doi:10.1111/evo.12291

# THE INTENSITY OF SELECTION ACTING ON THE *COUCH POTATO* GENE—SPATIAL–TEMPORAL VARIATION IN A DIAPAUSE CLINE

**Rodrigo Cogni,[1,2] Caitlin Kuczynski,[1] Spencer Koury,[1] Erik Lavington,[1] Emily L. Behrman,[3] Katherine R. O'Brien,[3] Paul S. Schmidt,[3] and Walter F. Eanes[1,4]**

[1]*Department of Ecology and Evolution, Stony Brook University, Stony Brook, New York*

[2]*Current address: Department of Genetics, University of Cambridge, Cambridge, United Kingdom*

[3]*Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania*

[4]*E-mail: walter.eanes@stonybrook.edu*

Received June 28, 2013
Accepted September 26, 2013

**Cosmopolitan populations of *Drosophila melanogaster* have co-opted a form of reproductive diapause to overwinter in northern populations. Polymorphism in the couch potato gene has been implicated in genetic variation for this diapause trait. Using a collection of 20 populations from Florida to Canada and 11 collections from 3 years in a Pennsylvania orchard, we estimated the allele frequencies for 15 single nucleotide polymorphisms (SNPs) in the couch potato gene. These include the specific polymorphism associated with diapause inducability. We find that the SNP polymorphism, 48034(A/T), is correlated with latitude and its frequencies are predicted by the incidence of diapause trait. We find that the clinal patterns for cpo SNPs sampled in 1997 are similar to the same SNPs sampled in 2009–2010. SNPs that show apparent associations with cpo expression are also clinal with the low-expression allele increasing in frequency, as would be predicted from functional knockout studies of cpo. Finally, we see a significant pattern where the frequency of the diapause-causing allele drops in frequency during the summer season, consistent with the drop in the incidence of the diapause trait. The selection required to drive this response is large, roughly 24% to 59% per generation depending on the degree of dominance.**

**KEY WORDS:** **Adaptation, gene flow, life-history evolution, physiology, population genetics, population structure.**

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

An important strategy used by organisms to temporarily overcome stressful and unfavorable environments is the physiological reallocation of energy to different life-history components (Zera and Harshman 2001). In many insects this is accomplished by the induction of a diapause where reproduction is temporarily terminated and activity suspended (Denlinger 2002). The genetics of diapause-like phenotypes in insects are complex and likely to involve different pathways that are idiosyncratic to the species (Emerson et al. 2009a). The control of diapause is likely to involve a cascade of events from environmental sensing, neuroendocrine signaling, and downstream physiology. Moreover, the association of diapause with resource reallocation from reproduction to somatic maintenance and life span extension make it important to a broader understanding of the genetic basis of aging, particularly as diapausing insects appear to arrest senescence (Tatar et al. 2001).

Aside from its obvious genetic and experimental advantages, *Drosophila melanogaster* is uniquely placed to understand the genetic architecture of diapause-like traits in an adaptive context. Its association with humans has required adaptation to temperate climates outside its original range in the African sub-Saharan tropics (David and Capy 1988). Recent estimates of an out-of-Africa colonization put the European and African divergence at about 16,000 years ago and east Asian separation from Europe at about 2500 years ago (Li and Stephan 2006; Thornton and Andolfatto 2006; Laurent et al. 2011). These ages are consistent

with favorable climatic conditions, the introduction and spread of agriculture and the increased concentration of human populations along with fermentation products and their storage (Barnard et al. 2011). This expansion is in contrast with populations in the New World and Australia that are founded much more recently; generally considered to have first appeared in mid-1800s or about 150 years ago in North America (Keller 2007). The ecology and natural history of *D. melanogaster* in temperate regions is central to understanding its recent adaptations (Reaume and Sokolowski 2006). In temperate regions, the *D. melanogaster* population is envisioned as a seasonal meta-population, where populations die back during the winter followed by local reestablishment of populations in the spring (Ives 1945, 1954; Shpak et al. 2010). In the spring, local populations expand along with increased interpopulation dispersal until the late fall when they once again collapse. This proposes a pattern where selection pressures vary geographically with climate in a clinal fashion and also vary seasonally in local populations.

*Drosophila melanogaster* is capable of a form of reproductive diapause (or dormancy) that has photoperiod and temperature dependence (Saunders 1987; Saunders et al. 1989; Emerson et al. 2009b). Early studies demonstrated that variation in the induction of this diapause has a significant genetic component and appears geographically variable (Williams and Sokolowski 1993). Schmidt et al. (2005) showed that southern populations in the United States are about 30% inducible to diapause, whereas females in northern populations can nearly all diapause; diapause incidence was strongly clinal. Furthermore, the diapause trait shows a pattern of seasonal change where it declines in incidence as the season progresses into the fall (Schmidt and Conde 2006). The diapause trait has many correlated associations with other fitness-related traits and the results of population cage experiments have shown that selection on lines possessing the potential to diapause can be quite strong (Schmidt and Conde 2006; Schmidt and Paaby 2008;).

Schmidt et al. (2005) showed that the genetic control of diapause was almost entirely determined by one or more factors on the third chromosome. Using a combination of QTL mapping, linkage disequilibrium association, and genetic knockout, Schmidt et al. (2008) identified allelic variation in the nearly 80 kb *couch potato* (*cpo*) gene as a major factor in controlling diapause and moreover identified a putative amino acid polymorphism (*I462K*) in exon 5 that appears primarily responsible for the variation of diapause in the Davis Peach Farm (DPF) population on Long Island, NY. Three other single nucleotide changes in the 3' end of exon 5, all in strong linkage disequilibrium with the *I462K* polymorphism, also showed significant association with diapause expression. The diapause-associated alleles in the *cpo* gene also increased in frequency with latitude.

Schmidt et al. (2008) proposed the *I462K* change to lie in the 3' end of exon 5 as defined by a splicing variant RH (see www.Flybase.org) that had been based on experimental and bioinformatic assessment in early investigations (Bellen et al. 1992). Alternative analysis in the modENCODE Project (Graveley et al. 2011) does not support the RH splice variant. That reassessment removes the apparent *I462K* change and replaces it as a intron SNP just outside exon 5, 38 bases into intron 6, in tight linkage disequilibrium in the reported amino acid polymorphisms in exon 5. This SNP possesses the highest association with diapause expression in the DPF lines and although the splicing remains equivocal, is henceforth referenced as SNP *48034(A/T)* using the start of exon 1 as a reference point.

In this report, we are interested in four questions. (1) How well is a cline in the diapause-associated SNP *48034(A/T)* predicted by the cline in diapause trait expression across the same range? (2) Are the clines in *cpo* SNPs repeatable on decadal scales? More specifically, are the clines discovered in 1997 collections found in collections made 12–13 years later? (3) Are SNPs with *cis*-acting *cpo* expression association clinal and in a fashion predicted by their direction of expression and its effect on diapause? (4) Given that diapause trait incidence cycles seasonally, does the *48034(A/T)* SNP show a predicted pattern of seasonal change and if so, can we estimate the intensity of selection required to affect this change across a season?

