MOLECULAR EVOLUTION IN *DROSOPHILA MELANOGASTER*: GENETIC ASPECTS OF THERMAL ADAPTATION

Dissertation

der Fakultät für Biologie

der Ludwig-Maximilians-Universität

München

vorgelegt von

Susanne Voigt

aus Gera

München, Juni 2015

Dekan: Prof. Dr. Heinrich Leonhardt

1. Gutachter: Prof. Dr. Wolfgang Stephan

2. Gutachter: Prof. Dr. Wolfgang Enard

Tag der Abgabe: 17. Juni 2015

Tag der mündlichen Prüfung: 21. Oktober 2015

ERKLÄRUNG:

Diese Dissertation wurde im Sinne von §12 der Promotionsordnung von Prof. Dr. Wolfgang Stephan betreut. Ich erkläre hiermit, dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

EIDESSTATTLICHE ERKLÄRUNG:

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 17. Juni 2015

Susanne Voigt

DECLARATION OF AUTHOR'S CONTRIBUTIONS

In this thesis, I present the results of my doctoral research from September 2010 to June 2015. The results are presented in three chapters. All of them resulted from collaborations with other scientists. The work of this doctoral thesis has resulted in two publications. Chapters 1 and 3 have been published. Chapter 2 is an unpublished manuscript.

In Chapter 1, I and Wolfgang Stephan conceived the study and its design. Population genetic analyses were performed by me, Stefan Laurent, and Maria Litovchenko. I performed the Sanger sequencing, molecular cloning, microinjections, tissue dissections, and expression analysis. The manuscript was written by me and Wolfgang Stephan. The final manuscript was published in:

Voigt, S., S. Laurent, M. Litovchenko, and W. Stephan, 2015 Positive selection at the *polyhomeotic* locus led to decreased thermosensitivity of gene expression in temperate *Drosophila melanogaster*. Genetics 200: 591-599.

In Chapter 2, I and Wolfgang Stephan conceived the study and its design. I conducted the tissue dissections and expression analyses. Anna Christina Erpf contributed to the expression analysis in adult flies. I performed the population genetic analyses.

In Chapter 3, Ricardo Wilches and Wolfgang Stephan conceived the study and its design. Ricardo Wilches performed the chill coma experiments and complementation tests. I performed the expression analysis and Sanger sequencing. Pablo Duchen and Stefan Laurent conducted the population genetic analyses. The manuscript was written by Ricardo Wilches and Wolfgang Stephan and published in:

Wilches, R., **S. Voigt**, P. Duchén, S. Laurent, and W. Stephan, 2014 Fine-mapping and selective sweep analysis of QTL for cold tolerance in *Drosophila melanogaster*. G3-Genes Genomes Genetics 4: 1635-1645.

SUMMARY

The genetic basis underlying adaptive evolution is still largely unknown. Adaptive evolution is facilitated by natural selection that acts on the genetic variation present in a population. Favoring some genetic variants over others, natural selection eventually produces adaptations that allow populations to survive in changing or new environments. Populations colonizing new habitats that differ from their original habitat are often confronted with a multitude of novel ecological constraints to which they need to adapt.

A well-annotated genome and a diverse genetic toolkit make the fruit fly *Drosophila melanogaster* an ideal model system for studying the genetics underlying adaptation. As a cosmopolitan species, *D. melanogaster* has adapted to a wide range of thermal environments. Despite having a tropical origin in southern-central Africa, it has successfully settled in temperate environments around the world. Thermal adaptations that have helped to deal with the greater range and variability in temperature as well as low-temperature extremes have been required to prosper in temperate environments.

Chromatin-based gene regulation is known to be disrupted by varying temperatures. Variation in the temperature, at which flies live, result in varying expression levels of Polycomb group (PcG) regulated genes with higher expression at lower temperatures. Chapter 1 and 2 of this thesis aim to answer the question whether this thermosensitivity of PcG regulation has been detrimental for colonizing temperate environments and thus needed to be buffered by natural selection. Thermosensitivity of PcG regulation was observed in different natural populations of *D. melanogaster*. A lower degree of thermosensitive expression was consistently found for populations from temperate climates when compared to those from the tropics. In Chapter 1, evidence is presented for positive selection acting on the

polyhomeotic (*ph*) gene region to reduce thermosensitivity of PcG regulation in temperate populations from Europe. The targets of selection appear to be single nucleotide polymorphisms (SNPs) in a relatively small *cis*-regulatory region between the two PcG target genes *polyhomeotic proximal* (*ph-p*) and *CG3835* that are highly differentiated between European and African populations. Using reporter gene assays, it was demonstrated that these SNPs influence gene expression and that the European alleles confer reduced thermosensitivity of expression in contrast to the African alleles. In Chapter 2, thermosensitivity of another PcG target gene, *vestigial* (*vg*), was investigated in six natural populations including four temperate populations from high-altitude Africa and central to high-latitude Europe, and two tropical populations from the ancestral species range. All four temperate populations exhibited a lower degree of thermosensitive expression than the two tropical populations. The underlying mechanisms of increased buffering, however, seem to differ between these temperate populations.

Thermal adaptation to temperate environments also includes dealing with low-temperature extremes. Severe cold stress is a main limiting factor imposed on *D. melanogaster* by temperate climates. Increased cold tolerance in temperate populations is thought to have evolved by natural selection. Cold tolerance is a quantitative trait that appears to be highly polygenic and has been mapped to different quantitative trait loci (QTL) in the genome. In Chapter 3, such a QTL region was fine-mapped to localize causal genes for increased cold tolerance in temperate flies. As a result, *brinker* (*brk*) was identified as a new candidate gene putatively involved in cold stress adaptation.

ZUSAMMENFASSUNG

Die genetische Grundlage adaptiver Evolution ist noch weitgehend unbeschrieben. Adaptive Evolution wird durch natürliche Selektion ermöglicht, die auf die genetische Variation einer Population einwirkt. Durch das Bevorzugen einiger genetischer Varianten gegenüber anderen führt natürliche Selektion letztendlich zu Anpassungen, welche Populationen erlauben in neuen oder sich verändernden Lebensräumen zu überleben. Wenn Populationen neue Lebensräume erschließen, die sich von ihrem ursprünglichen Lebensraum unterscheiden, werden sie meist mit vielerlei neuen ökologischen Bedingungen konfrontiert, an welche sie sich anpassen müssen.

Ein gutkartiertes Genom und diverse genetische Werkzeuge machen die Fruchtfliege *Drosophila melanogaster* zu einem idealen Modelsystem um die genetischen Grundlagen adaptiver Evolution zu erforschen. Als eine weltweit verbreitete Art hat sich *D. melanogaster* an Lebensräume verschiedenster Temperaturbedingungen angepasst. Trotz ihres tropischen Ursprungs im südlichen Zentralafrika, hat sie sich erfolgreich in Regionen des gemäßigten Klimas weltweit angesiedelt. Temperaturbedingte Anpassungen, die halfen mit größeren Temperaturspannen und -schwankungen sowie mit extremen Niedrigtemperaturen umzugehen, waren von Nöten um in gemäßigten Regionen zu gedeihen.

Es ist bekannt, dass schwankende Temperaturen chromatinbasierte Genregulation stören können. Schwankungen in der Umgebungstemperatur von Fruchtfliegen bewirken wiederum Schwankungen im Expressionslevel von der Polycomb group (PcG)-regulierten Genen, wobei eine höhere Expression bei niedrigeren Temperaturen auftritt. Kapitel 1 und 2 dieser Dissertation versuchen die Frage zu beantworten, ob diese Temperatursensitivität in der PcG Regulation nachteilig für die Besiedlung von gemäßigten Regionen war und aufgrund dessen durch natürliche

Selektion reduziert werden musste. Temperatursensitivität von PcG Regulation konnte in verschiedenen natürlichen Populationen von *D. melanogaster* beobachtet werden. Im Vergleich zu Populationen tropischer Herkunft wurde in Populationen aus gemäßigten klimatischen Bedingungen einheitlich ein geringer Grad an temperatursensitiver Expression gefunden. In Kapitel 1 werden Beweise erbracht, dass es aufgrund von positiver Selektion in der *polyhomeotic* (*ph*) Genregion zu einer Reduzierung von Temperatursensitivität in der PcG Regulation in Populationen aus dem gemäßigten Klima Europas kam. Die Ziele der Selektion scheinen Single-Nukleotid-Polymorphismen (SNPs) in einer relativ kleinen *cis*-regulatorischen Region zwischen den zwei PcG-regulierten Genen *polyhomeotic proximal* (*ph-p*) und *CG3835* zu sein, welche sich hochgradig zwischen europäischen und afrikanischen Populationen unterscheiden. Anhand von Reportergenexperimenten konnte nachgewiesen werden, dass diese SNPs die Genexpression beeinflussen und dass die europäischen Allele, im Gegensatz zu den afrikanischen Allelen, zu einer reduzierten Temperatursensitivität in der Expression führen. In Kapitel 2 wurde die Temperatursensitivität eines anderen PcG-regulierten Gens, *vestigial* (*vg*), in natürlichen Populationen untersucht. Dafür wurden sechs Populationen betrachtet, vier aus gemäßigten klimatischen Regionen des afrikanischen Hochgebirges und des mittel- bis nordeuropäischen Kontinents und zwei tropische Populationen aus der Ursprungsregion von *D. melanogaster*. Alle vier Populationen aus gemäßigten Regionen wiesen einen geringeren Grad an temperatursensitiver Expression gegenüber den zwei tropischen Populationen auf. Die der geringeren Sensitivität zugrundeliegende Mechanismen scheinen sich jedoch zwischen den Populationen gemäßigten Klimas zu unterscheiden.

Die Anpassung an die Temperaturverhältnisse von gemäßigten Regionen beinhaltet auch mit extremen Niedrigtemperaturen umzugehen. Starker und ausdauernder Kältestress ist ein entscheidender limitierender Faktor gemäßigten Klimas für *D. melanogaster*. Natürliche Selektion scheint für die Entwicklung erhöhter Kältetoleranz in Populationen gemäßigter Regionen verantwortlich zu sein.

