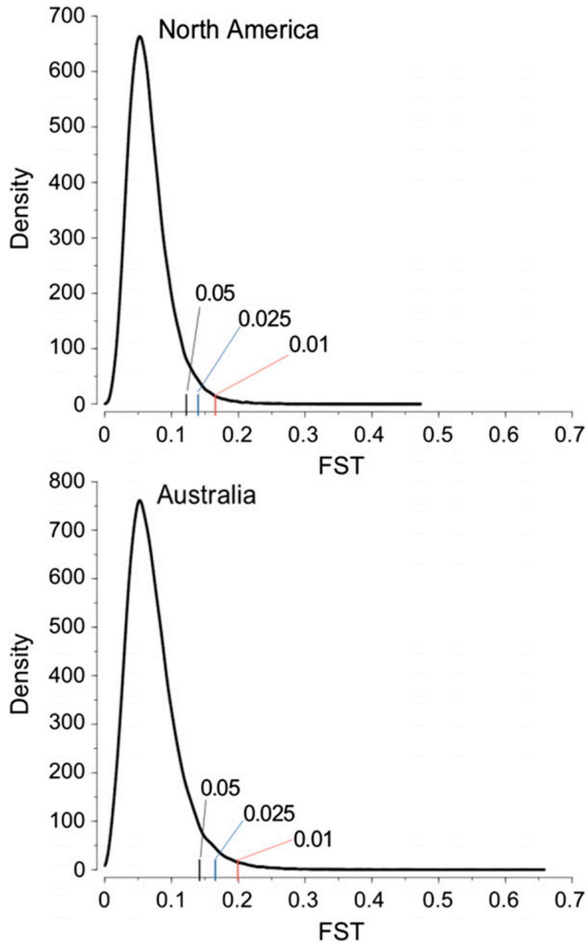

Figure 2 Density of F_{st} estimates from 1-kb windows in regions of normal recombination from the North American cline (top) and the Australian cline (bottom), respectively. The 5%, 2.5%, and 1% tail cutoffs are indicated in each population with colored hash marks.

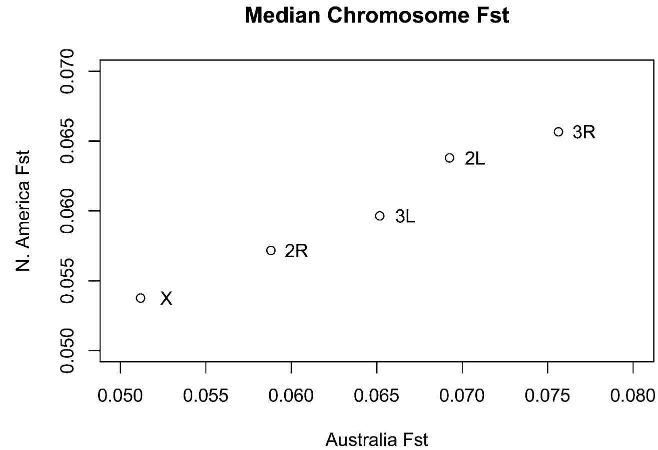

Figure 3 Median F_{st} of 1-kb windows in the recombining portion of each chromosome arm plotted in each clinal comparison. Note that while the scales of the axes are different, reflecting greater absolute differentiation among our Australia samples, rank order of chromosomal differentiation is identical among samples.

the normally recombining portion of the genome on each continent, from which the top 5%, 2.5%, and 1% tails of window F_{st}'s could be identified. Generally we considered differentiated windows to be in either the top 2.5% or the top 1% tail. Shared outlier regions were defined simply as the intersection of the windows occupying the tails of both continents. A second approach was to use individual SNP frequencies throughout the genome to identify candidate differentiated SNPs. This second analysis is useful for generating hypotheses on classes of SNPs that may be under selection or for investigating genome-wide properties of SNPs belonging to different classes (*e.g.*, CDS *vs*. intergenic).

Whole-genome analysis of parallel differentiation

If parallel differentiation were common even in genomic regions that were not in the shared outlier set, then F_{st} would be positively correlated between continents. We used simulations to determine whether the observed correlation (see *Results*) was significant. The null model of Coop *et al.* (2010) was used to simulate divergence from all four populations simultaneously. Shared population history in the model is represented through the covariance matrix associated with the multivariate normal random variable representing population allele frequencies. These simulations assume that the populations were independent (off-diagonal elements set to zero) with continental F_{st} equal to our observed median estimates ($\sim$0.06) and ancestral allele frequencies drawn from the standard neutral site frequency spectrum. Using this model, 10,000 SNPs were simulated and F_{st} at each SNP was computed. From these values a correlation coefficient of F_{st} between continents was calculated, representing one replicate of the simulation. The results reported here are from 100,000 simulation replicates.

We further investigated the question of parallel differentiation at the whole-genome scale by using a clustering analysis. The rationale for this analysis is that SNPs that are targets of spatially varying selection on both continents would lead to clustering of populations by latitude rather than by continent. To examine this question, we created distance trees of the four populations, using SNP frequencies in each population as the input. Three distance metrics were used: cord distance (Cavalli-Sforza and Edwards 1967), Nei's *D* (Nei 1972), and Euclidean distance. Trees were constructed using subsets of SNPs with varying levels of F_{st} as defined by the Australian comparison. Distance trees were constructed using neighbor joining (Saitou and Nei 1987). We were particularly interested in asking whether more differentiated SNPs (in Australia) reflect clustering of populations by environment (high latitude *vs.* low latitude) or by geography (Australia *vs.* North America) in the four-population tree. Support for the recovered distance tree at each F_{st} cutoff was assessed using nonparametric bootstrapping (1000 replicates), sampling SNPs with replacement from our entire data set.

SNP-level analysis of parallel outliers

We generated lists of candidate selected SNPs by focusing on those that were segregating on both continents and that

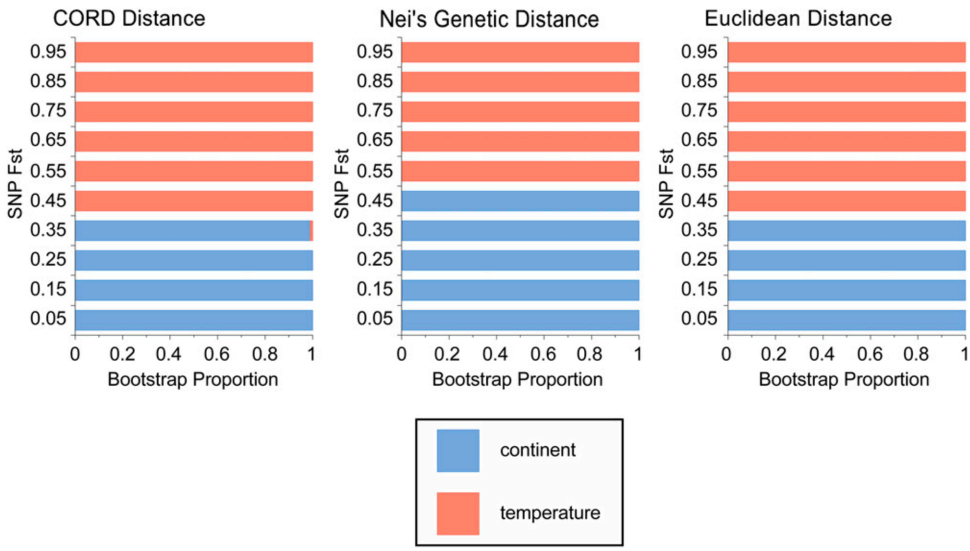

Figure 4 Results for cluster analysis of SNP frequencies. Shown are the proportions of bootstrap replicates that cluster our four sampled populations by either continent (blue) or temperature (red) broken down by the level (percentile) of F_{st} in the Australian cline. These results support convergent evolution of allele frequencies such that more highly differentiated sites cluster populations together on the basis of the environment in which they live rather than geography.

were strongly differentiated in the same direction (*e.g.*, at a site segregating A/T the T allele was at higher frequency in Tasmania and Maine). Mean F_{st} was calculated for each such SNP for both continents.

Functional annotation

We used the DAVID online functional annotation tools to compare enrichment for functional terms among groups of genes (Huang *et al.* 2009). DAVID's tools use a modified Fisher's exact test (the EASE score) to determine the extent of enrichment for a subset of genes compared to a specified background. We compared subsets to their backgrounds as described in *Results* and found the most enriched FAT Gene Ontology (GO) terms in each comparison (FAT GO annotation enriches for more specific GO terms, giving less weight to extremely broad terms). In addition, we also used hypergeometric tests in some cases to test for enrichment of specific GO terms over the expected background. These are the marginal inputs to the EASE score used by DAVID. Recently Pavlidis *et al.* (2012) pointed out that the structure of the genome, particularly the gene length distribution, may lead to spurious enrichments of GO categories among significant windows, even in the absence of true enrichment. While this is so, we present enriched GO categories for the sake of hypothesis generation rather than confirmation.

Results

After mapping and application of the postmapping filters described above, mean coverage from each of the populations was as follows: QUE mean = 26.2×, TAS mean = 26.0×, FLA mean = 36.3×, and MAI mean = 48.9×. The mean for the down-sampled data was 21.6× for each population.

Genomic patterns of polymorphism

Summary statistics of polymorphism in 1-kb windows from the four populations are shown in Table 1. Two patterns are immediately clear. First, high-latitude populations are less polymorphic than low-latitude populations on both continents, genome-wide and for each chromosome arm (all comparisons have *P*-values $<2.2e\text{-}16$). For example, mean $\Theta_\pi = 0.0038$ *vs.* $\Theta_\pi = 0.0045$ in Tasmania *vs.* Queensland, a reduction of ~16% in Tasmania. The reduction in high-latitude polymorphism is less extreme in North America, representing an ~10% decrease. The ratios of X-linked *vs.* autosomal polymorphism (Θ_π) range from 0.63 to 0.68 (Queensland = 0.639, Tasmania = 0.676, Florida = 0.632, and Maine = 0.684), similar to previous genome-wide estimates (Sackton *et al.* 2009; Langley *et al.* 2012; Mackay *et al.* 2012). The low-latitude populations have lower ratios of X- to autosomal-linked variation compared to high-latitude populations. This could be explained by differences in operational sex ratio between populations or by systematic differences in selection on X chromosomes and autosomes, which could result from the presence of clinally varying inversions on the autosomes and the absence of such inversions on the X chromosome.

A second trend is that high-latitude populations exhibit a greater skew toward high-frequency alleles than do low-latitude populations, as summarized by Fay and Wu's *H* statistic ($P < 2.2e\text{-}16$ for both continents). This pattern is consistent across chromosome arms in the Australian samples (all *P*-values $<1e\text{-}05$). The pattern is observed on three of five chromosomes arms in North America, the two exceptions being the X, which shows greater skew toward high-frequency alleles in Florida than in Maine, and 2R, which shows no significant difference between populations. Taken together, the reductions in polymorphism and greater skew in the frequency spectrum of high-latitude populations are highly suggestive of recurrent local adaptation in these populations. This supports previous investigation of latitudinal differentiation in this species (*e.g.*, Kolaczkowski *et al.* 2011).

To test whether segregating inversions affect the site frequency spectrum we calculated Fay and Wu's *H* in 1-kb

Table 3 Correlation coefficients for F_{st} *vs.* recombination rate (cM/Mb) for each continental comparison by chromosome arm as well as throughout the genome (GW)

Continent	Chromosome	Spearman's ρ (F_{st} *vs.* cM/Mb)	*P*-value
Australia	X	−0.0286	1.02*E*-04
	2L	0.0996	2.20*E*-16
	2R	−0.0966	2.20*E*-16
	3L	0.0314	3.49*E*-05
	3R	0.1320	2.20*E*-16
	GW	−0.0461	2.20*E*-16
North America	X	−0.0370	4.90*E*-07
	2L	−0.0039	0.5863
	2R	−0.0237	5.29*E*-03
	3L	−0.0351	3.52*E*-06
	3R	0.1700	2.20*E*-16
	GW	−0.0425	2.20*E*-16

windows that overlapped cosmopolitan inversions *In(3R)P*, *In(3R)Mo*, *In(3L)P*, *In(2L)t*, and *In(2R)NS* and in windows that did not overlap these inversions. We observed a skew toward high-frequency alleles within regions spanned by inversions relative to outside of such regions ($P < 2.2e\text{-}16$ for both continents). This is suggestive of inversions being a potent target of local adaptation (Hoffmann *et al.* 2004; Kirkpatrick and Barton 2006; Kolaczkowski *et al.* 2011; Fabian *et al.* 2012; Kirkpatrick and Kern 2012). We fitted a linear model to explain variation in 1-kb window *H*, which included an effect of geography (low latitude *vs.* high latitude), inversion status, and an interaction effect. Interestingly the interaction was significant along with the main effects, such that windows from high-latitude, inverted regions were significantly more skewed than those from low-latitude, inverted regions. Thus, the inversions within high-latitude populations are the main drivers of this result, suggesting that inversions within low-latitude populations do not harbor a comparatively skewed site frequency spectrum.

Genomic patterns of differentiation

Estimates of F_{st} on the two continents can be found in Table 2 and Figure 2. F_{st} is slightly (but significantly) higher between the Australian populations (mean 1-kb F_{st} = 0.0716) than between the North American populations (mean 1-kb F_{st} = 0.0657). Mean F_{st} from the Australian sample is considerably lower than observed in our previous analysis of the same Australian samples (Kolaczkowski *et al.* 2011, mean F_{st} = 0.112). We believe this effect is attributable to differences in the alignment procedure (ungapped *vs.* gapped) used between the two studies or to differences in the quality of the data generated given the technology available at the time of the Kolaczkowski *et al.* (2011) study.

Levels of differentiation are significantly heterogenous among chromosome arms for both continents. (Kruskal–Wallis rank sum test, $P < 2.2e\text{-}16$ for both continents). We calculated the median F_{st} for 1-kb windows that overlap or do not overlap cosmopolitan inversions as above. For both continents, we found that genomic regions overlapping cosmopolitan inversions are more differentiated (Australia inverted region F_{st} = 0.0631 *vs.* F_{st} = 0.0582 outside of inversions; North America inverted region F_{st} = 0.070 *vs.* F_{st} = 0.060 outside of inverted regions: Wilcoxon rank sum test *P*-value $<2.2e\text{-}16$ for both continents). In addition, we looked at a coarse level if 1-kb windows that overlap genes show greater differentiation among populations than nongenic windows. The results from this parsing of the data were not compelling: the difference is extremely small in F_{st} between genic and nongenic windows and leans slightly toward nongenic windows being slightly more differentiated (Australia genic window F_{st} = 0.0638 *vs.* F_{st} = 0.0649 in nongenic windows; North America genic window F_{st} = 0.0593 *vs.* F_{st} = 0.0610 in nongenic windows).

Given the evidence of parallel inversion clines on the two continents, perhaps it is not surprising that the rank order of chromosomal differentiation is the same on both continents (Figure 3). To further investigate this pattern we calculated a nonparametric correlation coefficient of F_{st} for 1-kb windows on the two continents, using the intersection of windows sampled in both clines from the normally recombining portion of the genome. We find the correlation to be surprisingly high, with a Spearman rank correlation $\rho = 0.2$ ($P < 2.2e\text{-}16$). Correlation coefficients were unchanged in windows that overlapped or did not overlap genes. To test whether this would be expected under a null model of divergence of two pairs of populations from a common ancestor we performed simulations according to the null model of Coop *et al.* (2010), tuned to represent the level of observed differentiation among populations. None of our simulation draws exceeded a correlation coefficient $>\rho = 0.05$, indicating that our result is unexpected under a model of independent divergence ($P < 1e\text{-}4$). These results support the idea that parallel differentiation is common and that the cosmopolitan inversions play a role in this parallelism. To further address this question and investigate whether the inverted regions drive the genomic observation, we estimated the correlation between 1-kb F_{st} estimates on the two continents for each chromosome arm and for regions spanned by the common inversions and those not spanned by the inversions. Although the correlation in windowed F_{st} is higher for inverted regions ($\rho = 0.195$, $P < 2.2e\text{-}16$) than for regions outside of inversions ($\rho = 0.143$, $P < 2.2e\text{-}16$), both are individually highly significant and unexpected given a model of independent divergence.

Classically, convergent evolution has been identified when similar phenotypes have been evolved in independent lineages. At the level of the genotype underlying convergent traits, we would expect to observe similar allele frequencies at sites responsible for adaptive differentiation on both continents. To extend our genomic analysis of parallel adaption to the SNP level we created distance trees for the four populations, using SNP frequencies in each population as our input. If F_{st} in the Australian cline were uncorrelated with F_{st} in the North American cline, there should be no bias in which pairs of populations cluster (*e.g.*, Queensland

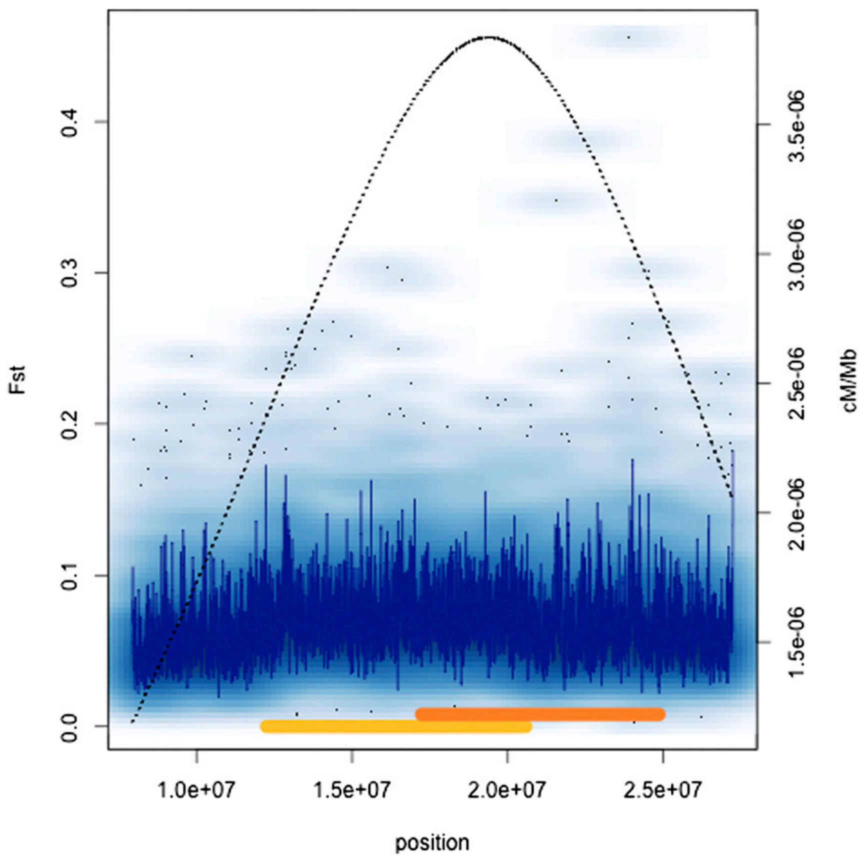

Figure 5 Density scatter plots of F_{st} across chromosome 3R in Australia. The dark blue line represents the running median estimate of F_{st}. The secondary *y*-axis (dashed black line) shows our estimates of recombination rate across the same region. The positions of *In(3R)P* and *In(3R)Mo* are shown with gold and orange bars, respectively. In these plots, density at a position is given by the intensity of the blue cloud, and finer black dots represent outliers.

should be paired with Florida or Maine in equal proportion). Figure 4 shows the results of the bootstrap replicates for the distance trees for each of our three metrics, where horizontal lines represent the proportion of replicates that cluster populations by continent (blue) or by temperature (red). The overwhelming result for this analysis is that for those sites that are most differentiated (roughly the top 60% of F_{st}), populations cluster by the environment in which they live rather than by their geographic location, although there is a clear signal separating North American from Australian populations in the least clinally differentiated SNPs. This result is also echoed in average pairwise patterns of F_{st} both at the chromosome arm level and genome-wide (see Table S24, Table S25, Table S26, Table S27, Table S28, and Table S29). These results provide strong evidence for convergent allele frequency differentiation on both continents and moreover suggest that there are substantial numbers of only moderately differentiated SNPs that are targets of spatially varying selection.

