Genetic variation in *senescence marker protein-30* is associated with natural variation in cold tolerance in *Drosophila*

KATIE J. CLOWERS[1]*, RICHARD F. LYMAN[2], TRUDY F. C. MACKAY[2,3] AND THEODORE J. MORGAN[1]†

[1] *The Division of Biology and The Ecological Genomics Institute, Kansas State University, Manhattan, KS 66506, USA*
[2] *Department of Genetics, North Carolina State University Raleigh, NC 27695, USA*
[3] *The W. M. Keck Center for Behavioral Biology, North Carolina State University Raleigh, NC 27695, USA*

Summary

A comprehensive understanding of the genetic basis of phenotypic adaptation in nature requires the identification of the functional allelic variation underlying adaptive phenotypes. The manner in which organisms respond to temperature extremes is an adaptation in many species. In the current study, we investigate the role of molecular variation in *senescence marker protein-30* (*Smp-30*) on natural phenotypic variation in cold tolerance in *Drosophila melanogaster*. *Smp-30* encodes a product that is thought to be involved in the regulation of Ca^{2+} ion homeostasis and has been shown previously to be differentially expressed in response to cold stress. Thus, we sought to assess whether molecular variation in *Smp-30* was associated with natural phenotypic variation in cold tolerance in a panel of naturally derived inbred lines from a population in Raleigh, North Carolina. We identified four non-coding polymorphisms that were strongly associated with natural phenotypic variation in cold tolerance. Interestingly, two polymorphisms that were in close proximity to one another (2 bp apart) exhibited opposite phenotypic effects. Consistent with the maintenance of a pair of antagonistically acting polymorphisms, tests of molecular evolution identified a significant excess of maintained variation in this region, suggesting balancing selection is acting to maintain this variation. These results suggest that multiple mutations in non-coding regions can have significant effects on phenotypic variation in adaptive traits within natural populations, and that balancing selection can maintain polymorphisms with opposite effects on phenotypic variation.

1. Introduction

For most organisms, the manner in which individuals cope with temperature extremes represents an ecologically and evolutionarily important phenotype that can influence many aspects of fitness, local adaptation and the range of habitats in which populations can persist (Leather *et al.*, 1993; Clarke, 1996; Zhen & Ungerer, 2008). Temperature extremes represent a major abiotic stressor that all organisms must be able to manage, either through adaptive genetic responses to environmental variation, plastic responses to the environment, or some combination of both genetic and plastic mechanisms (Hoffmann & Parsons, 1991; Ayrinhac *et al.*, 2004; Hoffmann *et al.*, 2005; Hoffmann & Willi, 2008). This is especially true for organisms that are isothermal with their environments (e.g. insects and plants). For such species, the presence of segregating genetic variation for thermal tolerance facilitates survival in a changing environment or expansion outside the historical habitat (Hoffmann *et al.*, 2003).

Drosophila melanogaster has had great success adapting to novel environments during its expansion to its current worldwide species distribution. *D. melanogaster* originated in central tropical Africa, expanded and colonized more temperate climates in Europe and Asia, and more recently has colonized the rest of the world (David & Capy, 1988). This

---

* Present address: Laboratory of Genetics, University of Wisconsin–Madison, 425-G Henry Mall, Madison, WI 53706, USA.
† Corresponding author. The Division of Biology, 116 Ackert Hall, Manhattan, KS 66502, USA. Tel: 785-532-6126. Fax: 785-532-6653. e-mail: tjmorgan@ksu.edu

movement out of Africa into more temperate regions of the world required *D. melanogaster* to cope with many novel thermal stresses. The manner in which populations respond to thermal extremes consists of responses to both high- and low-temperature extremes. Although both ends of the thermal spectrum are ecologically important, resistance to low temperatures appears to be a major climatic adaptation among *Drosophila* species (Gibert & Huey, 2001; Gibert *et al.*, 2001; David *et al.*, 2003) and populations (Gibert & Huey, 2001; Gibert *et al.*, 2001; Hoffmann *et al.*, 2002; Ayrinhac *et al.*, 2004; Schmidt & Paaby, 2008) in nature.

Many studies have shown that cold tolerance phenotypes represent an important adaptation in *Drosophila*; however, a comprehensive understanding of this adaptation requires a complete description of the genetic architecture influencing variation in the trait, as well as the identification of the functional allelic variants underlying such phenotypes. Our knowledge of the genetic architecture of cold tolerance has included studies focused on artificial selection (Tucic, 1979; Chen & Walker, 1993; Anderson *et al.*, 2005); quantifying genetic variation along cold tolerance clines (Gibert & Huey, 2001; Hoffmann *et al.*, 2002; Ayrinhac *et al.*, 2004; Kimura, 2004; Collinge *et al.*, 2006); associations of variation in cold tolerance with polymorphisms in candidate genes (Anderson *et al.*, 2003; Collinge *et al.*, 2008); quantum trait locus (QTL) analyses of cold tolerance phenotypes in recombinant inbred lines (RILs) (Morgan & Mackay, 2006; Norry *et al.*, 2007, 2008); identification of cold tolerant mutations (Takeuchi *et al.*, 2009) and genetic/genomic changes in gene expression in response to artificial selection on cold tolerance (Telonis-Scott *et al.*, 2009) or various cold stressors (Goto, 2000, 2001; Qin *et al.*, 2005; Sinclair *et al.*, 2007). These studies have clearly demonstrated that substantial heritable genetic variation exists for cold tolerance and that this variation is influenced by a large number of genes. Combining this data across studies exposes a number of promising candidate gene regions within the genome between cytological positions 73A-90B6. This region has been identified by QTL analyses of two different sets of RILs (Morgan & Mackay, 2006; Norry *et al.*, 2008). Morgan & Mackay (2006) localized a QTL between 76B and 87B, while Norry *et al.* (2008) identified a broader overlapping QTL between 73A and 90B6. Within the region between 73A and 90B6 many cold tolerance candidate genes exist including *desaturase 2* (87B; Greenberg *et al.*, 2003; Overgaard *et al.*, 2005; Coyne & Elwyn, 2006) as well as multiple cold responsive candidate genes including *Frost* (85E; Goto, 2001; Qin *et al.*, 2005; Sinclair *et al.*, 2007) and *senescence marker protein-30* (*Smp-30*) (88D; (Goto, 2000; Qin *et al.*, 2005; Sinclair *et al.*, 2007). It should be noted that although this broad QTL interval includes the left edge of the inversion, *In(3R)Payne* (Kennington *et al.*, 2006), all of the cold tolerance candidate loci discussed above are outside the inversion.