Kältetoleranz ist dem Anschein nach unter der Kontrolle vieler verschiedener Gene, so wurden mehrere Loci im Genom verantwortlich für die Ausprägung dieses quantitativen Merkmals, *quantitative trait loci* (QTL), gefunden. In Kapitel 3 wurde so eine QTL Region genauer untersucht um Gene ausfindig zu machen, die ursächlich für die erhöhte Kältetoleranz in Fruchtfliegen gemäßigter Regionen sind. Daraus resultierend konnte *brinker (brk)* als neues Kandidatengen involviert in der Anpassung an Kältestress identifiziert werden.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GENERAL INTRODUCTION

When a species is expanding its range, it may colonize habitats that differ from the ancestral habitat from which it originated. Populations settling in such novel environments therefore often have to deal with plentiful of new constraints ranging from abiotic factors, the physical aspects of the environment like temperature or light regimen, to biotic factors that result from interactions with other organisms such as predation pressure or interspecific competition. In order to permanently settle in the new habitat, populations need to adapt to the novel ecological constraints they encounter. Adaptive evolution favors phenotypic traits that increase the chances of an organism to survive and reproduce and is facilitated through natural selection. Natural selection, in turn, operates on the genetic variation underlying these phenotypic traits. Such genetic differences can be of diverse nature like single nucleotide polymorphisms (SNPs), small insertions and deletions (indels), or larger structural variations. Both, newly arisen mutations and mutations already present in the population, *i.e.* standing genetic variation, can be of advantage in the new habitat and can play a part in local adaptation. Through the action of positive selection, advantageous mutations are likely to rise in frequency and might eventually become fixed in the population. Such mutations can occur in the coding region of genes leading to changes in protein sequences such as amino acid substitutions, and thus to structural changes of the protein product. Otherwise, mutations could entail regulatory changes that alter transcription, splicing, transcript stability, or other regulatory processes. To which amount each of the two types of genetic changes contribute to adaptive evolution is still debated (Hoekstra and Coyne 2007; Wray 2007), yet many cases have been described demonstrating the importance of both. For instance, a single point mutation in the gene *acetylcholinesterase type 1* (*ace-1*) led to an amino acid substitution which allows

the mosquito *Culex pipiens* to resist organophosphate and carbamate insecticides. This resistance allele is present worldwide and has risen to high frequency in areas treated with those insecticides (Lenormand *et al.* 1998; Lenormand *et al.* 1999; Labbé *et al.* 2007). A famous example for an adaptive regulatory change involves the *lactase* (*LCT*) gene. It enables humans to digest the milk sugar lactose throughout adulthood. Selection appears to have acted on certain SNPs in a *cis*-regulatory element upstream of the gene that confer persistent expression of *LCT* in adults. Lactase persistence has evolved in pastoralist populations from Africa and Europe, though independently on each continent. The *cis*-regulatory SNPs associated with adult lactase expression differ between African and European populations which indicates convergent evolution (Bersaglieri *et al.* 2004; Tishkoff *et al.* 2007; Ingram *et al.* 2009). In *Drosophila melanogaster*, an insertion of an *Accord* transposable element in the upstream regulatory region of the cytochrome P450 gene *Cyp6g1* is associated with overexpression of *Cyp6g1* and resistance to insecticides like DDT. Insecticide resistance due to *Cyp6g1* overexpression is a trait that appears to have risen to high frequency in non-African populations due to recent positive selection favoring the *Accord* insertion (Daborn *et al.* 2002; Catania *et al.* 2004; Chung *et al.* 2007; Schmidt *et al.* 2010).

Many phenotypic differences among populations are thought to be derived from differences in gene expression. Recent advances in transcriptomic technologies have allowed to identify genes that are differentially expressed between populations on a genome-wide level. Abundant among-population variation in gene expression was observed for many species, and at least part of the variation seems to have evolved by natural selection (*e.g.*, Oleksiak *et al.* 2002; Townsend *et al.* 2003; Fay *et al.* 2004; Whitehead and Crawford 2006; Storey *et al.* 2007; Voolstra *et al.* 2007; Hutter *et al.* 2008; Catalán *et al.* 2012). DNA sequence polymorphisms at loci of differentially expressed genes are often used to infer natural selection. Recent strong positive selection, for example, can leave a specific pattern of polymorphism surrounding a recently selected locus. When a rare beneficial allele is driven to fixation in a

population due to directional selection, linked neutral or slightly deleterious variants *hitchhike* with it and also rise in frequency causing the depletion of variation around the selected site (*selective sweep*). Genetic variation increases as the recombinational distance from the selected site increases resulting in a valley of low polymorphism which is expected to be larger when recombination is lower and/or selection is stronger (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989; Stephan *et al.* 1992). Another feature caused by *genetic hitchhiking* concerns the site frequency spectrum (SFS), which summarizes the counts of derived variants in a genomic region. Under positive directional selection, the SFS is expected to be shifted due to an excess of both rare and high-derived variants around the target site of selection (Braverman *et al.* 1995; Fay and Wu 2000). A third signature of a selective sweep is due to nonrandom associations of alleles in the selected region that give rise to specific linkage disequilibrium (LD) patterns such as an elevated level of LD in the early phase of the hitchhiking process and a decay of LD, across the selected sited at the end of the selective phase (Kim and Nielsen 2004; Stephan *et al.* 2006).

Based on these characteristic sequence patterns, various population genetic tools were developed which can be applied to genetic data to infer recent positive selection (*e.g.*, Fu 1997; Fay and Wu 2000; Kim and Stephan 2002; Kim and Nielsen 2004; Nielsen *et al.* 2005; Pavlidis *et al.* 2013). In particular, with the advent of modern sequencing and genotyping technologies, it has become more and more feasible to scan genome-wide SNP data for signatures of positive selection (Nielsen *et al.* 2005; Pavlidis *et al.* 2010). Furthermore, genetic differentiation can be considered in the detection of genomic regions that have been under positive selection. When a locally beneficial allele rises in frequency in one population, it may lead to increased genetic differentiation between the population and other populations that have not been subjected to selection. Such allele frequency differences between populations can be measured by F_{ST} or related statistics, and exceptionally high levels of genetic differentiation may indicate the action of positive directional selection (Lewontin and Krakauer 1973; Beaumont and Balding 2004; Foll

and Gaggiotti 2008; Bonhomme *et al.* 2010; Günther and Coop 2013). Following this approach might be of advantage, when sweep patterns as described above are not observed in the selected region. For example, the sweep signal can have vanished over time due to the accumulation of recombination events in the region (Wollstein and Stephan 2015).

However, one major challenge of inferring selection based on polymorphism patterns is that other processes can leave similar footprints in the genome. Effects of demography, for example, can shape genetic variation in similar ways as selection (Pavlidis *et al.* 2008; Pavlidis *et al.* 2010; Stephan 2010). Natural populations often have complex migration and colonization histories that can obscure the detection of selective events. If the demographic history of a species is known, it can be incorporated into the inference of selection as a null hypothesis and selection can be tested against it (Thornton and Jensen 2007; Pavlidis *et al.* 2008; Pavlidis *et al.* 2010; Stephan 2010; Wollstein and Stephan 2015). The demography of a species, however, is in general unknown. In such cases, the null hypothesis can be derived from the genomic background. This follows the rationale that demographic effects and other non-selective processes, such as population structure, influence the whole genome, whereas positive selection should only affect individual loci (Nielsen *et al.* 2005; Stephan 2010; Wollstein and Stephan 2015).

Based on the described characteristic signatures a selective event can leave in the genome, positive selection can be inferred independent of phenotypes. A drawback of this bottom-up approach is that once loci under selection have been identified it can be challenging to study their function and relevance to fitness without phenotypic information. Top-down approaches like association and linkage studies avoid this problem by starting with phenotypes that are known to differ between environments or to have functional importance, and then go on to map the loci underlying these phenotypes. Such studies try to find correlations between the studied phenotypic trait and genotypic markers (*e.g.*, SNPs). The assumption is that

markers which show strong correlations with the phenotype are in LD with the causal loci (Mackay *et al.* 2009). After the identification of causal loci, this top-down approach can be complemented by the aforementioned bottom-up approach to determine whether selection has acted on the traits under investigation. The study of adaptive evolution of cryptic coat color in deer mice (*Peromyscus maniculatus*) is an example of combining both approaches. Deer mice inhabiting the pale soils of Nebraska Sand Hills have a lighter coat color than nearby ancestral populations of prairie habitats with darker soil. There seems to be a selective advantage of crypsis against predation, and lighter pigmentation has been linked to polymorphisms within the *Agouti* gene locus, which also exhibits molecular signatures of recent positive selection (Linnen *et al.* 2009; Vignieri *et al.* 2010; Linnen *et al.* 2013).

Linkage or association mapping is also the basis of quantitative trait locus (QTL) analysis that aims to unravel the genetic basis of complex traits (Mackay 2001; Mackay *et al.* 2009). Quantitative traits are controlled by a multitude of genes and show a continuous distribution of possible phenotypes within populations. Phenotypes might differ between populations due to adaptation to different environments. By dissecting the loci underlying quantitative traits, QTL mapping can be important to investigate the local adaptation of complex traits between populations of varying environments. However, identified QTL often contain numerous genes, and pinpointing causal genes by searching for footprints of positive selection might be challenging. The reason for this is that when positive selection acts on traits that are controlled by a large number of genes, the effect of any given gene is expected to be rather small, as is the resulting signature of selection (Mackay 2001; Berg and Coop 2014; Wollstein and Stephan 2015). Alternative strategies to determine causal genes in QTL regions can include high-resolution recombination mapping, quantitative complementation tests, or gene expression analysis (Mackay 2001; Mackay *et al.* 2009).

