Title: Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*

Running title: Demography, adaptation, and clines in Drosophila

Authors: Alan O. Bergland[1], Ray Tobler[1,2], Josefa González[1,3], Paul Schmidt[4], Dmitri Petrov[1]

1. Department of Biology, Stanford University, Stanford, CA 94305-5020
2. Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, Vienna A-1210, Austria
3. Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra). Passeig Maritim de la Barceloneta 37-49. 08003 Barcelona, Spain.
4. Department of Biology, The University of Pennsylvania, Philadelphia, PA 19104

Contact information for corresponding author:

Alan O. Bergland
Department of Biology
Stanford University
Stanford, CA 94305-5020

email: bergland@stanford.edu
phone: (650) 736 2249
fax: (650) 723 6132

Abstract

Populations arrayed along broad latitudinal gradients often show patterns of clinal variation in phenotype and genotype. Such population differentiation can be generated and maintained by a combination of demographic events and adaptive evolutionary processes. Here, we investigate the evolutionary forces that generated and maintain clinal variation genome-wide among populations of *Drosophila melanogaster* sampled in North America and Australia. We contrast patterns of clinal variation in these continents with patterns of differentiation among ancestral European and African populations. We show that recently derived North America and Australia populations were likely founded by both European and African lineages and that this admixture event generated genome-wide patterns of parallel clinal variation. The pervasive effects of admixture meant that only a handful of loci could be attributed to the operation of spatially varying selection using an F_{ST} outlier approach. Our results provide novel insight into a well-studied system of clinal differentiation and provide a context for future studies seeking to identify loci contributing to local adaptation in *D. melanogaster*.

Introduction

All species live in environments that vary through time and space. In many circumstances, such environmental heterogeneity can act as a strong selective force driving adaptive differentiation among populations. Thus, a major goal of evolutionary and ecological genetics has been to quantify the magnitude of adaptive differentiation among populations and to identify loci underlying adaptive differentiation in response to ecologically relevant environmental variation.

Phenotypic and genetic differentiation between populations has been examined in a variety of species. In some cases, patterns of differentiation are directly interpretable in the context of circumscribed environmental differences that occur over short spatial scales (1). For instance, differences in salinity experienced by freshwater and marine populations of sticklebacks has led to the identification of key morphological, physiological, and genetic differences between replicate pairs of populations (2, 3) . Similarly, pigmentation morph frequency closely tracks variation in substrate color for a variety of species (4) thereby providing an excellent opportunity to directly relate environmental variation to phenotypic and genetic differentiation.

Patterns of genetic and phenotypic variation have also been examined in species arrayed along broad geographical transects such as latitudinal clines (5). In this paradigm, the goal has often been to identify the phenotypic and genetic basis for adaptation to temperate environments. In certain cases it has been possible to directly relate latitudinal variation in specific environmental variables to aspects of phenotypic and genetic differentiation (e.g., photoperiod and critical photoperiod or flowering time; (6, 7)). In general, the collinearity of multiple ecological and environmental variables along latitudinal clines often complicates the direct relation of environmental variation to specific phenotypic and genetic differences. Nonetheless, because many genetically based phenotypic clines within species often mirror deeper phylogenetic differentiation between endemic temperate and tropical species, it is clear that populations distributed along latitudinal clines have adapted to aspects of temperate environments (8).

Latitudinal clines have been extensively studied in various drosophilid species, most notably *Drosophila melanogaster*. Parallel clines in morphological (9, 10), stress

tolerance (11, 12), and life-history traits (12, 13) have been identified in *D. melanogaster* populations distributed along multiple continents. These phenotypic clines demonstrate that flies from poleward locales are generally more hardy albeit less fecund, reflecting a classic trade-off between somatic maintenance and reproductive output (12) that would be alternately favored between populations exposed to harsh winters versus more benign tropical environments. Extensive clinal variation in various genetic markers has also been identified (14, 15). In some cases clinal genetic variants have been directly linked to clinally varying phenotypes (16-19), whereas in other cases parallel clinal variation at genetic markers has been documented across multiple continents (20-22). Taken as a whole, there is abundant evidence that local adaptation to spatially varying selection pressures associated with temperate environments has shaped clinal patterns of phenotypic and genetic variation in *D. melanogaster*.

Demographic forces can also heavily shape patterns of clinal variation (5) and the recent demographic history of *D. melanogaster* may be particularly germane to our understanding of clinal patters of genetic variation in this species. *D. melanogaster* is an Afro-tropical species (23) that has colonized the world in the wake of human migration. Population genetic inference suggests that *D. melanogaster* first migrated out of Africa to Eurasia approximately 15,000 years ago (24) and eventually migrated eastward across Asia, arriving to South East Asia approximately 2.5Kya (25). *D. melanogaster* invaded the Americas and Australia within the last several hundred years and likely colonized these continents in their entirety quickly (26, 27). Historical records suggest that *D. melanogaster* colonized North America and Australia each in a single event (26, 27). However, population genetic (28, 29) and morphological evidence (30) suggest that, for the Americas at least, there were multiple colonization events with some migrants coming from Africa and some from Europe. While there is less evidence that Australia experienced multiple waves of colonization by *D. melanogaster* such a scenario is plausible given the high rates of human migration and inter-continental travel during the 19th century.

Evidence of multiple waves of colonization of North America comes from morphological and genetic observations. It has been noted that Caribbean populations of *D. melanogaster* are more phenotypically similar to African populations than continental North American populations are (30). Moreover, population genetic evidence demonstrates that at least one mid-latitude North American population of *D. mleanogaster* is a mixture of European and African lineages (28). These observations suggest that African populations of *D. melanogaster* colonized the Caribbean and then southern North America while European populations colonized northern North America. If true, North America would represent a secondary contact zone between diverged European and African flies given the high degree of differentiation between these ancestral populations (31). A natural consequence of such a scenario is that many genetic variants would appear clinal, even in the absence of spatially varying selection pressures (5).

Investigating whether North America and Australia represent secondary contact zones is, therefore, crucial for our understanding of the extent of spatially varying selection operating on this species. Classic work in *Drosophila* population genetics has suggested that a large number of polymorphisms are clinal (e.g., 14) and recent genomic work has further confirmed that a large fraction of the genome is highly differentiated (21, 22, 32)

and clinal (33) between temperate and tropical populations both within North America and Australia. Analyses based on a limited number of markers suggested that there is limited population structure among *D. melanogaster* populations (34, 35). Consequently, patterns of clinal variation genome-wide are often thought to be generated and maintained by spatially varying selection (e.g., 21, 22, 32, 33). However, if secondary contact has occurred in both North America and Australia, clinality at many loci throughout the genome could have been generated by demographic forces due to the high divergence between European and African populations of flies (31).