Smp-30 is a particularly compelling candidate gene for cold tolerance as it has been shown to be highly responsive to cold stress (Goto, 2000; Qin *et al.*, 2005; Sinclair *et al.*, 2007). Goto (2000) found *Smp-30* was up-regulated after 1–2 days of acclimation to 15 °C, while both Qin *et al.* (2005) and Sinclair *et al.* (2007) found *Smp-30* was down-regulated after a shorter 2-h treatment at 0 °C. *Smp-30* contains three exons that encode a 33·3 kDa protein (Goto, 2000). Additionally, although *Smp-30* is not well studied in *D. melanogaster*, there are multiple biochemical studies in various mammalian cell lines suggesting that it has a role in Ca^{2+} ion homeostasis and signalling (Fujita *et al.*, 1998; Inoue *et al.*, 1999; Son *et al.*, 2008). This is physiologically significant because a shift in ion homeostasis (e.g. Ca^{2+}) can lead to a change in the electrochemical potential across cell and organelle membranes, which lead to cold-induced immobilization or injury via several mechanisms (Hochachka, 1986; Pullin & Bale, 1988; Kelty *et al.*, 1996; Kostal *et al.*, 2004, 2007; Takeuchi *et al.*, 2009). Finally, other physiological studies have found that calcium is required for rapid cold-hardening protection in the Antarctic midge (Teets *et al.*, 2008), as chilling causes changes in membrane properties that increase calcium flux, which, in turn, activates calcium-dependent protein kinases and transcription factors that regulate cold-specific proteins and genes.

Given the genetic data for *Smp-30* and the known physiological importance of Ca^{2+} and ion homeostasis and signalling in cold-stress phenotypes in insects, we sought to determine if natural molecular variation in *Smp-30* is associated with natural phenotypic variation in cold tolerance. To do this, we used an association mapping approach, which surveys sequence and phenotypic variation within a natural population of flies and statistically tests, if polymorphisms within the candidate gene explain the natural phenotypic variation in the trait. In this study, we are able to link molecular variation in the 2000 base pair coding and non-coding sequence of *Smp-30* with natural phenotypic variation in cold tolerance in *D. melanogaster*.

2. Materials and methods

(i) Drosophila *stocks*

The lines used in this experiment were inbred lines created from a sample of females collected at the Raleigh, NC Farmer's Market in the spring of 2002. Wild-caught females were used to initiate isofemale lines, which were subsequently inbred (via full-sib

mating) for 20 generations, resulting in a panel of 350 fully inbred yet naturally derived lines. This panel of lines has preserved the natural variation among lines, while eliminating variation within lines. In the current study, 257 randomly selected lines were used. Flies were reared on a standard cornmeal–agar–molasses medium at 25 °C and 70 % humidity on a 12 : 12 light : dark cycle.

(ii) *Phenotypic assays*

Cold tolerance was quantified by measuring the chill-coma recovery time developed by Morgan & Mackay (2006). Four replicates of 25 flies per line per sex were measured in a randomized design. Five-to-seven-day-old same-sex flies were subjected to 0 °C for 3 h. Upon removal from the cold, flies were returned to room temperature (23 °C) and the percent recovery from chill-coma was recorded after 22 min at 23 °C. This experimentally determined time point was chosen because this was the average 50 % recovery time point in a pilot study of 13 randomly chosen lines from this panel (Supplemental Figure 1). Cold tolerance was measured on each line during the summer of 2006.

(iii) *Quantitative genetic analyses*

Analyses of variance (ANOVA) were used to partition variance in cold tolerance among the inbred lines using the following model, $y = \mu + L + S + L \times S + \varepsilon$, where y is the percent recovery from chill coma, μ is the overall mean, L is the random effect of line, S is the fixed effect of sex and ε is the residual variance. We estimated the total genotypic variance (σ_G^2) among lines as $\sigma_G^2 = \sigma_L^2 + \sigma_{LS}^2$, where σ_L^2 is the among line variance component and σ_{LS}^2 is the variance caused by the interaction of line and sex. The total phenotypic variance was estimated as $\sigma_P^2 = \sigma_G^2 + \sigma_E^2$, where σ_E^2 is the environmental variance. The broad sense heritability of cold tolerance was estimated as $H^2 = \sigma_G^2/\sigma_P^2$. To compare the amount of genetic and environmental variance, the coefficient of variance was estimated. The coefficient of genetic variance was calculated as $CV_G = 100 \times \sqrt{\sigma_G^2}/\bar{x}$, where $\bar{x}$ is the average percent of chill-coma recovery and σ_G^2 is the genetic variation, while the coefficient of environmental variation was calculated as $CV_E = 100 \times \sqrt{\sigma_E^2}/\bar{x}$, where $\bar{x}$ is the average percent of chill-coma recovery and σ_E^2 is the environmental variance. All analyses were carried out using SAS in PROC GLM and VARCOMP procedures (SAS Institute).

(iv) *Sequencing of* Smp-30

DNA was extracted from single flies using the Puregene DNA isolation kit (Gentra Systems) in the summer of 2007. We sequenced approximately 2000 base pairs of the *Smp-30* gene region, including 750 base pairs upstream of the translation start site and 100 base pairs downstream of the end of the coding sequence for 125 lines (GenBank accession nos. HM034321–HM034445). This was done using three pairs of PCR primers that were designed to amplify three roughly 600 bp amplicons spanning the *Smp-30* gene region. Primers were designed using the published *D. melanogaster* sequence as a reference and the Primer3 program (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). The three amplicons were amplified per line using 30 μl PCR reactions, which were subsequently sequenced in both directions by the AGTC at the University of Kentucky. Sequences were edited and contigs assembled in Seqman Pro (DNA* Lasergene suite). Contigs were then aligned using MUSCLE (http://www.ebi.ac.uk/Tools/muscle/index.html) and the polymorphisms (both single nucleotide polymorphisms (SNPs) and insertion/deletions) were identified. For the remaining 132 lines, we genotyped the majority of these polymorphisms in each line. This genotyping was done via a combination of partial direct sequencing, SNP-Restriction fragment length polymorphism (Carbone *et al*., 2006), as well as allele-specific amplification of insertion deletion polymorphisms. Ultimately this resulted in a polymorphism dataset that contained on average 214 genotyped lines at each polymorphism. For the association mapping analysis, polymorphisms were used only if the minor allele frequency was at least 2 %.

(v) *Statistical analyses*

(a) *Linkage disequilibrium (LD)*

Patterns of LD and significance tests were assessed with TASSEL (http://www.maizegenetics.net). LD was quantified as $r^2 = (P_{ii}P_{jj} - P_{ij}P_{ji})^2/p_i(1-p_i)p_j(1-p_j)$, where p_i and p_j are the allele frequencies at the ith and jth locus, respectively, and P_{ij} are the haplotype frequencies at loci i and j. Significance of pairwise LD was determined using Fisher's exact test. Allele frequencies for each of the polymorphisms were estimated using PROC FREQ in SAS.