## Materials and Methods

### STOCKS AND COLLECTIONS

In 2009 and 2010, we collected from 18 populations across the eastern United States (Table 1). In one location, Linvilla Orchards, Media (PA), we sampled monthly from late spring to early fall (July 2009 to June 2011; Table 2). For the clinal study, the allele frequency at the Media (PA) population is presented as the average among all 11 seasonal collections. We also used samples from the *Drosophila* Genetic Reference Panel (DGRP), a population from Raleigh (NC) collected in 2005 for which there are whole genomic sequences (Mackay et al. 2012); we obtained 34 DGRP inbred lines from the Bloomington Stock Center and two more samples in 2007 and 2008 from Sudbury, ON, from Thomas Merritt. In statistical and modeling of the data, we use the effective number of chromosomes recovered as our sample size. Females collected were separated as soon possible in the field and upon return to the lab placed in individual vials and allowed to lay eggs. In each isofemale line, the progeny from the $F_1$ generation were preserved in EtOH for bulk DNA preparation. Two female progeny were collected from each preserved line and pooled with the collected males in the bulk sample for pyrosequencing. By sampling two progeny per line in the $F_1$ generation, we are

**Table 1.** Sample localities, collection dates, latitudes, and effective numbers of chromosomes sampled in the study.

| Locality | Month of collection | Latitude | Effective chromosomes[1] |
|---|---|---|---|
| Sudbury, ON[2] | Aug 2008–2009 | 46.40 | 108 |
| Bowdoin, ME | Oct 2009 | 44.03 | 172 |
| Shoreham, VT | Sep 2009 | 43.89 | 200 |
| Marion, NY | Sep 2009 | 43.16 | 184 |
| Newark, NY | Sep 2009 | 43.05 | 150 |
| Putney, VT | Oct 2009 | 43.00 | 200 |
| Harvard, MA | Sep 2009 | 42.50 | 192 |
| Middlefield, CT | Sep 2009 | 41.52 | 132 |
| Riverhead, NY | Aug 2009 | 40.92 | 54 |
| Princeton, NJ | Jun 2009 | 40.35 | 66 |
| Media, PA | Jun–Nov 2009–2011 | 39.92 | 113.7[4] |
| Churchville, MD | Oct 2010 | 39.56 | 280 |
| Charlottesville, VA | Sep 2010 | 38.03 | 206 |
| Raleigh, NC[3] | Aug 2005 | 35.77 | 37 |
| Smithfield, NC | Jul 2010 | 35.51 | 63 |
| Eutawville, SC | Jun 2010 | 33.39 | 74 |
| Hahira, GA | Nov 2010 | 30.99 | 54 |
| Jacksonville, FL | Oct 2010 | 30.32 | 110 |
| Venus, FL | Aug 2010 | 27.07 | 42 |
| Homestead, FL | Jul 2010 | 25.47 | 84 |

[1]See text for definition of number of effective chromosomes.
[2]Pooled data 2008, 2009.
[3]Inbred lines from the Drosophila Genetic Reference Panel.
[4]Average of 11 collections.

sampling 2–4 independent chromosomes from the population per line, with an average of three. The expected number of autosomes per sample is therefore three times the number of female lines plus twice the number of wild-collected males.

## PYROSEQUENCING

We used pyrosequencing in the bulk DNA preps to estimate SNP frequency (Doostzadeh et al. 2008; Lavebratt and Sengul 2006). Fifteen polymorphic sites, all in exons in *cpo*-RV transcript, were screened by bulk pyrosequencing (Fig. S1). These sites were reported in Zhu (2009) and also identified in the DGRP sequences. Except for exon 5, numbering starts at position 3R:13,754,593 in Release 5.52 from FlyBase. Exon 5 numbering is with respect to the first base in the first codon. These SNPs represent only the intermediate frequency variable sites in the small untranslated exons 1 and 3, as well as coding exon 5, and again the small coding exons 6–12. The two small amino acid repeat regions in exon 5 could not be screened. Bulk DNA purification was performed with Puregene Core Kit A (Qiagen) using 42 to 100 flies per preparation. We checked the reliability of the method by comparing the estimated frequency of each SNP by pyrosequencing to the expected frequency for the Raleigh population based on the DGRP genome sequences. The correlation is very high ($r = 0.99$).

## STATISTICAL ANALYSES

Allele frequency estimates from the pyrosequencing were arcsine transformed (Sokal and Rohlf 1981) and tested by linear regression against latitude or time of collection (days post May 14 or

**Table 2.** Samples used in the seasonal collections from Linvilla Orchard, Media, PA.

| Collection date | Effective number of chromosomes |
|---|---|
| Jul-09 | 116 |
| Aug-09 | 27 |
| Oct-09 | 57 |
| Nov-09 | 148 |
| Jun-10 | 27 |
| Jul-10 | 200 |
| Aug-10 | 120 |
| Sep-10 | 114 |
| Oct-10 | 68 |
| Nov-10 | 68 |
| Jun-11 | 140 |

month of collection). Individual probabilities for single tests are determined by random permutation of latitudes or collection date.

### EXPRESSION VARIATION

Expression variation was tested for each of the 15 individual SNPs using a nested analysis of variance (ANOVA) of the Affymetrix *Drosophila* 2.0 array expression data reported in Ayroles et al. (2009) and the original 37 sequences of the DGRP first released in 2009 (Mackay et al. 2012). The probe set 1624608_s_at was used to screen all 12 proposed transcripts. All the expression data were first masked to remove probes that overlapped any polymorphisms in the gene and this resulted in seven useful probes (Benovoy et al. 2008; Chen et al. 2009). Our model first removes sex-specific (fixed) effects and then the ANOVA is carried out on those residuals to estimate a SNP allele effect, a nested line in SNP effect, and nested vial within line effects ($Y_{ijkl} = \mu + A_i + B_{ij} + C_{ijk} + \varepsilon_{ijkl}$, where $Y_{ijkl}$ is an individual expression measure, $A_i$ is the fixed effect of the $i$th allele, $B_{ij}$ is the random effect of line $j$ within the $i$th SNP allele, and $C_{ijk}$ is the random effect of $k$th vial within line $j$ and the $i$ SNP allele and $\varepsilon_{ijkl}$ is the error term within vials). We ignore inter-SNP correlations caused by linkage disequilibrium among sites. However, it is the effect of marginal SNP expression across all backgrounds that will determine the response on SNPs if *cpo* expression is a target of selection on a contemporary time scale.

### ESTIMATION OF SELECTION ACROSS SEASON

We possess point allele frequencies and assume genotype frequencies in the first generation are in Hardy–Weinberg equilibrium. Using the elementary population genetics of deterministic selection, we start with frequencies $p^2$, $2pq$, and $q^2$ for the *48034(T/T)*, *48034(A/T)*, and *48034(A/A)* genotypes, respectively. Although diapause expression has dominance involved, we can make no assumption about the genotypic fitnesses, and assume fitnesses of 1, $1 + hs$, and $1 + s$. where $s$ is the selection coefficient and $h$ the degree of dominance (Crow and Kimura 1970). Most selection in this frequency range is expected to generate a near linear response in allele frequency change, irrespective of degree of dominance, $h$.

To obtain an estimate of $s$, we carry out an simple approach similar to Approximate Bayesian Computation (Beaumont et al. 2002; Csilléry et al. 2010). Our *summary statistic* of selection response is the slope of the transformed allele frequency change across generations (assumed here as collection month), $b$. In each of 10,000 simulations, the selection parameter, $s_i$ is randomly sampled from a uniform distribution with defined limits $-1$ to $+1$.