The fruit fly *Drosophila melanogaster* is one of the most intensively studied organisms across various fields of biology. Owing to its long history as a model organism in genetic studies, it is also an ideal species to study the genetic basis of adaptation with its well-annotated genome and diverse genetic toolkit. As a cosmopolitan species, *D. melanogaster* has successfully adapted to a wide range of different environments (Lachaise *et al.* 2007). The species demographic history is now largely known which is particularly useful when searching for molecular signatures of selection, as it can be used as a null model against which selection can be tested (Thornton and Jensen 2007; Pavlidis *et al.* 2008; Pavlidis *et al.* 2010; Stephan 2010; Wollstein and Stephan 2015). Knowing the demographic history of a species also provides the opportunity to compare derived populations that have undergone adaptations to novel environments and their presumably less adapted ancestors, which can be further helpful in finding genes and genetic changes that underlie adaptation. Like the other members of the *Drosophila melanogaster* species subgroup, *D. melanogaster* is thought to be of Afrotropical origin (Lachaise *et al.* 1988; David *et al.* 2007). Recent genome-wide analyses of DNA sequence variation in various African and non-African populations support a sub-Saharan origin of *D. melanogaster* and provided further insights into the species biogeographic and demographic history (Ometto *et al.* 2005; Li and Stephan 2006; Laurent *et al.* 2011; Pool *et al.* 2012; Duchén *et al.* 2013). The center of origin appears to be in southern-central Africa, since the highest genetic diversity was observed in populations from this region (*e.g.,* Zambia and Zimbabwe) (Pool *et al.* 2012). From there, it first spread throughout sub-Saharan Africa, and then expanded further reaching the Eurasian continent after the last glaciation around 10,000 years ago (Li and Stephan 2006; Stephan and Li 2007; Pool *et al.* 2012). The out-of-Africa migration was concomitant with a drastic decrease in population size (*bottleneck*) which resulted in a significant loss of genetic variation (Li and Stephan 2006; Stephan and Li 2007; Pool *et al.* 2012). The initial non-African population may have been established in the Middle East where early human settlements in the Fertile Crescent could have helped the human

commensal to thrive. Around 5000 years ago, the original non-African population seems to have split into populations that moved on to colonize Europe and Asia, probably in the wake of human settlements (Keller 2007; Laurent *et al.* 2011). The colonization of North America appears to be even more recent, around 200 years ago, and is thought to have involved migration from Europe and Africa (Duchén *et al.* 2013).

Despite their tropical origin, many non-African and also high-altitude African populations have successfully settled in regions of temperate climate. Temperature is an important environmental factor for ectotherms such as *D. melanogaster* with major impacts on physiology and fitness (Clarke 1996). Therefore, permanent settlement in such regions was probably facilitated by thermal adaptations that helped to tolerate a wide range of temperatures. The greater range and variability in temperature as well as freezing temperatures have probably been major challenges for *D. melanogaster* while colonizing temperate environments.

Variation in temperature, for example, is known to affect chromatin-based gene regulation (Fauvarque and Dura 1993; Gibert *et al.* 2011). Genes regulated by the Polycomb group (PcG) are known to vary in their transcriptional output due to changes in ambient temperature. Decreasing the temperature, at which flies are reared or held, results in higher expression of genes regulated by PcG proteins (Fauvarque and Dura 1993; Chan *et al.* 1994; Zink and Paro 1995; Bantignies *et al.* 2003; Gibert *et al.* 2011). PcG proteins are important regulators of development and cell differentiation and are present in all metazoans (Simon and Kingston 2009; Beisel and Paro 2011). Together with the Trithorax group (TrxG) of proteins, PcG proteins are involved in the regulation of several hundreds of genes in *Drosophila*. Many of the target genes encode transcription and signaling factors with important roles in development and cell-fate specification (Schuettengruber *et al.* 2009; Schwartz *et al.* 2010; Simon and Kingston 2009; Kassis and Brown 2013; Steffen and Ringrose 2014). Both groups of proteins act in large multiprotein complexes to control transcription

by altering the chromatin structure through histone modifications and other mechanisms (Kassis and Brown 2013; Steffen and Ringrose 2014). PcG and TrxG proteins function in an antagonistical manner to maintain repressed and activated transcription states, respectively. They are recruited to their target genes by *cis*-regulatory DNA elements called Polycomb response elements (PREs) that can preserve the memory of an activated or repressed state of their target genes over several cell generations. PREs consist of different binding sites for proteins that appear to be involved in PcG/TrxG recruitment (Kassis and Brown 2013; Steffen and Ringrose 2014). They regulate their target genes in combination with other regulatory DNA sequences (*i.e.,* enhancers) in a cell- or tissue-specific fashion. This interplay between the different types of regulatory DNA modifies transcription in such a way that enhancers initially determine the level of transcription which is then epigenetically maintained by PREs (Schwartz *et al.* 2010; Kassis and Brown 2013; Steffen and Ringrose 2014).

Since PcG regulation seems to be disrupted by lower temperatures, it was hypothesized that adaptation to temperate environments might have included the buffering of this thermosensitive regulatory process. A higher degree in expression plasticity due to a greater range and variability in temperature might have been detrimental, and natural selection might have acted to stabilize the transcriptional output of PcG-regulated genes in populations adapting to temperate climates (Levine and Begun 2008). Several studies (Harr *et al.* 2002; Levine and Begun 2008; Gibert *et al.* 2011) suggest that adaptation to temperate environments included selection acting on genes involved in PcG regulation. Furthermore, a genome-wide expression analysis found more genes that exhibit expression plasticity due to rearing temperature in tropical populations than in temperate populations (Levine *et al.* 2011). This indicates that thermosensitivity of gene expression has been detrimental for *D. melanogaster* while settling in temperate environments and has been reduced by selection.

The chapters 1 and 2 of this thesis present evidence for reduced thermosensitivity of expression of PcG-regulated genes in temperate populations of *D. melanogaster* at different gene loci. Both loci harbor known PcG target genes that were identified by genome-wide chromatin immunoprecipitation (ChIP) experiments (Schuettengruber *et al.* 2009; Schwartz *et al.* 2010), as well as experimentally validated and well-studied PREs (Fauvarque and Dura 1993; Okulski et al. 2011). Thermosensitivity of PcG target genes was observed in wild-type flies from different natural populations. A lower degree of temperature-induced expression plasticity was consistently detected in temperate populations when compared to tropical populations. The observations at both loci, therefore, seem to support the hypothesis that temperature-induced expression plasticity needed to be buffered when *D. melanogaster* settled in temperate climates (Begun and Levine 2008)

In chapter 1, evidence is presented for recent positive selection acting on the *polyhomeotic* gene region in populations from temperate environments. The gene region harbors the two known PcG target genes *polyhomeotic-proximal (ph-p)* and *CG3835* (Schuettengruber *et al.* 2009; Schwartz *et al.* 2010). While the function of *CG3835* is still unknown, the PcG target gene *ph-p* is in itself a PcG protein belonging to one of the major PcG protein complexes (Kassis and Brown 2013; Steffen and Ringrose 2014). Five SNPs were identified as the potential targets of positive selection that are highly differentiated between African and European populations. The SNPs are located in the intergenic region of *ph-p* and *CG3835* that also includes a PRE and the promoters of the two genes. Using transgenic reporter gene assays, these sequence variants could be linked to gene expression differences in response to rearing temperature with a reduced thermosensitivity for European variants.

In chapter 2, we provide further evidence that thermosensitivity of PcG regulation has been buffered in temperate populations of *D. melanogaster*. The PcG target gene *vestigial* (*vg*) encodes a transcription factor that plays an important role in development and patterning of the wing (Williams *et al.* 1991; Lindsley and Zimm

1992; Kim *et al.* 1996). Thermosensitive *vg* expression was observed in six different natural populations including four temperate populations from Europe and one from high-altitude Africa, as well as two tropical populations from the ancestral species range. The degree of temperature-induced expression plasticity was consistently lower in temperate than tropical populations.

Another major challenge for *D. melanogaster* while colonizing temperate environments have been freezing temperatures. Severe cold stress is one of the main limiting factors imposed on *D. melanogaster* by temperate climates, and thus increasing cold tolerance has probably played a dominant role in the adaptation to such climates. Temperate and tropical populations, indeed, differ in their tolerance to low-temperature extremes, and increased cold tolerance is thought to have evolved by natural selection (Hoffmann *et al.* 2002; Schmidt *et al.* 2005). Cold tolerance is a quantitative trait that appears to be highly polygenic, and that has been mapped to different QTL in the genome (Morgan and Mackay 2006; Norry *et al.* 2008; Svetec *et al.* 2011; Mackay *et al.* 2012). In Chapter 3, such a QTL region was fine-mapped in an attempt to localize candidate genes that contribute to increased cold tolerance. In a previous study, this X-chromosomal QTL region was found to be responsible for differences in cold tolerance between temperate European and tropical African populations (Svetec *et al.* 2011). The fine-mapping approach included quantitative complementation tests, gene expression analysis, as well as population genetic analysis; and yielded a new candidate gene putatively involved in cold stress adaptation.