(b) *Association mapping*

Statistical associations between polymorphisms and percent recovery from chill coma were determined by a set of two-way ANOVAs. For each of the 79 polymorphisms, the variation among sex-specific line-means was explained using the following model: $y = \mu + M + S + M \times S + \varepsilon$, where y is the sex-specific line-mean of percent recovery from chill coma, μ is the grand mean, M and S are the fixed effects of marker allele and sex, respectively and ε is the residual error. The terms of primary interest in an association

analysis are the main effect of marker (*M*) and the interaction effect of marker-by-sex (*M* × *S*), as these correspond to the effect of overall or sex-specific allelic variation on the phenotype of interest. In our analysis, we tested 79 polymorphisms, thus to control the type I error at an experiment wise level of less than or equal to 0·05, we used the highly conservative Bonferroni correction. For 79 polymorphisms, our significance threshold was equal to $P = 0·00063$ or $-\log(P) = 3·2007$.

In addition to investigating the individual main effects of markers, we also examined pairwise epistatic interactions among polymorphisms in the *Smp-30* gene region shown to be associated with percent recovery from chill coma. To do this, the variation among sex-specific line-means was explained using the following model: $y = \mu + M_i + M_j + S + M_i \times M_j + M_i \times S + M_j \times S + M_i \times M_j \times S + \varepsilon$, where y, μ, S and ε are identical to the analysis presented above, but now M_i and M_j are the fixed marker effects of significant polymorphisms identified above. In this epistatic analysis, the terms of primary interest are the inter-action effects between the two markers (i.e. $M_i \times M_j$ and $M_i \times M_j \times S$), as these correspond to the effect of overall or sex-specific epistatic effects on the phenotype of interest. This analysis of epistasis was applied only to markers with significant main effect. One limitation of the formal epistatic analysis is that to detect interactions among two biallelic markers, it is essential that all the four two-locus genotypes are represented in the dataset. Unfortunately, for polymorphisms in strong LD or close physical linkage this is not always the case. To deal with this problem, we also performed analyses that investigated the influence of haplotype variation on percent recovery in chill coma. To do this, we used a simple two-way ANOVA similar to the single marker analysis presented above, except rather than using the fixed effects of a marker and sex, we tested the fixed effects of haplotype and sex. Haplotypes were constructed by concatenating the two-, three- or four-way combinations of significant single markers into a haplotype variable. For the epistatic and haplotype analysis, we used a significance threshold of $P = 0·05$.

(c) *Molecular population genetic analyses*

We used DnaSP Version 5.0 (Librado & Rozas, 2009) to evaluate patterns of molecular variation within *Smp-30* and test for patterns of molecular evolution that are consistent with departures from neutrality. Molecular variation in *Smp-30* was quantified by the nucleotide diversity based on the number of segregating sites (θ_S; Watterson, 1975), the average number of nucleotide differences between pairs of sites (θ_π; Nei & Tajima, 1981), as well as the number of singleton sites (θ_η; Fu & Li, 1993). To test for depar-
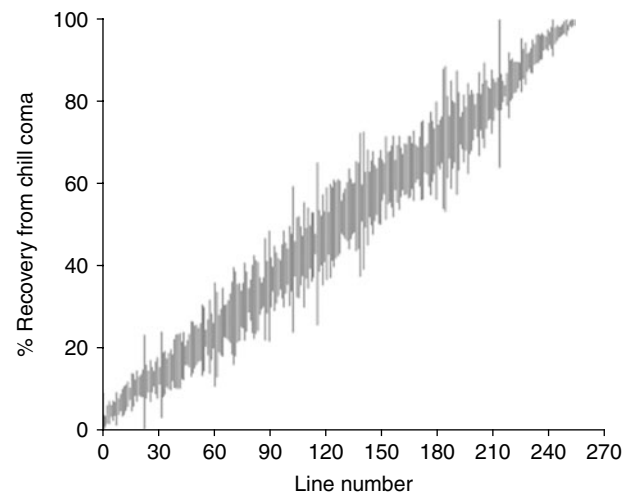

Fig. 1. Variation in percent recovery from chill coma among the inbred lines. The *x*-axis is the line number sorted in rank order based on percent recovery from chill coma, while the *y*-axis is the percentage of flies recovered from chill coma at an experimentally determined time point of 22 min at room temperature (see methods). Lines with 0 % recovery in 22 min are susceptible to cold, while lines with 100 % recovery in 22 min are resistant to cold.

ture from neutrality, we used Tajima's *D* (Tajima, 1989). Estimates of *D* were calculated for the entire gene as well as for sliding window intervals spanning the entire gene. For the sliding window analysis, we used 50 bp windows with a step size of 10 bp.

3. Results

(i) *Variation among lines*

(a) *Phenotypic variation*

We observed abundant natural phenotypic variation in cold tolerance among 255 natural inbred lines sampled from the Raleigh, North Carolina farmer's market (Ayroles *et al.*, 2009), by measuring the percentage of flies that recovered from chill coma during a 22 min observation period at room temperature after 3 h at 0 °C ($F_{255,255} = 8·89$; $P < 0·0001$; Fig. 1, Supplementary Figure 1). Variation among the lines ranged from lines that were extremely susceptible to cold (0 % recovered in the observation period) to lines that were extremely resistant to cold (100 % recovered in the observation period), and all intermediate cold tolerance phenotypes (Fig. 1). However, on average 49·96 % of the flies were recovered during the 22-min observation period, suggesting that our rapid percentage assay is near the 50 : 50 recovery point, which corresponds to the assay point with maximum statistical power.