The models of dual colonization and adaptive differentiation as evolutionary forces that generate and maintain clinal variation in North America and Australia are not mutually exclusive. Notably, one plausible model is that dual colonization of these continents generated patterns of clinal variation and spatially varying selection acting has slowed the rate of genetic homogenization among populations. Accordingly, we sought to investigate whether genome-wide patterns of clinal genetic variation in North America and Australia show signals of dual colonization and local adaptation. We find that both North American and Australian populations show several genomic signatures consistent with secondary contact and suggest that this demographic process is likely to have generated patterns of clinal variation at a large fraction of the genome in both continents. Despite this genome-wide signal of recent admixture, we find evidence that spatially varying selection has shaped patterns of allele frequencies at some loci along latitudinal clines. We discuss these findings in relation to the well-documented evidence of spatially varying selection acting on this species as well as the interpretation of patterns of genomic variation along broad latitudinal clines in general.

Results

Data. We examined genome-wide estimates of allele frequencies from ~30 populations of *D. melanogaster* sampled throughout North America, Australia, Europe and Africa (Fig. 1A). Our analyses largely focused on patterns of variation in North American and Australian populations and, consequently, we primarily focus on two sets of SNP markers. First, we utilized allele frequency estimates at ~500,000 high quality SNPs that segregate at intermediate frequency (MAF > 15%) in North America (36). The second set was composed of ~300,000 SNPs that segregate at intermediate frequency in Australia (32). For analyses that examine patterns of polymorphism in both North America and Australia, we examined SNPs that were at intermediate frequency in both continents, yielding a dataset of ~190,000 SNPs. Because of the low sequencing coverage in the Australian populations, it is unclear if the reduced polymorphism in that continent reflects the demographic history of those populations or experimental artifact. Although our analysis primarily focused on patterns of polymorphism in North America and Australia, we also examined allele frequency estimates at both sets of polymorphic SNPs in populations sampled in Europe (37, 38) and Africa (31).

Genomic signals of secondary contact. We performed a series of independent analyses to examine whether North America and Australia represent secondary contact zones of European and African populations of *D. melanogaster*. First, we constructed a neighbor-joining tree based on genome-wide allele frequency estimates from populations sampled

world-wide using 500 sets of 10,000 randomly sampled SNPs (Fig 1B). As expected, African populations exhibited the greatest diversity (31) and clustered at the base of the tree while European populations clustered at the tip. North American and Australian populations clustered between African and European populations (Fig. 1A), a pattern that supports the model (39) that both North American and Australian populations result from secondary contact of European and African ones.

Next, we calculated the proportion of African ancestry in North American and Australian populations by modeling these populations as a linear combination of African and European ancestry. Proportion African ancestry is negatively correlated with latitude in North America and Australia (Fig. 1C). Within the Pennsylvanian population, the proportion of African ancestry was not different between samples collected during the spring and the fall (Fig. 1C). Note, in this analysis the proportion of European ancestry is the inverse (i.e., 1-α; see Materials and Methods) of the proportion African ancestry; thus the proportion of European ancestry is positively correlated with latitude in North America and Australia.

Finally, we calculated the f_3 (40) statistic – a formal test of admixture – for each North American and Australian population using each sampled European and African population as a putative source population. We observe a significantly negative f_3 statistic (Table 1) for each North American population when using the Italian and Cameroonian populations as donor populations. Significantly negative f_3 statistics for subsets of North American populations were observed when using other European and African populations as donor sources (Data S1). A negative f_3 statistic can be taken as conclusive evidence of secondary contact between these, or closely related, donor populations (41). We did not observe significantly negative f_3 statistics for either Australian population using any combination of European or African source populations (Data S1).

We note that the absence of evidence of admixture using the f_3 statistic for the Australian populations should not be taken as evidence of the absence of admixture. Notably, the choice of donor populations can influence the value of the f_3 statistic. For instance, we do not observe significantly negative f_3 statistics for all North American populations when using alternate founder populations (Data S1). Therefore, we speculate that Australian populations were founded by European (likely British, see *Discussion*) and/or African populations that are not included in our dataset.

Taken together, these results support the view that both North America and Australia represent secondary contact zones between European and African lineages of *D. melanogaster*. Moreover, our results confirm an earlier model (23) that European *D. melanogaster* colonized high latitude locales in North American and Australia whereas African flies colonized low latitude locales in these continents. Genome-wide, low-latitude populations are more similar to African ones whereas high-latitude populations are more similar to European ones.

Under this dual-colonization scenario, we would expect that a large fraction of the genome varies clinally. Indeed, among North American populations of *D. melanogaster* approximately one third of all common SNPs, on the order of 10^5, are clinal (36). The vast extent of clinal variation in North America, then, is consistent with a dual colonization scenario which would generate patterns of clinal variation at a large fraction of the genome. However, these results do not preclude the existence of spatially varying selection that could also be acting among these populations which could explain patterns

of differentiation reported for some loci (e.g., 14, 15, 17) and could slow the rate of homogonization of allele frequencies at neutral polymorphisms throughout the genome among clinally distributed populations. We note that a similar analysis of the extent of clinality in Australia is not possible because we lack genome-wide allele frequency estimates from intermediate latitude populations in that continent.

Genomic signals of parallel local adaptation along latitudinal gradients. Our previous analysis supports the model that the demographic history of *D. melanogaster* has contributed to genome-wide patterns of differentiation among temperate and tropical populations of *D. melanogaster* living in North America and Australia. Regardless of this putative demographic history, multiple lines of evidence suggest that populations of flies living along broad latitudinal gradients have adapted to local environmental conditions that may be associated with aspects of temperate environments (see *Discussion*). Accordingly, we performed several tests to assess whether there is a strong, observable genomic signal of local adaptation.

First, we sought to identify F_{ST} outliers. We used a novel technique, OutFLANK (42), that attempts to identify SNPs subject to spatially varying selection while maintaining a low false positive rate. This method models the distribution of F_{ST} values after trimming the extreme tails under the assumption that the central portion (e.g., the 5th-95th quantiles) of the F_{ST} distribution largely reflects the demographic history of the sampled populations. Then, using the inferred F_{ST} distribution as a null distribution, OutFLANK seeks to identify SNPs that are more differentiated than expected by chance. At a false discovery rate (FDR) of 5%, we did not identify any SNPs with F_{ST} values significantly higher than expected by chance among temperate and tropical populations in Australia. At a similar FDR of 5%, we identified ~200 SNPs with F_{ST} values significantly higher than expected by chance (Data S2) among North American populations. Note that the genome-wide average F_{ST} among North American populations is lower than among Australian populations (0.025 vs. 0.08 respectively), suggesting that the lack of significantly elevated F_{ST} values in Australia is not due to a lack of population differentiation but rather a high genome-wide differentiation likely caused by recent secondary contact.

The exact number of SNPs with significantly elevated F_{ST} in any particular continent will be subject to a various of considerations including the number of sampled populations, the precision of allele frequency estimates, and the power of particular analytic methods to detect outlier F_{ST}. Some of these factors vary between our North American and Australian samples and thus our power to detect significant elevation of F_{ST} will vary between continents. Therefore, we investigated the general patterns of differentiation and parallelism between the sets of populations sampled in North America and Australia. In addition, we also examined patterns of differentiation and parallelism between these continents and populations sampled from the Old-World (i.e., Europe and Africa).