The genotype frequencies and corresponding allele frequencies for successive six generations are simulated from the deterministic selection trajectory associated with the sampled $s_i$. At each generation, we then resample point frequencies from the simulated trajectory assuming binomial sampling (using the empirical sample sizes) around each simulated point. A linear regression of arcsine transformed sampled frequencies *against month* (generation) of collection is carried out and the slope, $b_i$ determined for the sampled selection coefficient, $s_i$. This is carried out for $i = 10{,}000$ replicates and the biplot distribution of $b_i$ slopes for $s_i$ estimated. The marginal distribution of $s_i$ provides a likelihood distribution from which a maximum estimate and confidence limits of $s$ can be evaluated by weighed local-linear regression (Beaumont et al. 2002) where the rejection limits of the "accepted" slopes in the regression are based on the 97.5% confidence intervals of the empirically observed slope of the true data against *month of collection* (Fig. S2). We do this for fixed values of $h = 0$, 0.5, and 1.0.

## Results

### CONTEMPORARY *cpo* CLINES IN 2009–2010

Of the 15 SNP alleles in exons 1, 3, 5, 6, and 12, and intron 6 (Table S1), nine showed significant clines with latitude at $\alpha = 5\%$ under permutation (Table 3). These tests are not independent because of linkage disequilibrium among the closely linked SNPs within each exon, but the point frequency estimates from bulk pyrosequencing does not allow for permutation testing that carries over the correlation structure imposed by the disequilibrium. If we use a conservative Bonferroni correction ($\alpha = 0.0033$), then we still have 8 SNPs that are significant. Of the three SNPs screened in exon 1 (5′ UTR), SNP 3493 showed a significant cline ($r = 0.78$, $P < 3 \times 10^{-4}$). The single SNP 15311 in exon 3 (also a 5′ UTR) did not vary significantly with latitude. Of the six SNPs screened in exon 5, four are silent (540, 804, 849, 1134) and two (693, and 1040) involve amino acid changes. Three SNPs (540, 804, 1040) show significant latitudinal clines. The *48034(A/T)* polymorphism lying (35 bp) just outside exon 5 shows a strong cline ($r = 0.89$, $P = 0$ in 100,000 permutations) as does the *A347V* polymorphism at 1040 ($r = 0.90$, $P = 0$ in 100,000 permutations). We screened four more SNPs in exons 6–12 (78104, 78163, 79080, 82403), which span 4.3 kb and are separated from exon 5 by greater than 40 kb. Three of these SNPs (78104, 78163, 82403) show significant associations with latitude. SNPs 78104 and 78163 are only 59 bases apart and are in significant LD (78,104, $r = 0.65$, $P < 0.0018$; 78,163, $r = 0.76$, $P < 8.62 \times 10^{-5}$), but, 82,403 ($r = 0.72$, $P < 3 \times 10^{-4}$) is not in linkage disequilibrium with latter two.

The *48034(A)* SNP allele has the strongest diapause association, yet it is not possible to screen this SNP using restriction digests (Schmidt et al. 2008). We have now screened this SNP using bulk pyrosequencing. In Figure 1, we have superimposed the observed frequencies of both the *48034(/T)* (nondiapausing) and $Ala^{347}$ alleles in our collections against frequencies predicted

**Table 3.** Tests of clines with latitude and *cis*-expression effects of 15 SNPs in the *cpo* gene.

| | | Cline | | Expression | |
|---|---|---|---|---|---|
| SNP | Position[1] | $r^2$ | *P*-value[3] | $F_{1,37}$[4] | *P*-value |
| Exon 1 | | | | | |
| *3247* | 3,247 | 0.41 | 0.06465 | 9.1403 | *0.0034*[6] |
| *3458* | 3,458 | 0.34 | 0.12341 | 0.0232 | 0.8794 |
| *3493* | 3,493 | 0.78 | *0.00002* | 1.2037 | 0.2761 |
| Exon 3 | | | | | |
| *15311* | 15,311 | 0.15 | 0.55344 | 7.4368 | *0.0080* |
| Exon 5 | | | | | |
| *540* | 37,246 | 0.61 | *0.00453* | 18.5132 | *0.0001* |
| *693(H/Q)* | 37,399 | 0.35 | 0.13729 | 1.4619 | 0.2305 |
| *804* | 37,510 | 0.84 | *0.00001* | 2.2440 | 0.1384 |
| *849* | 37,552 | 0.42 | 0.08533 | 1.6698 | 0.2003 |
| *1040(A347V)* | 37,743 | 0.90 | *0.00000* | 0.7053 | 0.7845 |
| *1134* | 37,837 | 0.25 | 0.29434 | 0.2472 | 0.6205 |
| Intron 6 | | | | | |
| *48034(A/T)*[5] | 38,034 | 0.89 | *0.00000* | 0.4259 | 0.5160 |
| Exons 6,12 | | | | | |
| *78104* | 78,104 | 0.65 | *0.00019* | 0.0781 | 0.7807 |
| *78163* | 78,163 | 0.76 | *0.00020* | 4.5722 | *0.0358* |
| *79080* | 79,080 | 0.47 | *0.03326* | 0.4667 | 0.4966 |
| *82403* | 82,403 | 0.72 | *0.00031* | 0.0371 | 0.8479 |

[1]Position relative to start of *cpo* transcript RV.
[2]Product–moment correlation.
[3]Probability in 100,000 permutations.
[4]*F*-value for SNP effect in nested ANOVA (see text).
[5]Formerly *I463K* in Schmidt et al. (2008).
[6]Significant values in italics.

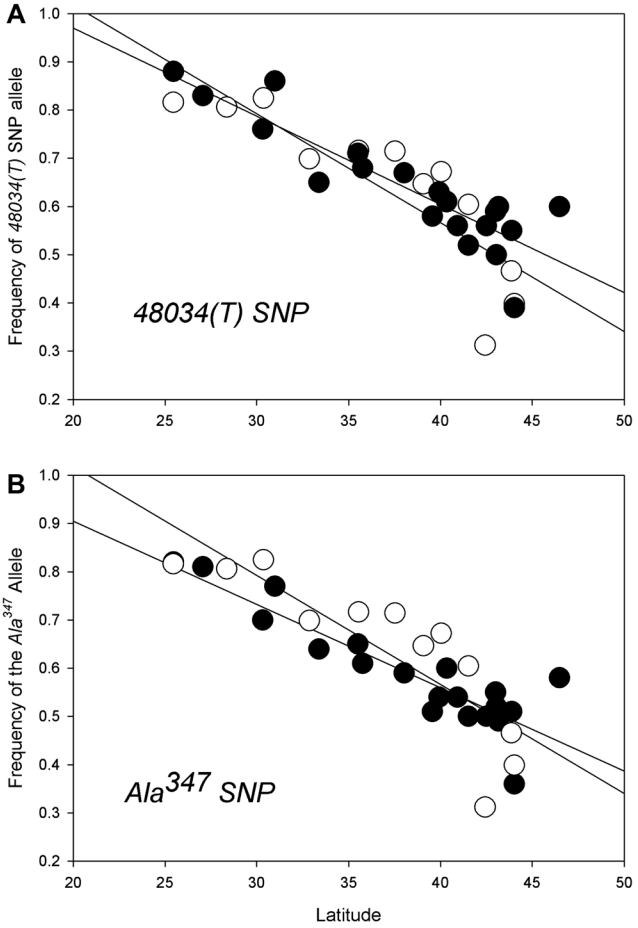


**Figure 1.** Clines for the nondiapausing *48034(T)* and *A347V* polymorphisms. (A) Frequency of the nondiapausing *48034(T)* allele plotted against latitude (black fill; intercept = 1.33 ± 0.086, slope = −0.0183 ± 0.0022; $F_{1,18} = 71.45$, $P < 0.0001$). (B) Frequency of the nondiapausing *A347V* in exon 5 plotted against latitude (black fill; intercept = 1.25 ± 0.078, slope = −0.0173 ± 0.002; $F_{1,18} = 71.45$, $P < 0.0001$). Predicted frequencies of the *48034(T)* and $Ala^{347}$ alleles (1.47 ± 0.17, slope = −0.0226 ± 0.0045; both no fill) determined from the incidence of diapause expression (square root of nondiapause frequency) in isofemale lines collected in 1997.