A significant portion of this phenotypic variation was the result of sex-specific genetic effects, as demonstrated by the highly significant *line-by-sex* interaction term ($F_{255,1471} = 2·23$; $P < 0·0001$). Finally,
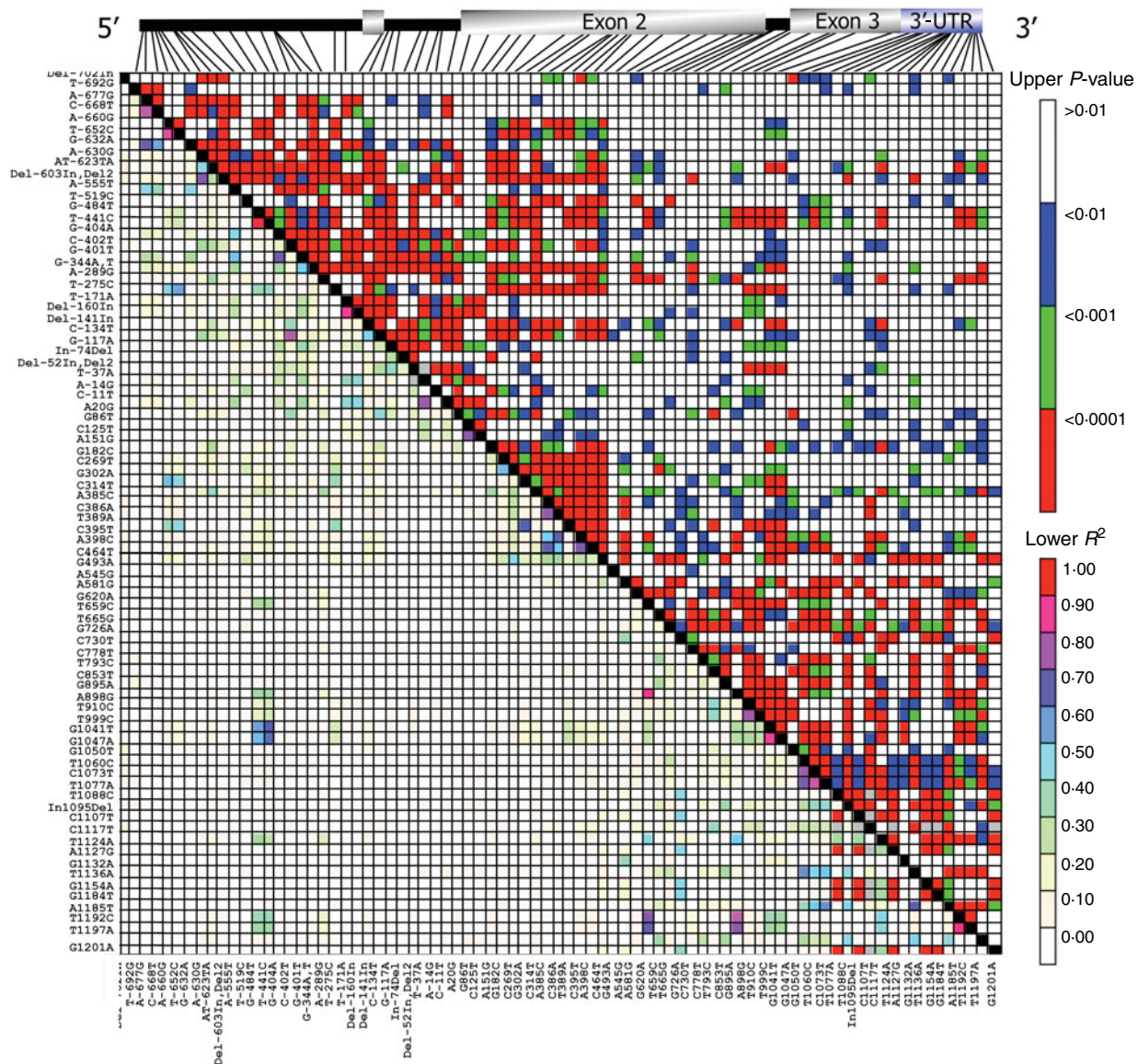
Fig. 2. *Smp-30* polymorphisms. The gene structure of *Smp-30* is depicted; the ATG translation start site is at the beginning of the 2nd exon. Seventy-nine polymorphisms were identified among the 255 lines. Patterns of LD are shown below the gene structure, with *P* values from Fisher's exact test above the diagonal and estimates of r^2 below the diagonal.

consistent with the significant line and line-by-sex terms in the ANOVA among lines, most of this variation in the percentage of flies recovered from cold was the result of genetic effects, as established by large broad-sense heritability estimates from pooled analyses, $H^2 = 0.707$ ($\sigma^2_G = 766.35$; $\sigma^2_P = 1083.77$) and sex-specific analyses ($H^2_{Male} = 0.710$; $H^2_{Female} = 0.705$). In addition, the pronounced genetic effect was consistent with the large magnitude of the coefficient of genetic variation ($CV_G = 55.40$), relative to the coefficient of environmental variation ($CV_E = 35.66$). Although all three of these estimates of variation (H^2, CV_G, and CV_E) are high for *Drosophila*, they are consistent with other studies of chill-coma recovery time in *D. melanogaster* (Anderson *et al.*, 2005; Morgan & Mackay, 2006).

(b) *Molecular variation in the* Smp-30 *gene region*

We sequenced approximately 2000 bp encompassing the *Smp-30* gene region in this set of wild-derived inbred lines, including the exons, introns and 5′-UTR (5′-untranslated region) and 3′-UTR, and 750 bp upstream of the translation start site. The gene region was highly polymorphic, with 79 polymorphisms that had minor allele frequency (MAF) of at least 2% (Fig. 2; Supplemental Table 1). Seventy-two of these polymorphisms were SNPs, while seven were insertion–deletion (indel) polymorphisms. The majority of these polymorphisms we located in non-coding regions, including the putative 5′ regulatory region, introns, and the 3′-UTR (Fig. 2). Consistent with the large number of polymorphisms, the estimates
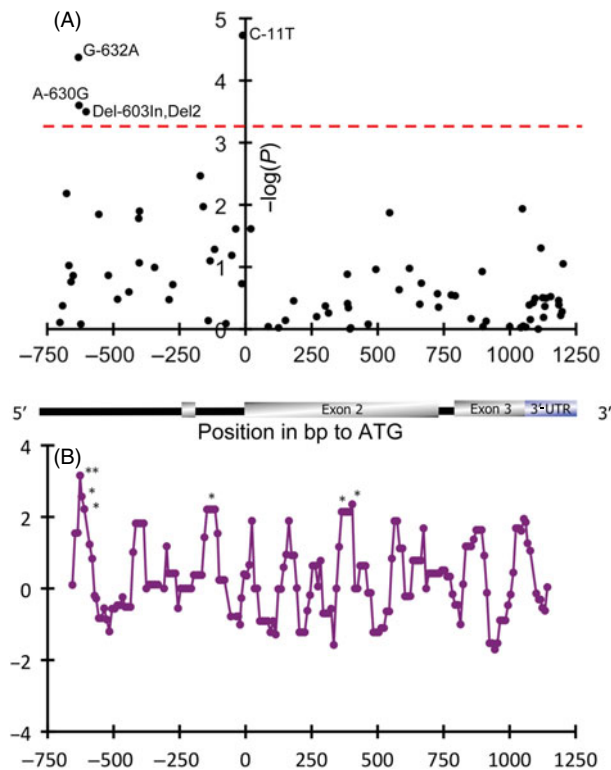
Fig. 3. Phenotypic associations and balancing selection in *Smp-30*. (*a*) Statistical associations between molecular variation in *Smp-30* and percent recovery from chill coma. The *x*-axis is the position in base pairs relative to the ATG start site (at the beginning of exon 2), while the *y*-axis is the significance of each marker on a –log scale. The red horizontal dashed line is the experiment-wise $P < 0.05$ threshold given by the Bonferroni correction. The marker names of the four highly significant polymorphisms are noted next to the corresponding data point. (*b*) Evolutionary analyses of molecular variation in *Smp-30* via a sliding window analysis of Tajima's *D*. * $P < 0.05$, ** $P < 0.01$.