We were also interested in examining the effect of rates of crossing over on F_{st}. Table 3 shows the Spearman rank correlation coefficients between F_{st} and recombination rate both for genome-wide 1-kb window comparisons and on a chromosome-by-chromosome basis. Rate of crossing over explains very little of the variation on a genome-wide basis; Spearman's $\rho \sim -0.04$ in both clinal samples. However, chromosome 3R showed strong positive correlations between recombination and differentiation for both continents. This is notable given that two cosmopolitan inversions, *In(3R)P* and *In(3R)Mo*, are known to exhibit geographic differentiation (*e.g.*, Stalker 1976; Voelker *et al.* 1978). Figure 5 and Figure 6 show F_{st} variation across 3R along with the positions of the rearrangements and our estimates of crossing over for Australian and North American samples, respectively. Figure S1 and Figure S2 show complementary numbers for each chromosome arm for both continents. The chromosome 3R inversions are located in regions of higher crossing in standard homokaryotypic chromosomes, despite the suppression that must occur in inversion heterozygotes. Thus, the correlation between recombination and F_{st} on 3R may be a spurious one driven by the presence of inversions in regions of high recombination in the St karyotype.

Outlier windows

The analyses presented in the previous section support the idea that parallel differentiation is common across the genome, but they provide little insight into the associated biology. To investigate biological patterns associated with spatially varying selection we focused on regions of the genome that fall in an empirical tail distribution of F_{st} on both continents.

Table 4 summarizes the numbers of outlier windows recovered at each empirical cutoff along with the intersection counts. Among outlier samples of the 4351 (4352 for North

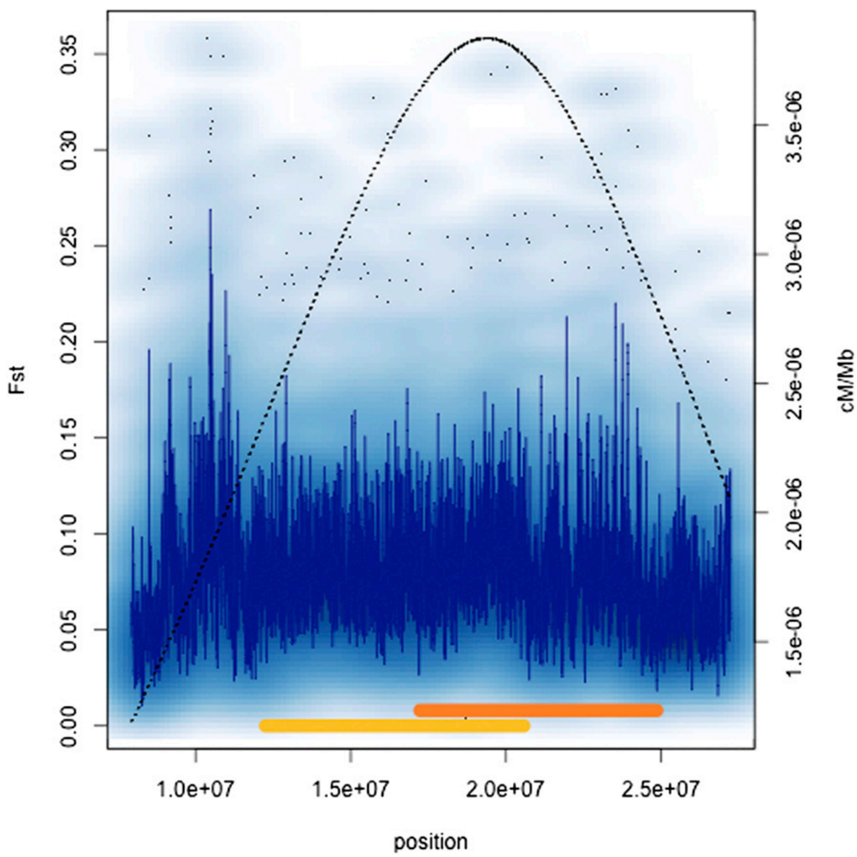

Figure 6 Density scatter plots of F_{st} across chromosome 3R in North America. See Figure 5 legend for details.

America) windows (top 5%) in each cline, we find overlap at 423 windows. This observation is highly unlikely by chance, using a hypergeometric test ($P < 2.2e\text{-}16$). Indeed, overlap between continents is statistically significant genome-wide at each empirical cutoff tested (Table 4). Focusing on the 5% tail, which given the larger number of windows compared to other cutoffs should provide the most power to detect parallel differentiation, we observe especially high levels of overlap for 3R and 2L, suggesting that cosmopolitan inversions play a significant role in the parallelism. However, the fact that this result is significant on the X chromosome, which harbors no common inversion in these populations, shows that inversions alone cannot explain the whole pattern. Table 5 shows contingency tables comparing the number of genes containing overlapping outlier windows on each continent and their intersection. For example, at the 1% cutoff there are 94 genes that overlap outlier windows on both continents. This is substantially more than expected under independence (Fisher's exact test, $P = 6.93e\text{-}27$) and further supports the idea that parallelism is common.

Annotation enrichments at all three empirical cutoffs for both Australia and North America are shown in Figure 7. Generally, we find strong enrichments in multiple functions on both continents but not a great degree of similarity among continents in rank order among annotations. For example, we observed strong enrichments in many classes of RNA genes [including microRNAs (miRNAs) and small nuclear RNAs (snRNAs)] in Australia. Australian outlier windows were also enriched for regulatory elements (Oreganno), CDS, and UTR sequences. In North America, annotation enrichments were generally weaker than those observed in Australia (see Figure 6), yet some functional similarities were present. Outlier windows in the North American sample parallel Australia in enrichments for miRNAs, snRNAs, transfer RNAs (tRNAs), UTRs, and CDS sequences. None of the tests for correlations in annotation enrichments rejected the null hypothesis: 5% tail Spearman's rank correlation $\rho = 0.049$ ($P = 0.89$), 2.5% tail $\rho = -0.055$ ($P = 0.88$), and 1.0% tail $\rho = 0.212$ ($P = 0.551$). Thus, while there is some evidence of parallel functional enrichment in outlier windows, the pattern is not sufficiently strong to yield a statistical signal.

To further investigate the biological properties of shared regions we examined the annotation enrichments for windows in the empirical tail of both continents (shared outliers, Figure 8; note that the y-axis has to be shown on a log scale as some of the enrichments observed are extremely strong). A few functional categories are clear outliers in this analysis: miRNAs, small nucleolar RNAs (snoRNAs), CDS, and 3′-UTRs are strongly enriched in the 1% tail windows. At the 5% and 2.5% cutoffs, miRNAs and CDS are still strongly enriched, with the addition of regulatory elements as annotated in the Oreganno set.

Table 4 Counts of outlier 1-kb windows in two clinal samples at three empirical cutoffs

Chromosome	Australian count	North American count	Both count	*P*-value
		5% tail		
X	727	976	59	0.00067
2L	892	992	93	4.12*E*-10
2R	477	458	38	5.15*E*-07
3L	747	711	47	0.00227
3R	1508	1215	186	<2.2*e*-16
GW	4351	4352	423	<2.2*e*-16
		2.5% tail		
X	395	544	16	0.1237
2L	423	464	20	0.00518
2R	274	227	15	4.78*E*-05
3L	339	354	18	0.00021
3R	745	587	41	0.00032
GW	2176	544	110	5.43*E*-12
		1% tail		
X	180	233	2	0.665095
2L	148	178	3	0.17059
2R	144	80	5	0.00147
3L	117	145	3	0.07455
3R	282	235	9	0.009
GW	871	871	22	9.24*E*-05

The total number of windows included in the intersection analysis was 86,216. This included 17,875 windows on chromosome X, 18,470 windows on chromosome 2L, 13,767 windows on chromosome 2R, 17,339 windows on chromosome 3L, and 18,765 windows on chromosome 3R. The number of outlier windows shared by both samples in given in the "Both count" column. *P*-values are from hypergeometric tests.

Table 5 Comparisons of the numbers of genes overlapped by outlier windows at various empirical cutoffs

	Australian NS genes	Australian significant genes
5% tail		
North American NS genes	7,736	1,272
North American significant genes	1,357	807
	P = 8.309*E*-120	OR = 3.62
2.5% tail		
North American NS genes	9,046	844
North American significant genes	934	348
	P = 1.62*E*-71	OR = 3.99
1.0% tail		
North American NS genes	10,152	435
North American significant genes	491	94
	P = 6.93*E*-27	OR = 4.47

P-values and associated odds ratios are from Fisher's exact test on the presented contingency tables. In each case, there is significantly more overlap in genes that contain outlier windows than one would expect under the null model of independence. NS, nonsignificant; OR, odds ratio.

Outlier windows overlapping protein-coding genes

We carried out two types of analysis to investigate the biology of parallel differentiation from a gene-centric perspective. First, we carried out a GO enrichment analysis for the protein coding genes hit by outlier windows for each continent separately. Second, we identified the set of genes that are located in the shared outlier windows and then tested them for GO enrichments. The results for each continent are in Table S1, Table S2, Table S3, Table S4, Table S5, Table S6, Table S7, Table S8, Table S9, Table S10, Table S11, Table S12, Table S13, Table S14, Table S15, Table S16, Table S17, and Table S18. There is substantial overlap among the significant GO terms found on each continent, which is not surprising given the excess of shared outlier windows. For example, in the 5% tail for North America and Australia, there are 237 and 199 significant biological process GO terms, respectively, after correction for multiple tests. Of these significant terms, 149 are shared among clines. Similar results were obtained using the 2.5% and 1% cutoffs. Shared significant GO terms point to enrichments in genes involved in such processes as transcription regulation, eye development, wing morphogenesis, and circadian rhythm. Indeed it is worth noting that both wing morphology (*e.g.*, McKechnie *et al.* 2010) and circadian rhythm phenotypes (reviewed in Kyriacou *et al.* 2008) are well-studied clinally varying phenotypes. Some interesting behavioral processes are also shared among continents, including learning and memory and olfactory learning. Similarly, high levels of overlapping significant terms are found for the molecular function and cellular component branches of the GO hierarchy. Pavlidis *et al.* (2012) recently pointed to potential problems with GO-type analyses in the *Drosophila* genome that may lead to false positives. While this is so, our results for shared significant GO terms are clearly consistent with known features of spatially varying selection in this species and thus are likely not artifactual.

To further examine biological patterns underling genes shared among outlier windows we performed a DAVID analysis. As input, we used the genes that were found in outlier windows (5% tail) on both continents. This list was compared to a background gene list corresponding to those genes found in all the windows in the genome that survived our filtering and were used in our analyses. We limited our analysis to those clusters with enrichment scores >1.301, corresponding to a geometric mean *P*-value <0.05 of the associated terms in the cluster. This truncation yields 89 significant annotation clusters as predicted by DAVID (see Supporting Information). The single most significant cluster (enrichment score = 11.74) includes terms such as wing disc morphogenesis, wing disc development, appendage development, and metamorphosis. Many of the same genes responsible for wing development appear to be influenced by selection on both continents. Other notable clusters from this analysis include axon guidance and neuronal projection/development (enrichment score = 10.83), eye/photoreceptor development (enrichment score = 8.62), oogenesis and follicle cell development (enrichment score = 8.15), transcriptional regulation (enrichment score = 6.16), *EGF* signaling (enrichment score = 5.19), adult locomotory behavior (enrichment score = 4.99), olfaction (enrichment score = 3.5), and growth (enrichment score = 3.42). Many of these clusters agree with previous observations (Kolaczkowski *et al.* 2011; Fabian *et al.* 2012).

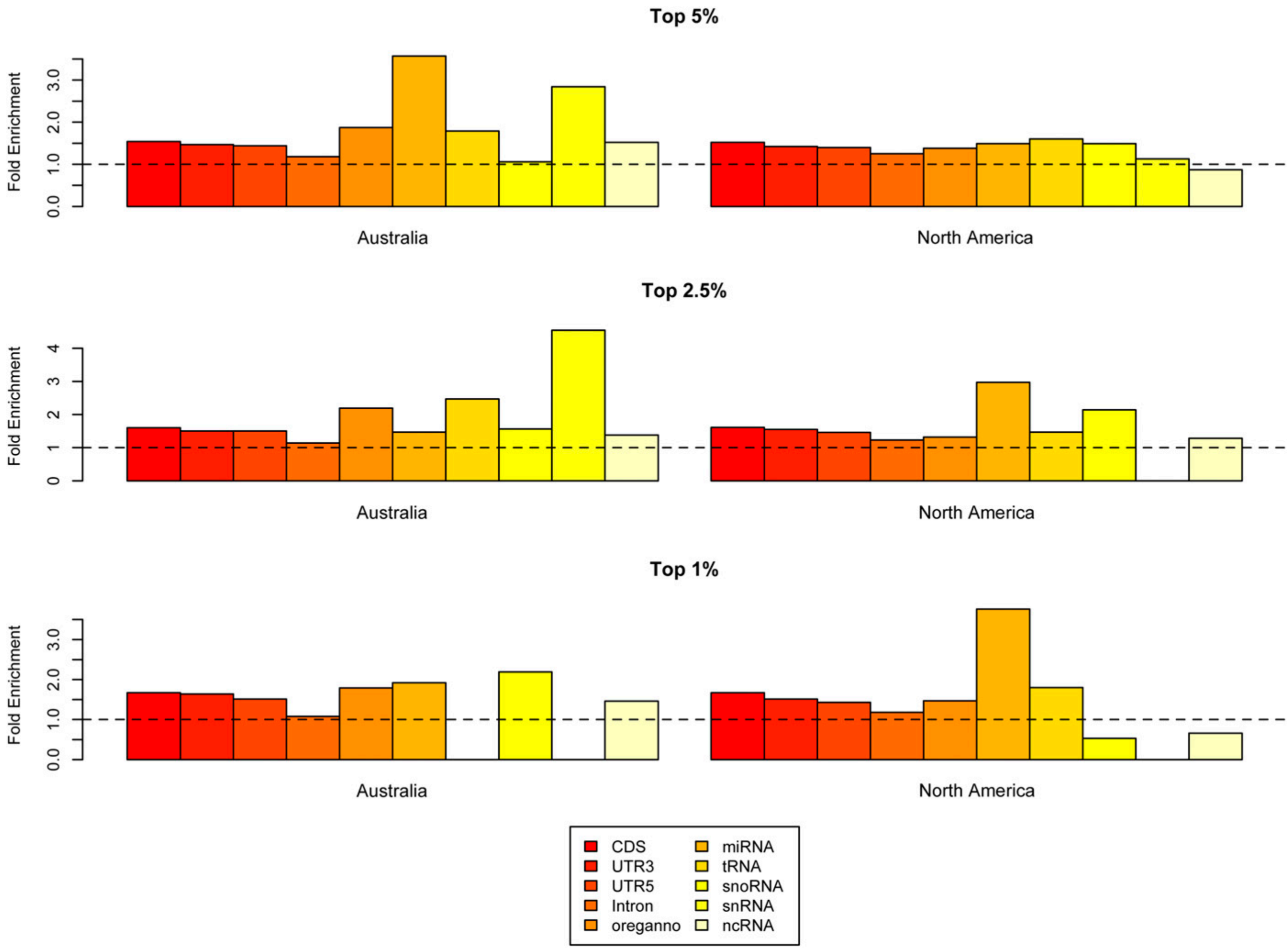

Figure 7 Annotation enrichments within outlier F_{st} windows for Australian and North American samples. The dashed line represents no enrichment over what is expected based on the number of bases sampled.

Convergent SNPs

We observed a total of 361,171 SNPs that were segregating on both continents, of which 64% (229,705) were convergent (defined as those SNPs that show the same direction of allele frequency change on both continents). Under the null hypothesis of drift we expect ~50% of SNPs to be convergent. The large excess of convergent SNPs is indicative of parallel adaptation and is consistent with the neighbor-joining analysis. Convergent SNPs (as defined above) were more differentiated than nonconvergent SNPs, which also strongly supports the inference of natural selection—in North America, mean $F_{st} = 0.091$ for convergent SNPs and mean $F_{st} = 0.075$ for nonconvergent SNPs (*t*-test: $P < 0.0001$), and in Australia, mean $F_{st} = 0.10$ for convergent SNPs and mean $F_{st} = 0.085$ for nonconvergent SNPs (*t*-test: $P < 0.0001$).

To further enrich this set of SNPs for targets of selection we identified convergent SNPs that were among the top 10% most differentiated individual SNPs on both continents, which corresponds to F_{st} of at least 0.248 in Australia and at least 0.213 in North America. We refer to these SNPs as strongly convergent SNPs. Strongly convergent SNPs represent ~1.8% of convergent SNPs (4038 SNPs, Table S19). These SNPs were not enriched for any type of annotation (synonymous, nonsynonymous, UTR, or intronic/intergenic, Table S19) compared either to all other convergent SNPs (chi-square test, $P = 0.71$) or to all other SNPs that vary on both continents (chi-square test, $P = 0.55$).