by the incidence of the diapause trait observed from the isofemale lines sampled in 1997 (Schmidt et al. 2005). We believe that diapause expression is associated with autosomal dominance for both the *48034(T)* and $Ala^{347}$ alleles (Schmidt et al. 2005), so that the frequency of each allele should be predicted by the incidence of diapause expression (i.e., the square root of nondiapause incidence assuming Hardy–Weinberg equilibrium frequencies). There is no significant difference in the slopes or intercepts of the linear regression lines (Fig. 1, legend) showing remarkably strong concordance between the predictions from the diapause expression cline from 1997 collections and the observed clines in the *48034(T)* and $Ala^{347}$ polymorphisms sampled in 2009 and 2010.

## REPEATABILITY OF CLINES FROM 1997

We were able to rescreen four of the original SNPs, as well as *48034A/T*, using our bulk pyrosequencing method. We compared these frequencies to those from 1997 (Fig. 2; Schmidt et al. 2008). Our repeated collections represent populations that span 12–13 years. Assuming a generation time of 3 weeks in the field and suitable growth seasons of 6 months in the north and 12 months in the south, we estimate the passage of 100 generations in the north to 216 generations in the south. We see the clines are effectively unchanged 12–13 years later. None of the quadratic parameters are significantly different between collections (Table S2).

## *cpo* EXPRESSION ASSOCIATIONS AND CLINES

Our knockout experiments of the *cpo* gene showed that lower expression of *cpo* increased expression of diapause (Schmidt et al. 2008). A prediction arising from this observation is that other SNP alleles associated with lower expression of *cpo* might also increase with latitude. Overall, four of 15 SNPs (540, 3247, 15311, and 78163) have significant associations with *cpo* expression in

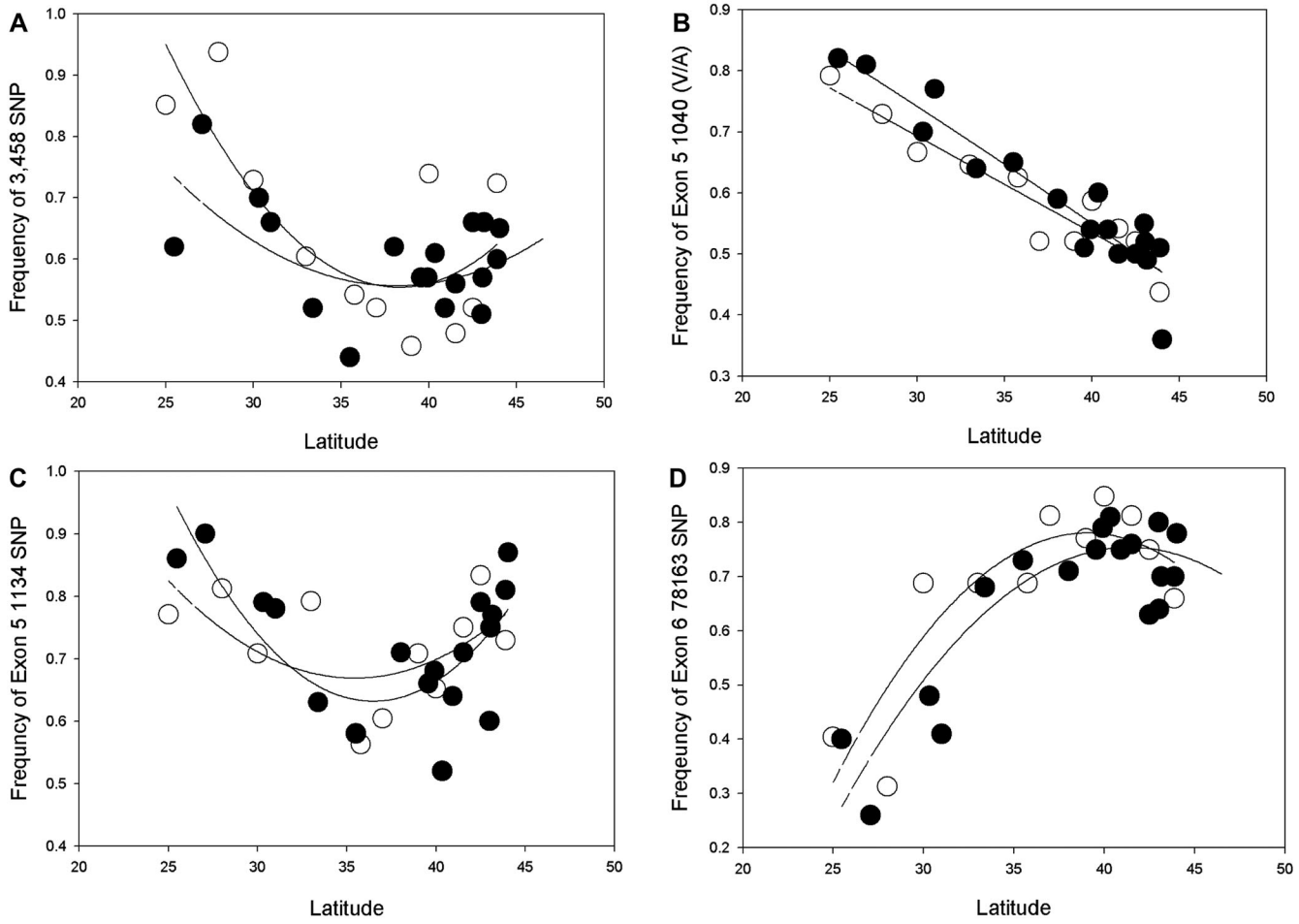


**Figure 2. Repeatability of *cpo* SNP clines. Comparison of geographic patterns of allele frequency in the ten 1997 samples (Schmidt et al. 2008) (no fill) screened with restriction enzyme digests to the 18 (2009–2010) collections screened by bulk pyrosequencing (black). Quadratic regression fits are applied to simplify visual pattern inspection. None of the quadratic parameters are significantly different between collections (Table S2). (A) Exon 1 SNP 3458–*TspR1*. (B) Exon 5 SNP 1040–*Afe1*. (C) Exon 5 SNP 1134–*BsiE1*. (D) Exon 6 SNP 78163–*Dde1*.**

the DGRP lines (Fig. 3 and Table 3). Two of these SNPs possess clines (540, 78163) and in both cases the lowest expressing allele increases with latitude. None of the four SNPs lacking clines show any *cpo* expression variation. Thus, two of four SNPs identified as expression Quantitative Trait Nucleotides (eQTNs) show changes where alleles with decreased expression increase in northern populations.