Table 1. *Results of significant genotype–phenotype associations*

Marker[a]	MAF[b]	N[c]	F	P	2a[d]
G-632A	0·17	199	17·15	0·00004	16·44
A-630G	0·44	199	13·65	0·00025	−11·04
Del-603In, Del2	0·43, 0·09	209	7·42	0·00032	−12·29
C-11T	0·08	246	18·57	0·00002	20·86

[a] Marker names correspond to the common allele, the position in base pairs relative to the ATG start site (i.e. the beginning of exon 2), prefixed by the common allele and followed by the minor allele. In cases where three allelic classes were discovered, the minor alleles are listed in order of descending allele frequency.
[b] Minor allele frequency.
[c] The sample size (*N*, number of lines) used in the ANOVA.
[d] Mean difference in cold tolerance (% recovery) between homozygous genotypes (common homozygote – rare homozygote).

(ii) *Associations between phenotypic variation and molecular polymorphism*

We tested for associations between cold tolerance and each of the 79 polymorphisms identified within the *Smp-30* gene region. Figure 3*a* summarizes the significance, on a −log(*P*) scale, of these single-marker analyses. We observed four markers that exhibited highly significant associations with cold tolerance; these include G-632A ($P = 0.00004$), A-630G ($P = 0.00025$), Del-603In,Del2 ($P = 0.00032$) and C-11T ($P = 0.00002$). All of these polymorphisms exceed the highly conservative experiment-wise Bonferroni correction ($P = 0.05/79 = 0.00063$; red dashed line in Fig. 3*a*). In addition, all of the markers are located in non-coding portions of the *Smp-30* gene region; three of the polymorphisms (G-632A, A-630G and Del-603In,Del2) are located within 30 base pairs of one another and are in the putative 5′ regulatory region of the gene (Goto, 2000), while the fourth (C-11T) resides in the first intron, 11 base pairs upstream of the translation start site (Fig. 3*a*). Two of the significant polymorphisms (G-632A and C-11T) have skewed allele frequencies, with frequencies of the common allele equal to 0·83 and 0·92, respectively, while the other two polymorphisms (A-630G and Del-603In,Del2) have common allele frequencies near 0·50 (Supplemental Table 1).

The mean difference in cold tolerance between the homozygous genotypes for the significant polymorphisms ranged from −12·29% to 20·86% (Table 1). However, the mean phenotypic differences attributable to each polymorphism should be interpreted with caution because the polymorphisms are not independent, due to LD (Fig. 2). Nearly all of the significant polymorphisms were in strong LD with each other ($P < 0.0001$; Fisher's exact test; Fig. 2). The single exception was the marker pair A-630G and
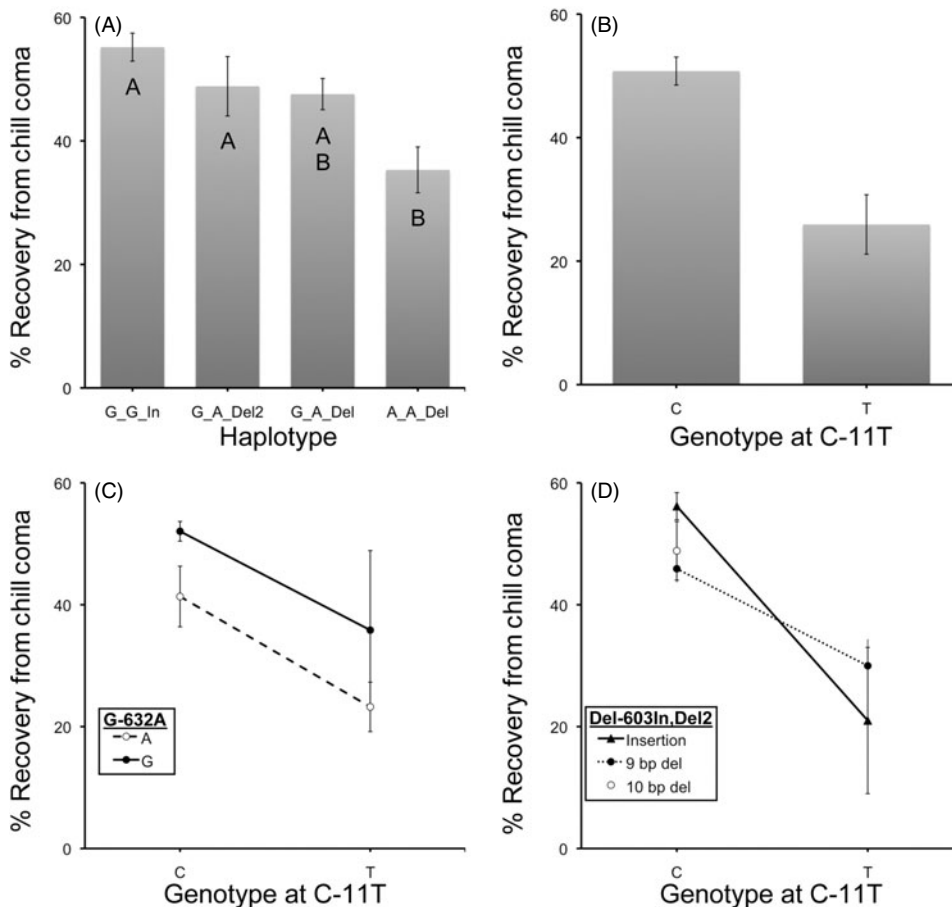
of nucleotide diversity were $\theta_S = 0.00976$, $\theta_\pi = 0.01213$ and $\theta_\eta = 0.00986$.

We calculated LD between all pairs of polymorphisms (Fig. 2). We observed significant LD among many polymorphisms; however, the actual association between alleles (i.e. the absolute values of LD) as measured by R^2 was quite small, with a few exceptions where markers in close physical proximity exhibited large R^2 values. Thus, the widespread significance of the LD is likely an effect of our large sample size rather than the presence of defined haplotype blocks. The majority of the significant LD was contained within an approximately 1000 bp block at the 5′ end of the gene and in an 800 bp block at the 3′ end of the gene (Fig. 2). Although there was some significant LD between these blocks, the overall frequency of significant LD declined with physical distance and thus LD was reduced between polymorphisms from the 5′ and 3′ ends of the gene.

Fig. 4. Phenotypic effects of the significantly associated regions on the *Smp-30* gene. (*a*) Highly significant variation in percent recovery from chill coma associated with the haplotypes in the 5′ upstream region of the gene containing significant markers: G-632A, A-630G and Del-603In,Del2 ($F_{3,390} = 7.48$; $P < 0.0001$). (*b*) Variation in percent recovery from chill coma associated with the marker C-11T. (*c* and *d*) Evidence for statistical independence of the effect of polymorphisms within the 5′ haplotype (i.e. G-632A, A-630G and Del-603In,Del2) and the C-11T polymorphism. Both figures show the main effects of each polymorphism but lack of significant pairwise interactions between polymorphisms. Panel (C) is for the marker pair G-632A by C-11T ($F_{1,390} = 0.02$; $P = 0.8947$), while panel (D) is for the marker pair Del-603In,Del2 by C-11T ($F_{1,408} = 0.79$; $P = 0.3739$). The same pattern holds for marker pair A-630G by C-11T ($F_{1,390} = 0.44$; $P = 0.5071$) graph not shown.