Positive Selection at the *Polyhomeotic* Locus Led to Decreased Thermosensitivity of Gene Expression in Temperate *Drosophila melanogaster*

Susanne Voigt,[1] Stefan Laurent, Maria Litovchenko, and Wolfgang Stephan
Section of Evolutionary Biology, Department of Biology II, University of Munich, 82152 Planegg-Martinsried, Germany

ABSTRACT *Drosophila melanogaster* as a cosmopolitan species has successfully adapted to a wide range of different environments. Variation in temperature is one important environmental factor that influences the distribution of species in nature. In particular for insects, which are mostly ectotherms, ambient temperature plays a major role in their ability to colonize new habitats. Chromatin-based gene regulation is known to be sensitive to temperature. Ambient temperature leads to changes in the activation of genes regulated in this manner. One such regulatory system is the Polycomb group (PcG) whose target genes are more expressed at lower temperatures than at higher ones. Therefore, a greater range in ambient temperature in temperate environments may lead to greater variability (plasticity) in the expression of these genes. This might have detrimental effects, such that positive selection acts to lower the degree of the expression plasticity. We provide evidence for this process in a genomic region that harbors two PcG-regulated genes, *polyhomeotic proximal* (*ph-p*) and *CG3835*. We found a signature of positive selection in this gene region in European populations of *D. melanogaster* and investigated the region by means of reporter gene assays. The target of selection is located in the intergenic fragment between the two genes. It overlaps with the promoters of both genes and an experimentally validated Polycomb response element (PRE). This fragment harbors five sequence variants that are highly differentiated between European and African populations. The African alleles confer a temperature-induced plasticity in gene expression, which is typical for PcG-mediated gene regulation, whereas thermosensitivity is reduced for the European alleles.

KEYWORDS positive selection; gene regulation; environmental sensitivity; polycomb group

DROSOPHILA melanogaster is a species that has colonized all continents on Earth. Now a cosmopolitan species with a worldwide distribution, it started its global spread from its sub-Saharan ancestral range relatively recently. Its origin is thought to be in southern-central Africa from which it first expanded through Africa and finally reached the Eurasian continent on the order of 10,000 years ago (David and Capy 1988; Lachaise and Silvain 2004; Li and Stephan 2006; Stephan and Li 2007; Pool *et al.* 2012). This settlement was accompanied by a severe population size bottleneck and involved a significant loss of genetic variation (Li and Stephan 2006; Stephan and Li 2007; Pool *et al.* 2012). The colonization of Europe and Asia from its original source population appears to be even more recent since European and Asian populations share a most recent common ancestor (MRCA) ~5000 years ago (Laurent *et al.* 2011).

For insects, which are mostly ectotherms, differences in temperature are one of the most important environmental variables that influence the distribution of species in nature (Clarke 1996). In the temperate climate of Europe, the range of possible temperatures is probably one of the major challenges *D. melanogaster* was confronted with during colonization.

Chromatin-based gene regulation is known to be sensitive to temperature (Fauvarque and Dura 1993; Gibert *et al.* 2011). In the case of the Polycomb group (PcG)-mediated gene regulation, it is known that genes under the control of this group of proteins have a higher transcriptional output when flies are reared or held at lower temperatures than at higher ones (Fauvarque and Dura 1993; Chan *et al.* 1994; Zink and Paro 1995; Bantignies *et al.* 2003; Gibert *et al.*

doi: 10.1534/genetics.115.177030
Manuscript received February 25, 2015; accepted for publication April 2, 2015; published Early Online April 8, 2015.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177030/-/DC1.
[1]Corresponding author: Section of Evolutionary Biology, Department of Biology II, University of Munich, Grosshaderner Str. 2, 82152 Planegg-Martinsried, Germany. E-mail: voigt@biologie.uni-muenchen.de

2011). The PcG and another group of proteins, the Trithorax group (TrxG), act antagonistically to epigenetically maintain repressed and activated transcription states, respectively. They act through *cis*-regulatory DNA elements called Polycomb response elements (PREs), which recruit the proteins of the two groups to their target genes. PREs regulate their target genes in combination with other regulatory DNA sequences (*i.e.*, enhancers) in a cell- or tissue-specific manner. This interplay modifies the expression of PcG-regulated genes in such a way that enhancers initially determine the level of transcriptional output, which is subsequently epigenetically maintained by PREs (Schwartz *et al.* 2010; Kassis and Brown 2013; Steffen and Ringrose 2014).

Several recent studies (Harr *et al.* 2002; Levine and Begun 2008; Gibert *et al.* 2011) explored the question of whether the temperature-induced expression plasticity of PcG-regulated genes may have been detrimental to *D. melanogaster* while settling in temperate environments. These studies suggest that adaptation included selection acting to buffer this thermosensitive process in temperate populations.

In this study, we provide evidence for selection acting in *cis* to buffer the temperature-induced expression plasticity of PcG regulation in populations adapted to temperate environments. We carried out population genetic analyses to show that a DNA sequence region between the two PcG-regulated genes *polyhomeotic proximal* (*ph-p*) and *CG3835* has been the target of a selective sweep in European populations of *D. melanogaster*. Furthermore, using transgenic reporter gene assays, we demonstrate that sequence variation in this 5-kb selected fragment mediates differences in gene expression between European and African sequence variants. Temperature-sensitive expression is observed in the case of the African alleles but not in the European ones. These results are consistent with positive selection favoring *cis*-regulatory polymorphisms that led to decreased thermosensitivity of gene expression in temperate populations.

Materials and Methods

Fly lines and sequence data

Assembled full genome sequences were taken from the *Drosophila* Population Genomics Project (DPGP) (http://www.dpgp.org), including those from 133 sub-Saharan African lines [among them the Zambian population sample (Siavonga) of 27 lines] and a French population sample (Lyon, 8 lines). We additionally analyzed a Dutch population from Leiden consisting of 10 lines and two Malaysian samples from Kuala Lumpur and Kota Kinabalu consisting of 7 and 16 lines, respectively. Full genomes for the Dutch and Malaysian samples were assembled following the approach of Pool *et al.* (2012) and are available at http://evol.bio.lmu.de/downloads. Nucleotides with known admixture or identity-by-descent according to Pool *et al.* (2012) were replaced with missing value labels in the analysis. The same was done for sites exhibiting heterozygosity since heterozygotes are not expected in genome data from haploid embryos (Pool *et al.* 2012). Additionally, 12 lines of the aforementioned Dutch population and 12 lines of one from Zimbabwe (Lake Kariba) were fully sequenced between positions 2,030,513 and 2,059,036 on the X chromosome (FlyBase release 5), applying the Sanger method (Sanger *et al.* 1977). This method was also used to sequence fragments containing the polymorphisms of interest in population samples from Siavonga, Zambia (10 lines); Munich, Germany (12 lines); Umea, Sweden (14 lines); and Kuala Lumpur, Malaysia (11 lines). All Sanger sequences were deposited in GenBank (accession nos. KR024038–KR024162).

Population genetic analysis

To analyze DNA sequence polymorphisms in a 73-kb region around the *ph* locus, sequences of the Zambia, Dutch, and French population samples generated by the DPGP were used. Nucleotide diversity was estimated in terms of π (Tajima 1989) and divergence was calculated against a *D. simulans* sequence (Hu *et al.* 2013). The composite-likelihood ratio (CLR) test of positive selection was performed, applying the software SweeD (Pavlidis *et al.* 2013). It computes the CLR between a selective sweep model and a neutral model based on the background genomic patterns of polymorphism (Kim and Stephan 2002). We ran the program on the complete X chromosome and calculated the significance threshold (95th quantile) by generating neutral coalescent simulations, using the demographic model of Laurent *et al.* (2011). To improve the power of the test statistic, the European sample was extended by adding the French population sample (Pavlidis *et al.* 2010, 2013) and two additional site classes of the site-frequency spectrum (SFS) consisting of sites that are monomorphic in the European sample and polymorphic in the Zambian one (Nielsen *et al.* 2005). Polarization was done against *D. simulans* (Hu *et al.* 2013).

Outlier analyses were performed using BayeScan version 2.1 (Foll and Gaggiotti 2008), a Bayesian method based on a logistic regression model that separates locus-specific effects of selection from population-specific effects of demography. F_{ST} coefficients (Beaumont and Balding 2004) are estimated and decomposed into a population-specific component (β) and a locus-specific one (α). Departure from neutrality at a given SNP locus is assumed when α is significantly different from zero. Positive values of α suggest positive directional selection, whereas negative α-values indicate balancing selection. BayeScan runs were carried out using default parameters for a 300-kb genomic window around the *ph* locus with sequences from seven European and African populations. These included samples from The Netherlands (10 lines), France (8 lines), Cameroon (10 lines), Gabon (9 lines), Ethiopia (8 lines), Rwanda (27 lines), and Zambia (27 lines).

Expression analysis in whole adult flies

Gene expression was analyzed in whole adult flies from the aforementioned population samples from The Netherlands and Zimbabwe. Flies were reared on a standard cornmeal–molasses medium at ~28° and 18° with a 14/10-hr light/dark

cycle. Expression was measured in 11 fly lines per population. For each line, RNA was extracted from five males and five females (aged 4–6 days). RNA extraction including DNase I digestion was performed using the MasterPure RNA Purification Kit (Epicentre, Madison, WI; http://www.epibio.com). RNA purity was assessed via the ratio of absorbances at 260 and 280 nm ($A_{260}/A_{280} > 1.8$). It was then reverse transcribed into complementary DNA (cDNA), using random primers and SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, CA; http://www.lifetechnologies.com). RT-qPCR reactions were run with iQ SYBR Green Supermix (Bio-Rad, Hercules, CA; http://www.bio-rad.com) on a CFX96 real-time PCR cycler (Bio-Rad). Primers for target and reference genes were designed, applying the QuantPrime software (Arvidsson *et al.* 2008). Per fly line, two biological replicates were run in duplicates. No template controls (NTCs) were included to control contamination and primer specificity was confirmed by melting-curve analysis. Relative expression was calculated using the qBase relative quantification framework (Hellemans *et al.* 2007). Both reference genes (*RpL32* and *RpS20*) were stably expressed across samples. This was assessed by calculating the coefficient of variation and the *M* stability parameter according to Hellesmans *et al.* (2007). Log-transformed normalized relative quantities were subjected to a paired *t*-test to test for statistically significant expression differences.