## SEASONAL PATTERN

We possess 11 monthly collections from the late spring to late fall from Linvilla Orchards (Media, PA) from June 2009 to November 2010, and including June 2011 (Table 2; see also Table S3). In the New Hope, PA and Princeton, NJ, orchard populations screened in 2003 and 2004, the incidence of the diapause trait is highest early in the season (∼80%) and diminishes through the fall (to ∼40%) only to rebound the following spring (Schmidt et al. 2005). Given the pattern of autosomal dominance for the control of diapause expression by *cpo* (that explains the geographic change), we predict that the *48034(A)* allele should drop from about 55% in the spring to about 25% in the fall (Fig. 4). Among the eight SNPs in *cpo* with latitudinal clines, the silent sites 804 and 1040, the *48034(A/T)* intron polymorphism, and the exon 1 polymorphism at 3459 show significant associations with date of collection (days post May 14). The 804 SNP is in sufficient linkage disequilibrium to have hitchhiked along with the *48034(A)* allele. The *48034(A/T)* polymorphism shows the strongest correlation with collection date ($r = 0.693$, $P < 0.018$ by two-tailed permutation). As hypothesized, in each case the northern-biased diapause-associated allele starts out higher in frequency and decreases with the season. Figure 4 shows the linear relationship between *48034(A)* frequency (with 95% confidence limits; intercept $= 0.5113 \pm 0.054$, slope $= -0.0013 \pm 0.0004$) against sample collection day after May 14 of each year. The relationship remains significant even if the smaller June 2010 sample is

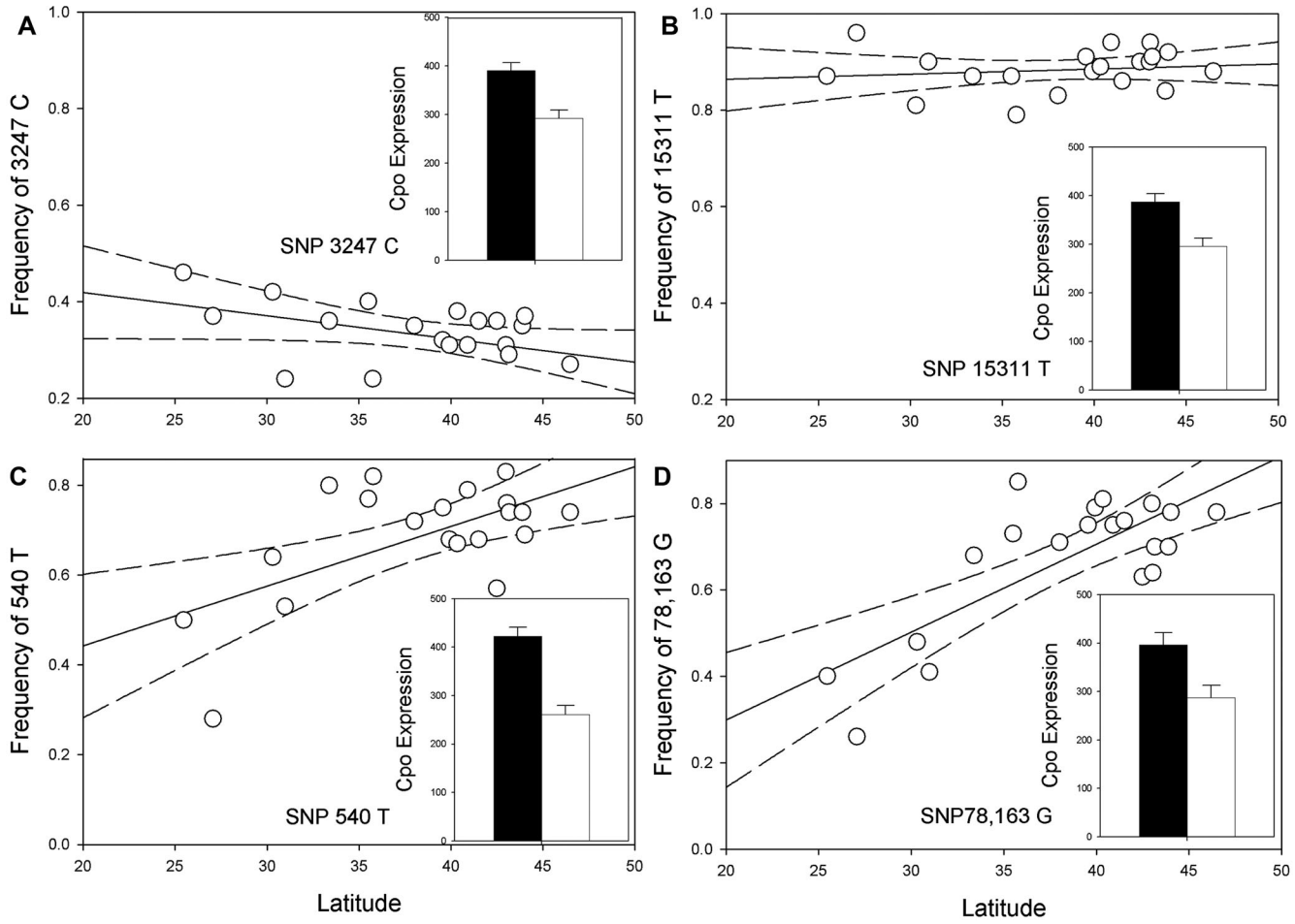


**Figure 3. Four SNPs with significant associations with line *cpo* transcript expression variation in the DGRP. Insert bar diagrams are expression levels for the SNPs and the frequency of the lower expression SNP is plotted against latitude (regressions shown). Clockwise (A) 3247 in exon 1, Expression $F_{1,35} = 9.14$, $P < 0.0034$. (B) SNP 15311 in exon 3, $F_{1,35} = 7.43$, $P < 0.008$ SNP. (C) SNP 540 in exon 5; $F_{1,35} = 18.51$, $P < 0.0001$. (D) SNP 78163 in exon 6; $F_{1,35} = 4.57$, $P < 0.036$.**

removed (bulked sample $n = 27$, $P < 0.028$) or the other smaller August sample ($n = 27$) is removed as well ($P < 0.0038$), albeit the slopes are more shallow. The change is also significant when regressed against month of collection (with 95% confidence limits; intercept $= 0.02812 \pm 0.1204$, slope $= -0.04061 \pm 0.0139$; $P < 0.0172$). The prediction from the regression is that the *48034(A)* frequency is ∼47% in June decreasing to ∼27% by November. This is similar to the *48034(A)* allele frequencies predicted from the change of the diapause trait in 2003–2004, which are 58% and 26%, respectively (the slopes and intercepts are nonsignificantly different between the two regressions). The lack of concordance might the attributable to the fact that both represent samples from near-by orchards and different years of collection.