C-11T, which exhibited significant but a less robust level of LD ($P < 0.01$; Fig. 2). Regardless of the lack of independence caused by LD among the significant polymorphisms, on a striking pattern that emerges from this data is the finding that two of the significant polymorphisms (G-632A and A-630G), which are located only 2 bp apart in the 5′ putative regulatory region of *Smp-30* (Goto, 2000), have opposite effects on chill-coma recovery (Table 1). Marker G-632A has a difference in chill–coma recovery between the homozygous alleles [G ($p = 0.83$); A ($q = 0.17$)] of 16·44%, while marker A-630G has a difference in chill–coma recovery between the homozygous alleles [A ($p = 0.56$); G ($q = 0.44$)] of -11.04%. This pattern is noteworthy because it is unlikely that a pair of polymorphisms in such a close physical linkage would have opposite effects on the same trait unless there was significant balancing selection acting to maintain the pair of antagonistic polymorphisms.

The three significant polymorphisms in the 5′ putative regulatory region of *Smp-30* (Goto, 2000) were in strong LD with one another and formed only four haplotypes (Fig. 4*a*). This three-marker haplotype was significantly associated with percent recovery from chill coma ($F_{3,390} = 7.48$; $P < 0.0001$; Fig. 4*a*). To test for epistastis among the significant polymorphisms, we separated the polymorphisms into two regions: region 1 (Fig. 4*a*) contained the set of three physically linked markers (i.e. G-632A, A-630G and Del-603In,Del2) in the 5′ putative regulatory region of *Smp-30* (Goto, 2000), while region 2 was the C-11T marker (Fig. 4*b*). We tested for epistasis between the three possible pairwise combinations of markers in region 1 and the single marker in region 2 and did not detect any significant epistatic interactions in any analysis (Fig. 4*c*, *d*). This can be easily seen by the lack of significant interaction effects between markers: G-632A by C-11T ($F_{1,390} = 0.02$; $P = 0.8947$; Fig. 4*c*),

Del-603In,Del2 by C-11T ($F_{1,408}=0\cdot79$; $P=0\cdot3739$; Fig. 4*d*) and A-630G by C-11T ($F_{1,390}=0\cdot44$; $P=0\cdot5071$; graph not shown).

(iii) *Molecular population genetics of* Smp-30

We tested for evidence of non-neutral molecular evolution by estimating Tajima's *D* (Tajima, 1989). Tajima's *D* was estimated using the set of 137 complete sequences for the entire gene region and with an analysis of 50 bp sliding windows (Fig. 3*b*). Tajima's *D* over the entire gene region was positive ($D=0\cdot6752$), but not significantly different from zero. However, when we performed a sliding window analysis, we identified three regions where Tajima's *D* is significantly positive (Fig. 3*b*), indicating a pattern of molecular variation consistent with the maintenance of excess variation by balancing selection. The region with the most significant Tajima's *D* values occurred in the putative 5′ regulatory region of the gene (Goto, 2000) between sites −581 and −650 (Fig. 3*b*). Interestingly, this region overlaps the region containing three polymorphisms (G-632A, A-630G and Del-603In,Del2) that are significantly associated with phenotypic variation in cold tolerance (Fig. 3*a*; Table 1), which are in strong LD, but that have phenotypic effects that are in opposite directions based on differences between the homozygous genotypes at each polymorphism. The common allele at marker G-632A increases the percentage of flies recovered from chill coma, while the common allele at markers A-630G and Del-603In,Del2 decrease the percentage of flies recovered from chill coma.

4. Discussion

One of the primary goals of evolutionary genetics is to identify the causal allelic variants that contribute to phenotypic adaptation in nature (Lewontin, 1974; Hoekstra & Coyne, 2007). In *Drosophila*, cold tolerance is a major climatic adaptation that has been documented by the presence of phenotypic clines on multiple continents (Gibert *et al.*, 2001; Hoffmann *et al.*, 2002; Ayrinhac *et al.*, 2004; Schmidt & Paaby, 2008) and also significantly influences the divergence and climate matching among *Drosophila* species (Gibert & Huey, 2001; Gibert *et al.*, 2001). As a result of these recently documented adaptive patterns, many regions of the genome have also been identified that influence cold tolerance in *Drosophila*, including the candidate gene *Smp-30* (Goto, 2000; Qin *et al.*, 2005; Sinclair *et al.*, 2007), which encodes a protein product involved in Ca^{2+} ion homeostasis and signalling. However, to understand cold adaptation in nature, we must link polymorphisms in functionally important loci like *Smp-30* with ecologically relevant natural phenotypic variation. To date, no study has performed a detailed analysis of sequence variation within *Smp-30* and cold tolerance variation within a *natural* population (but see Rako *et al.*, 2007) to identify the segregating allelic variants influencing cold tolerance in *nature*.

Our study confirms that *Smp-30* is a cold tolerance gene (Goto, 2000; Qin *et al.*, 2005; Sinclair *et al.*, 2007) and more importantly that there is a set of four polymorphisms (G-632A, A-630G, Del-603In,Del2 and C-11T), three of which are in the putative 5′ regulatory region of the gene (Goto, 2000) and one that is just upstream of the ATG start site, which are strongly associated with variation in cold tolerance. All of these polymorphisms are at appreciable frequencies within our population (MAF ≥ 0·08) and all were in significant LD with one another. This significant LD makes it difficult to disentangle the individual influence of each polymorphism on variation in cold tolerance. However, if we use the most conservative (i.e. the smallest; A-630G) estimate for the difference in the cold tolerance between homozygous marker classes, we observe a change of 11% between the homozygous genotypes, which is a dramatic effect on the cold tolerance.

LD throughout the 2-kb *Smp-30* gene region obstructs our ability to identify the quantitative trait nucleotides (Long *et al.*, 1998; Carbone *et al.*, 2006) responsible for the observed variation in cold tolerance. Thus, it is possible that G-632A, A-630G, Del-603In,Del2 and C-11T are not the causal polymorphisms, but rather are in LD with other true causal regulatory polymorphism upstream of these markers. This is because markers G-632A, A-630G and Del-603In,Del2 all reside within the first 200 base pairs of our sequencing reads; therefore it is possible that by only sequencing 750 base pairs upstream of the ATG start site we are missing true regulatory polymorphisms upstream of these strongly associated markers. Future studies should extend the sequencing sufficiently far in the 5′ and 3′ directions outside of the coding sequence to observe the ultimate decay of LD between our significant polymorphisms and other possible unidentified causal polymorphisms further upstream. One other study has evaluated the association between a variation in 5′ regulatory region *Smp-30* and natural variation in cold tolerance in an Australian population (Rako *et al.*, 2007). In this study, Rako *et al.* (2007) used a pair of primers that span our three significantly associated polymorphisms, G-632A, A-630G and Del-603In,Del2, but found no significant association between their four different length variant (330–342 bp) alleles. Rako *et al.* (2007) used a somewhat atypical association analysis experimental design; however, they were able to detect multiple associations between other temperature stress genes and thermal phenotypes, suggesting that their approach was robust. Thus, in order