Reporter gene assays

The genomic region between *ph-p* and *CG3835* reaching from 2,030,598 to 2,035,598 on the X chromosome in FlyBase release 5 (Pierre *et al.* 2014) was PCR amplified from one Dutch and one Zimbabwean strain (NL01 and ZK186, respectively), using the primers 5′-GCCACAGTCACAGCACTAAGT-3′ and 3′-CCTTTCATCCATAAGTCAGTG-5′. The PCR products were cloned directly into the "*pCR4Blunt-TOPO*" vector (Invitrogen). The insert was then excised as a *Hin*dIII/*Not*I fragment and cloned into the "*placZ-2attB*" integration vector (Bischof *et al.* 2007). The identity and orientation of the cloned fragments were confirmed by restriction analysis and sequencing. Integration vector DNA was purified with the QIAprep Spin Miniprep Kit (QIAGEN, Hilden, Germany; http://www.qiagen.com) and used for microinjection of early-stage embryos of the *ΦX86Fb* strain (*attP* site at cyological band 86Fb). This strain includes a stable source of ΦC31 integrase on the X chromosome. The integration site used was selected by the criteria of no binding of PcG/TrxG proteins and no occurrence of their specific histone marks within a window of ±5 kb around the site. Following microinjection, viable flies were crossed to a "white$^-$" strain to remove the integrase and establish stable lines. Resulting offspring were screened for red eye color as a marker of successful transformants.

Reporter gene expression was measured in brains and midguts of third instar larvae via RT-qPCR. Flies containing one copy of the inserted construct were grown on a standard cornmeal–molasses medium at 28° and 17° with a 14/10-hr light/dark cycle. Five females and five males were allowed to mate and oviposit for 3 and 7 days at 28° and 17°, respectively. Tissue of the resulting progeny was dissected and immediately stored in RNAlater (QIAGEN). RNA extraction and RT-qPCR were performed as described above. Primer sequences for the *lacZ* reporter gene were taken from Zhang *et al.* (2013). For normalization the two aforementioned reference genes (*RpL32* and *RpS20*) were used. Three biological replicates per construct in the particular tissue at the particular rearing temperature were run in triplicates. Negative controls included NTCs and no reverse-transcription controls (NRTs) to exclude contamination. Furthermore, negative controls also consisted of midgut and brain dissections of larvae reared at the two different temperatures of an "empty" *ΦX86Fb* strain without any integrated constructs. Both reference genes were stably expressed across samples. Log-transformed normalized relative quantities were calculated as described above and subjected to a Welch two-sample *t*-test to test for statistically significant expression differences between the different rearing temperatures and transgenic constructs. False discovery rate (FDR) was controlled using the multiple-testing correction method of Benjamini and Hochberg (1995).

Results

DNA sequence polymorphism in the ph region

Full-genome data provided by the *Drosophila* Population Genomics Project were used to analyze a 73-kb genomic region of intermediate recombination rates (Fiston-Lavier *et al.* 2010) that is located on the X chromosome between positions 1,990,000 and 2,063,000 (FlyBase release 5). In an African population sample from Siavonga, Zambia, a reduction of variation in the region is observed (Figure 1A) that overlaps with the valley of low polymorphism detected in previous studies in a Zimbabwean population sample from Lake Kariba (Beisswanger *et al.* 2006; Beisswanger and Stephan 2008). In these previous studies, evidence was presented that this reduction of polymorphism originated most likely from the action of positive directional selection in the recent past causing a selective sweep in the ancestral species range. As was shown before (Beisswanger *et al.* 2006), the Dutch population sample from Leiden harbors an even more pronounced valley of low polymorphism that spans >60 kb (Figure 1A).

Likelihood analysis of selective sweeps in the European population

The 73-kb region shown in Figure 1A was submitted to a composite-likelihood-ratio test that is based on the site-frequency spectrum used by SweeD (Pavlidis *et al.* 2013). Since a larger sample size may lead to more accurate results in distinguishing selective sweeps from demographic events and inferring the genomic position of sweeps (Pavlidis *et al.* 2010, 2013), the French population sample from DPGP (Pool *et al.* 2012) was added to the Dutch sample to obtain a larger European data set. SweeD was run on the complete

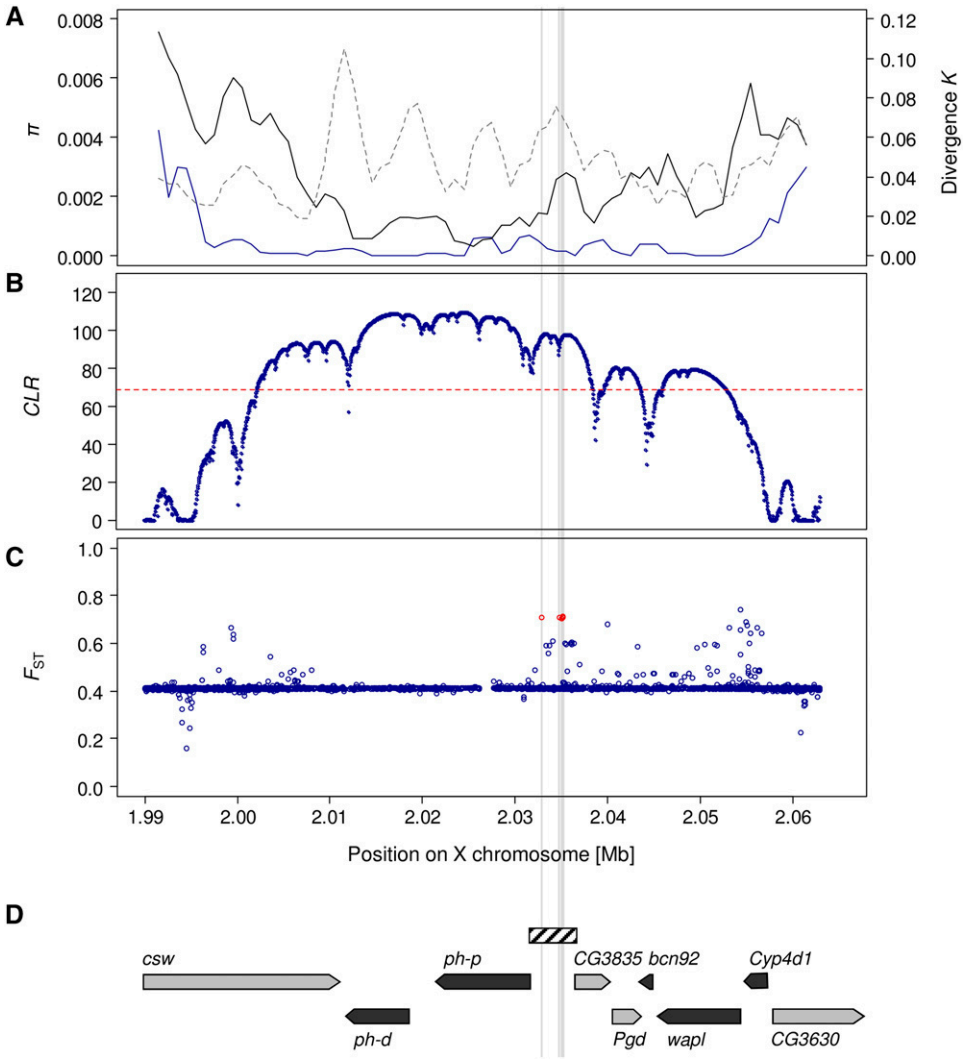


Figure 1 Evidence of positive selection in the *Drosophila polyhomeotic* (*ph*) genomic region. (A) Nucleotide variability (π) in European population samples from The Netherlands (blue line) and an African sample from Zambia (black line), as well as divergence (K) to *D. simulans* (dashed line), is shown in a sliding window across the region (window size = 3000 bp, step size = 1000 bp). (B) Composite-likelihood-ratio (CLR) analysis of a selective sweep using the site-frequency-based test statistic of SweeD (Pavlidis *et al.* 2013) for a European sample consisting of the Dutch and the French lines. The significance threshold of the test statistic is given by the dashed line. (C) BayeScan F_{ST} coefficients (Foll and Gaggiotti 2008) averaged over seven population samples from Europe and Africa (see *Materials and Methods*). BayeScan results based on a 300-kb window surrounding the *ph* locus (of which 73 kb are shown here) reveal a total of six outlier SNPs (FDR = 0.05) within the selective sweep region that was identified by the SweeD test. One of these is already segregating in the African samples, in contrast to the other five outlier SNPs that are located in the *ph-p*/*CG3835* intergenic region. The latter are highlighted in red and by gray lines across panels. (D) The positions of the genes contained within the region. The arrowheads indicate the direction of transcription. The hatched box corresponds to the portion used for reporter gene analysis.

X chromosome and the CLR profile of the region of interest is shown in Figure 1B. The SweeD test provided a likelihood profile that is much broader than the valley of reduced variation in Africa (Beisswanger *et al.* 2006; Beisswanger and Stephan 2008) and spans almost the entire region of very low polymorphism found in the European sample (Figure 1, A and B).

Genetic differentiation between the European and African population samples in the ph region

Because a large fraction of the region of low variation in Europe contains no or very few SNPs, the CLR test cannot be used to identify the targets of selection. Instead, following Wilches *et al.* (2014), we utilized genetic differentiation between African and European populations to obtain model-based F_{ST} coefficients for each SNP (Foll and Gaggiotti 2008; Riebler *et al.* 2008). BayeScan analyses (Foll and Gaggiotti 2008) were run on an X-chromosomal window of 300 kb surrounding the *ph* locus between positions 1,900,000 and 2,200,000 (FlyBase release 5). SNP data from seven European and African populations were considered, which included samples from The Netherlands, France, Cameroon, Ethiopia, Gabon, Rwanda, and Zambia. Including all seven population samples with a total of 11,894 SNPs, BayeScan yielded 22 significant outlier SNPs (FDR = 0.05) with positive α-values, suggesting that these SNPs are targets of positive directional selection (Supporting Information, Table S2).