## ESTIMATION OF SELECTION

We see changes in SNP frequencies consistent with natural selection favoring the nondiapause *48034(T)* allele during the summer expansion and the dipausing *48034(A)* allele during the winter population collapse. The change in allele frequency offers the unique opportunity to estimate the strength of selection acting over approximately six generations of selection in the summer. We can estimate the distribution of selection coefficients that would be required to generate the observed change in allele frequencies. These allele frequency changes are large and we assume that genetic drift plays a minor role (even during the winter bottleneck) throughout the seasonal cycle; the frequency change is deterministic. The major source of variation in the observed trajectories is therefore the sampling around the trajectory and we integrate that source of uncertainty into the estimation of the selection (see Materials and Methods).

In our simulations, our *summary statistic* of selection response is the slope of the allele frequency change across *collection months*, $b$, and assuming 1 month per generation. For each value of the selection parameter, $s_i$, sampled from a uniform distribution (from $-1$ to $+1$), allele frequencies for successive six

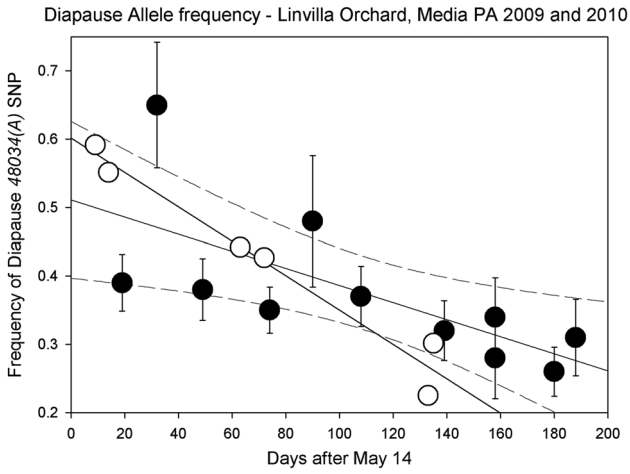


**Figure 4. Plot of *48034(A)* frequency against collection day in the Linvilla, Media, PA orchard (black fill). The expected frequency of the *48034(A)* allele predicted from the incidence of diapause in the seasonal isofemale collections from nearby orchards in 2003–2004 samples (no fill; intercept = 0.602 ± 0.0217, slope = −0.0025 ± 0.0002, $t = -10.7$, $P < 0.0005$; Schmidt and Conde 2006). The confidence bars are one standard error.**

generations are simulated from the deterministic selection trajectory. The degree of dominance is considered at $h = 0$, 0.5, and 1, and we assume it is constant through the season. We then resample point frequencies based on our sample sizes and assuming binomial sampling from the simulated trajectory, and calculate $b_i$, the slope of a linear regression of simulated allele frequencies per generation. The results of our simulations and estimation are shown in Figure S1.

For additive allele effects on fitness ($h = 0.5$), we estimate that the average per generation selection during the summer–fall response against the *48034(A/A)* homozygote is $s = -0.345$. In the case of complete dominance for fitness ($h = 1$) of the diapause allele, we excluded sampling values of $s_i > 0.75$ because the strong selection against the diapause allele results in very rapid and unrealistic fixation of the *48034(T)* allele (when $s = 1$ and $h = 1$, the *48034(A)* allele is a dominant lethal). In this case, we estimate $s = -0.241$. If the diapause allele is recessive with respect to fitness, even stronger selection is required, $s = -0.597$. These estimates have large confidence limits (Fig. S2), and $s$ values as low as $-0.12$ would appear to be sufficient to drive this observed response under $h = 1$, or the full dominance of the diapausing allele.

## Discussion

This study shows that, as a determinant of genetic variation in diapause expression, strong selection appears acting on *cpo* and the *48034(A/T)* polymorphism, both across a decade and on a broad geographic scale and moreover even within seasons in a local Pennsylvania orchard population. The frequencies of the SNP *48034(A/T)* predicted from the clinal incidence of the diapause trait in 1997 collections match well the observed frequencies in our 2009–2010 latitudinal clines.

The clines in the set of *cpo* SNPs first reported in Schmidt et al. (2008) are repeatable 12–13 years later (100–216 generations depending on location in the cline). In *Drosophila*, other than chromosomal inversions (Levitan 2003; Balanya et al. 2006), and *Adh* (Umina et al. 2005), there is little data on temporal changes in genes over similar time intervals (Bradshaw and Holzapfel 2006). This pattern is in contrast with *Adh*, for which the position of latitudinal clines has changed over the past 25 years (Umina et al. 2005). This difference between genes may simply be because of the shorter time scale here, a different genetic architecture of the phenotypes, the different effects of temperature versus other environmental parameters on gene function, and a host of other differences; unfortunately there is no genome-wide, longitudinal, and temporal context of variation in which to place the patterns observed at *Adh* and *cpo*. What is clear is that although latitudinal clines are often interpreted as adaptive responses to climate, not all clines appear to be shifting. Whether this is unique to *cpo* or typical of most variation in the eastern populations will require genome-wide estimates with sufficient depth to test the hypothesis (Fabian et al. 2012).

We also identified four SNP alleles among the 15 (one each in exon 1, 3, 5, and 6) that are eQTNs for *cpo* transcript levels. These SNPs span nearly the entire 75 kb of the *cpo* gene. Two of the eQTNs are nonclinal and two show significantly increasing frequency of the low-expression allele with latitude. This is consistent with our experimental observation using knockouts to reduced expression and favor diapause induction and suggests that the propensity to diapause may be associated with reduced expression of the *cpo* gene in natural populations. Studies on *D. montana* and *Culex pipiens* have also shown diapause-induced changes (increases) in *cpo* expression, albeit in a complex fashion, and these are not knockout studies, so causality is difficult to ascertain (Zhang and Denlinger 2011; Kankare et al. 2012).

The frequency of the diapause trait is highest in the spring followed by a reduction in the fall (Schmidt and Conde 2006). Given this we would have expected that the frequency of the *48034(A)* allele would follow that seasonal progression if this is a general feature of temperate populations, as those in eastern Pennsylvania. More specifically, from the change in mean diapause potential we proposed that the frequency of the *48034(A)* allele would drop from 55% to 25%. A similar frequency change is observed in our field samples taken across seasons in a single orchard (albeit a different locality from the earlier diapause study in

2003–2004) in Pennsylvania from 2009 to 2011. This strong shift is perhaps expected given the rapid responses of lines possessing the diapause trait to stress-imposed selection in laboratory population cages (Schmidt and Conde 2006). It might be postulated that this seasonal change is alone the cause of the latitudinal clines, because our latitudinal collections were made throughout the season from June to November and regional collections share common collection dates. However, when collection month is included as a covariate it has no significant effect on frequency and the latitudinal clines remain highly significant after its effect is removed. Furthermore, it should be noted that to track local abundance and facilitate sampling southern collections were made early in the season, whereas northern collections were made later (Table 1). This temporal sampling, associated with time of local abundance, would instead tend to weaken the latitudinal clines, especially because northern collections would have experienced postseason selection. Finally, although *cpo* lies inside the polymorphic and clinal inversion *In(3R)Payne* and this confounds interpretation of clines in Australia (Lee et al. 2011), we do not feel that this contributes to the seasonal pattern in the PA population. We cannot screen this inversion by pyrosequencing, but although the frequency of *In(3R)Payne* is clinal in the eastern United States, the inversion is uncommon ($< 5\%$) at this latitude (Sezgin et al. 2004; Fabian et al. 2012).