For these analyses, we first examined whether SNPs that were highly differentiated among one set of populations were also differentiated in another set (hereafter, 'co-differentiated'). To perform this analysis, we calculated the odds ratio (see *Materials and Methods)* that SNPs fell above a particular quantile threshold of the F_{ST} distribution in any two sets of populations (Fig 2A). We performed this analysis for SNPs that fell either

within or outside of the large cosmopolitan inversions (43). We find that SNPs that are highly differentiated in North America are also highly differentiated in Australia. In addition, we find that SNPs that are highly differentiated in either North America or Australia are also highly differentiated between Europe and Africa. Although patterns of co-differentiation are higher among SNPs within the large, cosmopolitan inversion than for SNPs outside the inversions, the qualitative patterns remain the same for either SNP class suggesting that clinal variation in inversions *per se* does not drive the observed high levels of co-differentiation.

SNPs that are co-differentiated among temperate and tropical populations in North America, Australia, or the Old-World can be differentiated in a parallel way or at random among each geographic region. We show here that there is a high degree of parallelism at the SNP level, genome-wide, among polymorphisms that are highly differentiated in any two sets of populations (Fig 2B). Patterns of parallelism at highly co-differentiated SNPs are similar among SNPs within or outside the large cosmopolitan inversions again suggesting that clinal variation in inversions are not driving genome-wide patterns of parallelism.

High rates of co-differentiation and parallelism among temperate and tropical populations sampled throughout the world can be interpreted in two ways. On the one hand, these patterns could be taken as evidence of parallel adaptation to aspects of temperate environments. On the other hand, these patterns are consistent with the model presented above that North American and Australian populations are the result of recent secondary contact between European and African lineages of flies (see *Results: Genomic signals of secondary contact)*.

To differentiate these alternative interpretations, we estimated rates enrichment of highly co-differentiated SNPs and rates of parallelism at highly co-differentiated SNPs among classes of polymorphisms that that we expect, *a priori*, to be more or less likely to contribute to local adaptation. We reasoned that SNPs falling in short-introns, which have been previously shown to evolve neutrally (44), would be the least likely to contribute to local adaptation. In contrast, SNPs in other functional classes (e.g., coding, UTR, intron) might be more likely to contribute to local adaptation along latitudinal clines (21). We contrasted rates of co-differentiation and parallelism at these putatively functional SNP classes with rates at the short-intron (hereafter 'neutral') SNPs and at control SNPs matched to each class by several important biological and experimental features. These comparisons take into account the spatial distribution of SNPs along the chromosome (see *Materials and Methods*). We reasoned that if parallel adaptive processes have contributed to genome-wide signals of co-differentiation and parallelism in Australia and North America, (1) some functional SNP classes would show a higher rate of co-differentiation and parallelism than neutral SNPs, (2) functional SNPs would show a higher rate of co-differentiation and parallelism than their control SNPs, and (3) neutral SNPs would show a lower rate of co-differentiation and parallelism than their control SNPs.

We find little evidence that various functional classes show differences in rates of co-differentiation or parallelism than either neutral SNPs or their matched controls (Fig. 3). Moreover, neutral SNPs show similar rates of co-differentiation and parallelism as their matched controls (Fig. 3). There is suggestive evidence that SNPs falling in 5' UTRs show greater of co-differentiation than expected by chance, but this comparison is not

significant after correcting for multiple tests (see $F_{ST} > 95\%$ Fig. 3A; $p_{naive} = 0.01$; $p_{corrected} = 0.24$). Moreover, highly co-differentiated SNPs in 5'UTR are not more likely to be parallel than expected by chance (Fig. 3B), suggesting that the observed excess of co-differentiation may be a statistical artifact. All other tests of excess co-differentiation or parallelism at different SNP classes were not significantly different from expectation ($p > 0.05$).

Taken together, the tests we performed to identify strong genomic signals of parallel adaptation along latitudinal clines were equivocal. We show that there were a modest number of F_{ST} outliers among North American populations sampled along a broad latitudinal cline and no observable F_{ST} outliers among Australian populations. While the outlier detection method we used is highly conservative, the fact that so few outliers were detected suggests that the bulk of the F_{ST} distribution is determined by the demographic history of this species. We show that SNPs with high F_{ST} among any one set of populations are likely to have high F_{ST} among other sets of population. Furthermore, SNPs that are highly co-differentiated are likely to vary in a parallel fashion among geographic regions. While this result could suggest parallel adaptation, it is also consistent with the dual colonization model we present above. Finally, we show that rates of co-differentiation and parallelism at highly co-differentiated SNPs are similar between functional SNPs, neutral SNPs, and their matched control SNPs suggesting that the evolutionary forces shaping allele frequencies along latitudinal clines are similar across SNPs that are more- or less-likely to contribute to local adaptation.

Discussion. Herein we report results from a series of analyses that (1) examine whether populations of *D. melanogaster* sampled throughout North America and Australia show signatures of recent secondary contact between European and African lineages, and (2) examine whether there is a genomic signal of spatially varying selection acting along latitudinal gradients. We find that both North America and Australia show several signatures of secondary contact (Fig. 1BC, Table 1). Notably, high latitude populations are closely related to European populations, whereas low latitude populations are more closely related to African ones. This result implies that a large portion of clinal variation within these continents could, in principal, be generated by the dual colonization of both North America and Australia (Fig. 1A). Consistent with this view, SNPs that are highly differentiated between temperate and tropical locales in either North America or Australia are also highly likely to be differentiated in a parallel way between Europe and Africa (Fig. 2AB). In addition, we report that genome-wide scans for significantly differentiated polymorphisms identified a limited number of outlier loci. Taken together, our results support the model that recent secondary contact in North America and Australia has generated clinal variation at a large fraction of polymorphisms genome-wide and that spatially varying selection acting at a moderate number of loci acts to slow the rate of genomic homogenization between geographically separated populations.

Secondary contact and the generation of clinal variation in allele frequencies. Recent secondary contact between formerly (semi-) isolated populations is a potent force that can generate clinal variation genome-wide (5). In *D. melanogaster*, high levels of genetic differentiation have been observed between temperate and tropical populations sampled in North America and Australia (21, 22, 32, 33). In North America at least, most of these

highly differentiated SNPs vary clinally (i.e., in a roughly monotonic fashion along latitudinal gradients;(36)). Moreover, surveys of allele frequencies along latitudinal clines in both North America and Australia at allozymes (14), SNPs (14, 36, 45), microsatellites (46), and transposable elements (20) have repeatedly demonstrated that approximately one third of all surveyed polymorphisms are clinal in either continent. At face value the high proportion of clinal polymorphisms throughout *D. melanogaster*'s genome suggests that demographic processes such as secondary contact have contributed to the generation of clinal variation in this species among recently colonized locales (26, 27).

Accordingly, we tested if newly derived populations of *D. melanogaster* show signatures of recent secondary contact. Using a variety of tests, we show that genome-wide patterns of genetic variation from populations sampled in North America and Australia are consistent with recent secondary contact (Fig. 1BC, Table 1). While historical records from North America (26) and Australia (27) suggest a single point of colonization of *D. melanogaster*, results from morphological, behavioral, and genetic studies reported here and elsewhere (28-30) suggest that a dual colonization scenario is more likely. At least for the Americas active trade between Europe and western Africa supports the model that North America represents a secondary contact zone. Australia did not experience the same types of trade with the Old World and throughout the 19th century intercontinental travel to Australia was primarily restricted to British ships. However, British ships traveling to Australia ported in South Africa and India then, after the opening of the Suez Canal in East Africa (47). This raises the possibility that secondary contact between European and African fruit fly lineages could have occurred immediately prior to the successful colonization of Australia by *D. melanogaster* in the mid 19th century (27). Under this mixed-lineage, single colonization scenario, rapid ecological sorting of colonizing lineages to temperate and tropical niches (48) may have created a gradient where European flies were initially predominant at high latitudes and African flies predominant at low latitudes within Australia.