To determine how much SNP convergence was associated with larger spatial scale effects, we determined the proportion of the strongly convergent SNPs that fell within previously defined convergent windows (the 5% "overlap tail" presented above); 5.8% of strongly convergent SNPs fell into these windows. While this represents a small proportion of strongly convergent SNPs, it is a >10-fold excess compared to expectation (only 0.42% of the genome is represented in these windows; chi-square test, $P < 0.001$). Eleven percent of strongly convergent SNPs fell within the 5-kb region surrounding the 1-kb 5% overlap tail windows, which also represents a significant excess (compared to 2.08% of the genome; chi-square test, $P < 0.0001$).

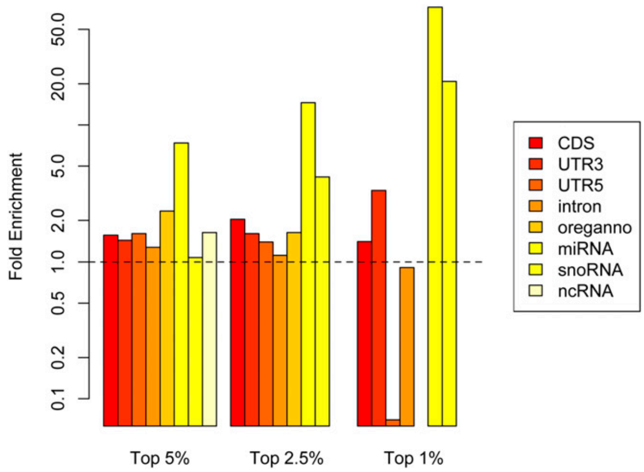

Figure 8 Annotation enrichments within intersection windows (shared outlier widows among clines). Note the *y*-axis is shown on a log scale.

A priori, it seems likely that strongly convergent, synonymous SNPs in particular should usually be found within strongly differentiated windows, as it is generally assumed that synonymous SNPs are relatively unlikely to be direct targets of strong spatially varying selection. Indeed, we found that strongly convergent synonymous SNPs were more likely than other (*e.g.*, nonsynonymous) strongly convergent SNPs to be found in an overlap tail window (7.8% compared to 5.5%; Pearson's chi-square test, $P = 0.039$). However, this still leaves many synonymous SNPs that do not appear to associate directly with a strongly differentiated window. While some of these SNPs may be strongly convergent due to chance, it seems more probable that these SNPs are associated with differentiated windows that are not sufficiently differentiated to be in the 5% overlap tail. Overall, 137/377 of the 5% overlap tail windows contained at least one strongly convergent SNP, and 181/377 windows had a SNP within 5 kb surrounding the window (2 kb to either side of the window). Thus, the SNP-based analysis shows overlap with the window-based analysis, as expected. However, it also appears that there is substantial information contained in the SNPs that is not contained in the windows, likely because outlier windows are associated with a somewhat larger physical scale of differentiation compared to the set of all parallel outlier SNPs.

One goal of the analysis at the single-nucleotide level is to identify SNPs playing a role in adaptation to high-latitude environments. Thus, we investigated whether genes that carry strongly convergent SNPs share any common biological features. For this analysis, we considered genes that carried either a nonsynonymous or a UTR polymorphism, as these categories of SNP may be more likely to be the direct targets of selection than other categories of SNPs. We performed a DAVID analysis to compare genes carrying at least one strongly convergent UTR or nonsynonymous SNP (303 genes) to a background list of 4893 genes containing at least one UTR or nonsynonymous SNP shared by both continents. We also compared to all *Drosophila* genes in DAVID's database. GO categories that were significantly enriched at a false discovery rate <0.10 are shown in Table S20. The comparison to all *Drosophila* genes recaptured these GO terms as well as showing enrichment for other terms associated with cell membrane function and metal binding. Six nonsynonymous and two 5′-UTR SNPs were among the top 100 most differentiated SNPs in the analysis (these were in the top 1% of F_{st} on both continents). We performed a detailed annotation of these SNPs (Table S21) and found that many of them are associated with genes of known function.

Different nonsynonymous SNPs in the same genes

Convergent evolution can be defined at multiple levels. For example, one might observe adaptive divergence on both continents but at different sites in the same gene. In principle, such convergent changes would support the idea that strong selection on the same gene could have different population genetic outcomes on the two continents, either because of differences in the details of selection (at a focal gene or in the genetic background) or because of differences relating to stochastic events associated with founding events associated with colonization. We generated a list of genes showing high F_{st} nonsynonymous SNPs on both continents but for which the nonsynonymous SNPs are different. We then prioritized them as candidates for convergent evolution (at the gene level) based on the degree of conservation of the corresponding residues in sequenced outgroup species. Our top five candidate genes, *hkl*, *otk*, *ana1*, *chm*, and *trp* (summary statistics for and locations of the candidate SNPs in each gene are provided in Table S22), are promising targets of future experimental work.

Continent-specific differentiation

Although the main issue addressed in this article is the degree of shared differentiation on two continents, there are many interesting differences between continents as well. An unknown fraction of such differences could represent false positives or false negatives. However, in a number of cases continent-specific differentiation is characterized by good coverage in all four population samples and substantial physical distances; these are therefore likely to be real. Three excellent examples of strong differentiation in Australia but not in North America are the *upd2* region [X chromosome (Kolaczkowski *et al.* 2011)], the tip of the X chromosome, and the *Cyp6g1* region [chromosome 2R (Daborn *et al.* 2002; Schmidt *et al.* 2010)] (Figure S3, Figure S4, and Figure S5).

Discussion

Phenotypic convergence of independent lineages across similar environments has been a fundamental, ubiquitous observation of evolutionary biology ever since Darwin. However, the degree to which convergent evolution at the phenotypic level is determined by convergent genetic changes is largely unknown (see Losos 2011 for a recent review). Our data suggest that the parallel phenotypic latitudinal clines in *D. melanogaster* are reflected in substantial convergent evolution at the genomic level (Table 4, Table 5, and Figure 4). This strongly supports the idea that selection on ancestral variation underlies much of the latitudinal differentiation on multiple continents in this species (Turner *et al.* 2008; Fabian *et al.* 2012). The alternative explanation, recurrent selected mutations occurring on multiple continents, not only is unlikely given the short timescales for population divergence in the United States and Australia, but also predicts substantial amounts of large physical scale differentiation, which is not evident. Similar patterns of parallel selection on ancestral variation have also been observed in replicated populations of freshwater sticklebacks (Hohenlohe *et al.* 2010; Jones *et al.* 2012). We observed substantial enrichment of convergent allele frequency changes in the United States and Australia, even for SNPs that are not extremely differentiated. This suggests that our empirical tail cutoffs are conservative and that dense sampling of several populations along latitudinal transects on both continents will reveal that a significant fraction of the genome exhibits parallel clines. While inversions are clearly important for latitudinal adaptation in this species, the strong patterns of parallelism observed at the SNP level are not strictly associated with regions spanned by inversions. Better sampling of genomes, populations, and continents will be necessary to quantitatively evaluate alternate models of selection on standing variation. Additional modeling of selection occurring on standing variation in the context of specific demographic models and underlying genomic patterns of linkage disequilibrium will also be important.

Nevertheless, inspection of Table 4 and Table 5 reveals a subset of protein-coding genes that show differentiation on each continent, but in different portions of the genes. For instance, at the 5% tail 807 genes are hit by at least one outlier window on each continent, but only 423 outlier windows are shared. If we ask how many genes in common are hit by *only* different outlier windows in both clines (*i.e.*, exclude any gene that contains intersecting windows), we find that 374 genes are differentiated over different portions of the gene on each continent. This general pattern was observed at other tail cutoffs (not shown). Thus, while there is an abundance of exact convergence at the molecular level, there may also be adaptive divergence at the same gene that is proceeding through different changes. This situation mirrors what has been seen in other systems, including other *Drosophila* species, where a large portion of convergent evolution seems to be the result of nonidentical genetic changes (Gompel and Prud'homme 2009; Kopp 2009; Christin *et al.* 2010).

Acknowledgments

A.D.K. was supported in part by Rutgers University and National Science Foundation award MCB-1161367. D.J.B. was supported by National Institutes of Health grant GM084056. C.D.J. acknowledges the support of the University Cancer Research Fund.

Literature Cited

Akey, J. M., G. Zhang, K. Zhang, L. Jin, and M. D. Shriver, 2002 Interrogating a high-density SNP map for signatures of natural selection. Genome Res. 12(12): 1805–1814.

Beaumont, M. A., and R. A. Nichols, 1996 Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. B Biol. Sci. 263(1377): 1619–1626.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature 365: 548–550.

Cavalli-Sforza, L. L., and A. W. Edwards, 1967 Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 19(3 Pt 1): 233.

Christin, P. A., D. M. Weinreich, and G. Besnard, 2010 Causes and evolutionary significance of genetic convergence. Trends Genet. 26(9): 400–405.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard, 2010 Using environmental correlations to identify loci underlying local adaptation. Genetics 185: 1411–1423.

Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil *et al.*, 2002 A single P450 allele associated with insecticide resistance in Drosophila. Science 297(5590): 2253–2256.

David, J. R., and P. Capy, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. 4(4): 106–111.

Duchen, P., D. Živković, S. Hutter, W. Stephan, and S. Laurent, 2013 Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. Genetics 193: 291–301.

Fabian, D. K., M. Kapun, V. Nolte, R. Kofler, P. S. Schmidt *et al.*, 2012 Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America. Mol. Ecol. 21(19): 4748–4769.

Gompel, N., and B. Prud'homme, 2009 The causes of repeated genetic evolution. Dev. Biol. 332(1): 36–47.

Hoffmann, A. A., and A. R. Weeks, 2007 Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. Genetica 129(2): 133–147.

Hoffmann, A. A., C. M. Sgrò, and A. R. Weeks, 2004 Chromosomal inversion polymorphisms and adaptation. Trends Ecol. Evol. 19 (9): 482–488.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson *et al.*, 2010 Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6(2): e1000862.

Huang, D. W., B. T. Sherman, X. Zheng, J. Yang, T. Imamichi *et al.*, 2009 Extracting biological meaning from large gene lists with DAVID. Curr. Protoc. Bioinformatics Chap. 13: Unit 13.11.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli *et al.*, 2012 The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484(7392): 55–61.

Keller, A., 2007 *Drosophila melanogaster*'s history as a human commensal. Curr. Biol. 17(3): R77–R81.

Knibb, W. R., 1982 Chromosome inversion polymorphisms in Drosophila melanogaster II. Geographic clines and climatic associations in Australasia, North America and Asia. Genetica 58(3): 213–221.

Kirkpatrick, M., and N. Barton, 2006 Chromosome inversions, local adaptation and speciation. Genetics 173(1): 419–434.

Kirkpatrick, M., and A. Kern, 2012 Where's the money? Inversions, genes, and the hunt for genomic targets of selection. Genetics 190(4): 1153–1155.
Kolaczkowski, B., A. D. Kern, A. K. Holloway, and D. J. Begun, 2011 Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. Genetics 187: 245–260.
Kopp, A., 2009 Metamodels and phylogenetic replication: a systematic approach to the evolution of developmental pathways. Evolution 63(11): 2771–2789.
Kyriacou, C. P., A. A. Peixoto, F. Sandrelli, R. Costa, and E. Tauber, 2008 Clines in clock genes: fine-tuning circadian rhythms to the environment. Trends Genet 24: 124–132.
Lachaise, D., M. L. Cariou, J. R. David, F. Lemeunier, L. Tsacas *et al.*, 1988 Historical biogeography of the Drosophila melanogaster species subgroup. Evol. Biol. 22: 159–225.
Langley, C. H., K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider *et al.*, 2012 Genomic variation in natural populations of *Drosophila melanogaster*. Genetics 192: 533–598.
Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14): 1754–1760.
Losos, J. B., 2011 Convergence, adaptation, and constraint. Evolution 65(7): 1827–1840.
Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The Drosophila melanogaster genetic reference panel. Nature 482(7384): 173–178.
McKechnie, S. W., M. J. Blacket, S. V. Song, L. Rako, X. Carroll *et al.*, 2010 A clinally varying promoter polymorphism associated with adaptive variation in wing size in Drosophila. Mol. Ecol. 19: 775–784.
Nei, M., 1972 Genetic distance between populations. Am. Nat. 106: 283–292.
Pavlidis, P., J. D. Jensen, W. Stephan, and A. Stamatakis, 2012 A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Mol. Biol. Evol. 29(10): 3237–3248.
Pickrell, J. K., G. Coop, J. Novembre, S. Kudaravalli, J. Z. Li *et al.*, 2009 Signals of recent positive selection in a worldwide sample of human populations. Genome Res. 19(5): 826–837.
Sackton, T. B., R. J. Kulathinal, C. M. Bergman, A. R. Quinlan, E. B. Dopman *et al.*, 2009 Population genomic inferences from sparse high-throughput sequencing of two populations of Drosophila melanogaster. Genome Biol. Evol. 1: 449.
Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4(4): 406–425.
Schmidt, J. M., R. T. Good, B. Appleton, J. Sherrard, G. C. Raymant *et al.*, 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. PLoS Genet. 6(6): e1000998.
Singh, R. S., and A. D. Long, 1992 Geographic variation in *Drosophila*: from molecules to morphology and back. Trends Ecol. Evol. 7(10): 340–345.
Stalker, H. D., 1976 Chromosome studies in wild populations of *D. melanogaster*. Genetics 82: 323–347.
Stephan, W., and H. Li, 2007 The recent demographic and adaptive history of Drosophila melanogaster. Heredity 98(2): 65–68.
Teshima, K. M., G. Coop, and M. Przeworski, 2006 How reliable are empirical genomic scans for selective sweeps? Genome Res. 16(6): 702–712.
Turner, T. L., M. T. Levine, M. L. Eckert, and D. J. Begun, 2008 Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. Genetics 179: 455–473.
Voelker, R. A., C. C. Cockerham, F. M. Johnson, H. E. Schaffer, T. Mukai *et al.*, 1978 Inversions fail to account for allozyme clines. Genetics 88: 515–527.
Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard, 2006 A map of recent positive selection in the human genome. PLoS Biol. 4(3): e72.

Communicating editor: W. Stephan

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161463/-/DC1

Parallel Geographic Variation in *Drosophila melanogaster*

Josie A. Reinhardt, Bryan Kolaczkowski, Corbin D. Jones, David J. Begun, and Andrew D. Kern