Moreover, we estimated the required selection coefficient associated with change in *48034(A/A)* genotype across the season. Not unexpectedly these large frequency changes, require selection coefficients that are large ($s = 0.241$–$0.59$) depending on the degree of dominance. Clearly, this spring–fall response is compensated for by favorable large selection for diapause over the winter. This a unique observation because there are few direct appraisals of selection on molecular polymorphism in the wild. Haldane (1924) first estimated selection intensity on the classic industrial melanism polymorphism in the peppered moth *Biston betularia*. Recent estimates from the long-term decline of melanistic alleles place a 15–19% disadvantage assuming a constant fitness model (Grant et al. 1996; Cook and Turner 2008). Mullen and Hoekstra (2008) estimated selection on a number of morph phenotypes based on migration–selection balance across a cline in *Peromyscus*. Selection coefficients varied widely across phenotypes and assumed models, but values as high as 21% were estimated.

In estimating selection from the change in allele frequency, we assumed that genetic drift plays little role in the change in frequency of the *48034(A)* allele. We believe that even with substantial seasonal reductions, the local population sizes are probably large. This belief is based on the effective sizes of local seasonal populations that are estimated from the rates of lethal allelism in *Drosophila* done in many populations. Lethal allelism is dependent on recent local population size and can be very high early in season and drop notably. The Amherst, MA, populations, studied for many years by Ives and Band (Band and Ives 1963; Ives and Band 1986), produced average early season estimates of *local* $N_e$ of $\sim$1400 that then increase to 4,700 by Oct–Nov. Season-wide summaries for the Raleigh, NC, population put local $N_e$ as greater than 10,000 (Mukai and Yamaguchi 1974). Thus, we assume the allele frequency trajectory here is best simply predicted by deterministic selection.

There are, however, alternative interpretations to the action of strictly strong selection in the spring–fall recovery of nondiapause genotypes. Although the shift in allele frequency must entail hard selection in the orchard with winter die back, we cannot reject the possibility that long-distance dispersal from southern populations or even migration out of local refugia restores some of the allele frequency during the spring-fall rebound. A more nuanced view of local structure would be one of a shifting local mosaic of selection with seasonal dilution of the diapause frequencies established during the winter in the orchard. Schmidt and Conde (2006) showed that although diapause incidence was high in rural Pennsylvania orchards ($\sim$80%) and showed seasonal shifts, it was notably lower ($\sim$25%) in an urban indoor fruit market in Philadelphia, approximately 60-km away. The incidence of diapause suggests a 17% frequency of the diapause allele in this "protected niche," which is close to the FL frequencies in the cline. The urban fruit market also showed a similar seasonal cycle in diapause potential, but much shallower. A monthly replacement rate of the outdoor population by about 33% from the indoor populations would result in the same trajectory of frequency response. It is possible that some of the recovery we see in the Linvilla Orchard is either long-distance migration from sources thousands of kilometers south, or from an extensive contribution from indoor-niche source populations. The fact that other *cpo* SNPs with clines do not show the same seasonal pattern does not support a simple long-distance contribution. Nevertheless, the dramatic increase in diapause incidence and the high *48034(A)* frequency at the start of the season clearly requires selection favoring winter survival of diapause genotypes in the orchard and this selection must at least be at the level operating in the summer, even if some of the apparent summer selection is attributable to migration from other sources. Otherwise the latitudinal cline should disappear.

This study emphasizes the importance of studying clinal variation in intersecting contexts. There are many reports of clines in the literature, but most of these are ends-unto-themselves, and not part of a total set of observations that support and consistently reinforce predictions from phenotypic associations, life-history variation, or functional characterizations. None of these studies have introduced a seasonal or temporal axis to the variation with its expectations about the action of natural selection. Aside from the observations on the seasonal change in diapause incidence

(Schmidt and Conde 2006), we might expect that if climatic selection is operating both geographically and locally in a seasonal fashion, as in *D. melanogaster* populations, then a genotype that tightly increases with latitude should also be locally more common in the early season and its frequency should diminish as the season progresses. This study bears out that expectation and is unique in that it tightens the connection between an important life-history trait and the apparent molecular polymorphism responsible for its variation.

Finally, the value of geographic information in studying natural selection in the genome is best introduced when coupled with ancillary evidence, as we demonstrate for the *cpo* polymorphism. This is because clines can be established by a number of phenomena unrelated to selection on the suspect gene. Both historical processes, as well as the problem of statistical replication and the coupling of migration, make population samples not independent (Vasemägi 2006). Other problems involve population admixture and separating linkage with other loci. Certainly, the role of admixture in clinal variation in *D. melanogaster* remains an open and important question (Yukilevich et al. 2010; Duchen et al. 2013). The detailed use of haplotype structure surrounding a polymorphism of interest was first introduced by Berry and Kreitman (1993) through the simultaneous study of many linked and unlinked polymorphism as a way to tease apart the targets of selection in clines—addressing the question of hitchhiking. The ability to screen multiple SNPs across a gene and region allow a dissection of this question on a regular basis in this era. In the case here, the *cpo* SNP and region with strongest mapping association with diapause also shows the strongest response to latitude, as well as the strongest seasonal change. With respect to the *couch potato* gene and its allelic variation there is certainly now ample evidence for its association with reproductive diapause and ample evidence for genetic variation in the gene responding to strong selection. The next step here will be deciphering the developmental or functional role of *cpo* and the genes or pathways it may be interacting with in driving the cascade of effects that lead to the numerous pleiotropic effects on fitness-related traits that are also under strong life-history selection.

## ACKNOWLEDGMENTS

The authors thank T. Merritt for supplying the lines from Sudbury, ON, and J. True and J. Lachance for additional collections from New York. This study was funded by Collaborative National Science Foundation grants DEB0543050 and DEB0921372 to WFE and DEB0542859 and DEB0921307 to PSS.

## LITERATURE CITED

Ayroles, J. F., M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman et al. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. Nat. Genet. 41:299–307.

Balanya, J., J. M. Oller, R. B. Huey, G. W. Gilchrist, and L. Serra. 2006. Global genetic change tracks global climate warming in *Drosophila subobscura*. Science 313:1773–1775.

Band, H. T., and P. T. Ives. 1963. Genetic structure of populations. I. On the nature of the genetic load in the South Amherst Population of *Drosophila melanogaster*. Evolution 17:198–215.

Barnard, H., A. N. Dooley, G. Areshian, B. Gasparyan, and K. F. Faull. 2011. Chemical evidence for wine production around 4000 BCE in the Late Chalcolithic Near Eastern highlands. J. Archaeol. Sci. 38:977–984.

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Bellen, H. J., H. Vaessin, E. Bier, A. Kolodkin, D. D'Evelyn et al. 1992. The Drosophila couch potato gene: an essential gene required for normal adult behavior. Genetics 131:365–375.

Benovoy, D., T. Kwan, and J. Majewski. 2008. Effect of polymorphisms within probe-target sequences on olignonucleotide microarray experiments. Nucleic Acids Res. 36:4417–4423.

Berry, A., and M. Kreitman. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. Genetics 134:869–893.

Bradshaw, W. E., and C. M. Holzapfel. 2006. Climate change–evolutionary response to rapid climate change. Science 312:1477–1478.

Chen, L., G. P. Page, T. Mehta, R. Feng, and X. Q. Cui. 2009. Single nucleotide polymorphisms affect both cis- and trans-eQTLs. Genomics 93:501–508.