Although secondary contact is capable of generating patterns of clinal variation genome-wide, clines generated through this demographic process are transient. As admixed populations approach migration-selection equilibrium, clines at neutral loci should attenuate. Moreover, once at equilibrium, neutral differentiation should be minimal (49) for species such as *D. melanogaster* where Nm has been estimated to be on the order of ~1 (50, 51) and long-distance dispersal is believed to be frequent (52).

Thus, the critical question in determining whether the vast amount of clinal variation in North American and Australian flies has been generated by demography or selection is whether or not this species is at migration-selection equilibrium in these continents. There are several reasons why we suspect this species is not at equilibrium. First, *D. melanogaster* appeared in North America and Australia in the mid- to late 19th century (26, 27), or on the order of 1000 generations ago, assuming approximately 10 generations per year. Estimates of local, demic N are on the order of 10^4 (53-55) implying that m is on the order of 10^{-4} (if $Nm \sim 1$). If these estimates are accurate to the order of magnitude, it would take approximately 2500 generations to get about half way to equilibrium (56) or ~10,000 generations to fully approach equilibrium (57). Thus, from a simple demographic perspective, it would seem unlikely that *D. melanogaster* has reached migration-selection-drift equilibrium.

A second piece of evidence that *D. melanogaster* have yet to reach equilibrium come from contrasting patterns of migration and differentiation between drosophilid species. Singh and Rhomberg (51) contrasted estimates of population differentiation and *Nm* between North American populations of *D. melanogaster* and *D. pseudoobscura*. They note that both species have similar estimates of *Nm* (~1) but that *D. melanogaster* shows higher levels of differentiation than *D. pseudoobscura*. These authors suggested the discrepancy between levels of differentiation between these species is a function of their ecology and adaptive evolutionary dynamics. They argued that high levels of differentiation among *D. melanogaster* populations results from local adaptation driven by the varied selection pressures associated with human commensalism, whereas low levels of differentiation in *D. pseudoobscura* might result from habitat selection. Both species, however, appear to rapidly evolve to subtle shifts in selection pressures experienced in the field (36, 58). Therefore, we conclude that differences in patterns of differentiation between these species reflects the fact that *D. pseudoobscura* is a Nearctic endemic and is thus likely to be closer to equilibrium than emigrant populations of *D. melanogaster*.

Finally, it is worth noting that others have suggested that non-African populations of *D. melanogaster* are not at equilibrium. In general, non-African populations of *D. melanogaster* show a reduction in diversity coupled with an excess of rare variants (59). This genome-wide patterns is consistent with a population bottleneck during colonization followed by population expansion. Others have noted that non-African populations of *D. melanogaster* also have higher levels of linkage-disequilibrium (LD) than expected under the standard neutral model (60-62) whereas LD in African populations is more consistent with neutrality (62 cf. 63). Although genome-wide elevation of LD could be caused by various factors including pervasive positive- or negative-selection, admixture would also possibly generate this signal.

Previous studies examining departure from equilibrium models in *D. melanogaster* have concluded that caution should be taken when conducting genome-wide scans for positive-selection given the non-equilibrium nature of this species (60). Notably, demographic forces such as population bottlenecks can, in principal, mimic many of the signatures left by some types of adaptive evolution. A complimentary approach to quantify the magnitude of adaptive evolution and to identify loci subject to selection is to identify polymorphisms that are differentiated between populations that are subject to divergent selection pressures. However, results presented here demonstrate that, for *D. melanogaster* at least, signatures of adaptive evolution from genome-wide patterns of differentiation along latitudinal clines in newly derived populations in North America and Australia should be taken with a similar or even greater degree of caution as traditional scans for recent, positive selection.

Spatially varying selection and the maintenance of clinal variation in allele frequencies. Whereas secondary contact is capable of generating clinal variation, spatially varying selection is required for its long-term maintenance. There is little doubt that populations of *D. melanogaster* living along broad latitudinal clines in temperate environments have adapted to spatially varying selection pressures. Support for the idea of local adaptation along latitudinal clines comes from three main lines of evidence.

First, certain phenotypes show repeatable clines along latitudinal and altitudinal gradients that mirror deeper phylogenetic variation among temperate and tropical species. For instance, aspects of body size vary clinaly in North America (64) and Australia (65) as well as along altitudinal/latitudinal clines in India (66) and altitudinal clines within Africa (67, 68). Given such patterns of parallelism within and among continents, including within the ancestral African range, the most plausible explanation is that parallel selection pressures have generated these patterns of latitudinal and altitudinal variation. These intraspecific clines mimic interspecific patterns among temperate and tropical endemic drosophilids following Bergmann's rule (69, 70) again implicating that natural selection has shaped these patterns of genetically based, phenotypic variation.

Second, certain genetic and phenotypic clines in *D. melanogaster* have shifted over decadal scales. Shifts in these clines are consistent with adaptation to aspects of global climate change wherein alleles common in low-latitude populations have become more prevalent in high-latitude ones over the last 20 years (15).

Finally, using a conservative outlier detection approach (42), we identify several hundred polymorphisms in North America that are significantly differentiated (see *Results* and Data S2). Although the function of many of these polymorphisms is presently unknown, several are within the genes known to affect life-history traits that vary among temperate and tropical populations. For instance, one significantly differentiated SNP in North America (3R:17433977) resides within the first intron of the *Insulin receptor* gene (*InR*). Natural polymorphisms in *InR* have been recently shown to contribute to local adaptation between temperate and tropical populations of flies (16, 17). Two additional significantly differentiated SNPs (3R:13749473 and 3R:13894182) reside within introns of *couch-potato* (*cpo),* a gene which has been shown to be associated with diapause incidence in natural populations (13 cf. 18). Intriguingly, the SNP in *cpo* that has been previously associated with diapause incidence (3R: 13793588) is not among the significantly differentiated SNPs within North America. In our dataset, this SNP has an observed F_{ST} of 0.1 among North American populations falling in the upper 1.5% of the F_{ST} distribution. However, the associated false discovery rate of this SNP under the model used in OutFLANK is 80%. Similarly, the extensively studied threonine/lysine polymorphism (53, 71) that encodes the Fast and Slow allozyme variants at *Alchohol dehydrogenase* (*Adh,* 2L:14617051) falls in the upper 3.5% quantile of the North American F_{ST} distribution (FDR 99%).