DOI: 10.1534/genetics.114.161463

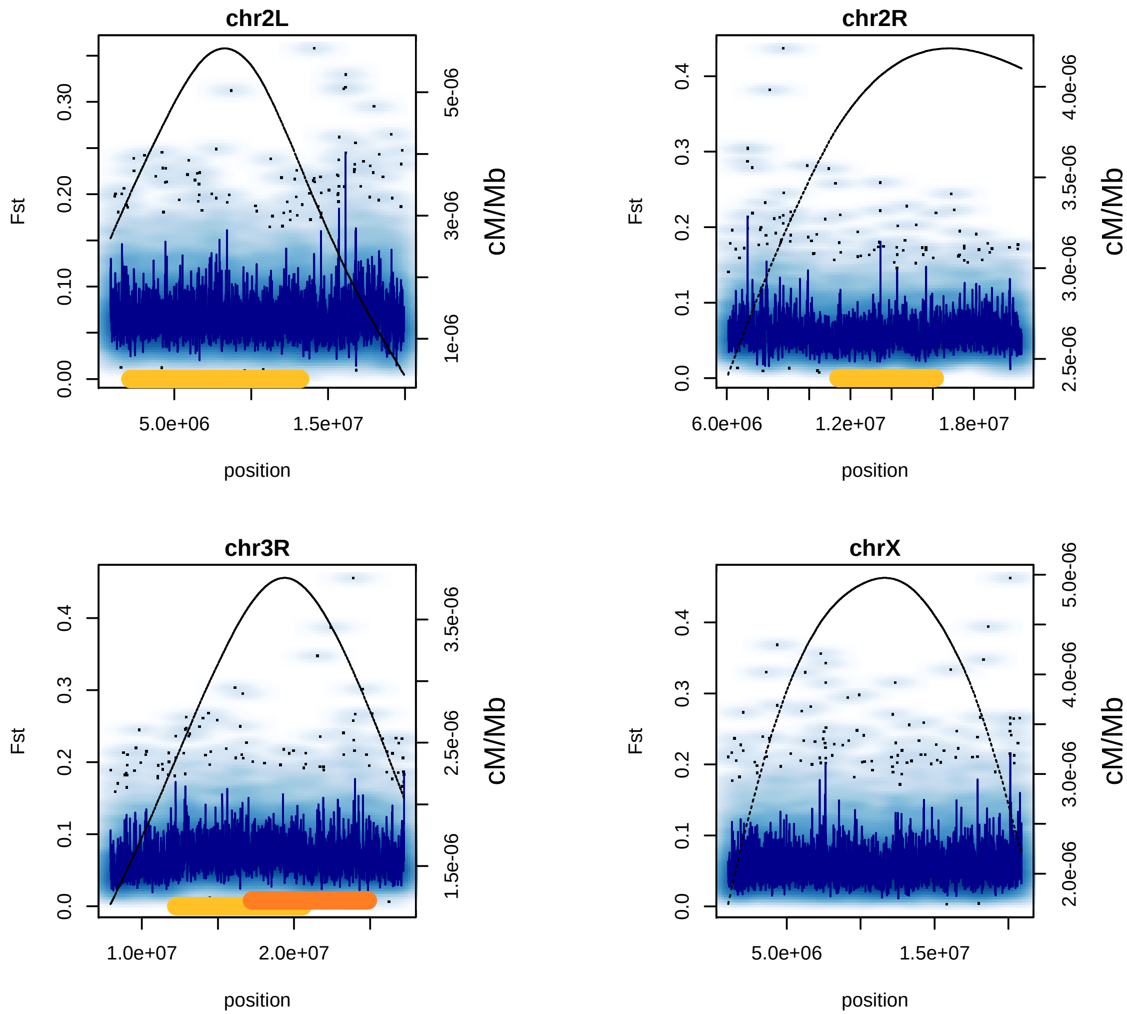

Figure S1 Australian cline Fst across chromosome arms

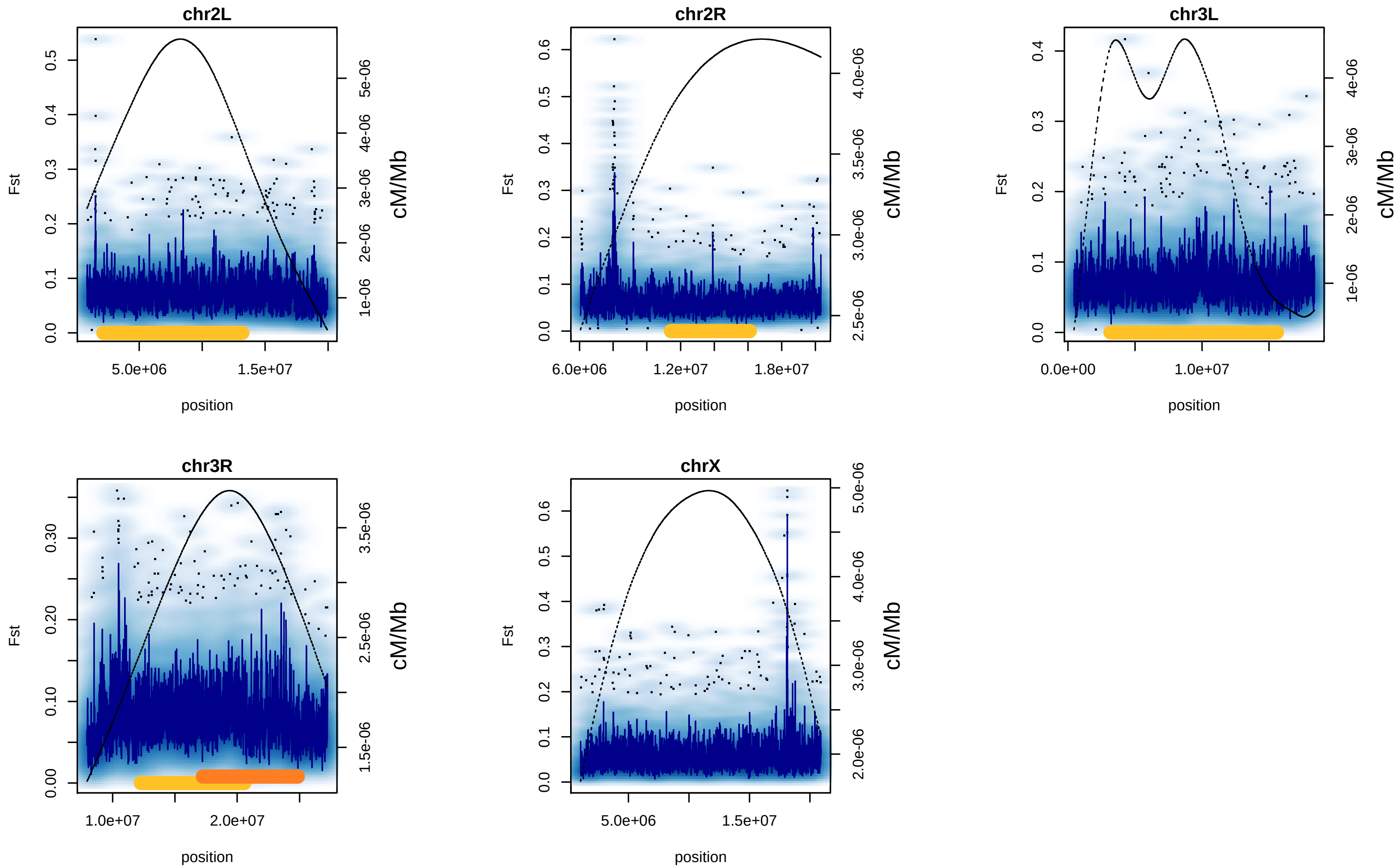

Figure S2 North American cline Fst across chromosome arms

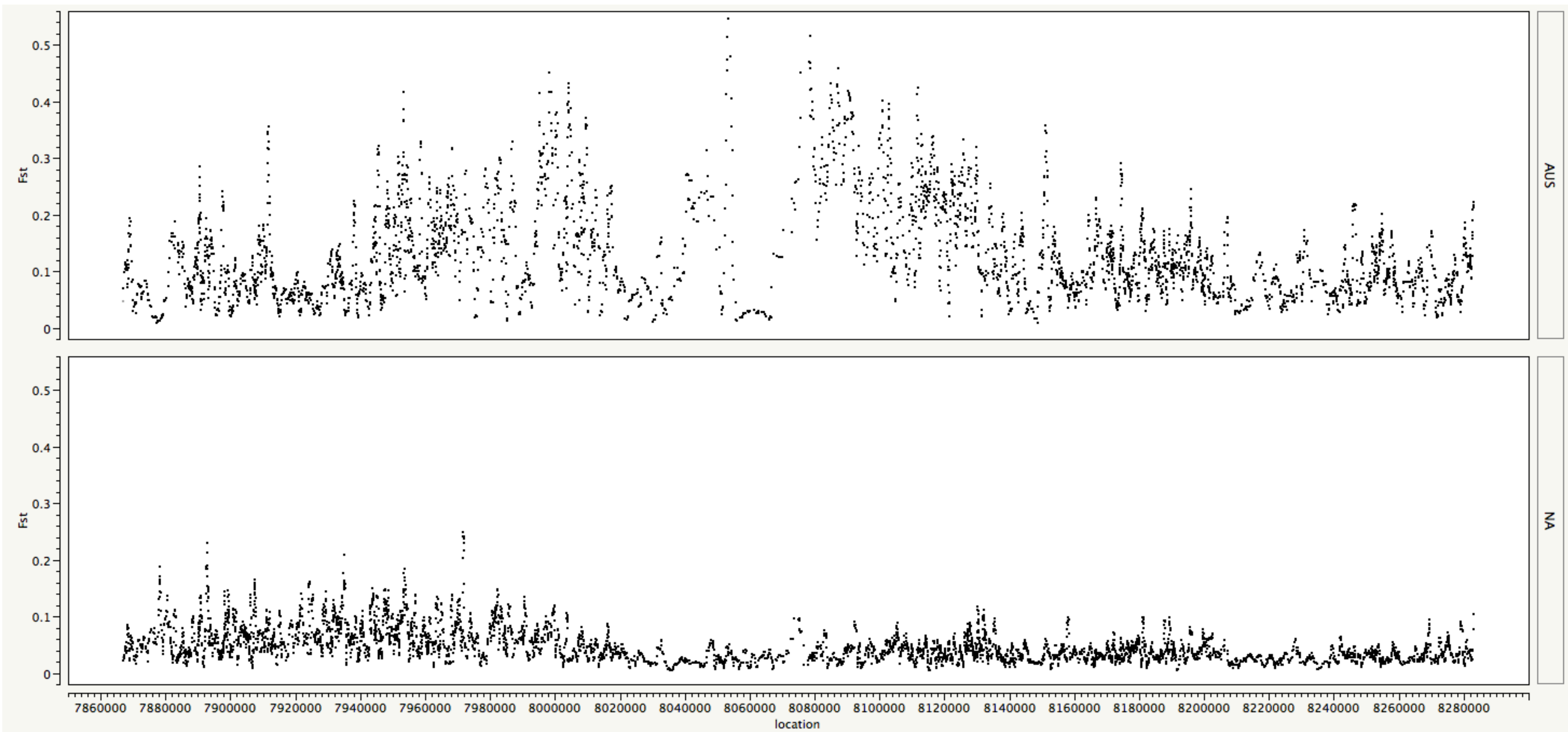

Figure S3 Comparison of Fst between continents in Cyp6G1 region

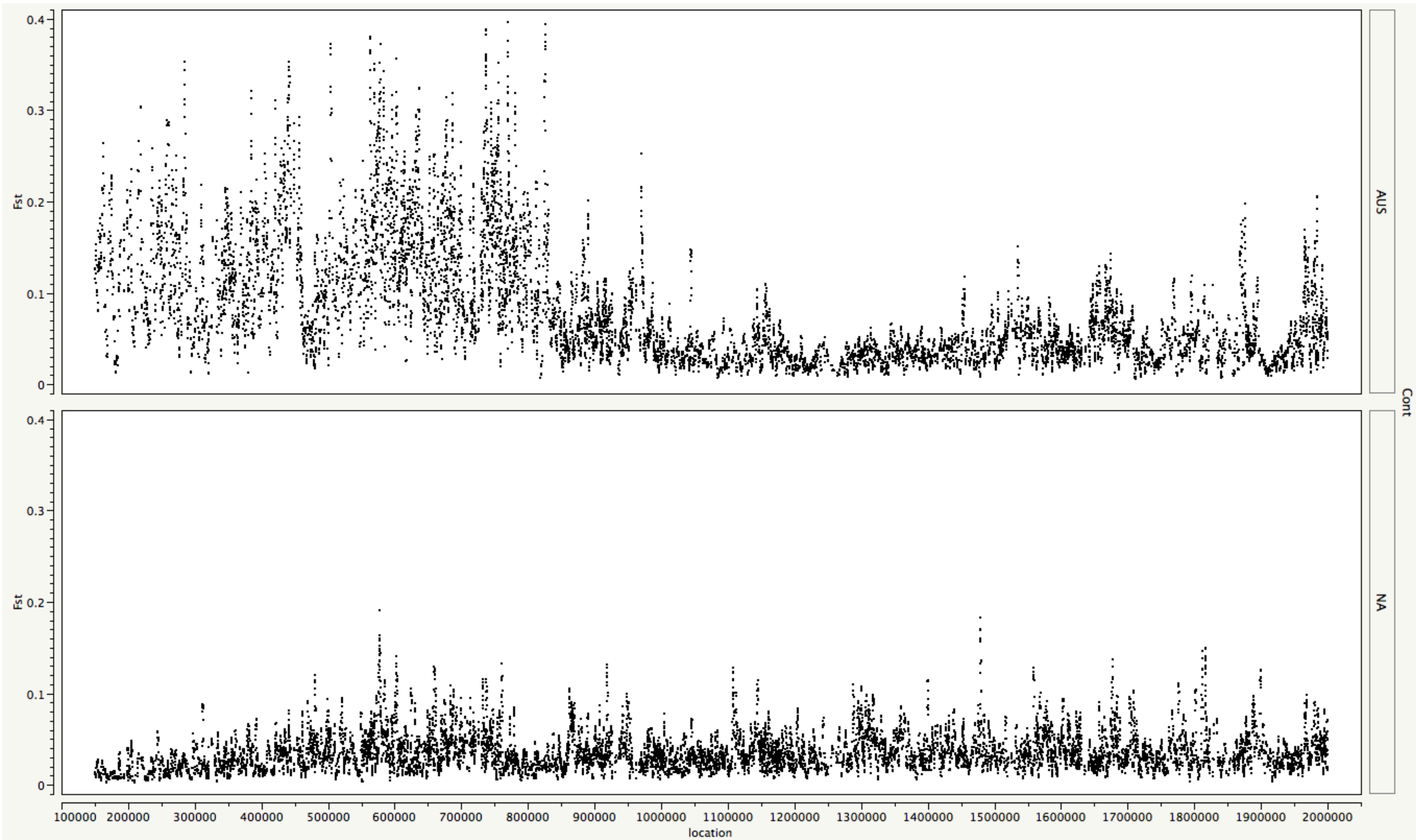

Figure S4 Comparison of Fst between continents at tip of chromosome X

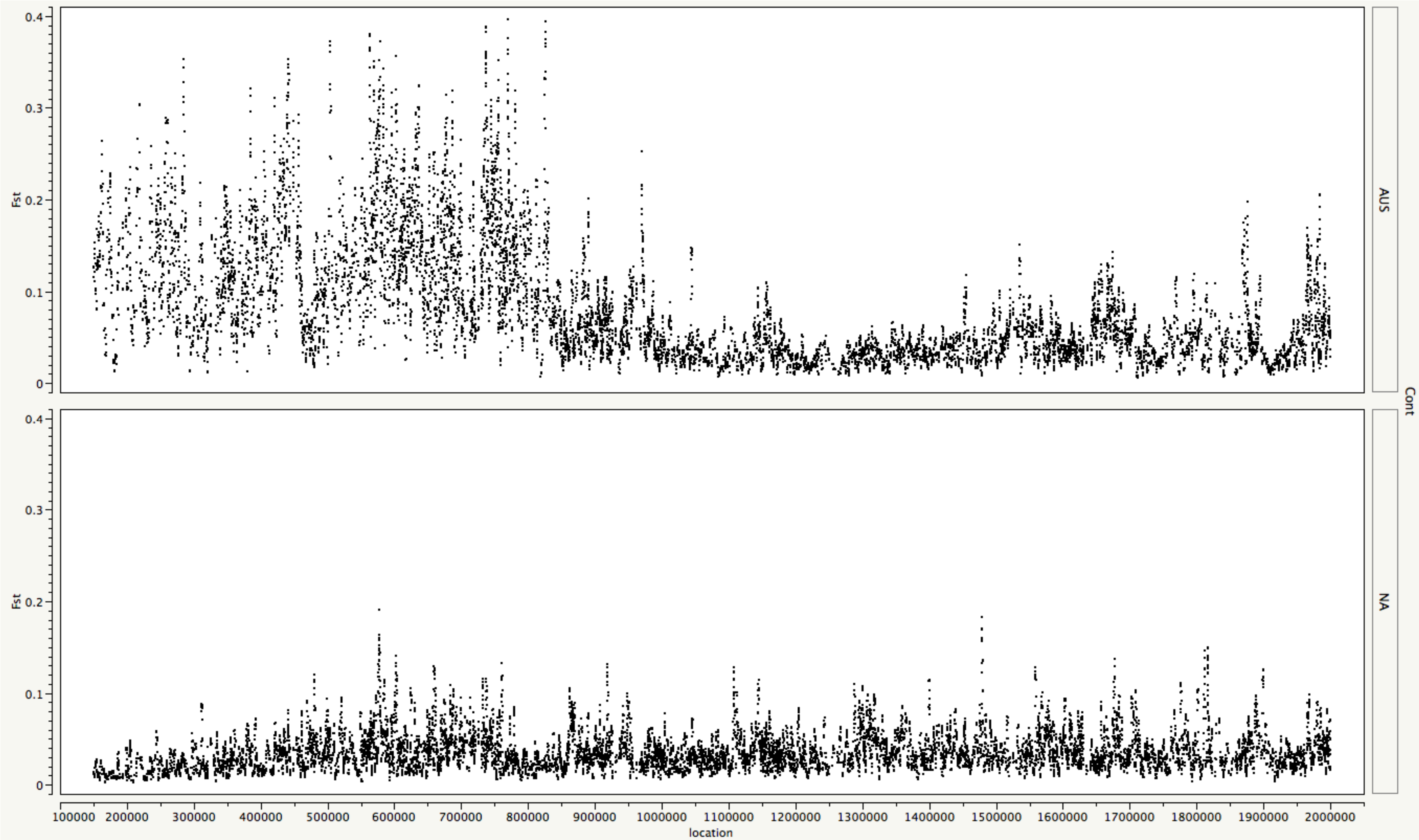

Figure S5 Comparison of Fst between continents in Upd2 region

Table S1 GO term enrichment for biological process in the Australian cline for genes that overlap windows in the top 1% tail of Fst.

	1 % Tail Biological Process	
GO term	q-value	description
GO:0008355	4.99E-06	olfactory learning
GO:0007611	9.11E-06	learning or memory
GO:0001738	9.68E-05	morphogenesis of a polarized epithelium
GO:0007156	9.68E-05	homophilic cell adhesion
GO:0008285	9.68E-05	negative regulation of cell proliferation
GO:0007165	9.68E-05	signal transduction
GO:0007306	1.74E-04	eggshell chorion assembly
GO:0007507	2.37E-04	heart development
GO:0007427	2.68E-04	epithelial cell migration, open tracheal system
GO:0051276	3.00E-04	chromosome organization
GO:0008104	3.08E-04	protein localization
GO:0017148	3.71E-04	negative regulation of translation
GO:0007623	4.62E-04	circadian rhythm
GO:0007155	6.08E-04	cell adhesion
GO:0006811	6.12E-04	ion transport
GO:0002009	7.72E-04	morphogenesis of an epithelium
GO:0007424	8.38E-04	open tracheal system development
GO:0048477	9.93E-04	oogenesis
GO:0007391	1.23E-03	dorsal closure
GO:0007435	1.23E-03	salivary gland morphogenesis
GO:0045475	1.36E-03	locomotor rhythm
GO:0006355	1.37E-03	regulation of transcription, DNA-dependent
GO:0007275	1.37E-03	multicellular organismal development
GO:0006468	1.73E-03	protein amino acid phosphorylation
GO:0007476	4.39E-03	imaginal disc-derived wing morphogenesis
GO:0002121	6.46E-03	inter-male aggressive behavior
GO:0006813	6.60E-03	potassium ion transport
GO:0048749	9.21E-03	compound eye development
GO:0045449	1.17E-02	regulation of transcription
GO:0007411	1.24E-02	axon guidance
GO:0016319	1.38E-02	mushroom body development
GO:0007298	1.40E-02	border follicle cell migration
GO:0001700	1.87E-02	embryonic development via the syncytial blastoderm
GO:0007015	2.28E-02	actin filament organization
GO:0001745	2.73E-02	compound eye morphogenesis
GO:0006814	2.94E-02	sodium ion transport
GO:0008360	4.19E-02	regulation of cell shape
GO:0055085	4.19E-02	transmembrane transport
GO:0007269	4.73E-02	neurotransmitter secretion
GO:0016192	4.95E-02	vesicle-mediated transport

Table S2 GO term enrichment for biological process in the Australian cline for genes that overlap windows in the top 2.5% tail of Fst.