to evaluate disagreement between our studies we placed their reverse primer sequence (5′-TTAAAAG GTATAAAACAATCAACACTG) on our alignment (at position −491). We were not able to place their forward primer sequence as it occurs approximately 100 bp upstream of the start of our sequence reads. By scanning the region flanked by their reverse primer and the start of our sequence reads, we were able to construct four plausible length variant alleles from our sequence data. Although we cannot definitely identify the molecular basis of Rako et al.'s (2007) 12 bp length variation (i.e. 330–342 bp) PCR polymorphism, we did observe two indel polymorphisms within the region that would have been amplified by their primers. The first 3 bp indel (Del-702In) was not associated with cold tolerance, while the second 9–10 bp indel (Del-603In, Del2) was significantly associated with cold tolerance (Table 1; Fig. 3). If we combine the length variation attributable to each of these indel markers and create a composite polymorphism, we generate six possible length variant alleles, because marker Del-603In,Del2, has a common 9 bp deletion allele, an insertion allele and a rare 10 bp deletion allele (Supplemental Table 1). Thus the alleles at this composite marker in order of length from shortest to longest are 13, 12, 10, 9, 3 and 0 bp deletion alleles. Four of these alleles are present in our data (13, 12, 9 and 3 bp). If we repeat our association analysis and test for the association between this composite length polymorphism and percent recovery from chill coma, we observe a nominally significant association [$F_{3,310}=3\cdot90$; $P=0\cdot0093$; $-\log(P)=2\cdot03$]; this does not exceed our multiple testing criteria [$P=0\cdot0006$; $-\log(P)=3\cdot20$; Fig. 3]. Thus, this analysis of our composite length variant allele suggests that the lack of association observed by Rako et al. (2007) is likely not in disagreement with our data. However, by decoupling this length variation by direct sequencing, we are able to detect a significant association between cold tolerance and marker Del,-603In,Del2 as well as two other SNPs (G-632A and A-630G) which could not be detected via the PCR product length polymorphism.

This study shows that Smp-30 is associated with natural phenotypic variation in cold tolerance; however, the most complete picture of this association emerges when the results of the association, LD, and molecular evolutionary analyses are considered together. The presence of multiple highly significant polymorphisms in close proximity to one another (2–30 bp apart), which are also in strong LD with one another, but display opposite phenotypic effects (i.e. G-632A and A-630G) suggests that the allelic variation at these polymorphisms is likely being maintained by balancing selection. This argument becomes even more compelling when it is coupled, the Tajima's D analysis that identifies this same region as a region

of the gene with the maintenance of an excess variation (i.e. a positive Tajima's D) suggesting that this putative 5′ regulatory region (Goto, 2000) between sites −581 and −650 is not only significantly associated with natural phenotypic variation in chill coma but is also being maintained by balancing selection. These results suggest that multiple mutations in non-coding regions can have significant effects on phenotypic variation on adaptive traits within natural populations, and that balancing selection can maintain physically linked polymorphisms with opposite effects on phenotypic variation in nature.

References

Anderson, A. R., Collinge, J. E., Hoffmann, A. A., Kellett, M. & McKechnie, S. W. (2003). Thermal tolerance trade-offs associated with the right arm of chromosome 3 and marked by the *hsr-omega* gene in *Drosophila melanogaster*. *Heredity* **90**, 195–202.

Anderson, A. R., Hoffmann, A. A. & McKechnie, S. W. (2005). Response to selection for rapid chill-coma recovery in *Drosophila melanogaster*: physiology and life-history traits. *Genetical Research* **85**, 15–22.

Ayrinhac, A., Debat, V., Gibert, P., Kister, A. G., Legout, H., Moreteau, B., Vergillno, R. & David, J. R. (2004). Cold adaptation in geographical populations of *Drosophila melanogaster*: phenotypic plasticity is more important than genetic variability. *Functional Ecology* **18**, 700–706.

Ayroles, J. F., Carbone, M. A., Stone, E. A., Jordan, K. W., Lyman, R. F., Magwire, M. M., Rollmann, S. M., Duncan, L. H., Lawrence, F., Anholt, R. R. & MacKay, T. F. (2009). Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics* **41**, 299–307.

Carbone, M. A., Jordan, K. W., Lyman, R. F., Harbison, S. T., Leips, J., Morgan, T. J., DeLuca, M., Awadalla, P. & MacKay, T. F. (2006). Phenotypic variation and natural selection at *Catsup*, a pleiotropic quantitative trait gene in *Drosophila*. *Current Biology* **16**, 912–919.

Chen, C. P. & Walker, V. K. (1993). Increase in cold-shock tolerance by selection of cold resistant lines in *Drosophila melanogaster*. *Ecological Entomology* **18**, 184–190.

Clarke, A. (1996). The influence of climate change of the distribution and evolution of organisms. In *Animals and Temperature: Phenotypic and Evolutionary Adaptation*, (ed I. A. Johnston & A. F. Bennett), pp. 377–407. Cambridge: Cambridge University Press.

Collinge, J. E., Anderson, A. R., Weeks, A. R., Johnson, T. K. & McKechnie, S. W. (2008). Latitudinal and cold-tolerance variation associate with DNA repeat-number variation in the *hsr-omega* RNA gene of *Drosophila melanogaster*. *Heredity* **101**, 260–270.

Collinge, J. E., Hoffmann, A. A. & McKechnie, S. W. (2006). Altitudinal patterns for latitudinally varying traits and polymorphic markers in *Drosophila melanogaster*

from eastern Australia. *Journal of Evolutionary Biology* **19**, 473–482.

Coyne, J. A. & Elwyn, S. (2006). Does the *desaturase-2* locus in *Drosophila melanogaster* cause adaptation and sexual isolation? *Evolution* **60**, 279–291.

David, J. R. & Capy, P. (1988). Genetic variation of *Drosophila melanogaster* natural populations. *Trends in Genetics* **4**, 106–111.

David, J. R., Gibert, P., Moreteau, B., Gilchrist, G. W. & Huey, R. B. (2003). The fly that came in from the cold: geographic variation of recovery time from low-temperature exposure in *Drosophila subobscura*. *Functional Ecology* **17**, 425–430.

Fu, Y.-X. & Li, W.-H. (1993). Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

Fujita, T., Inoue, H., Kitamura, T., Sato, N., Shimosawa, T. & Maruyama, N. (1998). Senescence marker protein-30 (SMP30) rescues cell death by enhancing plasma membrane Ca^{2+} pumping activity in Hep G2 cells. *Biochemical and Biophysical Research Communications* **250**, 374–380.

Gibert, P. & Huey, R. B. (2001). Chill-coma temperature in *Drosophila*: effects of developmental temperature, latitude, and phylogeny. *Physiological and Biochemical Zoology* **74**, 429–434.