The identification of significantly differentiated SNPs within North America can be taken as evidence of local adaptation to spatially varying selection pressures. However, the observation that two SNPs (one in *cpo* and one in *Adh*) that each likely contribute to local adaptation fall in an upper, but not extreme, tail of the F_{ST} distribution suggests that there are many more ecologically relevant and functional polymorphisms that have contributed to local adaptation in *D. melanogaster*. However, the signal of high differentiation caused by spatially varying selection at these SNPs is likely masked by recent admixture that has contributed to a high level of differentiation genome-wide. In light of these results, we suggest that scans for local adaptation based on patterns of genetic differentiation in *D. melanogaster* are an important first step in identifying adaptively differentiated clinal polymorphisms but that additional evidence, such as functional validation (17, 18), should be gathered before concluding that differentiation is caused by adaptive processes.

Conclusions. It has long been recognized that genetic differentiation among populations can be caused by both adaptive and demographic (neutral) processes (72). Due to *D. melanogaster*'s large effective population size (73) and high migration rate (52), others concluded that differentiation among populations sampled along latitudinal gradients is primarily caused by spatially varying selection. Work presented here supports the notion that spatially variable selection does contribute to some differentiation among populations (see *Results: Spatially varying selection…*). However, several genome-wide signatures presented here (Fig 1BC, Table 1) and elsewhere (28, 29, 36) indicate that populations of flies in North America and Australia result from admixture of European and African lineages. High-latitude (temperate) populations in North America and Australia are more closely related to European populations whereas low-latitude (tropical) populations are more closely related to African ones (Fig 1BC) suggesting that admixture occurred along a latitudinal gradient and that this demographic event generated clinal genetic variation at roughly 1/3 of all common SNPs (36). These colonizing lineages of flies were likely already differentially adapted to the temperate and tropical conditions that they encountered in North America and Australia. Consequently the recent demographic history of this species in North America and Australia is collinear with both local adaptation within these newly colonized continents and among the ancestral ranges.

One practical consequence of the collinearity of demography and adaptation is that the identification of clinality at any particular locus cannot be taken exclusively as evidence of spatially varying selection. We propose that an alternative approach to identify loci that contribute to aspects of adaptation to temperate environments in *D. melanogaster* is to identify alleles that vary over spatial and seasonal gradients (36) that are orthogonal to the demographic history of this species.

Materials and Methods

Genome-wide allele frequency estimates. We utilized novel and publically available genome-wide estimates of allele frequencies of *D. melanogaster* populations sampled world-wide (Figure 1A, Table S1). Allele frequency estimates of six North American populations are described in Bergland *et al*. (36). Allele frequency estimates of three European populations are described in Bastide *et al*. (37) and Tobler *et al*. (38). Allele frequency estimates from 22 African populations are described in Pool *et al*. (31). Allele frequency estimates of two Australian populations are described in Kolaczkowski *et al*. (32). Allele frequency estimates from an additional two Australian populations are reported here for the first time. Allele frequency estimates from these additional Australian populations were made by pooling ten individuals from each of 22 isofemale lines originating from Innisfail (17°S) or Yering Station (37°S), Australia (isofemale lines kindly provided by A. Hoffmann). Sequencing libraries and mapping followed methods outlined in (36). Because Australian data were low coverage (~10X per sample, on average), we combined the two northern populations and two southern populations into two new, synthetic populations which we refer to as 'tropical' and 'temperate,' respectively.

We performed SNP quality filtering similar to the methods presented in Bergland et al (36). Briefly, we excluded SNPs within 5bp of polymorphic indels, SNPs within

repetitive regions, SNPs with average minor allele frequency less than 15% in both North America and Australia, SNPs with low (<5) or excessively high read depth (>2 times median read depth) and SNPs not present in the Drosophila Genetic Reference Panel (59). African samples were not quality filtered for read depth because allele frequency estimates from these samples were derived from sequenced haplotypes and not pooled samples. Regions of inferred admixture (31) in African samples (i.e., introgression of European haplotypes back to African populations) were removed from analysis.

Estimation of the population tree. We calculated Nei's genetic distance (74) between each pair of populations and generated a population tree using the neighbor-joining algorithm implemented in the R (75) package *ape* (76). To generate bootstrap values for each node, we randomly sampled 10,000 SNPs 500 times.

Estimation of the proportion African and European ancestry. We obtained maximum-likelihood estimates of the proportion of African and European ancestry in North American and Australian populations. For these estimates, we modeled each North American and Australian population as a linear combination of African and European populations. Note, for the Pennsylvanian population, we pooled together flies collected over the course of three years in either the spring or fall.

To estimate the proportion of African ancestry in each sampled population, we maximized the likelihood,

$$\log \mathrm{Lik}(\alpha; x_{ij}, nEff_{ij}, fq_i^{Af}, fq_i^{Eu}) = \sum_{i=1}^{n} \ln\left[\alpha \cdot Bin(x_{ij}, nEff_{ij}, fq_i^{Af}) + (1-\alpha) \cdot Bin(x_{ij}, nEff_{ij}, fq_i^{Eu})\right],$$

where α is the estimated proportion African ancestry (and 1-α the proportion of European ancestry); x_{ij} is the count of alternate (non-reference) reads at SNP i in North American or Australian population j, $nEff_{ij}$ is the number of effective reads at SNP i in population j; fq_i^{Af} and fq_i^{Eu} are the observed allele frequencies of SNP i averaged over all sub-populations in each of the African and European continents, respectively; among the n SNPs under investigation. We define $nEff_{ij}$, the effective number of reads at SNP i in population j as,

$$floor\left(\frac{rd_i \cdot nChr_j - 1}{rd_i + nChr_j}\right),$$

where rd_i is the number of reads covering the ith SNP and $nChr_j$ is the number of chromosomes sampled from the jth population. We use the effective number of reads rather than read depth to account for the double binomial sampling that can occur during pooled resequencing which can lead to inflated precision unless this correction is applied. α was estimated by maximizing this likelihood function using the *optimize* procedure in R. In order to generate confidence intervals of α, we performed bootstrap resampling by randomly sampling 500 sets of 10,000 SNPs.

Formal tests of admixture. We used the f_3 statistic (40) to test if North American and Australian populations show signatures of admixture between African and European populations. For each North American and Australian population, we calculated f_3 using

each European population as one putative donor population and each African population with more than 5 haplotypes as the other putative donor population. f_3 statistics were calculated using *TreeMix* version 1.13 (77) with 500 bootstrap replicates sub-sampling one of every 500 SNPs.

F_{ST} outlier test. We used OutFLANK (42) to test for the presence of polymorphisms with higher F_{ST} than expected by chance. Under an island model, the classic Lewontin-Krakauer test for F_{ST} outliers assumes that the distribution of F_{ST} is proportional to a χ^2 distribution with degrees of freedom equal to one less the number of populations examined (78). The assumption underlying this model have been criticized (79) and OutFLANK seeks to identify F_{ST} outliers by inferring the degrees of freedom of the observed F_{ST} distribution after trimming the distribution of high and low F_{ST} SNPs. This method has been shown to have a low false positive rate. We used OutFLANK to identify F_{ST} outliers among either North American populations (average pair-wise F_{ST} between samples collected in FL, GA, SC, NC, PA, and ME) or Australian populations (temperate vs. tropical) by trimming the top and bottom 5% of the observed F_{ST} distribution.