	2.5 % Tail Biological Process	
GO term	q-value	description
GO:0006355	2.24E-07	regulation of transcription, DNA-dependent
GO:0007611	1.26E-06	learning or memory
GO:0006468	1.26E-06	protein amino acid phosphorylation
GO:0008543	3.05E-06	fibroblast growth factor receptor signaling pathway
GO:0007476	3.11E-06	imaginal disc-derived wing morphogenesis
GO:0008284	3.49E-06	positive regulation of cell proliferation
GO:0007424	5.44E-06	open tracheal system development
GO:0008355	1.11E-05	olfactory learning
GO:0007156	1.56E-05	homophilic cell adhesion
GO:0007435	1.87E-05	salivary gland morphogenesis
GO:0007155	1.87E-05	cell adhesion
GO:0007275	1.87E-05	multicellular organismal development
GO:0007507	1.87E-05	heart development
GO:0046331	1.87E-05	lateral inhibition
GO:0007280	4.39E-05	pole cell migration
GO:0007513	4.39E-05	pericardial cell differentiation
GO:0007427	4.39E-05	epithelial cell migration, open tracheal system
GO:0008406	4.76E-05	gonad development
GO:0048190	4.98E-05	wing disc dorsal/ventral pattern formation
GO:0007411	5.20E-05	axon guidance
GO:0048477	7.54E-05	oogenesis
GO:0007623	1.42E-04	circadian rhythm
GO:0048749	1.42E-04	compound eye development
GO:0008104	1.70E-04	protein localization
GO:0007391	2.08E-04	dorsal closure
GO:0007419	2.14E-04	ventral cord development
GO:0007398	3.20E-04	ectoderm development
GO:0045475	3.20E-04	locomotor rhythm
GO:0001751	3.20E-04	compound eye photoreceptor cell differentiation
GO:0001738	3.82E-04	morphogenesis of a polarized epithelium
GO:0008345	3.82E-04	larval locomotory behavior
GO:0007509	4.07E-04	mesoderm migration
GO:0007498	4.07E-04	mesoderm development
GO:0035225	4.07E-04	determination of genital disc primordium
GO:0008354	4.07E-04	germ cell migration
GO:0007617	4.07E-04	mating behavior
GO:0006811	4.92E-04	ion transport
GO:0007219	4.92E-04	Notch signaling pathway

GO:0007428	5.97E-04	primary branching, open tracheal system
GO:0007274	6.14E-04	neuromuscular synaptic transmission
GO:0008356	6.14E-04	asymmetric cell division
GO:0051225	6.14E-04	spindle assembly
GO:0045433	7.31E-04	male courtship behavior, veined wing generated song production
GO:0045187	7.31E-04	regulation of circadian sleep/wake cycle, sleep
GO:0010552	7.31E-04	positive regulation of gene-specific transcription from RNA polymerase II promoter
GO:0016477	8.59E-04	cell migration
GO:0008285	8.59E-04	negative regulation of cell proliferation
GO:0007480	8.59E-04	imaginal disc-derived leg morphogenesis
GO:0007165	8.59E-04	signal transduction
GO:0002121	8.89E-04	inter-male aggressive behavior
GO:0042127	1.20E-03	regulation of cell proliferation
GO:0048675	1.20E-03	axon extension
GO:0008407	1.20E-03	bristle morphogenesis
GO:0007622	1.20E-03	rhythmic behavior
GO:0016079	1.20E-03	synaptic vesicle exocytosis
GO:0008344	1.24E-03	adult locomotory behavior
GO:0001709	1.24E-03	cell fate determination
GO:0008347	1.24E-03	glial cell migration
GO:0007494	1.64E-03	midgut development
GO:0002009	1.64E-03	morphogenesis of an epithelium
GO:0035023	1.64E-03	regulation of Rho protein signal transduction
GO:0045449	1.71E-03	regulation of transcription
GO:0007268	1.77E-03	synaptic transmission
GO:0016360	1.77E-03	sensory organ precursor cell fate determination
GO:0007390	1.77E-03	germ-band shortening
GO:0007475	1.77E-03	apposition of dorsal and ventral imaginal disc-derived wing surfaces
GO:0000902	1.77E-03	cell morphogenesis
GO:0055085	1.77E-03	transmembrane transport
GO:0008049	2.57E-03	male courtship behavior
GO:0007431	2.57E-03	salivary gland development
GO:0030707	2.66E-03	ovarian follicle cell development
GO:0008340	3.61E-03	determination of adult lifespan
GO:0042052	3.61E-03	rhabdomere development
GO:0007314	3.61E-03	oocyte anterior/posterior axis specification
GO:0007298	3.68E-03	border follicle cell migration
GO:0007306	4.11E-03	eggshell chorion assembly
GO:0051017	4.11E-03	actin filament bundle formation
GO:0001745	4.11E-03	compound eye morphogenesis
GO:0008587	4.11E-03	imaginal disc-derived wing margin morphogenesis
GO:0016055	4.73E-03	Wnt receptor signaling pathway
GO:0030162	4.75E-03	regulation of proteolysis

GO:0030718	4.82E-03	germ-line stem cell maintenance
GO:0035277	5.70E-03	spiracle morphogenesis, open tracheal system
GO:0008103	5.70E-03	oocyte microtubule cytoskeleton polarization
GO:0006914	5.70E-03	autophagy
GO:0035147	5.70E-03	branch fusion, open tracheal system
GO:0007269	6.05E-03	neurotransmitter secretion
GO:0007367	6.05E-03	segment polarity determination
GO:0007619	6.05E-03	courtship behavior
GO:0008586	7.37E-03	imaginal disc-derived wing vein morphogenesis
GO:0051276	7.37E-03	chromosome organization
GO:0007616	7.37E-03	long-term memory
GO:0009408	7.37E-03	response to heat
GO:0006816	7.37E-03	calcium ion transport
GO:0019991	7.37E-03	septate junction assembly
GO:0007455	7.37E-03	eye-antennal disc morphogenesis
GO:0007552	7.37E-03	metamorphosis
GO:0045892	7.50E-03	negative regulation of transcription, DNA-dependent
GO:0006813	7.95E-03	potassium ion transport
GO:0016319	8.34E-03	mushroom body development
GO:0008293	9.71E-03	torso signaling pathway
GO:0017148	9.77E-03	negative regulation of translation
GO:0042078	9.77E-03	germ-line stem cell division
GO:0033227	1.23E-02	dsRNA transport
GO:0007426	1.31E-02	tracheal outgrowth, open tracheal system
GO:0046843	1.35E-02	dorsal appendage formation
GO:0007417	1.48E-02	central nervous system development
GO:0045893	1.51E-02	positive regulation of transcription, DNA-dependent
GO:0006897	1.51E-02	endocytosis
GO:0042067	1.67E-02	establishment of ommatidial polarity
GO:0007444	1.88E-02	imaginal disc development
GO:0007242	1.98E-02	intracellular signaling cascade
GO:0007399	2.06E-02	nervous system development
GO:0006325	2.10E-02	chromatin organization
GO:0007422	2.11E-02	peripheral nervous system development
GO:0008360	2.18E-02	regulation of cell shape
GO:0006357	2.48E-02	regulation of transcription from RNA polymerase II promoter
GO:0048666	2.70E-02	neuron development
GO:0006470	3.05E-02	protein amino acid dephosphorylation
GO:0007317	3.14E-02	regulation of pole plasm oskar mRNA localization
GO:0007224	3.14E-02	smoothened signaling pathway
GO:0008045	3.14E-02	motor axon guidance
GO:0042048	3.15E-02	olfactory behavior
GO:0001700	3.33E-02	embryonic development via the syncytial blastoderm
GO:0007423	3.49E-02	sensory organ development

GO:0007474	3.69E-02	imaginal disc-derived wing vein specification
GO:0007163	3.69E-02	establishment or maintenance of cell polarity
GO:0007349	3.69E-02	cellularization
GO:0007015	4.26E-02	actin filament organization
GO:0008595	4.39E-02	determination of anterior/posterior axis, embryo
GO:0008063	4.39E-02	Toll signaling pathway

Table S3   GO term enrichment for biological process in the Australian cline for genes that overlap windows in the top 5% tail of Fst.

	5 % Tail Biological Process	
GO term	q-value	description
GO:0046331	0.00E+00	lateral inhibition
GO:0016198	0.00E+00	axon choice point recognition
GO:0006355	4.83E-10	regulation of transcription, DNA-dependent
GO:0007156	5.44E-08	homophilic cell adhesion
GO:0007411	5.39E-07	axon guidance
GO:0006468	5.53E-07	protein amino acid phosphorylation
GO:0048190	2.06E-06	wing disc dorsal/ventral pattern formation
GO:0007611	2.06E-06	learning or memory
GO:0007476	2.67E-06	imaginal disc-derived wing morphogenesis
GO:0008284	6.75E-06	positive regulation of cell proliferation
GO:0042051	8.05E-06	compound eye photoreceptor development
GO:0008543	9.53E-06	fibroblast growth factor receptor signaling pathway
GO:0007165	1.17E-05	signal transduction
GO:0048477	3.51E-05	oogenesis
GO:0008355	4.04E-05	olfactory learning
GO:0007398	4.34E-05	ectoderm development
GO:0007424	4.98E-05	open tracheal system development
GO:0007507	5.28E-05	heart development
GO:0007498	7.16E-05	mesoderm development
GO:0001709	7.16E-05	cell fate determination
GO:0048749	7.16E-05	compound eye development
GO:0002009	7.48E-05	morphogenesis of an epithelium
GO:0007415	8.57E-05	defasciculation of motor neuron axon
GO:0007513	1.10E-04	pericardial cell differentiation
GO:0045570	1.10E-04	regulation of imaginal disc growth
GO:0007552	1.54E-04	metamorphosis
GO:0008104	1.73E-04	protein localization
GO:0007155	2.35E-04	cell adhesion
GO:0007419	2.35E-04	ventral cord development
GO:0008045	2.35E-04	motor axon guidance
GO:0007427	2.35E-04	epithelial cell migration, open tracheal system
GO:0055085	2.42E-04	transmembrane transport
GO:0007623	2.42E-04	circadian rhythm
GO:0007391	2.42E-04	dorsal closure
GO:0008406	2.46E-04	gonad development
GO:0008356	2.62E-04	asymmetric cell division
GO:0048096	2.66E-04	chromatin-mediated maintenance of transcription
GO:0035286	2.66E-04	leg segmentation
GO:0046667	2.66E-04	compound eye retinal cell programmed cell death
GO:0048814	2.66E-04	regulation of dendrite morphogenesis

GO:0045433	2.80E-04	male courtship behavior, veined wing generated song production
GO:0007275	3.42E-04	multicellular organismal development
GO:0008586	3.43E-04	imaginal disc-derived wing vein morphogenesis
GO:0007314	4.38E-04	oocyte anterior/posterior axis specification
GO:0042052	4.38E-04	rhabdomere development
GO:0008285	4.38E-04	negative regulation of cell proliferation
GO:0007413	4.87E-04	axonal fasciculation
GO:0008407	4.87E-04	bristle morphogenesis
GO:0008354	5.13E-04	germ cell migration
GO:0007435	5.78E-04	salivary gland morphogenesis
GO:0006816	6.54E-04	calcium ion transport
GO:0046620	6.77E-04	regulation of organ growth
GO:0006811	6.97E-04	ion transport
GO:0001751	7.25E-04	compound eye photoreceptor cell differentiation
GO:0042127	7.25E-04	regulation of cell proliferation
GO:0007444	7.78E-04	imaginal disc development
GO:0048100	8.18E-04	wing disc anterior/posterior pattern formation
GO:0007617	8.18E-04	mating behavior
GO:0035225	8.18E-04	determination of genital disc primordium
GO:0015012	8.18E-04	heparan sulfate proteoglycan biosynthetic process
GO:0007509	8.18E-04	mesoderm migration
GO:0051017	8.81E-04	actin filament bundle formation
GO:0030206	9.69E-04	chondroitin sulfate biosynthetic process
GO:0010002	9.69E-04	cardioblast differentiation
GO:0009408	9.69E-04	response to heat
GO:0045449	9.76E-04	regulation of transcription
GO:0007274	1.01E-03	neuromuscular synaptic transmission
GO:0007619	1.08E-03	courtship behavior
GO:0045475	1.08E-03	locomotor rhythm
GO:0001738	1.23E-03	morphogenesis of a polarized epithelium
GO:0008345	1.23E-03	larval locomotory behavior
GO:0007298	1.23E-03	border follicle cell migration
GO:0007280	1.50E-03	pole cell migration
GO:0008347	1.50E-03	glial cell migration
GO:0008344	1.50E-03	adult locomotory behavior
GO:0035147	1.50E-03	branch fusion, open tracheal system
GO:0007480	1.64E-03	imaginal disc-derived leg morphogenesis
GO:0007219	1.77E-03	Notch signaling pathway
GO:0045187	1.77E-03	regulation of circadian sleep/wake cycle, sleep
GO:0001745	2.13E-03	compound eye morphogenesis
GO:0007428	2.37E-03	primary branching, open tracheal system
GO:0007268	2.43E-03	synaptic transmission
GO:0007390	2.45E-03	germ-band shortening
GO:0016331	2.45E-03	morphogenesis of embryonic epithelium

GO:0009649	2.45E-03	entrainment of circadian clock
GO:0048149	2.45E-03	behavioral response to ethanol
GO:0016332	2.45E-03	establishment or maintenance of polarity of embryonic epithelium
GO:0016360	2.45E-03	sensory organ precursor cell fate determination
GO:0002121	2.45E-03	inter-male aggressive behavior
GO:0030162	2.48E-03	regulation of proteolysis
GO:0007422	2.50E-03	peripheral nervous system development
GO:0048675	3.39E-03	axon extension
GO:0007297	3.39E-03	ovarian follicle cell migration
GO:0007622	3.39E-03	rhythmic behavior
GO:0007479	3.39E-03	leg disc proximal/distal pattern formation
GO:0008340	3.46E-03	determination of adult lifespan
GO:0007010	3.46E-03	cytoskeleton organization
GO:0007474	3.50E-03	imaginal disc-derived wing vein specification
GO:0008587	3.72E-03	imaginal disc-derived wing margin morphogenesis
GO:0007520	3.72E-03	myoblast fusion
GO:0016339	3.72E-03	calcium-dependent cell-cell adhesion
GO:0007242	3.86E-03	intracellular signaling cascade
GO:0048666	4.48E-03	neuron development
GO:0006357	4.91E-03	regulation of transcription from RNA polymerase II promoter
GO:0035172	5.04E-03	hemocyte proliferation
GO:0007549	5.04E-03	dosage compensation
GO:0008105	5.75E-03	asymmetric protein localization
GO:0007475	5.75E-03	apposition of dorsal and ventral imaginal disc-derived wing surfaces
GO:0007399	5.75E-03	nervous system development
GO:0000902	5.75E-03	cell morphogenesis
GO:0048066	5.75E-03	pigmentation during development
GO:0007525	5.75E-03	somatic muscle development
GO:0045197	5.75E-03	establishment or maintenance of epithelial cell apical/basal polarity
GO:0042048	5.87E-03	olfactory behavior
GO:0006914	5.87E-03	autophagy
GO:0035071	6.05E-03	salivary gland cell autophagic cell death
GO:0007269	6.60E-03	neurotransmitter secretion
GO:0006813	6.78E-03	potassium ion transport
GO:0008293	7.58E-03	torso signaling pathway
GO:0048813	7.58E-03	dendrite morphogenesis
GO:0007186	7.93E-03	G-protein coupled receptor protein signaling pathway
GO:0016055	8.13E-03	Wnt receptor signaling pathway
GO:0007455	8.84E-03	eye-antennal disc morphogenesis
GO:0045167	8.84E-03	asymmetric protein localization involved in cell fate determination
GO:0007632	8.84E-03	visual behavior
GO:0035215	8.84E-03	genital disc development
GO:0010552	8.84E-03	positive regulation of gene-specific transcription from RNA polymerase II promoter

GO:0045610	8.84E-03	regulation of hemocyte differentiation
GO:0007362	9.17E-03	terminal region determination
GO:0007494	1.20E-02	midgut development
GO:0035023	1.20E-02	regulation of Rho protein signal transduction
GO:0016319	1.21E-02	mushroom body development
GO:0002168	1.31E-02	instar larval development
GO:0008049	1.31E-02	male courtship behavior
GO:0051225	1.31E-02	spindle assembly
GO:0007431	1.31E-02	salivary gland development
GO:0042078	1.31E-02	germ-line stem cell division
GO:0006812	1.38E-02	cation transport
GO:0006897	1.38E-02	endocytosis
GO:0030718	1.38E-02	germ-line stem cell maintenance
GO:0007400	1.38E-02	neuroblast fate determination
GO:0048542	1.38E-02	lymph gland development
GO:0007446	1.38E-02	imaginal disc growth
GO:0007306	1.38E-02	eggshell chorion assembly
GO:0045186	1.40E-02	zonula adherens assembly
GO:0030536	1.40E-02	larval feeding behavior
GO:0048800	1.40E-02	antennal morphogenesis
GO:0007519	1.40E-02	skeletal muscle tissue development
GO:0016079	1.40E-02	synaptic vesicle exocytosis
GO:0007173	1.47E-02	epidermal growth factor receptor signaling pathway
GO:0007291	1.52E-02	sperm individualization
GO:0007602	1.59E-02	phototransduction
GO:0008360	1.70E-02	regulation of cell shape
GO:0007179	1.74E-02	transforming growth factor beta receptor signaling pathway
GO:0007426	1.74E-02	tracheal outgrowth, open tracheal system
GO:0016477	1.74E-02	cell migration
GO:0051726	1.89E-02	regulation of cell cycle
GO:0008103	1.96E-02	oocyte microtubule cytoskeleton polarization
GO:0035277	1.96E-02	spiracle morphogenesis, open tracheal system
GO:0008286	1.96E-02	insulin receptor signaling pathway
GO:0007224	2.01E-02	smoothened signaling pathway
GO:0016318	2.01E-02	ommatidial rotation
GO:0030707	2.02E-02	ovarian follicle cell development
GO:0007517	2.04E-02	muscle organ development
GO:0001700	2.04E-02	embryonic development via the syncytial blastoderm
GO:0048102	2.04E-02	autophagic cell death
GO:0007315	2.08E-02	pole plasm assembly
GO:0007613	2.08E-02	memory
GO:0007293	2.08E-02	germarium-derived egg chamber formation
GO:0030307	2.08E-02	positive regulation of cell growth
GO:0042067	2.30E-02	establishment of ommatidial polarity

GO:0007303	2.72E-02	cytoplasmic transport, nurse cell to oocyte
GO:0019991	2.72E-02	septate junction assembly
GO:0007015	2.98E-02	actin filament organization
GO:0007528	3.04E-02	neuromuscular junction development
GO:0007304	3.04E-02	chorion-containing eggshell formation
GO:0035099	3.04E-02	hemocyte migration
GO:0045941	3.04E-02	positive regulation of transcription
GO:0006325	3.04E-02	chromatin organization
GO:0007367	3.04E-02	segment polarity determination
GO:0007309	3.04E-02	oocyte axis specification
GO:0051056	3.04E-02	regulation of small GTPase mediated signal transduction
GO:0006366	3.28E-02	transcription from RNA polymerase II promoter
GO:0045893	3.28E-02	positive regulation of transcription, DNA-dependent
GO:0009790	3.42E-02	embryonic development
GO:0007409	3.50E-02	axonogenesis
GO:0030097	3.53E-02	hemopoiesis
GO:0017148	3.53E-02	negative regulation of translation
GO:0007286	3.53E-02	spermatid development
GO:0007254	3.53E-02	JNK cascade
GO:0007310	3.53E-02	oocyte dorsal/ventral axis specification
GO:0055059	3.90E-02	asymmetric neuroblast division
GO:0000381	3.99E-02	regulation of alternative nuclear mRNA splicing, via spliceosome
GO:0046843	4.17E-02	dorsal appendage formation
GO:0040014	4.18E-02	regulation of multicellular organism growth
GO:0006378	4.18E-02	mRNA polyadenylation
GO:0007417	4.31E-02	central nervous system development
GO:0006470	4.32E-02	protein amino acid dephosphorylation
GO:0007350	4.68E-02	blastoderm segmentation

Table S4 GO term enrichment for molecular function in the Australian cline for genes that overlap windows in the top 1% tail of Fst.