Cook, L. M., and J. R. G. Turner. 2008. Decline of melanism in two British moths: spatial, temporal and inter-specific variation. Heredity 101:483–489.

Crow J. F., and M. Kimura. 1970. *Introduction to population genetics theory*. Burgess Publishing Co., Minneapolis, MN.

Csilléry, K., Blum, M. G., Gaggiotti, O. E. and O. Francois. 2010. Approximate Bayesian Computation (ABC) in practice. Trends Ecol. Evol. 25:410–418.

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

Denlinger, D. L. 2002. Regulation of diapause. Annu. Rev. Entomol. 47:93–122.

Doostzadeh, J., S. Shokralla, F. Absalan, R. Jalili, S. Mohandessi et al. 2008. High throughput automated allele frequency estimation by pyrosequencing. PLoS ONE. 3:e2693.

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.

Emerson, K. J., W. E. Bradshaw, and C. M. Holzapfel. 2009a. Complications of complexity: integrating environmental, genetic and hormonal control of insect diapause. Trends Genet. 25:217–225.

Emerson, K. J., A. M. Uyemura, K. L. McDaniel, P. S. Schmidt, W. E. Bradshaw et al. 2009b. Environmental control of ovarian dormancy in natural populations of *Drosophila melanogaster*. J. Comp. Physiol. A Sens. Neural Behav. Physiol. 195:825–829.

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. 21:4748–4769.

Grant, B. S., D. F. Owen, and C. A. Clarke. 1996. Parallel rise and fall of melanic peppered moths in America and Britain. J. Hered. 87:351–357.

Graveley, B. R., A. N. Brooks, J. Carlson, M. O. Duff, J. M. Landolin et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. Nature 471:473–479.

Haldane, J. B. S. 1924. A mathematical theory of natural and artificial selection. Trans. Camb. Philos. Soc. 23:19–41.

Ives, P. T. 1945. Genetic structure of American populations of *Drosophila melanogaster*. Genetics 30:167–196.

———. 1954. Genetic changes in American populations of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S.A. 40:87–92.

Ives, P. T., and H. T. Band. 1986. Continuing studies on the south Amherst *Drosophila melanogaster* natural population during the 1970's and 1980's. Evolution 40:1289–1302.

Kankare, M., T. S. Salminen, H. Lampinen, and A. Hoikkala. 2012. Sequence variation in couch potato and its effects on life-history traits in a northern malt fly, *Drosophila montana*.J. Insect Physiol. 58:256–264.

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

Laurent, S. J. Y., A. Werzner, L. Excoffier, and W. Stephan. 2011. Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. Mol. Biol. Evol. 28:2041–2051.

Lavebratt, C., and S. Sengul. 2006. Single nucleotide polymorphism (SNP) allele frequency estimation in DNA pools using pyrosequencing. Nat Protoc. 1:2573–2582.

Lee, S. F., C. M. Sgro, J. Shirriffs, C. W. Wee, L. Rako et al. 2011. Polymorphism in the couch potato gene clines in eastern Australia but is not associated with ovarian dormancy in *Drosophila melanogaster*. Mol. Ecol. 20:2973–2984.

Levitan, M. 2003. Climatic factors and increased frequencies of 'southern' chromosome forms in natural populations of *Drosophila robusta*. Evol. Ecol. Res. 5:597–604.

Li, H., and W. Stephan. 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. PLoS Genet. 2:e166.

Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles et al. 2012. The *Drosophila melanogaster* genetic reference panel. Nature 482:173–178.

Mukai, T., and O. Yamaguchi. 1974. The genetic structure of natural populations of *Drosophila melanogaster*. XI. Genetic variability in a local population. Genetics 76:339–366.

Mullen, L. M., and H. E. Hoekstra. 2008. Natural selection along an environmental gradient: a classic cline in mouse pigmentation. Evolution 62:1555–1569.

Reaume, C. J., and M. B. Sokolowski. 2006. The nature of *Drosophila melanogaster*. Curr. Biol. 16:R623–R628.

Saunders, D. S. 1987. Photoperiodism and the hormonal control of insect diapause. Sci. Prog. 71:51–69.

Saunders, D. S., V. C. Henrich, and L. I. Gilbert. 1989. Induction of diapause in *Drosophila melanogaster*: photoperiodic regulation and the impact of arrhythmic clock mutations on time measurement. Proc. Natl. Acad. Sci. U.S.A. 86:3748–3752.

Schmidt, P. S., and A. B. Paaby. 2008. Reproductive diapause and life-history clines in north American populations of *Drosophila melanogaster*. Evolution 62:1204–1215.

Schmidt, P. S., L. Matzkin, M. Ippolito, and W. F. Eanes. 2005. Geographic variation in diapause incidence, life-history traits, and climatic adaptation in *Drosophila melanogaster*. Evolution 59:1721–1732.

Schmidt, P. S., C. T. Zhu, J. Das, M. Batavia, L. Yang et al. 2008. An amino acid polymorphism in the *couch potato* gene forms the basis for climatic adaptation in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. U.S.A. 105:16207–16211.

Schmidt, P. S., and D. R. Conde. 2006. Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in *Drosophila melanogaster*. Evolution 60:1602–1611.

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

Shpak, M., J. Wakeley, D. Garrigan, and R. C. Lewontin. 2010. A structured coalescent process for seasonally fluctuating populations. Evolution. 64:1395–1409.

Sokal R. R., and F. J. Rohlf. 1981. Biometry. Freeman, New York.

Tatar, M., S. A. Chien, and N. K. Priest. 2001. Negligible senescence during reproductive dormancy in *Drosophila melanogaster*. Amer. Nat. 158:248–258.

Thornton, K., and P. Andolfatto. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics. 172:1607–1619.

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

Vasemägi, A. 2006. The adaptive hypothesis of clinal variation revisited: single-locus clines as a result of spatially restricted gene flow. Genetics 173:2411–2414.

Williams, K. D., and M. B. Sokolowski. 1993. Diapause in *Drosophila melanogaster* females: a genetic analysis. Heredity 71: 312–317.

Yukilevich, R., T. L. Turner, F. Aoki, S. V. Nuzhdin, and J. R. True. 2010. Patterns and processes of genome-wide divergence between North American and African *Drosophila melanogaster*. Genetics 186:219–239.

Zera, A. J., and L. G. Harshman. 2001. The physiology of life history trade-offs in animals. Annu. Rev. Ecol. Syst. 32:95–126.

Zhang, Q. R., and D. L. Denlinger. 2011. Elevated couch potato transcripts associated with adult diapause in the mosquito *Culex pipiens*.J. Insect Physiol. 57:620–627.

Zhu, C.-T. 2009. The genetic mapping of reproductive diapause in *Drosophila melanogaster* (Diptera: Drosophilidae). Ph.D. Diss. Stony Brook University.

Associate Editor: P. Andolfatto

## *Supporting Information*

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Figure S1.** Cartoon of the *cpo* gene and the positions of the 15 SNPs.

**Figure S2.** Estimation of the selection coefficient using an approximate Bayesian approach.

**Table S1.** SNP frequencies for the populations.

**Table S2.** Results of tests for significant differences between collection year sets in the parameters of the quadratic fit ($y = y_0 + aX + bX_2$), of allele frequency $y$ to latitude $X$.

**Table S3.** SNP frequencies in the seasonal samples.