Differentiation and rates of parallelism at various SNP classes. We tested if SNPs at neutral sites (short-introns) or various functional categories (Fig. 3) were more likely than expected by chance to be co-differentiated or show parallel changes in allele frequency between temperate and tropical locales in both North America and Australia conditional on them being co-differentiated. To assess rates of co-differentiation we calculated the odds that SNPs fell above one of three F_{ST} quantile thresholds (85, 90, 95%) in both North America and Australia. We compared this value to the odds of co-differentiation from 500 sets of randomly selected SNPs that were matched to the focal SNPs by recombination rate (80), chromosome, inversion status (at the large, cosmoplitan inversions *In(2L)t, In(2R)NS, In(3L)Payne, In(3R)K, In(3R)Payne, In(3R)Mo, In(X)A,* and *In(X)Be*), average read depth in North America and Australia, and heterozygosity in both continents. To control for the possible autocorrelation in signal along the chromosome, we divided the genome into non-overlapping 50Kb blocks and randomly sampled, with replacement, one SNP per block. To assess rates of parallelism, we calculated the fraction of SNPs that were significantly co-differentiated and varied in a parallel fashion between North America and Australia for each SNP class and their matched, genomic controls, again controlling for the spatial distribution of SNPs along the chromosome. We report the difference in rates of parallelism (Fig. 3B). Standard deviations of the $\log_2$(odds-ratio) of co-differentiation and for differences in the rates of parallelism are calculated as in (36).

Acknowledgements. We thank Joyce Kao, Heather Machado, and Annalise Paaby for insightful comments on earlier versions of this manuscript. We thank David Lawrie for providing lists of neutrally evolving SNPs and Ary Hoffmann for graciously providing isofemale lines from Australia. AOB was supported by an NIH National Service Research Award (F32 GM097837). JG is a Ramon y Cajal fellow (RYC-2010-07306) supported by grants from the European Commission (Marie Curie CIG PCIG-GA-2011-293860) and from the Spanish Government (Fundamental Research Projects Grant BFU-

2011-24397). This work was supported by NSF DEB 0921307 (awarded to PS) and NIH R01GM089926 (awarded to PS and DP).

References

1. Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution* 29:165–176.
2. Jones FC et al. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.
3. Kitano J et al. (2010) Adaptive divergence in the thyroid hormone signaling pathway in the stickleback radiation. *Current Biology* 20:2124–2130.
4. Gray SM, McKinnon JS (2007) Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution* 22:71–79.
5. Endler JA (1993) *Geographic Variation, Speciation, and Clines* (Princeton University Press, Princeton, NJ).
6. Bradshaw WE, Lounibos LP (1977) Evolution of dormancy and its photoperiodic control in pitcher-plant mosquitoes. *Evolution* 31:546.
7. Stinchcombe JR et al. (2004) A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *PNAS* 101:4712–4717.
8. Gibert P, Moreteau B, Pétavy G, Karan D, David JR (2001) Chill-coma tolerance, a major climatic adaptation among drosophila species. *Evolution* 55:1063–1068.
9. Telonis Scott M, Hoffmann AA, Sgró CM (2011) The molecular genetics of clinal variation: a case study of ebony and thoracic trident pigmentation in *Drosophila melanogaster* from eastern Australia. *Molecular Ecology* 20:2100–2110.
10. Rajpurohit S, Parkash R, Ramniwas S, Singh S (2008) Variations in body melanisation, ovariole number and fecundity in highland and lowland populations of *Drosophila melanogaster* from the Indian subcontinent. *Insect Science* 15:553–561.
11. Hoffmann AA, Anderson A, Hallas R (2002) Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. *Ecology Letters* 5:614–618.
12. Schmidt PS, Matzkin L, Ippolito M, Eanes WF (2005) Geographic variation in diapause incidence, life-history traits, and climatic adaptation in *Drosophila melanogaster*. *Evolution* 59:1721–1732.
13. Lee SF et al. (2011) Polymorphism in the couch potato gene clines in eastern Australia but is not associated with ovarian dormancy in *Drosophila melanogaster*. *Molecular Ecology* 20:2973–2984.
14. Sezgin E, Duvernell DD, Matzkin LM, Duan Y, Zhu CT (2004) Single locus latitudinal clines in metabolic genes, derived alleles, and their relationship to temperate adaptation in *Drosophila melanogaster*. *Genetics* 168:923–931.
15. Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* 129:133–147.
16. Paaby AB, Bergland AO, Behrman EL, Schmidt PS (2014) An amino acid polymorphism in the Drosophila insulin receptor demonstrates pleiotropic and

adaptive function in life history traits. *bioRxiv:* 008193.
17. Paaby AB, Blacket MJ, Hoffmann AA, Schmidt PS (2010) Identification of a candidate adaptive polymorphism for *Drosophila* life history by parallel independent clines on two continents. *Molecular Ecology* 19:760–774.
18. Schmidt PS et al. (2008) An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *PNAS* 105:16207–16211.
19. Lee SF et al. (2013) Polymorphism in the neurofibromin gene, *Nf1*, is associated with antagonistic selection on wing size and development time in *Drosophila melanogaster*. *Molecular Ecology* 22:2716–2725.
20. González J, Karasov TL, Messer PW, Petrov DA (2010) Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*. *PLoS Genet* 6:e1000905.
21. Reinhardt JA, Kolaczkowski B, Jones CD, Begun DJ, Kern AD (2014) Parallel geographic variation in *Drosophila melanogaster*. *Genetics* 197:361–373.
22. Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* 179:455–473.
23. David J, Capy P (1988) Genetic variation of *Drosophila melanogaster* natural populations. *Trends in Genetics* 4:106–111.
24. Li H, Stephan W (2006) Inferring the demographic history and tate of adaptive substitution in *Drosophila*. *PLoS Genet* 2:e166.
25. Laurent SJY, Werzner A, Excoffier L, Stephan W (2011) Approximate Bayesian Analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. *Mol Biol Evol* 28:2041–2051.
26. Keller A (2007) *Drosophila melanogaster*'s history as a human commensal. *Current Biology* 17:R77–R81.
27. Bock IR, Parsons PA (1981) in eds Ashburner M, Carson HL, Thompson JN (Genetics and biology of Drosophila, London), pp 291–306.
28. Duchen P, Živković D, Hutter S, Stephan W, Laurent S (2013) Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* 193:291–301.
29. Caracristi G, Schlötterer C (2003) Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol Biol Evol* 20:792–799.
30. Rouault J-D, Marican C, Wicker-Thomas C, Jallon J-M (2004) Relations between cuticular hydrocarbon (hc) polymorphism, resistance against desiccation and breeding temperature; a model for hc evolution in *D. melanogaster* and *D. simulans*. *Genetica* 120:195–212.
31. Pool JE et al. (2012) Population genomics of Sub-Saharan *Drosophila melanogaster*: African diversity and Non-African admixture. *PLoS Genet* 8:e1003080.
32. Kolaczkowski B, Kern AD, Holloway AK, Begun DJ (2011) Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics* 187:245–260.
33. Fabian DK et al. (2012) Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology*