	1 % Tail Molecular Function	
GO term	q-value	description
GO:0005515	1.79E-06	protein binding
GO:0000166	9.09E-03	nucleotide binding
GO:0003700	1.17E-02	transcription factor activity
GO:0004674	1.17E-02	protein serine/threonine kinase activity
GO:0004672	1.17E-02	protein kinase activity
GO:0004871	1.26E-02	signal transducer activity
GO:0043565	1.50E-02	sequence-specific DNA binding
GO:0004930	1.81E-02	G-protein coupled receptor activity
GO:0005509	3.32E-02	calcium ion binding
GO:0003779	3.37E-02	actin binding
GO:0003729	3.60E-02	mRNA binding
GO:0042623	3.60E-02	ATPase activity, coupled
GO:0005524	4.08E-02	ATP binding

Table S5 GO term enrichment for molecular function in the Australian cline for genes that overlap windows in the top 2.5% tail of Fst.

	2.5 % Tail Molecular Function	
GO term	q-value	description
GO:0005515	5.44E-10	protein binding
GO:0003700	3.20E-09	transcription factor activity
GO:0004672	9.65E-06	protein kinase activity
GO:0004674	3.88E-05	protein serine/threonine kinase activity
GO:0043565	3.88E-05	sequence-specific DNA binding
GO:0005509	9.56E-05	calcium ion binding
GO:0008270	2.82E-04	zinc ion binding
GO:0004871	6.54E-04	signal transducer activity
GO:0004702	6.54E-04	receptor signaling protein serine/threonine kinase activity
GO:0004725	1.60E-03	protein tyrosine phosphatase activity
GO:0005524	1.60E-03	ATP binding
GO:0003702	1.98E-03	RNA polymerase II transcription factor activity
GO:0005089	2.78E-03	Rho guanyl-nucleotide exchange factor activity
GO:0004714	3.11E-03	transmembrane receptor protein tyrosine kinase activity
GO:0003676	4.02E-03	nucleic acid binding
GO:0008188	4.02E-03	neuropeptide receptor activity
GO:0005096	4.02E-03	GTPase activator activity
GO:0005102	4.13E-03	receptor binding
GO:0008026	5.10E-03	ATP-dependent helicase activity
GO:0003713	5.26E-03	transcription coactivator activity
GO:0005249	5.58E-03	voltage-gated potassium channel activity
GO:0004713	8.82E-03	protein tyrosine kinase activity
GO:0004930	8.82E-03	G-protein coupled receptor activity
GO:0003704	8.82E-03	specific RNA polymerase II transcription factor activity
GO:0016563	1.37E-02	transcription activator activity
GO:0004872	1.37E-02	receptor activity
GO:0003779	1.37E-02	actin binding
GO:0000166	1.39E-02	nucleotide binding
GO:0016566	1.42E-02	specific transcriptional repressor activity
GO:0008236	1.47E-02	serine-type peptidase activity
GO:0004222	1.60E-02	metalloendopeptidase activity
GO:0042803	1.68E-02	protein homodimerization activity
GO:0008083	1.81E-02	growth factor activity
GO:0019992	2.47E-02	diacylglycerol binding
GO:0004888	2.47E-02	transmembrane receptor activity
GO:0005516	3.85E-02	calmodulin binding
GO:0008017	4.56E-02	microtubule binding
GO:0046983	4.56E-02	protein dimerization activity
GO:0003729	4.85E-02	mRNA binding

Table S6 GO term enrichment for molecular function in the Australian cline for genes that overlap windows in the top 5% tail of Fst.

	5 % Tail Molecular Function	
GO term	q-value	description
GO:0003700	3.39E-14	transcription factor activity
GO:0005515	1.99E-13	protein binding
GO:0005524	8.48E-09	ATP binding
GO:0043565	3.00E-07	sequence-specific DNA binding
GO:0005509	3.00E-07	calcium ion binding
GO:0004672	1.97E-06	protein kinase activity
GO:0004674	2.91E-06	protein serine/threonine kinase activity
GO:0008188	7.43E-05	neuropeptide receptor activity
GO:0008270	1.12E-04	zinc ion binding
GO:0003702	2.61E-04	RNA polymerase II transcription factor activity
GO:0003779	6.72E-04	actin binding
GO:0004871	9.13E-04	signal transducer activity
GO:0005245	1.64E-03	voltage-gated calcium channel activity
GO:0005096	1.64E-03	GTPase activator activity
GO:0004702	1.67E-03	receptor signaling protein serine/threonine kinase activity
GO:0004725	2.56E-03	protein tyrosine phosphatase activity
GO:0004889	3.99E-03	nicotinic acetylcholine-activated cation-selective channel activity
GO:0032027	3.99E-03	myosin light chain binding
GO:0003774	4.34E-03	motor activity
GO:0003707	5.47E-03	steroid hormone receptor activity
GO:0005249	6.08E-03	voltage-gated potassium channel activity
GO:0004386	9.04E-03	helicase activity
GO:0004190	9.04E-03	aspartic-type endopeptidase activity
GO:0003777	9.04E-03	microtubule motor activity
GO:0008026	9.04E-03	ATP-dependent helicase activity
GO:0005102	1.01E-02	receptor binding
GO:0004222	1.01E-02	metalloendopeptidase activity
GO:0003676	1.16E-02	nucleic acid binding
GO:0004872	1.25E-02	receptor activity
GO:0003704	1.25E-02	specific RNA polymerase II transcription factor activity
GO:0003713	1.47E-02	transcription coactivator activity
GO:0005089	1.57E-02	Rho guanyl-nucleotide exchange factor activity
GO:0004888	1.57E-02	transmembrane receptor activity
GO:0019992	1.57E-02	diacylglycerol binding
GO:0000166	1.79E-02	nucleotide binding
GO:0008239	1.82E-02	dipeptidyl-peptidase activity
GO:0004879	1.82E-02	ligand-dependent nuclear receptor activity
GO:0042626	1.82E-02	ATPase activity, coupled to transmembrane movement of substances
GO:0003746	1.82E-02	translation elongation factor activity
GO:0016566	1.87E-02	specific transcriptional repressor activity

GO:0004930	2.17E-02	G-protein coupled receptor activity
GO:0005085	2.17E-02	guanyl-nucleotide exchange factor activity
GO:0017111	2.55E-02	nucleoside-triphosphatase activity
GO:0004714	2.55E-02	transmembrane receptor protein tyrosine kinase activity
GO:0005516	2.55E-02	calmodulin binding
GO:0003677	2.55E-02	DNA binding
GO:0005215	2.55E-02	transporter activity
GO:0008017	2.97E-02	microtubule binding
GO:0042623	3.27E-02	ATPase activity, coupled
GO:0016563	3.55E-02	transcription activator activity
GO:0050839	3.62E-02	cell adhesion molecule binding
GO:0035091	3.62E-02	phosphoinositide binding
GO:0008083	3.86E-02	growth factor activity
GO:0004713	4.10E-02	protein tyrosine kinase activity
GO:0003714	4.16E-02	transcription corepressor activity
GO:0032183	4.36E-02	SUMO binding

Table S7 GO term enrichment for cellular component in the Australian cline for genes that overlap windows in the top 1% tail of Fst.

	1 % Tail Cellular Component	
GO term	q-value	description
GO:0005886	1.97E-06	plasma membrane
GO:0045211	2.47E-04	postsynaptic membrane
GO:0042600	6.03E-04	chorion
GO:0005887	6.03E-04	integral to plasma membrane
GO:0005578	2.51E-03	proteinaceous extracellular matrix
GO:0043234	2.51E-03	protein complex
GO:0016020	2.51E-03	membrane
GO:0005622	1.45E-02	intracellular
GO:0005634	2.82E-02	nucleus
GO:0016021	3.70E-02	integral to membrane
GO:0005875	3.70E-02	microtubule associated complex

Table S8 GO term enrichment for cellular component in the Australian cline for genes that overlap windows in the top 2.5% tail of Fst.

2.5 % Tail Cellular Component		
GO term	q-value	description
GO:0005886	2.11E-13	plasma membrane
GO:0005634	1.38E-06	nucleus
GO:0005887	5.67E-05	integral to plasma membrane
GO:0005918	1.44E-04	septate junction
GO:0045211	9.38E-04	postsynaptic membrane
GO:0005912	1.42E-03	adherens junction
GO:0035003	1.42E-03	subapical complex
GO:0016327	1.42E-03	apicolateral plasma membrane
GO:0016324	1.42E-03	apical plasma membrane
GO:0045202	2.26E-03	synapse
GO:0016021	2.37E-03	integral to membrane
GO:0008076	3.27E-03	voltage-gated potassium channel complex
GO:0016459	5.74E-03	myosin complex
GO:0008021	6.02E-03	synaptic vesicle
GO:0045179	8.90E-03	apical cortex
GO:0043234	8.90E-03	protein complex
GO:0005667	1.15E-02	transcription factor complex
GO:0042600	1.48E-02	chorion
GO:0005622	1.68E-02	intracellular
GO:0016020	1.76E-02	membrane
GO:0005578	1.92E-02	proteinaceous extracellular matrix
GO:0005739	2.91E-02	mitochondrion
GO:0005875	3.68E-02	microtubule associated complex
GO:0005938	4.85E-02	cell cortex

Table S9 GO term enrichment for cellular component in the Australian cline for genes that overlap windows in the top 5% tail of Fst.

5 % Tail Cellular Component		
GO term	q-value	description
GO:0005886	1.98E-14	plasma membrane
GO:0005634	4.17E-11	nucleus
GO:0005887	2.97E-08	integral to plasma membrane
GO:0016021	2.72E-06	integral to membrane
GO:0016324	2.27E-04	apical plasma membrane
GO:0005913	3.46E-04	cell-cell adherens junction
GO:0045211	6.54E-04	postsynaptic membrane
GO:0035003	8.56E-04	subapical complex
GO:0005912	8.62E-04	adherens junction
GO:0005918	1.31E-03	septate junction
GO:0005737	1.76E-03	cytoplasm
GO:0045179	2.09E-03	apical cortex
GO:0005892	2.41E-03	nicotinic acetylcholine-gated receptor-channel complex
GO:0043195	2.41E-03	terminal button
GO:0016459	2.41E-03	myosin complex
GO:0005622	2.41E-03	intracellular
GO:0045202	2.41E-03	synapse
GO:0016323	2.41E-03	basolateral plasma membrane
GO:0008076	2.41E-03	voltage-gated potassium channel complex
GO:0016327	3.14E-03	apicolateral plasma membrane
GO:0005875	3.83E-03	microtubule associated complex
GO:0016020	4.18E-03	membrane
GO:0009986	4.52E-03	cell surface
GO:0008021	5.40E-03	synaptic vesicle
GO:0005578	6.13E-03	proteinaceous extracellular matrix
GO:0005604	8.43E-03	basement membrane
GO:0005938	8.74E-03	cell cortex
GO:0043190	1.26E-02	ATP-binding cassette (ABC) transporter complex
GO:0043025	1.36E-02	cell soma
GO:0005705	1.36E-02	polytene chromosome interband
GO:0030425	3.37E-02	dendrite
GO:0015629	3.37E-02	actin cytoskeleton
GO:0005667	3.98E-02	transcription factor complex

Table S10 GO term enrichment for biological process in the North American cline for genes that overlap windows in the top 1% tail of Fst.

	1 % Tail Biological Process	
GO term	q-value	description
GO:0007411	8.92E-06	axon guidance
GO:0006355	8.92E-06	regulation of transcription, DNA-dependent
GO:0007455	1.28E-04	eye-antennal disc morphogenesis
GO:0048749	1.30E-04	compound eye development
GO:0007165	1.54E-04	signal transduction
GO:0007155	2.44E-04	cell adhesion
GO:0007476	2.44E-04	imaginal disc-derived wing morphogenesis
GO:0007424	2.78E-04	open tracheal system development
GO:0007611	3.13E-04	learning or memory
GO:0030707	4.18E-04	ovarian follicle cell development
GO:0035317	4.18E-04	imaginal disc-derived wing hair organization
GO:0007530	4.18E-04	sex determination
GO:0007156	4.60E-04	homophilic cell adhesion
GO:0007422	5.26E-04	peripheral nervous system development
GO:0007350	7.08E-04	blastoderm segmentation
GO:0006811	1.16E-03	ion transport
GO:0006468	1.25E-03	protein amino acid phosphorylation
GO:0008355	1.77E-03	olfactory learning
GO:0006350	1.80E-03	transcription
GO:0016310	1.80E-03	phosphorylation
GO:0008104	1.82E-03	protein localization
GO:0006813	2.37E-03	potassium ion transport
GO:0007391	2.37E-03	dorsal closure
GO:0008407	2.75E-03	bristle morphogenesis
GO:0007623	2.85E-03	circadian rhythm
GO:0007420	2.85E-03	brain development
GO:0007186	3.70E-03	G-protein coupled receptor protein signaling pathway
GO:0007399	3.70E-03	nervous system development
GO:0045449	4.03E-03	regulation of transcription
GO:0048477	4.67E-03	oogenesis
GO:0048813	6.08E-03	dendrite morphogenesis
GO:0007242	8.01E-03	intracellular signaling cascade
GO:0008360	8.37E-03	regulation of cell shape
GO:0002121	8.37E-03	inter-male aggressive behavior
GO:0007417	9.30E-03	central nervous system development
GO:0007507	1.13E-02	heart development
GO:0007017	1.75E-02	microtubule-based process
GO:0006470	2.98E-02	protein amino acid dephosphorylation

Table S11 GO term enrichment for biological process in the North American cline for genes that overlap windows in the top 2.5% tail of Fst.

	2.5 % Tail Biological Process	
GO term	q-value	description
GO:0007476	1.63E-07	imaginal disc-derived wing morphogenesis
GO:0048749	1.63E-07	compound eye development
GO:0006355	1.63E-07	regulation of transcription, DNA-dependent
GO:0007411	3.96E-07	axon guidance
GO:0007350	1.62E-06	blastoderm segmentation
GO:0007155	4.43E-06	cell adhesion
GO:0007513	5.83E-06	pericardial cell differentiation
GO:0045449	8.22E-06	regulation of transcription
GO:0016203	1.10E-05	muscle attachment
GO:0007435	1.10E-05	salivary gland morphogenesis
GO:0007424	1.55E-05	open tracheal system development
GO:0006468	2.18E-05	protein amino acid phosphorylation
GO:0007156	2.62E-05	homophilic cell adhesion
GO:0007611	3.07E-05	learning or memory
GO:0035023	3.11E-05	regulation of Rho protein signal transduction
GO:0048477	8.61E-05	oogenesis
GO:0008355	8.63E-05	olfactory learning
GO:0007472	9.07E-05	wing disc morphogenesis
GO:0007399	9.14E-05	nervous system development
GO:0007242	1.17E-04	intracellular signaling cascade
GO:0007614	1.17E-04	short-term memory
GO:0007619	1.86E-04	courtship behavior
GO:0007165	2.23E-04	signal transduction
GO:0030097	2.46E-04	hemopoiesis
GO:0007268	2.87E-04	synaptic transmission
GO:0048800	3.69E-04	antennal morphogenesis
GO:0016318	4.06E-04	ommatidial rotation
GO:0008360	4.06E-04	regulation of cell shape
GO:0007417	4.06E-04	central nervous system development
GO:0016055	4.70E-04	Wnt receptor signaling pathway
GO:0007507	4.82E-04	heart development
GO:0035277	4.82E-04	spiracle morphogenesis, open tracheal system
GO:0007474	5.54E-04	imaginal disc-derived wing vein specification
GO:0007186	5.87E-04	G-protein coupled receptor protein signaling pathway
GO:0007617	6.47E-04	mating behavior
GO:0035172	6.47E-04	hemocyte proliferation
GO:0035225	6.47E-04	determination of genital disc primordium
GO:0007455	7.31E-04	eye-antennal disc morphogenesis
GO:0008407	7.66E-04	bristle morphogenesis
GO:0007428	1.01E-03	primary branching, open tracheal system

GO:0007420	1.03E-03	brain development
GO:0006811	1.05E-03	ion transport
GO:0008104	1.05E-03	protein localization
GO:0008356	1.07E-03	asymmetric cell division
GO:0007391	1.12E-03	dorsal closure
GO:0045433	1.15E-03	male courtship behavior, veined wing generated song production
GO:0007632	1.15E-03	visual behavior
GO:0001745	1.22E-03	compound eye morphogenesis
GO:0008586	1.25E-03	imaginal disc-derived wing vein morphogenesis
GO:0007400	1.43E-03	neuroblast fate determination
GO:0040014	1.43E-03	regulation of multicellular organism growth
GO:0016477	1.47E-03	cell migration
GO:0007422	1.57E-03	peripheral nervous system development
GO:0006813	1.80E-03	potassium ion transport
GO:0030707	1.83E-03	ovarian follicle cell development
GO:0045475	1.85E-03	locomotor rhythm
GO:0040018	1.85E-03	positive regulation of multicellular organism growth
GO:0048675	1.85E-03	axon extension
GO:0007173	1.99E-03	epidermal growth factor receptor signaling pathway
GO:0045165	1.99E-03	cell fate commitment
GO:0008286	1.99E-03	insulin receptor signaling pathway
GO:0030162	1.99E-03	regulation of proteolysis
GO:0007498	2.01E-03	mesoderm development
GO:0007017	2.07E-03	microtubule-based process
GO:0008105	2.68E-03	asymmetric protein localization
GO:0007442	2.68E-03	hindgut morphogenesis
GO:0007475	2.68E-03	apposition of dorsal and ventral imaginal disc-derived wing surfaces
GO:0006916	2.68E-03	anti-apoptosis
GO:0007443	2.68E-03	Malpighian tubule morphogenesis
GO:0030307	2.68E-03	positive regulation of cell growth
GO:0035317	2.68E-03	imaginal disc-derived wing hair organization
GO:0048149	2.68E-03	behavioral response to ethanol
GO:0006816	2.68E-03	calcium ion transport
GO:0008345	2.68E-03	larval locomotory behavior
GO:0007623	2.82E-03	circadian rhythm
GO:0001737	3.83E-03	establishment of imaginal disc-derived wing hair orientation
GO:0002168	3.83E-03	instar larval development
GO:0016199	3.83E-03	axon midline choice point recognition
GO:0030713	3.83E-03	ovarian follicle cell stalk formation
GO:0007369	3.83E-03	gastrulation
GO:0008049	3.83E-03	male courtship behavior
GO:0007431	3.83E-03	salivary gland development
GO:0042051	3.83E-03	compound eye photoreceptor development
GO:0007629	3.83E-03	flight behavior