Six of these 22 outliers are located in the region of significant CLR values. While one of those 6 is already segregating in the African samples, the other 5 are monomorphic in the population samples from Africa. The former is also identified as an outlier SNP (FDR = 0.07), when the European samples are excluded from the BayeScan analysis (position on X chromosome, 2,039,998; see Table S2). These results suggest that the differentiation of this SNP started in Africa, whereas the differentiation of the other 5 SNPs occurred outside the ancestral species range. The 5 SNPs are located in the intergenic region between *ph-p* and *CG3835* (Figure 1). Except for 1 of the 5, in which case no outgroup sequence was available, derived sequence variants are observed for all lines of the two European samples and ancestral variants for all lines of the five African samples.

Thus, the 5 SNPs mark two distinct haplotype groups, a derived one and an ancestral one. Complementing the DPGP data with Sanger sequencing of lines from populations of Europe, Africa, and Asia, it is observed that the derived haplotype group is in very high frequency in Europe (44 of 46 lines) and the ancestral one is in very high frequency in Africa (143 of 144 lines). Interestingly, in the two Asian population samples from Malaysia a third haplotype group is quite abundant (15 of 32 lines) (Figure 2). This group is a recombinant between the derived and the ancestral haplotypes where the second and third sequence variants are identical with those of the derived group and the rest with those of the ancestral one. In addition to the recombinant haplotype group, 8 lines of the derived group and 9 lines of the ancestral one make up the Southeast Asian samples (Figure 2). Thus, in contrast to the European population samples, in which the derived variants of these 5 SNPs are near fixation, the derived alleles occur in intermediate frequencies in the Asian samples.

Taken together, our observations suggest that the *ph* region was hit not only by a selective event causing a sweep in the ancestral African region, but also by another sweep that may have occurred outside the ancestral range, leading to the high frequency of the derived haplotype in Europe.

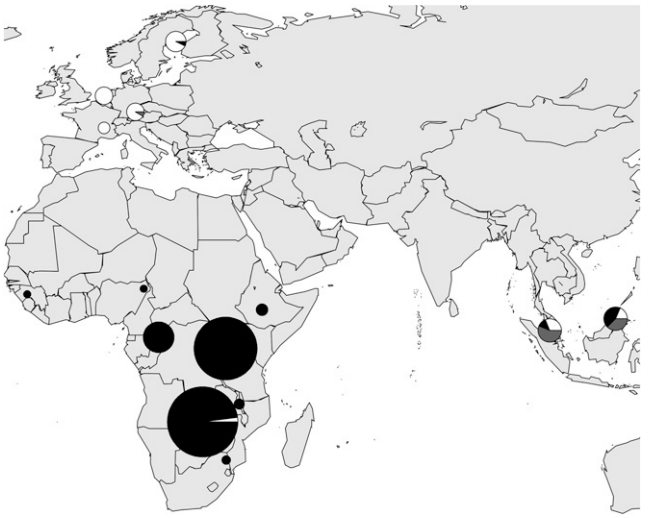


Figure 2 Geographical distribution of haplotype groups in the intergenic region between *ph-p* and *CG3835*. The derived, ancestral, and recombinant haplotype groups are shown in white, black, and gray, respectively. The area of the pie chart is proportional to the size of the population sample of a given region.

Sanger sequencing of the European sweep region

Accurate detection of short insertions and deletions is still difficult using next-generation sequencing. Therefore, to exclude insertion/deletion polymorphisms as possible targets of selection in Europe, we additionally sequenced the region of interest in population samples from Europe and Africa, using the Sanger method (see *Materials and Methods*). Since the target of selection in the European samples appears to be located in the upstream half of the valley of reduced variation, these 30 kb were fully sequenced in the Dutch population sample from Leiden and a sample from the ancestral range of *D. melanogaster* from Lake Kariba, Zimbabwe (Pool *et al.* 2012). Sanger sequencing supports the results of the full-genome data set in that the highest genetic differentiation is observed in the intergenic region of *ph-p* and *CG3835* and no highly differentiated insertions/deletions were found between the European and African samples.

Expression analysis and reporter gene assays

The intergenic region between *ph-p* and *CG3835* contains a PRE and the promoters of the two genes. This PRE was experimentally validated by using reporter gene assays and different PcG mutant backgrounds (Fauvarque and Dura 1993). The fragment exhibiting PRE activity as demonstrated by Fauvarque and Dura (1993) spans nearly the whole intergenic region and overlaps with the promoters of both genes. Since PREs function in an orientation-independent fashion (Busturia *et al.* 1997; Americo *et al.* 2002; Kozma *et al.* 2008), which was shown in particular for this specific PRE (Fauvarque and Dura 1993), and *ph-p* and *CG3835* are known PcG target genes as seen by chromatin immunoprecipitation (ChIP) experiments (Schuettengruber *et al.* 2009; Schwartz *et al.* 2010), it is likely that both genes are under the control of the PRE residing in the region between them. Thermosensitivity in expression as often found for genes regulated by PcG proteins was observed for *ph-p* in its natural genetic environment. In a rather crude experiment using whole adult flies reared at different temperatures, we observed a significantly higher expression when temperature was lower. However, this effect was significant only for the Zimbabwean population sample, not for the Dutch one (Figure 3).

To further test whether the five SNPs that define the derived and ancestral haplotype groups have an effect on gene expression, four reporter gene constructs were created in which the 5-kb intergenic region from either a Dutch (derived sequence variants) or a Zimbabwean (ancestral sequence variants) strain was fused to the *Escherichia coli lacZ* gene. The *lacZ* reporter gene was driven by either the *ph-p* promoter or the one of *CG3835* (Figure 4). Reporter gene constructs were inserted into a common genetic background using the site-specific ΦC31 integration system (Bischof *et al.* 2007), allowing the comparison of the expression of the different constructs at the same genomic position in an otherwise identical genetic background. It was also checked that the selected integration site was not located in a PcG-regulated genomic region (see *Materials and Methods*).

To explore whether a temperature-sensitive pattern can be observed in the regulation of the two PcG target genes, flies were reared at 17° and 28° and messenger RNA (mRNA) expression levels of the *lacZ* reporter were quantified in the different transgenic lines via RT-qPCR. Since *ph-p* is highly expressed in the brain and *CG3835* in the midgut of third instar larvae (Chintapalli *et al.* 2007), these tissues were dissected for the expression analysis. To measure *lacZ* mRNA

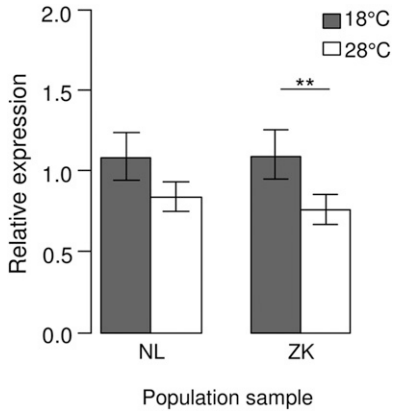


Figure 3 Expression of *ph-p* in population samples from The Netherlands (NL) and Zimbabwe (ZK). mRNA abundance in adult flies reared at 18° (shaded bars) and 28° (open bars) was measured via RT-qPCR. Error bars represent the 95% confidence interval. $^{**}P < 0.01$.

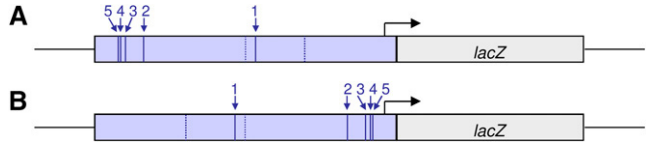


Figure 4 Schematic representation of the reporter gene constructs used. (A and B) *ph-p* promoter (A) and *CG3835* promoter (B) are driving the *E. coli lacZ* gene. Blue lines give the relative location of the sequence variants in the 5-kb intergenic region containing the PRE and the promoters of *ph-p* and *CG3835*. The solid lines and numbers correspond to the five highly differentiated candidate SNPs that mark the different haplotype groups. Dashed lines indicate the other two sequence variants that differ between the African and European strains from which the intergenic region incorporated in the constructs originated.

levels, when expression of both genes is low, RT-qPCRs were also run for lines in which the reporter gene is controlled by the *ph-p* promoter on samples from the larval midgut and in those with *lacZ* driven by the *CG3835* promoter on samples from the larval brain.

As expected from the endogenous expression, when the *ph-p* promoter was driving the reporter gene, a higher *lacZ* expression was observed in the brain than in the midgut, and vice versa in constructs with *lacZ* under the control of the *CG3835* promoter (Table S1). For all tissues and treatments, *lacZ* expression due to the *ph-p* promoter was higher than expression due to the *CG3835* promoter (Figure 5). Constructs carrying the ancestral sequence variants exhibited a temperature-sensitive expression pattern in the midgut while no such temperature-dependent expression difference was detected in the brain and for those constructs with the derived sequence variants (Figure 5). In the case of the ancestral sequence variants, midgut expression was approximately twofold higher when larvae were reared at 17° than at 28°. This difference due to temperature in *lacZ* expression was highly significant for the *ph-p* promoter, whereas no significance was reached for the promoter of *CG3835* (Figure 5, C and D, and Table S1). For the constructs with the derived sequence variants and *lacZ* under the control of the *ph-p* promoter, a significantly higher reporter gene expression at 28°, compared to that of constructs with the ancestral variants, led to a buffering of the thermosensitivity (Figure 5, C and D).