21:4748–4769.
34. Berry A, Kreitman M (1993) Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. *Genetics* 134:869–893.
35. Gravot E, Huet M, Veuille M (2004) Effect of breeding structure on population genetic parameters in *Drosophila*. *Genetics* 166:779–788.
36. Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA (2013) Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *arXiv:* 1303.5044.
37. Bastide H et al. (2013) A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. *PLoS Genet* 9:e1003534.
38. Tobler R et al. (2014) Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Mol Biol Evol* 31:364–375.
39. Kopelman NM, Stone L, Gascuel O (2013) The behavior of admixed populations in neighbor-joining inference of population trees. *Pacific Symposium on Biocomputing*:273–284.
40. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461:489–494.
41. Patterson N et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093.
42. Whitlock MC, Lotterhos KE OutFLANK. Available at: https://github.com/whitlock/OutFLANK.
43. Corbett-Detig RB, Hartl DL (2012) Population genomics of inversion polymorphisms in *Drosophila melanogaster*. *PLoS Genet* 8:e1003056.
44. Lawrie DS, Messer PW, Hershberg R, Petrov DA (2013) Strong purifying delection at synonymous sites in *D. melanogaster*. *PLoS Genet* 9:e1003527.
45. Lavington E et al. (2014) A small system—high-resolution study of metabolic adaptation in the central metabolic pathway to temperate climates in *Drosophila melanogaster*. *Mol Biol Evol* 31:2032–2041.
46. Gockel J, Kennington WJ, Hoffmann A, Goldstein DB, Partridge L (2001) Nonclinality of molecular variation implicates selection in maintaining a morphological cline of *Drosophila melanogaster*. *Genetics* 158:319–323.
47. Bach J (1976) *A Maritime History of Australia* (Thomas Nelson Limited, Melbourne).
48. Agosta SJ, Klemens JA (2008) Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. *Ecology Letters* 11:1123–1134.
49. Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787–792.
50. Yamazaki T, Choo J-K, Watanabe TK, Takahata N (1986) Gene flow in natural populations of *Drosophila melanogaster* with special reference to lethal allelism rates and protein variation. *Genetics* 113:73–89.
51. Singh RS, Rhomberg LR (1987) A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. *Genetics* 115:313–322.

52. Coyne JA, Milstead B (1987) Long-distance migration of *Drosophila*. 3. Dispersal of *D. melanogaster* alleles from a Maryland orchard. *American Naturalist* 130:70–82.
53. Powell J (1997) *Progress and Prospects in Evolutionary Biology : The Drosophila Model* (Oxford University Press).
54. McInnis DO, Schaffer HE, Mettler LE (1982) Field dispersal and population sizes of native *Drosophila* from North Carolina. *American Naturalist*.
55. McKenzie JA (1980) An ecological study of the alcohol dehydrogenase (*Adh*) polymorphism of *Drosophila melanogaster*. *Australian Journal of Zoology* 28:709–716.
56. Whitlock MC (1992) Temporal fluctuations in demographic parameters and the genetic variance among populations. *Evolution* 46:608–615.
57. Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: FST≠1/(4Nm+1). *Heredity* 82:117–125.
58. Wright S, Dobzhansky T (1946) Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. *Genetics* 31:125.
59. Mackay TFC et al. (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482:173–178.
60. Andolfatto P, Przeworski M (2000) A genome-wide departure from the standard neutral model in natural populations of *Drosophila*. *Genetics* 156:257–268.
61. Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005) Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res* 15:790–799.
62. Langley CH et al. (2012) Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* 192:genetics.112.142018–598.
63. Andolfatto P, Wall JD (2003) Linkage disequilibrium patterns across a recombination gradient in African *Drosophila melanogaster*. *Genetics*.
64. Coyne JA, Beecham E (1987) Heritability of two morphological characters within and among natural populations of *Drosophila melanogaster*. *Genetics*.
65. Kennington WJ, Gockel J, Partridge L (2003) Testing for asymmetrical gene flow in a *Drosophila melanogaster* body-size cline. *Genetics* 165:667–673.
66. Bhan V, Parkash R, Aggarwal D (2014) Effects of body-size variation on flight-related traits in latitudinal populations of *Drosophila melanogaster*. *Journal of genetics*.
67. Pitchers W, Pool JE, Dworkin I (2013) Altitudinal clinal variation in wing size and shape in African *Drosophila melanogaster*: one cline or many? *Evolution* 67:438–452.
68. Klepsatel P, Gáliková M, Huber CD, Flatt T (2014) Similarities and differences in altitudinal versus latitudinal variation for morphological traits in *Drosophila melanogaster*. *Evolution* 68:1385–1398.
69. Blanckenhorn WU, Demont M (2004) Bergmann and converse Bergmann latitudinal clines in arthropods: two ends of a continuum? *Integrative and Comparative Biology* 44:413–424.
70. Shelomi M (2012) Where are we now? Bergmann's rule *sensu lato* in insects. *The American Naturalist* 180:511–519.

71. Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* 304:412–417.
72. Wright S (1943) Isolation by distance. *Genetics* 28:114–1398.
73. Karasov T, Messer PW, Petrov DA (2010) Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet* 6:e1000924.
74. Nei M (1972) Genetic distance between populations. *American Naturalist* 106:283–292.
75. R Core Team R: A Language and Environment for Statistical Computing. Available at: http://www.R-project.org/.
76. Paradis E, Claude J, Strimmer K APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
77. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8:e1002967.
78. Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175–195.
79. Robertson A (1975) Remarks on the Lewontin-Krakauer test. *Genetics* 80:396.
80. Comeron JM, Ratnappan R, Bailin S (2012) The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet* 8:e1002905.

Figure legends.

Figure 1. (A) Map of collection locales (squares) and proposed colonization routes of *D. melanogaster* (arrows). Colors of the arrows indicate that populations locally adapted to either tropical environments (red) colonized similarly tropical locales, whereas populations locally adapted to temperate environments (blue) colonized similarly temperate locales. (B) Neighbor-joining tree of sampled populations. (C) Proportion of African ancestry among sampled populations. Note, the proportion of European ancestry is equal to one minus the proportion of African ancestry. The red (blue) point represents the proportion African ancestry of the Pennsylvanian samples collected in the fall (spring).

Figure 2. Patterns of co-differentiation and parallelism between North American, Australian, and Old-world populations. (A) $\log_2$ odds-ratio that SNPs fall above the F_{ST} quantile cut-off (x-axis) in both sets of populations (NA: North America; AUS: Australia; OW: Old-World). (B) Proportion of SNPs that vary in a parallel way given that they fall above the F_{ST} quantile cut-off in both sets of populations. Confidence bands represent 95% confidence intervals.

Figure 3. Patterns of (A) co-differentiation and (B) parallelism among various classes of SNPs relative to their matched controls. Vertical lines represent 95% confidence intervals. Horizontal dotted lines represents the null expectations. See *Materials and Methods* for details.

Table 1. **Formal admixture analysis**

	Population	f_3 †
N. America	Maine	**-5.43*****
	Pennsylvania	**-16.33*****
	North Carolina	**-4.45*****
	South Carolina	-3.10***
	Georgia	**-17.24*****
	Florida	**-24.00*****
Aus.	Temperate	8.40^{NS}
	Tropical	0.58 ^{NS}

† f_3 values calculated between listed population and allele frequency estimates made from the Italian and Congolese population. *** indicates p-value < 0.001. Bold indicates significance at $p < 0.05$ after Bonferroni correction.