GO:0007528	3.83E-03	neuromuscular junction development
GO:0008045	4.17E-03	motor axon guidance
GO:0007419	4.17E-03	ventral cord development
GO:0007427	4.17E-03	epithelial cell migration, open tracheal system
GO:0007269	5.13E-03	neurotransmitter secretion
GO:0007446	5.54E-03	imaginal disc growth
GO:0051017	5.54E-03	actin filament bundle formation
GO:0007416	5.54E-03	synapse assembly
GO:0001736	5.54E-03	establishment of planar polarity
GO:0007298	6.19E-03	border follicle cell migration
GO:0007140	6.53E-03	male meiosis
GO:0042067	6.53E-03	establishment of ommatidial polarity
GO:0030178	6.53E-03	negative regulation of Wnt receptor signaling pathway
GO:0007398	6.53E-03	ectoderm development
GO:0045893	6.86E-03	positive regulation of transcription, DNA-dependent
GO:0007349	6.86E-03	cellularization
GO:0001709	7.60E-03	cell fate determination
GO:0008344	7.60E-03	adult locomotory behavior
GO:0048190	7.89E-03	wing disc dorsal/ventral pattern formation
GO:0007494	8.46E-03	midgut development
GO:0007520	8.46E-03	myoblast fusion
GO:0001700	8.54E-03	embryonic development via the syncytial blastoderm
GO:0007530	9.93E-03	sex determination
GO:0001708	9.93E-03	cell fate specification
GO:0002121	9.93E-03	inter-male aggressive behavior
GO:0007390	9.93E-03	germ-band shortening
GO:0051276	9.93E-03	chromosome organization
GO:0000226	9.93E-03	microtubule cytoskeleton organization
GO:0019991	9.93E-03	septate junction assembly
GO:0055059	1.06E-02	asymmetric neuroblast division
GO:0007015	1.16E-02	actin filament organization
GO:0007517	1.19E-02	muscle organ development
GO:0007602	1.19E-02	phototransduction
GO:0045944	1.28E-02	positive regulation of transcription from RNA polymerase II promoter
GO:0007224	1.31E-02	smoothened signaling pathway
GO:0007317	1.31E-02	regulation of pole plasm oskar mRNA localization
GO:0016310	1.31E-02	phosphorylation
GO:0007274	1.31E-02	neuromuscular synaptic transmission
GO:0007163	1.68E-02	establishment or maintenance of cell polarity
GO:0007480	1.72E-02	imaginal disc-derived leg morphogenesis
GO:0006915	2.03E-02	apoptosis
GO:0030036	2.03E-02	actin cytoskeleton organization
GO:0006357	2.05E-02	regulation of transcription from RNA polymerase II promoter
GO:0030718	2.05E-02	germ-line stem cell maintenance

GO:0008595	2.05E-02	determination of anterior/posterior axis, embryo
GO:0001751	2.13E-02	compound eye photoreceptor cell differentiation
GO:0007346	2.13E-02	regulation of mitotic cell cycle
GO:0040007	2.13E-02	growth
GO:0006350	2.20E-02	transcription
GO:0006470	2.33E-02	protein amino acid dephosphorylation
GO:0007444	2.44E-02	imaginal disc development
GO:0008340	2.44E-02	determination of adult lifespan
GO:0002009	2.66E-02	morphogenesis of an epithelium
GO:0048813	3.11E-02	dendrite morphogenesis
GO:0006911	3.11E-02	phagocytosis, engulfment
GO:0006836	3.30E-02	neurotransmitter transport
GO:0008298	4.08E-02	intracellular mRNA localization
GO:0016321	4.24E-02	female meiosis chromosome segregation
GO:0016567	4.99E-02	protein ubiquitination
GO:0046843	4.99E-02	dorsal appendage formation

Table S12 GO term enrichment for biological process in the North American cline for genes that overlap windows in the top 5% tail of Fst.

	5 % Tail Biological Process	
GO term	q-value	description
GO:0006355	1.48E-08	regulation of transcription, DNA-dependent
GO:0007411	2.10E-08	axon guidance
GO:0007476	1.13E-07	imaginal disc-derived wing morphogenesis
GO:0048477	1.13E-07	oogenesis
GO:0035023	4.71E-07	regulation of Rho protein signal transduction
GO:0048749	9.49E-07	compound eye development
GO:0007350	5.78E-06	blastoderm segmentation
GO:0007399	5.78E-06	nervous system development
GO:0007513	1.22E-05	pericardial cell differentiation
GO:0045449	1.99E-05	regulation of transcription
GO:0007156	4.21E-05	homophilic cell adhesion
GO:0006468	4.21E-05	protein amino acid phosphorylation
GO:0007474	4.21E-05	imaginal disc-derived wing vein specification
GO:0007482	4.21E-05	haltere development
GO:0007455	4.38E-05	eye-antennal disc morphogenesis
GO:0007435	4.75E-05	salivary gland morphogenesis
GO:0008586	4.75E-05	imaginal disc-derived wing vein morphogenesis
GO:0007472	5.30E-05	wing disc morphogenesis
GO:0007155	5.88E-05	cell adhesion
GO:0007398	6.34E-05	ectoderm development
GO:0042051	8.75E-05	compound eye photoreceptor development
GO:0007419	8.76E-05	ventral cord development
GO:0007391	9.25E-05	dorsal closure
GO:0001745	9.25E-05	compound eye morphogenesis
GO:0007424	9.42E-05	open tracheal system development
GO:0010002	1.14E-04	cardioblast differentiation
GO:0048800	1.46E-04	antennal morphogenesis
GO:0035225	1.46E-04	determination of genital disc primordium
GO:0016203	1.46E-04	muscle attachment
GO:0007480	1.46E-04	imaginal disc-derived leg morphogenesis
GO:0035172	1.46E-04	hemocyte proliferation
GO:0006813	1.85E-04	potassium ion transport
GO:0007417	1.96E-04	central nervous system development
GO:0051276	2.00E-04	chromosome organization
GO:0007507	2.44E-04	heart development
GO:0007400	2.66E-04	neuroblast fate determination
GO:0007416	2.66E-04	synapse assembly
GO:0046331	3.30E-04	lateral inhibition
GO:0035286	3.30E-04	leg segmentation
GO:0035224	3.30E-04	genital disc anterior/posterior pattern formation

GO:0008356	3.30E-04	asymmetric cell division
GO:0046667	3.30E-04	compound eye retinal cell programmed cell death
GO:0008406	3.30E-04	gonad development
GO:0007425	3.30E-04	epithelial cell fate determination, open tracheal system
GO:0035309	3.30E-04	wing and notum subfield formation
GO:0007628	3.30E-04	adult walking behavior
GO:0007473	3.30E-04	wing disc proximal/distal pattern formation
GO:0030097	3.30E-04	hemopoiesis
GO:0007440	3.53E-04	foregut morphogenesis
GO:0035215	3.53E-04	genital disc development
GO:0007422	3.73E-04	peripheral nervous system development
GO:0035088	3.79E-04	establishment or maintenance of apical/basal cell polarity
GO:0007165	5.86E-04	signal transduction
GO:0016477	6.23E-04	cell migration
GO:0007498	6.67E-04	mesoderm development
GO:0008104	6.73E-04	protein localization
GO:0008407	7.59E-04	bristle morphogenesis
GO:0045893	7.59E-04	positive regulation of transcription, DNA-dependent
GO:0007479	8.97E-04	leg disc proximal/distal pattern formation
GO:0007442	8.97E-04	hindgut morphogenesis
GO:0007298	9.07E-04	border follicle cell migration
GO:0016055	9.07E-04	Wnt receptor signaling pathway
GO:0030162	1.05E-03	regulation of proteolysis
GO:0048100	1.08E-03	wing disc anterior/posterior pattern formation
GO:0042048	1.08E-03	olfactory behavior
GO:0008284	1.08E-03	positive regulation of cell proliferation
GO:0048813	1.08E-03	dendrite morphogenesis
GO:0007268	1.08E-03	synaptic transmission
GO:0007420	1.08E-03	brain development
GO:0008045	1.08E-03	motor axon guidance
GO:0007076	1.20E-03	mitotic chromosome condensation
GO:0007415	1.23E-03	defasciculation of motor neuron axon
GO:0007614	1.23E-03	short-term memory
GO:0008355	1.24E-03	olfactory learning
GO:0048190	1.36E-03	wing disc dorsal/ventral pattern formation
GO:0007186	1.39E-03	G-protein coupled receptor protein signaling pathway
GO:0008049	1.39E-03	male courtship behavior
GO:0007431	1.39E-03	salivary gland development
GO:0007494	1.49E-03	midgut development
GO:0046843	1.49E-03	dorsal appendage formation
GO:0007619	1.49E-03	courtship behavior
GO:0050770	1.61E-03	regulation of axonogenesis
GO:0030307	1.61E-03	positive regulation of cell growth
GO:0008360	1.86E-03	regulation of cell shape

GO:0001709	2.03E-03	cell fate determination
GO:0035277	2.03E-03	spiracle morphogenesis, open tracheal system
GO:0045165	2.03E-03	cell fate commitment
GO:0042052	2.13E-03	rhabdomere development
GO:0048854	2.13E-03	brain morphogenesis
GO:0001964	2.13E-03	startle response
GO:0045494	2.13E-03	photoreceptor cell maintenance
GO:0045433	2.13E-03	male courtship behavior, veined wing generated song production
GO:0009952	2.13E-03	anterior/posterior pattern formation
GO:0010552	2.13E-03	positive regulation of gene-specific transcription from RNA polymerase II promoter
GO:0045167	2.13E-03	asymmetric protein localization involved in cell fate determination
GO:0031987	2.13E-03	locomotion involved in locomotory behavior
GO:0008354	2.14E-03	germ cell migration
GO:0055059	2.14E-03	asymmetric neuroblast division
GO:0030707	2.20E-03	ovarian follicle cell development
GO:0030713	2.86E-03	ovarian follicle cell stalk formation
GO:0016199	2.86E-03	axon midline choice point recognition
GO:0040001	2.88E-03	establishment of mitotic spindle localization
GO:0007051	2.88E-03	spindle organization
GO:0016331	2.88E-03	morphogenesis of embryonic epithelium
GO:0007242	2.97E-03	intracellular signaling cascade
GO:0006816	3.07E-03	calcium ion transport
GO:0048149	3.07E-03	behavioral response to ethanol
GO:0006916	3.07E-03	anti-apoptosis
GO:0007443	3.07E-03	Malpighian tubule morphogenesis
GO:0007140	3.17E-03	male meiosis
GO:0007346	3.17E-03	regulation of mitotic cell cycle
GO:0007611	3.17E-03	learning or memory
GO:0007517	3.22E-03	muscle organ development
GO:0002121	3.26E-03	inter-male aggressive behavior
GO:0040018	3.94E-03	positive regulation of multicellular organism growth
GO:0030261	3.94E-03	chromosome condensation
GO:0046620	3.94E-03	regulation of organ growth
GO:0048675	3.94E-03	axon extension
GO:0007297	3.94E-03	ovarian follicle cell migration
GO:0008587	4.51E-03	imaginal disc-derived wing margin morphogenesis
GO:0035220	4.51E-03	wing disc development
GO:0040014	4.51E-03	regulation of multicellular organism growth
GO:0048542	4.51E-03	lymph gland development
GO:0007446	4.51E-03	imaginal disc growth
GO:0002009	4.65E-03	morphogenesis of an epithelium
GO:0006811	5.37E-03	ion transport
GO:0006979	5.47E-03	response to oxidative stress

GO:0008361	5.61E-03	regulation of cell size
GO:0045570	5.61E-03	regulation of imaginal disc growth
GO:0045466	5.61E-03	R7 cell differentiation
GO:0007447	5.61E-03	imaginal disc pattern formation
GO:0046329	5.61E-03	negative regulation of JNK cascade
GO:0007301	5.61E-03	ovarian ring canal formation
GO:0007617	5.61E-03	mating behavior
GO:0001700	5.93E-03	embryonic development via the syncytial blastoderm
GO:0008105	6.60E-03	asymmetric protein localization
GO:0007304	6.60E-03	chorion-containing eggshell formation
GO:0045197	6.60E-03	establishment or maintenance of epithelial cell apical/basal polarity
GO:0008345	6.60E-03	larval locomotory behavior
GO:0006836	6.60E-03	neurotransmitter transport
GO:0007379	6.60E-03	segment specification
GO:0009880	6.60E-03	embryonic pattern specification
GO:0007017	6.61E-03	microtubule-based process
GO:0007280	6.88E-03	pole cell migration
GO:0007548	6.88E-03	sex differentiation
GO:0007173	7.43E-03	epidermal growth factor receptor signaling pathway
GO:0030036	7.58E-03	actin cytoskeleton organization
GO:0007444	8.37E-03	imaginal disc development
GO:0016318	9.32E-03	ommatidial rotation
GO:0007427	9.32E-03	epithelial cell migration, open tracheal system
GO:0007157	1.00E-02	heterophilic cell adhesion
GO:0001751	1.00E-02	compound eye photoreceptor cell differentiation
GO:0007349	1.00E-02	cellularization
GO:0045610	1.00E-02	regulation of hemocyte differentiation
GO:0007632	1.00E-02	visual behavior
GO:0022416	1.00E-02	bristle development
GO:0030178	1.00E-02	negative regulation of Wnt receptor signaling pathway
GO:0016458	1.00E-02	gene silencing
GO:0016360	1.00E-02	sensory organ precursor cell fate determination
GO:0007552	1.00E-02	metamorphosis
GO:0007528	1.00E-02	neuromuscular junction development
GO:0007428	1.00E-02	primary branching, open tracheal system
GO:0019991	1.00E-02	septate junction assembly
GO:0007390	1.00E-02	germ-band shortening
GO:0007369	1.00E-02	gastrulation
GO:0007307	1.00E-02	eggshell chorion gene amplification
GO:0001708	1.00E-02	cell fate specification
GO:0000226	1.00E-02	microtubule cytoskeleton organization
GO:0035050	1.00E-02	embryonic heart tube development
GO:0055085	1.07E-02	transmembrane transport
GO:0007423	1.12E-02	sensory organ development

GO:0007163	1.19E-02	establishment or maintenance of cell polarity
GO:0006357	1.26E-02	regulation of transcription from RNA polymerase II promoter
GO:0006325	1.34E-02	chromatin organization
GO:0045944	1.41E-02	positive regulation of transcription from RNA polymerase II promoter
GO:0048666	1.41E-02	neuron development
GO:0002168	1.46E-02	instar larval development
GO:0016319	1.46E-02	mushroom body development
GO:0006955	1.46E-02	immune response
GO:0018991	1.54E-02	oviposition
GO:0007067	1.54E-02	mitosis
GO:0030718	1.54E-02	germ-line stem cell maintenance
GO:0008595	1.54E-02	determination of anterior/posterior axis, embryo
GO:0051533	1.54E-02	positive regulation of NFAT protein import into nucleus
GO:0007259	1.54E-02	JAK-STAT cascade
GO:0030536	1.54E-02	larval feeding behavior
GO:0007519	1.54E-02	skeletal muscle tissue development
GO:0007501	1.54E-02	mesodermal cell fate specification
GO:0042059	1.54E-02	negative regulation of epidermal growth factor receptor signaling pathway
GO:0060857	1.54E-02	establishment of glial blood-brain barrier
GO:0006915	1.64E-02	apoptosis
GO:0007269	1.64E-02	neurotransmitter secretion
GO:0007015	1.83E-02	actin filament organization
GO:0007623	2.08E-02	circadian rhythm
GO:0007409	2.08E-02	axonogenesis
GO:0035147	2.20E-02	branch fusion, open tracheal system
GO:0008344	2.20E-02	adult locomotory behavior
GO:0006914	2.20E-02	autophagy
GO:0008286	2.20E-02	insulin receptor signaling pathway
GO:0008103	2.20E-02	oocyte microtubule cytoskeleton polarization
GO:0008347	2.20E-02	glial cell migration
GO:0016567	2.23E-02	protein ubiquitination
GO:0051726	2.23E-02	regulation of cell cycle
GO:0009987	2.24E-02	cellular process
GO:0007475	2.29E-02	apposition of dorsal and ventral imaginal disc-derived wing surfaces
GO:0007062	2.29E-02	sister chromatid cohesion
GO:0007525	2.29E-02	somatic muscle development
GO:0001738	2.29E-02	morphogenesis of a polarized epithelium
GO:0007613	2.29E-02	memory
GO:0045746	2.29E-02	negative regulation of Notch signaling pathway
GO:0009312	2.29E-02	oligosaccharide biosynthetic process
GO:0008293	2.29E-02	torso signaling pathway
GO:0007088	2.61E-02	regulation of mitosis
GO:0040007	2.61E-02	growth
GO:0042067	2.61E-02	establishment of ommatidial polarity

GO:0042127	2.61E-02	regulation of cell proliferation
GO:0035317	2.96E-02	imaginal disc-derived wing hair organization
GO:0007530	2.96E-02	sex determination
GO:0006917	2.96E-02	induction of apoptosis
GO:0045475	2.96E-02	locomotor rhythm
GO:0001736	2.96E-02	establishment of planar polarity
GO:0016311	2.96E-02	dephosphorylation
GO:0035071	3.01E-02	salivary gland cell autophagic cell death
GO:0007169	3.32E-02	transmembrane receptor protein tyrosine kinase signaling pathway
GO:0007629	3.32E-02	flight behavior
GO:0001737	3.32E-02	establishment of imaginal disc-derived wing hair orientation
GO:0007362	3.32E-02	terminal region determination
GO:0051056	3.32E-02	regulation of small GTPase mediated signal transduction
GO:0006897	3.80E-02	endocytosis
GO:0006396	4.00E-02	RNA processing
GO:0016321	4.00E-02	female meiosis chromosome segregation
GO:0007274	4.00E-02	neuromuscular synaptic transmission
GO:0007310	4.00E-02	oocyte dorsal/ventral axis specification
GO:0017148	4.00E-02	negative regulation of translation
GO:0051017	4.73E-02	actin filament bundle formation
GO:0006865	4.80E-02	amino acid transport
GO:0016481	4.97E-02	negative regulation of transcription

Table S13 GO term enrichment for molecular function in the North American cline for genes that overlap windows in the top 1% tail of Fst.