Therefore, we may conclude that nucleotide differences between the European and African sequences in the intergenic region have led to differences in gene expression. However, which of these differences confer the observed expression differences is currently unknown. In addition to the five candidate SNPs, there are two other sites in the 5-kb insert that differ between the two fly strains (NL01 and ZK186) from which the fragment was taken (Figure 4). One of these two sites is upstream of the first candidate SNP and the other one is downstream of this SNP (Figure 4B). The latter harbors a derived variant in the African line that is rare in Africa (6 of 139 lines) and not found in Europe, whereas the former one is also highly differentiated between Africa and Europe with the derived variant in high frequency in Europe (42 of 42 lines) and rare in Africa (7 of 138 lines). The seven sites that differ between the African and European reporter gene constructs are all candidates responsible for the observed differences in *lacZ* expression (Figure 4 and Figure 5). However, only the highly differentiated SNPs are expected to be causative if selection is responsible for the observed expression differences. Each of these SNPs has the potential to insert or delete a transcription factor binding site (TFBS) motif or change its binding affinity (Hauenschild *et al.* 2008), located either in the PRE or in any other regulatory element in the *ph-p/CG3835* intergenic region. Interestingly, for the fifth of the candidate SNPs (Figure 4), the derived variant creates the Grh consensus sequence experimentally identified by Blastyák *et al.* (2006) and a Dsp1 consensus sequence that was demonstrated to be important in PcG recruitment (Déjardin *et al.* 2005). The derived state of the aforementioned additional highly differentiated SNP upstream of the first candidate SNP leads to the insertion of a motif, a GTGT sequence, which was shown to be functional in PRE activity in a number of studies (Kassis and Brown 2013).

Discussion

As was shown before (Beisswanger *et al.* 2006; Beisswanger and Stephan 2008), the genomic region around the *ph* locus exhibits a strong reduction in nucleotide polymorphism in *D. melanogaster* populations from Africa and Europe. Thus, the data suggest positive directional selection acting at this locus, leading to a selective sweep. In the previous studies, however, the question remained whether the sweep in the European population is independent of the African one or a result of a *trans*-population sweep that arose in Africa before the colonization of Europe (Beisswanger *et al.* 2006; Beisswanger and Stephan 2008). The much more pronounced reduction in nucleotide diversity in Europe could just be a product of the severe population size bottleneck *D. melanogaster* underwent during its migration out of Africa, and this bottleneck could

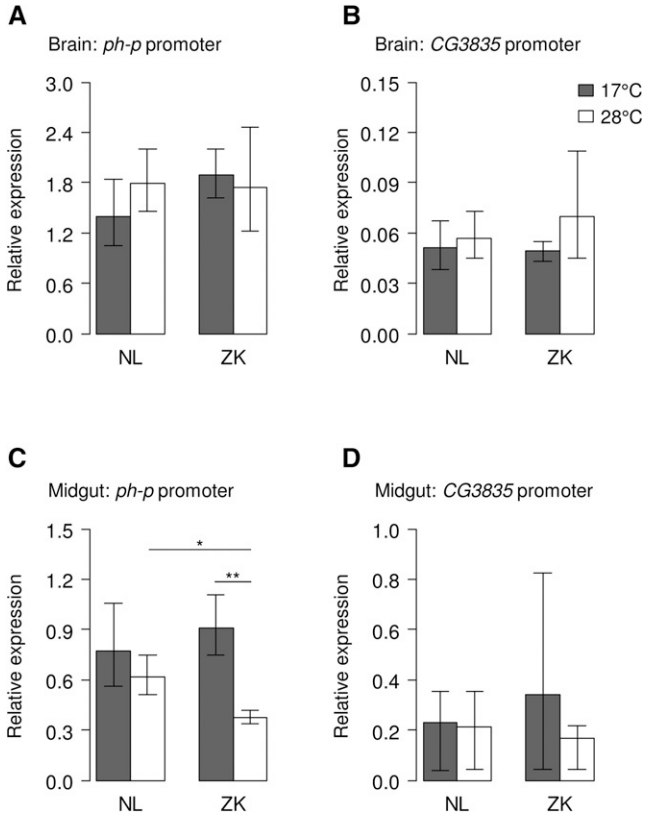


Figure 5 Reporter gene expression in larval tissues at two different rearing temperatures. mRNA abundance was measured via RT-qPCR in brains (A and B) and midguts (C and D) from third instar larvae reared at either 17° (shaded bars) or 28° (open bars). In A and C the *ph-p* promoter is driving *lacZ* expression whereas in B and D *lacZ* is under the control of the CG3835 promoter. The intergenic region containing the PRE and the promoters was taken from either a European strain from The Netherlands (NL) or an African one from Zimbabwe (ZK); that is, constructs were carrying either the derived or the ancestral sequence variants, respectively. Error bars represent the 95% confidence interval. $*P < 0.05$, $**P < 0.01$ (FDR = 0.05).

also be the cause for the very high differentiation of the genomic region between *ph-p* and *CG3835*. However, since Asian and European populations share a MRCA after the out-of-Africa bottleneck (Laurent *et al.* 2011), and populations from Asia show a high genetic diversity in the aforementioned intergenic region, it is unlikely that the bottleneck is responsible for the high frequency of the derived sequence variants found in Europe.

The genes *ph-p* and *CG3835* flanking this highly differentiated region are known PcG target genes (Schuettengruber *et al.* 2009; Schwartz *et al.* 2010) and harbor an experimentally validated PRE between each other (Fauvarque and Dura 1993). PcG-regulated genes are temperature sensitive in their expression; *i.e.*, their transcriptional output is higher when flies are reared or held at lower temperatures than at higher ones (Fauvarque and Dura 1993; Chan *et al.* 1994; Zink and Paro 1995; Bantignies *et al.* 2003; Gibert *et al.* 2011). This phenomenon prompted the hypothesis that if cold temperatures disrupt PcG regulation, then adaptation to temperate environments should include the buffering of this expression plasticity (Levine and Begun 2008). Natural selection would then act to stabilize the transcriptional output, leading to a lower degree of gene expression plasticity in response to varying temperatures. As a consequence, thermosensitivity of PcG target gene expression would be reduced by limiting the influence of the environment. A genome-wide expression analysis indeed identified more genes with expression plasticity due to rearing temperature in tropical compared to temperate populations (Levine *et al.* 2011). Our study also supports the reduced thermosensitivity of PcG target gene expression in temperate populations. The data suggest temperature sensitivity of PcG target gene expression in African populations that was selected against in populations from Europe to stabilize the transcriptional output across temperatures. At the locus under study, this was observed for the expression of *ph-p* in the natural genetic background and for larval midguts, using reporter gene assays. The reporter gene analysis linked the SNPs that were detected as likely targets of positive selection in Europe to the European stabilized gene expression. Rearing temperature had no effect on gene expression in larval brains. Possible explanations could be that brain expression is under a greater selective pressure against gene expression variability, and so the expression level is already less environmentally sensitive in the African populations, or that there is no thermosensitivity of expression in the larval brain. In addition, for the *ph-p* promoter-driven expression, the data indicate that higher transcriptional output at lower temperatures is not in itself detrimental. The greater variability in expression due to a higher degree of variation in temperature in temperate climates, however, seems to have been disadvantageous and needed to be reduced by the action of selection during the colonization of Europe.

More recent studies focused only on selection stabilizing the temperature-sensitive transcriptional output by directly acting on the proteins of the PcG system in populations from temperate environments (Harr *et al.* 2002; Levine and Begun 2008; Gibert *et al.* 2011). In this study, we present evidence for selection acting on *cis*-regulatory sequences to reduce the temperature sensitivity of PcG-regulated gene expression.

Since PRE function is highly dependent on the genomic location, one drawback of our study may be that reporter gene assays were only done at one integration site in the genome (Kassis and Brown 2013; Steffen and Ringrose 2014). This position effect is mainly due to regulatory elements in the vicinity of the integration site that can have an influence on the function of the transgenic PRE. However, it is likely that redoing the study at an additional integration site would yield similar results to those reported here. One reason for this is that studies observing this position effect mainly looked at smaller PRE sequences (Kassis and Brown 2013). Our inserted fragment is ~5 kb in length and likely contains other regulatory sequences (*i.e.*, enhancers) in addition to the two promoters and the PRE. Furthermore, we could reproduce the endogenous expression pattern of the different tissues in the transgenic lines with a higher expression

of *ph-p* in larval brains than in midguts and vice versa for *CG3835*.

Here, we report a *cis*-regulatory change mediating a decreased thermosensitivity of PcG regulation at a specific locus. The question arises of whether selection against temperature-sensitive expression variability in temperate populations is a global phenomenon, *i.e.*, PcG target genes in general exhibit such a buffering, or whether it is specific for the locus examined in this study. The former is supported by other studies that have shown greater expression plasticity in tropical populations than in temperate ones (Levine *et al.* 2011) and given evidence for spatially varying selection targeting proteins of the Polycomb group (Harr *et al.* 2002; Levine and Begun 2008). It would then also be of interest to which amount either of both, *cis*-regulatory and *trans*-regulatory changes, contributes to the reduced thermosensitivity in temperate populations and whether one can observe other PcG target genes with *cis*-regulatory changes.

The buffering of the temperature-induced expression plasticity due to the derived sequence variants is likely to be explained by changes in TFBS motifs. There are two possibilities for how this could have happened. First, changes in TFBS motifs occurred in enhancer sequences, altering the strength of the enhancer, resulting in a change in the transcriptional output that is then maintained by the associated PRE. Second, the PRE could have been directly targeted by selection and TFBSs of PcG proteins and associated factors could have been modified, leading, *e.g.*, to changes in PcG recruitment and therefore to differences in the expression level that is maintained by the PRE (Schwartz *et al.* 2010; Steffen and Ringrose 2014). For enhancers as well as for PREs, it is well documented that small changes in their sequences can have large effects on the expressed phenotype and both *cis*-regulatory elements are known to evolve rapidly (Hauenschild *et al.* 2008). Therefore, it seems likely that a change in sequence of one of them (or both) is responsible for the expression differences described in this study. To find the causative sequence variant(s) and the associated TFBS(s), further experimental studies are needed.

Acknowledgments

We acknowledge S. Lange and A. Steincke for providing technical assistance. We thank Charles Langley (University of California, Davis) for providing us with the assembled genomes for the Dutch and Malaysian populations. We are also very grateful for constructive comments from the two reviewers and David Begun. This work was supported by the Deutsche Forschungsgemeinschaft Research Unit grants STE 325/12-1 and 12-2.