Figure 1.

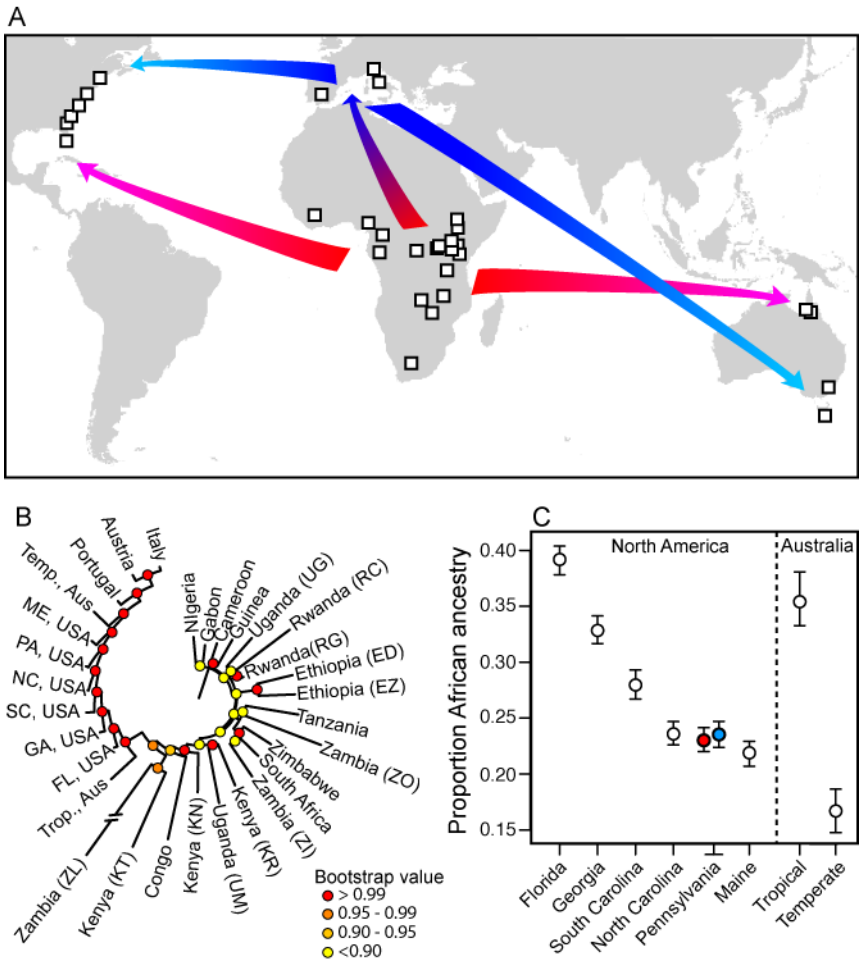

A
B
Italy
Austria
Portugal
Temp., Aus
ME, USA
PA, USA
NC, USA
SC, USA
GA, USA
FL, USA
Trop., Aus
Zambia (ZL)
Kenya (KT)
Congo
Kenya (KN)
Uganda (UM)
Kenya (KR)
Zambia (ZI)
South Africa
Zimbabwe
Zambia (ZO)
Tanzania
Ethiopia (EZ)
Ethiopia (ED)
Rwanda(RG)
Rwanda (RC)
Uganda (UG)
Guinea
Cameroon
Gabon
Nigeria
Bootstrap value
> 0.99
0.95 - 0.99
0.90 - 0.95
<0.90
C
North America
Australia
Proportion African ancestry
0.40
0.35
0.30
0.25
0.20
0.15
Florida
Georgia
South Carolina
North Carolina
Pennsylvania
Maine
Tropical
Temperate

Figure 2.

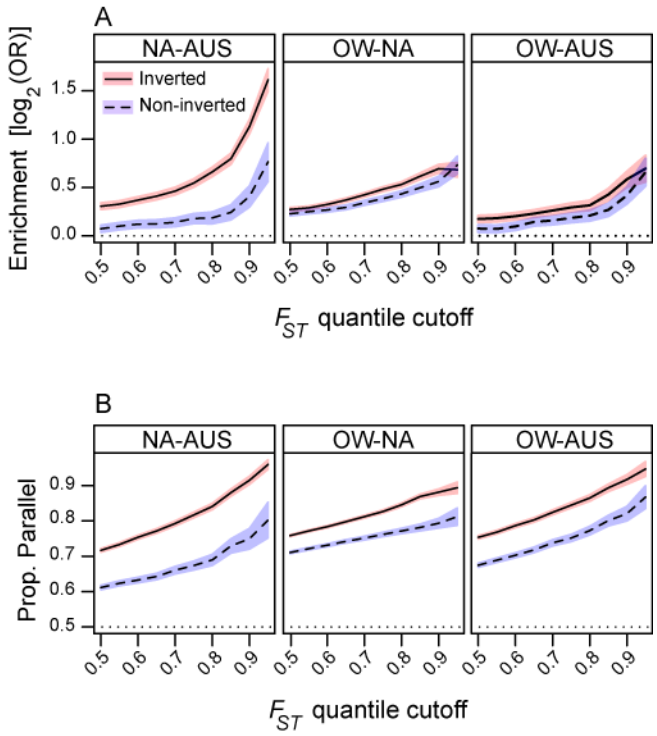
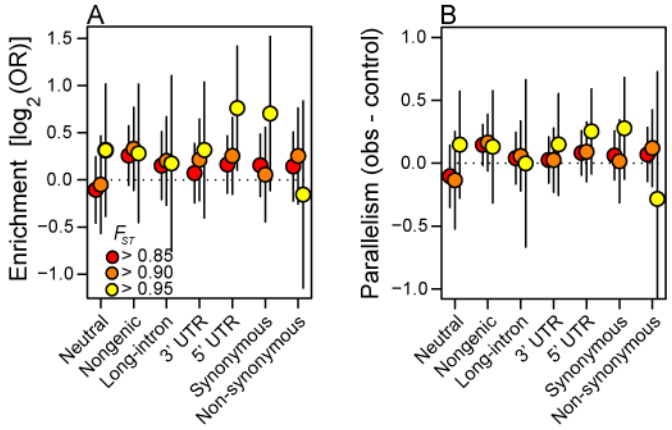

A
NA-AUS
OW-NA
OW-AUS
Inverted
Non-inverted
Enrichment [log2(OR)]
1.5
1.0
0.5
0.0
0.5
0.6
0.7
0.8
0.9
FST quantile cutoff
B
NA-AUS
OW-NA
OW-AUS
Prop. Parallel
0.9
0.8
0.7
0.6
0.5
0.5
0.6
0.7
0.8
0.9
FST quantile cutoff

Figure 3.


A
B
Enrichment [log2(OR)]
Parallelism (obs - control)
1.5
1.0
0.5
0.0
−0.5
−1.0
FST
> 0.85
> 0.90
> 0.95
Neutral
Nongenic
Long-intron
3' UTR
5' UTR
Synonymous
Non-synonymous