	1 % Tail Molecular Function	
GO term	q-value	description
GO:0003700	3.80E-07	transcription factor activity
GO:0043565	2.48E-05	sequence-specific DNA binding
GO:0003704	5.27E-05	specific RNA polymerase II transcription factor activity
GO:0004672	2.96E-04	protein kinase activity
GO:0005515	5.34E-04	protein binding
GO:0000166	5.34E-04	nucleotide binding
GO:0019992	2.01E-03	diacylglycerol binding
GO:0016563	3.32E-03	transcription activator activity
GO:0004674	7.09E-03	protein serine/threonine kinase activity
GO:0003702	7.18E-03	RNA polymerase II transcription factor activity
GO:0004725	7.68E-03	protein tyrosine phosphatase activity
GO:0008188	7.68E-03	neuropeptide receptor activity
GO:0008270	1.08E-02	zinc ion binding
GO:0005509	1.64E-02	calcium ion binding
GO:0042803	1.89E-02	protein homodimerization activity
GO:0003729	1.89E-02	mRNA binding
GO:0004713	2.04E-02	protein tyrosine kinase activity
GO:0003677	2.07E-02	DNA binding
GO:0003779	2.20E-02	actin binding
GO:0003774	2.55E-02	motor activity
GO:0008017	3.42E-02	microtubule binding
GO:0020037	3.96E-02	heme binding
GO:0003723	4.02E-02	RNA binding
GO:0042623	4.75E-02	ATPase activity, coupled
GO:0004930	4.94E-02	G-protein coupled receptor activity

Table S14 GO term enrichment for molecular function in the North American cline for genes that overlap windows in the top 2.5% tail of Fst.

GO term	2.5% Tail Molecular Function q-value	description
GO:0003700	4.16E-10	transcription factor activity
GO:0043565	1.61E-08	sequence-specific DNA binding
GO:0005515	2.23E-05	protein binding
GO:0005089	4.38E-05	Rho guanyl-nucleotide exchange factor activity
GO:0003702	1.09E-04	RNA polymerase II transcription factor activity
GO:0004714	1.09E-04	transmembrane receptor protein tyrosine kinase activity
GO:0004871	1.62E-04	signal transducer activity
GO:0004672	1.87E-04	protein kinase activity
GO:0008270	2.93E-04	zinc ion binding
GO:0003707	3.28E-04	steroid hormone receptor activity
GO:0004879	3.28E-04	ligand-dependent nuclear receptor activity
GO:0008227	4.83E-04	amine receptor activity
GO:0003713	5.28E-04	transcription coactivator activity
GO:0008134	7.22E-04	transcription factor binding
GO:0000166	7.22E-04	nucleotide binding
GO:0003729	1.66E-03	mRNA binding
GO:0004713	1.68E-03	protein tyrosine kinase activity
GO:0003779	1.76E-03	actin binding
GO:0005509	1.89E-03	calcium ion binding
GO:0008188	1.89E-03	neuropeptide receptor activity
GO:0004674	1.90E-03	protein serine/threonine kinase activity
GO:0003704	2.08E-03	specific RNA polymerase II transcription factor activity
GO:0019992	2.68E-03	diacylglycerol binding
GO:0008092	2.68E-03	cytoskeletal protein binding
GO:0004888	2.68E-03	transmembrane receptor activity
GO:0005096	4.22E-03	GTPase activator activity
GO:0003714	4.22E-03	transcription corepressor activity
GO:0004725	5.05E-03	protein tyrosine phosphatase activity
GO:0005524	5.99E-03	ATP binding
GO:0016563	7.45E-03	transcription activator activity
GO:0005102	7.82E-03	receptor binding
GO:0003677	8.76E-03	DNA binding
GO:0042803	9.25E-03	protein homodimerization activity
GO:0005085	1.00E-02	guanyl-nucleotide exchange factor activity
GO:0008017	1.18E-02	microtubule binding
GO:0004930	1.23E-02	G-protein coupled receptor activity
GO:0003774	1.87E-02	motor activity
GO:0003723	1.95E-02	RNA binding
GO:0008083	2.42E-02	growth factor activity
GO:0032183	2.42E-02	SUMO binding

GO:0042623	2.42E-02	ATPase activity, coupled
GO:0003676	4.34E-02	nucleic acid binding
GO:0003777	4.95E-02	microtubule motor activity

Table S15 GO term enrichment for molecular function in the North American cline for genes that overlap windows in the top 5% tail of Fst.

	5% Tail Molecular Function	
GO term	q-value	description
GO:0019904	0.00E+00	protein domain specific binding
GO:0003700	9.43E-13	transcription factor activity
GO:0043565	5.69E-09	sequence-specific DNA binding
GO:0005515	6.05E-09	protein binding
GO:0005089	1.47E-07	Rho guanyl-nucleotide exchange factor activity
GO:0003704	1.19E-06	specific RNA polymerase II transcription factor activity
GO:0005509	6.53E-06	calcium ion binding
GO:0004871	1.82E-05	signal transducer activity
GO:0008270	4.25E-05	zinc ion binding
GO:0003702	5.31E-05	RNA polymerase II transcription factor activity
GO:0016563	6.32E-05	transcription activator activity
GO:0017147	7.70E-05	Wnt-protein binding
GO:0004674	2.20E-04	protein serine/threonine kinase activity
GO:0004714	3.05E-04	transmembrane receptor protein tyrosine kinase activity
GO:0005267	3.09E-04	potassium channel activity
GO:0008188	3.09E-04	neuropeptide receptor activity
GO:0003779	4.30E-04	actin binding
GO:0008017	5.96E-04	microtubule binding
GO:0004879	8.37E-04	ligand-dependent nuclear receptor activity
GO:0003707	8.37E-04	steroid hormone receptor activity
GO:0004672	8.90E-04	protein kinase activity
GO:0005041	1.27E-03	low-density lipoprotein receptor activity
GO:0004888	1.63E-03	transmembrane receptor activity
GO:0003713	1.63E-03	transcription coactivator activity
GO:0004713	1.88E-03	protein tyrosine kinase activity
GO:0008227	2.12E-03	amine receptor activity
GO:0008134	2.31E-03	transcription factor binding
GO:0005326	2.31E-03	neurotransmitter transporter activity
GO:0004889	3.22E-03	nicotinic acetylcholine-activated cation-selective channel activity
GO:0005524	3.50E-03	ATP binding
GO:0032183	4.08E-03	SUMO binding
GO:0003714	5.58E-03	transcription corepressor activity
GO:0005096	5.58E-03	GTPase activator activity
GO:0005102	7.22E-03	receptor binding
GO:0003677	7.22E-03	DNA binding
GO:0004725	1.73E-02	protein tyrosine phosphatase activity
GO:0008092	1.97E-02	cytoskeletal protein binding
GO:0005112	2.38E-02	Notch binding
GO:0042803	2.51E-02	protein homodimerization activity
GO:0042623	2.97E-02	ATPase activity, coupled

GO:0005085	3.07E-02	guanyl-nucleotide exchange factor activity
GO:0030528	3.28E-02	transcription regulator activity
GO:0004930	3.32E-02	G-protein coupled receptor activity
GO:0000166	4.61E-02	nucleotide binding
GO:0000287	4.61E-02	magnesium ion binding
GO:0004653	4.93E-02	polypeptide N-acetylgalactosaminyltransferase activity
GO:0050839	4.93E-02	cell adhesion molecule binding
GO:0003729	4.93E-02	mRNA binding
GO:0019992	4.98E-02	diacylglycerol binding

Table S16 GO term enrichment for cellular component in the North American cline for genes that overlap windows in the top 1% tail of Fst.

	1 % Tail Cellular Component	
GO term	q-value	description
GO:0005634	6.87E-09	nucleus
GO:0005622	4.74E-03	intracellular
GO:0005886	4.74E-03	plasma membrane
GO:0016021	2.47E-02	integral to membrane
GO:0005875	2.47E-02	microtubule associated complex
GO:0005938	2.47E-02	cell cortex
GO:0005887	2.47E-02	integral to plasma membrane
GO:0016020	2.93E-02	membrane
GO:0005737	4.09E-02	cytoplasm
GO:0005794	4.95E-02	Golgi apparatus

Table S17 GO term enrichment for cellular component in the North American cline for genes that overlap windows in the top 2.5% tail of Fst.

	2.5% Tail Cellular Component	
GO term	q-value	description
GO:0005634	2.94E-12	nucleus
GO:0005886	8.90E-10	plasma membrane
GO:0005737	1.72E-04	cytoplasm
GO:0045211	8.60E-04	postsynaptic membrane
GO:0016021	8.60E-04	integral to membrane
GO:0005622	8.60E-04	intracellular
GO:0005578	8.60E-04	proteinaceous extracellular matrix
GO:0045179	8.60E-04	apical cortex
GO:0005887	1.11E-03	integral to plasma membrane
GO:0045202	2.75E-03	synapse
GO:0005918	3.83E-03	septate junction
GO:0005875	4.55E-03	microtubule associated complex
GO:0016459	6.83E-03	myosin complex
GO:0005794	1.15E-02	Golgi apparatus
GO:0005938	1.15E-02	cell cortex
GO:0005576	2.29E-02	extracellular region
GO:0005912	2.30E-02	adherens junction
GO:0005700	2.37E-02	polytene chromosome
GO:0016020	4.28E-02	membrane
GO:0008021	4.90E-02	synaptic vesicle

Table S18 GO term enrichment for cellular component in the North American cline for genes that overlap windows in the top 5% tail of Fst.

	5% Tail Cellular Component	
GO term	q-value	description
GO:0005634	6.54E-12	nucleus
GO:0005886	5.50E-10	plasma membrane
GO:0005622	5.70E-07	intracellular
GO:0005737	2.14E-04	cytoplasm
GO:0045179	2.14E-04	apical cortex
GO:0005578	4.05E-04	proteinaceous extracellular matrix
GO:0016021	4.33E-04	integral to membrane
GO:0005887	9.71E-04	integral to plasma membrane
GO:0035003	9.71E-04	subapical complex
GO:0005938	1.01E-03	cell cortex
GO:0009986	1.09E-03	cell surface
GO:0005918	1.47E-03	septate junction
GO:0045211	2.63E-03	postsynaptic membrane
GO:0005875	3.15E-03	microtubule associated complex
GO:0005892	3.25E-03	nicotinic acetylcholine-gated receptor-channel complex
GO:0016324	3.57E-03	apical plasma membrane
GO:0016459	1.42E-02	myosin complex
GO:0005768	1.42E-02	endosome
GO:0005794	1.64E-02	Golgi apparatus
GO:0005795	1.90E-02	Golgi stack
GO:0045172	2.14E-02	germline ring canal
GO:0030135	2.14E-02	coated vesicle
GO:0000775	2.40E-02	chromosome, centromeric region
GO:0005813	3.04E-02	centrosome
GO:0045177	3.04E-02	apical part of cell
GO:0005868	3.04E-02	cytoplasmic dynein complex
GO:0005912	3.39E-02	adherens junction
GO:0045202	3.69E-02	synapse
GO:0005874	3.69E-02	microtubule
GO:0043190	4.14E-02	ATP-binding cassette (ABC) transporter complex
GO:0005576	4.25E-02	extracellular region
GO:0016020	4.75E-02	membrane

Table S19 Annotations of single nucleotide polymorphisms present on both continents

	All shared SNPs	All convergent SNPs	Strongly convergent SNPs (top 10% FST on both continents)
All annotations	361171	229705	4038
Three Prime UTR	11399	6552	107
Five prime UTR	8684	5107	99
Nonsynonymous	13526	8811	147
Synonymous	45398	29450	499
Intronic	150302	97021	1754
Intergenic	131799	82754	1432

Table S20 Functional annotation of strongly convergent UTR and nonsynonymous SNPs

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.161463/-/DC1

Table S21 Functional annotation of extremely differentiated convergent SNPs

Chr	Position	SNP classification	Fbgn	Gene Name	Functional Summary	AUFST	NAFST	Codon change	AA change
3R	15229365	five prime UTR	FBgn0038704	Aprataxin-like protein	DNA repair, metal binding	5.24E-01	6.31E-01	NA	NA
3R	12842274	five prime UTR	FBgn0004577	Dmel_CG15337	Unknown	5.26E-01	5.22E-01	NA	NA
X	8054656	nonsynonymous	FBgn0030014	Dmel_CG31091	lipid and ester catabolism	5.12E-01	4.65E-01	ATG->TTG	M->L
2R	7013061	nonsynonymous	FBgn0053503	Dmel_CG4218	Unknown	8.23E-01	6.75E-01	TTA->TCA	L->S
2L	14547899	nonsynonymous	FBgn0250844	Dmel_CG6454	calcium channel membrane targeting	5.12E-01	4.65E-01	AAG->ATG	K->M
3R	20766471	nonsynonymous	FBgn0029157	Peroxidase	response to oxidative stress /peroxide metabolism	8.23E-01	5.46E-01	TTG->TTC	L->F
3R	20177828	nonsynonymous	FBgn0039187	Probable cytochrome P450 12d1	oxidation/reduction, metal binding	6.44E-01	6.45E-01	ACT->AGT	T->S
3R	21727544	nonsynonymous	FBgn0051091	slingshot	eye development/mushroom body development	7.39E-01	7.55E-01	TCC->GCC	S->A

Table S22 Candidate convergent nonsynonymous SNPs

Position	N1	N2	Fst	Codon		AA	mel and outgroup ref sequence state
hkl							
AUS 19,045,921	20	29	0.34	aac/gac	N/D	D= mel,sim, sech, yak, ere; Q = ana	
AUS 19,045,936	20	30	0.21	aaa/gaa	K/E	many changes from E to K and reverse	
NA 19,045,921	not segregating						
NA 19,045,936	not segregating						
NA 19,046,788	28	34	0.35	aca/cca	T/P	P = mel,sim, sech, yak, ere, ana; other species not T	
NA 19,046,828	29	27	0.48	aac/agc	N/S	S=mel,sim,sech,ana,pse,pers,vir,moj,grim; T = ere, L=will	
AUS 19,046,788	not segregating						
AUS 19,046,828	not segregating						
otk							
AUS 7,890,468	29	22	0.38	gtg/ctg	V/L	L = all species	
AUS 7, 890,646	18	26	0.33	gat/gag	D/E	D = mel; E = other species	
AUS 7,901,503	36	43	0.26	ttg/tcg		L/S	L = all species
NA 7,890,468		not segregating					
NA 7, 890,646	not segregating						
NA 7,901,503		not segregating					
NA 7,892,834				0.71	tct/gct		S/A
NA 7,892,839				0.70	agc/aac	S/N	
AUS 7,892,834	not segregating						
AUS 7,892,839	not segregating						

Position	N1	N2	Fst	Codon		AA	mel and outgroup ref sequence state
ana1							
AUS 20,359,353	31	37	0.36	gag/gcg	E/A	E=mel;A=sim,sech,yak,ere	
AUS 20,359,385	33	36	0.45	cga/gga	R/G	R=mel,sim,yak,ere,ana	
NA 20,359,353	15	49	0.007				
NA 20,359,385	not segregating						

NA 20,360,921	36	52	0.33	atg/ctg	M/L	L=all flies except virilis; virilis=M			
NA 20,361,171	36	39	0.38	cag/ctg	Q/L	Q= mel, si, sech,yak, pse, pers. will, moj; R=ere;H=ana;L=vir, gri			
AUS 20,360,921	43	28	0.007						
AUS 20,361,171	35	27	0.06						
Chm									
AUS 7,413,693	23	29	0.52	cag/gag	Q/E	Q= all species			
NA 7,413,693		not segregating							
NA 7,412,813		30	56	0.24	gat/tat		D/Y	Y=mel;D=si, sech, yak, ere, ana, wil, vir, moj, gri;E=pse,pers	
NA 7,413,132		23	36	0.22	cca/cga	P/R	R=all species; no data for willistoni		
AUS 7,412,813	not segregating								
AUS 7,413,132	not segregating								
Trp									
AUS 25,740,476	20	24	0.35	ggc/agc	G/S	S=mel;G=sim,sech,yak,ere,ana,;A=pse;T=will, other no data			
AUS 25,741,376	36	27	0.22	tca/cca	S/P	P=mel,si,sech,ya,ere,ana,pse,pers,will;T=vir,gri;others ND			
NA 25,740,476	38	52	0.02						
NA 25,741,376	36	56	0.06						
NA 25,741,665	27	28	0.34	ttc/tcc		F/S	S=mel,sim,sech,yak,ere;A=ana,pse.pers.will.moj.gri		
NA 25,743,316	32	40	0.24	cac/aac	H/N	N=ana; H=all other species			
AUS 25,741,665	not segregating								
AUS 25,743,316	27	18	0.10						

CANDIDATE NORTH AMERICAN SPECIFIC DIFFERENTIATED NONSYNONYMOUS SNPS

Gene	Position	N1	N2	Fst	codon		residue	references sequence states	
Lsp2 (3L)	12,122,543	37	36	0.41	atg/gtg	M/V		V=mel,sim,sech,yak,ere,ana,pse,pers,grim;	
	Not segregating in AUS								L=wil,vi,moj
Oatp33Ea (2L)	12,442,835	31	65	0.50	tgt/ggt		C/G		L=all species
In AUS		29	37	0.08					

Table S23.

Seed length	Average coverage		Number of fixed or polymorphic bases		Number of singleton bases	
	30	45	30	45	30	45
k1n3	25.4	25.1	15801	15336	74753	72716
k2n5	30	29.9	19737	19522	149263	148169
k3n7	32	32	22457	22306	157904	157193
k4n9	32.4	32.4	23891	23744	160718	160294

s = seed length   k = number of errors allowed in the seed

n = number of errors allowed in the entire read

Table S24 Genome-wide average pairwise 1KB window Fst

	FLA	MAINE	QUE	TAZ
FLA	-	0.066	0.063	0.077
MAINE	-	-	0.066	0.067
QUE	-	-	-	0.072
TAZ	-	-	-	-

Table S25 ChrX average pairwise 1KB window Fst

	FLA	MAINE	QUE	TAZ
FLA	-	0.061	0.061	0.071
MAINE	-	-	0.058	0.069
QUE	-	-	-	0.06
TAZ	-	-	-	-

Table S26 Chr2L average pairwise 1KB window Fst

	FLA	MAINE	QUE	TAZ
FLA	-	0.069	0.063	0.083
MAINE	-	-	0.068	0.067
QUE	-	-	-	0.075
TAZ	-	-	-	-

Table S27 Chr2R average pairwise 1KB window Fst

	FLA	MAINE	QUE	TAZ
FLA	-	0.062	0.064	0.073
MAINE	-	-	0.062	0.067
QUE	-	-	-	0.067
TAZ	-	-	-	-

Table S28 Chr3L average pairwise 1KB window Fst

	FLA	MAINE	QUE	TAZ
FLA	-	0.065	0.064	0.077
MAINE	-	-	0.064	0.069
QUE	-	-	-	0.072
TAZ	-	-	-	-

Table S29 Chr3R average pairwise 1KB window Fst

	FLA	MAINE	QUE	TAZ
FLA	-	0.071	0.064	0.081
MAINE	-	-	0.076	0.06
QUE	-	-	-	0.083
TAZ	-	-	-	-