Gibert, P., Moreteau, B., Petavy, D., Karan, D. & David, J. R. (2001). Chill-coma tolerance, a major climatic adaptation among *Drosophila* species. *Evolution* **55**, 1063–1068.

Goto, S. G. (2000). Expression of *Drosophila* homologue of *senescence marker protein-30* during cold acclimation. *Journal of Insect Physiology* **46**, 1111–1120.

Goto, S. G. (2001). A novel gene that is up-regulated during recovery from cold shock in *Drosophila melanogaster*. *Gene* **270**, 259–264.

Greenberg, A. J., Moran, J. R., Coyne, J. A. & Wu, C. I. (2003). Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science* **302**, 1754–1757.

Hochachka, P. W. (1986). Defense strategies against hypoxia and hypothermia. *Science* **231**, 234–241.

Hoekstra, H. E. & Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016.

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

Hoffmann, A. A. & Parsons, P. (1991). *Evolutionary Genetics of Environmental Stress*. Oxford: Oxford university press.

Hoffmann, A. A., Shirriffs, J. & Scott, M. (2005). Relative importance of plastic vs genetic factors in adaptive differentiation: geographical variation for stress resistance in *Drosophila melanogaster* from eastern Australia. *Functional Ecology* **19**, 222–227.

Hoffmann, A. A., Sorensen, J. G. & Loeschcke, V. (2003). Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *Journal of Thermal Biology* **28**, 175–216.

Hoffmann, A. A. & Willi, Y. (2008). Detecting genetic responses to environmental change. *Nature Reviews Genetics* **9**, 421–432.

Inoue, H., Fujita, T., Kitamura, T., Shimosawa, T., Nagasawa, R., Inoue, R., Maruyama, N. & Nagasawa, T. (1999). Senescence marker protein-30 (SMP30) enhances the calcium efflux from renal tubular epithelial cells. *Clinical and Experimental Nephrology* **3**, 261–267.

Kelty, J. D., Killian, K. A. & Lee, R. E. (1996). Cold shock and rapid cold-hardening of pharate adult flesh flies (*Sarcophaga crassipalpis*): effects on behaviour and neuromuscular function following eclosion. *Physiological Entomology* **21**, 283–288.

Kennington, W. J., Partridge, L. & Hoffmann, A. A. (2006). Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster* are indicative of coadaptation. *Genetics* **172**, 1655–1663.

Kimura, M. T. (2004). Cold and heat tolerance of drosophilid flies with reference to their latitudinal distributions. *Oecologia* **140**, 442–449.

Kostal, V., Renault, D., Mehrabianova, A. & Bastl, J. (2007). Insect cold tolerance and repair of chill-injury at fluctuating thermal regimes: role of ion homeostasis. *Comparative Biochemistry and Physiology A-Molecular and Integrative Physiology* **147**, 231–238.

Kostal, V., Vambera, J. & Bastl, J. (2004). On the nature of pre-freeze mortality in insects: water balance, ion homeostasis and energy charge in the adults of *Pyrrhocoris apterus*. *Journal of Experimental Biology* **207**, 1509–1521.

Leather, S. R., Walters, K. F. A. & Bale, J. S. (1993). *The Ecology of Insect Overwintering*. Cambridge [England]; New York, NY, USA: Cambridge University Press.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.

Long, A. D., Lyman, R. F., Langley, C. H. & Mackay, T. F. C. (1998). Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**, 999–1017.

Morgan, T. J. & Mackay, T. F. (2006). Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity* **96**, 232–242.

Nei, M. & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145–163.

Norry, F. M., Gomez, F. H. & Loeschcke, V. (2007). Knockdown resistance to heat stress and slow recovery from chill coma are genetically associated in a quantitative trait locus region of chromosome 2 in *Drosophila melanogaster*. *Molecular Ecology* **16**, 3274–3284.

Norry, F. M., Scannapieco, A. C., Sambucetti, P., Bertoli, C. I. & Loeschcke, V. (2008). QTL for the thermotolerance effect of heat hardening, knockdown resistance to heat and chill-coma recovery in an intercontinental set of recombinant inbred lines of *Drosophila melanogaster*. *Molecular Ecology* **17**, 4570–4581.

Overgaard, J., Sorensen, J. G., Petersen, S. O., Loeschcke, V. & Holmstrup, M. (2005). Changes in membrane lipid composition following rapid cold hardening in *Drosophila melanogaster*. *Journal of Insect Physiology* **51**, 1173–1182.

Pullin, A. S. & Bale, J. S. (1988). Cause and effects of pre-freeze mortality in aphids. *Cryo-Letters* **9**, 102–113.

Qin, W., Neal, S. J., Robertson, R. M., Westwood, J. T. & Walker, V. K. (2005). Cold hardening and transcriptional change in *Drosophila melanogaster*. *Insect Molecular Biology* **14**, 607–613.

Rako, L., Blacket, M. J., McKechnie, S. W. & Hoffmann, A. A. (2007). Candidate genes and thermal phenotypes: identifying ecologically important genetic variation for thermotolerance in the Australian *Drosophila melanogaster* cline. *Molecular Ecology* **16**, 2948–2957.

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

Sinclair, B. J., Gibbs, A. G. & Roberts, S. P. (2007). Gene transcription during exposure to, and recovery from, cold and desiccation stress in *Drosophila melanogaster*. *Insect Molecular Biology* **16**, 435–443.

Son, T. G., Kim, S. J., Kim, K., Kim, M., Chung, H. Y. & Lee, J. (2008). Cryoprotective roles of senescence marker protein 30 against intracellular calcium elevation and oxidative stress. *Archives of Pharmacal Research* **31**, 872–877.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Takeuchi, K., Nakano, Y., Kato, U., Kaneda, M., Aizu, M., Awano, W., Yonemura, S., Kiyonaka, S., Mori, Y., Yamamoto, D. & Umeda, M. (2009). Changes in temperature preferences and energy homeostasis in dystroglycan mutants. *Science* **323**, 1740–1743.

Teets, N. M., Elnitsky, M. A., Benoit, J. B., Lopez-Martinez, G., Denlinger, D. L. & Lee, R. E. (2008). Rapid cold-hardening in larvae of the Antarctic midge *Belgica antarctica*: cellular cold-sensing and a role for calcium. *American Journal of Physiology–Regulatory Integrative and Comparative Physiology* **294**, R1938–R1946.

Telonis-Scott, M., Hallas, R., McKechnie, S. W., Wee, C. W. & Hoffmann, A. A. (2009). Selection for cold resistance alters gene transcript levels in *Drosophila melanogaster*. *Journal of Insect Physiology* **55**, 549–555.

Tucic, N. (1979). Genetic capacity for adaptation to cold resistance at different developmental stages of *Drosophila melanogaster*. *Evolution* **33**, 350–358.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–276.

Zhen, Y. & Ungerer, M. C. (2008). Clinal variation in freezing tolerance among natural accessions of *Arabidopsis thaliana*. *New Phytologist* **177**, 419–427.