Literature Cited

Americo, J., M. Whiteley, J. L. Brown, M. Fujioka, J. B. Jaynes *et al.*, 2002 A complex array of DNA-binding proteins required for pairing-sensitive silencing by a polycomb group response element from the *Drosophila engrailed* gene. Genetics 160: 1561–1571.

Arvidsson, S., M. Kwasniewski, D. M. Riaño-Pachón, and B. Mueller-Roeber, 2008 QuantPrime–a flexible tool for reliable high-throughput primer design for quantitative PCR. BMC Bioinformatics 9: 465.

Bantignies, F., C. Grimaud, S. Lavrov, M. Gabut, and G. Cavalli, 2003 Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. Genes Dev. 17: 2406–2420.

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

Beisswanger, S., and W. Stephan, 2008 Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated *polyhomeotic* genes in *Drosophila*. Proc. Natl. Acad. Sci. USA 105: 5447–5452.

Beisswanger, S., W. Stephan, and D. De Lorenzo, 2006 Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. Genetics 172: 265–274.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57: 289–300.

Bischof, J., R. K. Maeda, M. Hediger, F. Karch, and K. Basler, 2007 An optimized transgenesis system for *Drosophila* using germ-line-specific ϕC31 integrases. Proc. Natl. Acad. Sci. USA 104: 3312–3317.

Blastyák, A., R. K. Mishra, F. Karch, and H. Gyurkovics, 2006 Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. Mol. Cell. Biol. 26: 1434–1444.

Busturia, A., C. D. Wightman, and S. Sakonju, 1997 A silencer is required for maintenance of transcriptional repression throughout *Drosophila* development. Development 124: 4343–4350.

Chan, C. S., L. Rastelli, and V. Pirrotta, 1994 A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. EMBO J. 13: 2553.

Chintapalli, V. R., J. Wang, and J. A. Dow, 2007 Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. 39: 715–720.

Clarke, A., 1996 The influence of climate chance of the distribution and evolution of organisms, pp. 377–407 in *Animals and Temperature: Phenotypic and Evolutionary Adaptation*, edited by I. A. Johnston, and A. F. Bennett. Cambridge University Press, Cambridge, UK.

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

Déjardin, J., A. Rappailles, O. Cuvier, C. Grimaud, M. Decoville *et al.*, 2005 Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1. Nature 434: 533–538.

Fauvarque, M. O., and J. M. Dura, 1993 *Polyhomeotic* regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in *Drosophila*. Genes Dev. 7: 1508–1520.

Fiston-Lavier, A.-S., N. D. Singh, M. Lipatov, and D. A. Petrov, 2010 *Drosophila melanogaster* recombination rate calculator. Gene 463: 18–20.

Foll, M., and O. Gaggiotti, 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Gibert, J.-M., F. Karch, and C. Schlötterer, 2011 Segregating variation in the polycomb group gene *cramped* alters the effect of temperature on multiple traits. PLoS Genet. 7: e1001280.

Harr, B., M. Kauer, and C. Schlötterer, 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 99: 12949–12954.

Hauenschild, A., L. Ringrose, C. Altmutter, R. Paro, and M. Rehmsmeier, 2008 Evolutionary plasticity of polycomb/trithorax response elements in *Drosophila* species. PLoS Biol. 6: e261.

Hellemans, J., G. Mortier, A. De Paepe, F. Speleman, and J. Vandesompele, 2007 qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 8: R19.

Hu, T. T., M. B. Eisen, K. R. Thornton, and P. Andolfatto, 2013 A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. Genome Res. 23: 89–98.

Kassis, J. A., and J. L. Brown, 2013 Polycomb group response elements in *Drosophila* and vertebrates, pp. 83–118 in *Advances in Genetics*, edited by J. C. Dunlap, S. F. Goodwin, and T. Friedmann. Academic Press, New York/London/San Diego.

Kim, Y., and W. Stephan, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160: 765–777.

Kozma, G., W. Bender, and L. Sipos, 2008 Replacement of a *Drosophila* Polycomb response element core, and in situ analysis of its DNA motifs. Mol. Genet. Genomics 279: 595–603.

Lachaise, D., and J.-F. Silvain, 2004 How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. Genetica 120: 17–39.

Laurent, S. J., 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.

Levine, M. T., and D. J. Begun, 2008 Evidence of spatially varying selection acting on four chromatin-remodeling loci in *Drosophila melanogaster*. Genetics 179: 475–485.

Levine, M. T., M. L. Eckert, and D. J. Begun, 2011 Whole-genome expression plasticity across tropical and temperate *Drosophila melanogaster* populations from Eastern Australia. Mol. Biol. Evol. 28: 249–256.

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

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for selective sweeps using SNP data. Genome Res. 15: 1566–1575.

Pavlidis, P., J. D. Jensen, and W. Stephan, 2010 Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. Genetics 185: 907–922.

Pavlidis, P., D. Živković, A. Stamatakis, and N. Alachiotis, 2013 SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. 30: 2224–2234.

Pierre, S. E. S., L. Ponting, R. Stefancsik, and P. McQuilton The FlyBase Consortium, 2014 FlyBase 102—advanced approaches to interrogating FlyBase. Nucleic Acids Res. 42: D780–D788.

Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno *et al.*, 2012 Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet. 8: e1003080.

Riebler, A., L. Held, and W. Stephan, 2008 Bayesian variable selection for detecting adaptive genomic differences among populations. Genetics 178: 1817–1829.

Sanger, F., S. Nicklen, and A. R. Coulson, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463–5467.

Schuettengruber, B., M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek *et al.*, 2009 Functional anatomy of polycomb and trithorax chromatin landscapes in *Drosophila* embryos. PLoS Biol. 7: e1000013.

Schwartz, Y. B., T. G. Kahn, P. Stenberg, K. Ohno, R. Bourgon *et al.*, 2010 Alternative epigenetic chromatin states of polycomb target genes. PLoS Genet. 6: e1000805.

Steffen, P. A., and L. Ringrose, 2014 What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat. Rev. Mol. Cell Biol. 15: 340–356.

Stephan, W., and H. Li, 2007 The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity 98: 65–68.

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

Wilches, R., S. Voigt, P. Duchén, S. Laurent, and W. Stephan, 2014 Fine-mapping and selective sweep analysis of QTL for cold tolerance in *Drosophila melanogaster*. G3 4: 1635–1645.

Zhang, Y., S. Arcia, B. Perez, P. Fernandez-Funez, and D. E. Rincon-Limas, 2013 pΔ TubHA4C, a new versatile vector for constitutive expression in *Drosophila*. Mol. Biol. Rep. 40: 5407–5415.

Zink, D., and R. Paro, 1995 *Drosophila* Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target DNA. EMBO J. 14: 5660.

Communicating editor: D. J. Begun

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.177030/-/DC1

Positive Selection at the *Polyhomeotic* Locus Led to Decreased Thermosensitivity of Gene Expression in Temperate *Drosophila melanogaster*

Susanne Voigt, Stefan Laurent, Maria Litovchenko, and Wolfgang Stephan

DOI: 10.1534/genetics.115.177030

Table S1 Fold-changes in *lacZ* reporter gene expression

Promoter	Comparison		Fold-change
	between	within	
ph-p	BR/MG	17°C & NL	2.67
		17°C & ZK	3.08 **
		28°C & NL	4.28 **
		28°C & ZK	6.81 **
	17°C/28°C	BR & NL	0.77
		BR & ZK	1.09
		MG & NL	1.24
		MG & ZK	2.41 **
	NL/ZK	17°C & BR	0.74
		17°C & MG	0.85
		28°C & BR	1.03
		28°C & MG	1.64*
CG3835	BR/MG	17°C & NL	0.33 **
		17°C & ZK	0.21 *
		28°C & NL	0.61 *
		28°C & ZK	0.40 *
	17°C/28°C	BR & NL	0.89
		BR & ZK	0.70
		MG & NL	1.08
		MG & ZK	2.03
	NL/ZK	17°C & BR	1.04
		17°C & MG	0.67
		28°C & BR	0.82
		28°C & MG	1.27

mRNA abundance was measured via RT-qPCR in brains (BR) and midguts (MG) of third instar larvae reared at 17°C or 28°C. *lacZ* reporter gene expression was either driven by the promoter of *ph-p* or the *CG3835* promoter. Promoter and adjacent regulatory regions were either derived from a European strain from the Netherlands (NL) or an African one from Zimbabwe (ZK). Statistical testing included t-tests and correction for multiple testing. *P<0.05, **P<0.01 (FDR=0.05).

Table S2 *BayeScan* outlier SNPs

Sample	Number of Populations	Position on X chromosome	Posterior Probability	Posterior Odds $\log_{10}$(PO)	alpha	F_{ST}
Europe-Africa	7	2,054,375	1.000	3.699	1.834	0.739
		2,035,164	**0.926**	**1.095**	**1.700**	**0.712**
		2,173,611	0.928	1.111	1.677	0.709
		2,035,208	**0.917**	**1.042**	**1.676**	**0.708**
		2,034,789	**0.919**	**1.054**	**1.665**	**0.706**
		2,032,933	**0.921**	**1.064**	**1.661**	**0.706**
		2,035,090	**0.920**	**1.059**	**1.641**	**0.702**
		2,054,995	0.996	2.442	1.502	0.687
		2,039,998	0.990	1.996	1.449	0.678
		2,055,053	0.988	1.930	1.438	0.676
		1,999,363	0.976	1.617	1.391	0.668
		2,053,119	0.976	1.617	1.379	0.666
		2,056,220	0.972	1.537	1.369	0.664
		2,054,347	0.966	1.448	1.337	0.658
		1,907,955	0.955	1.329	1.248	0.643
		2,056,678	0.933	1.141	1.248	0.642
		2,055460	0.911	1.012	1.246	0.642
		1,905,431	0.955	1.329	1.231	0.640
		1,999,517	0.950	1.275	1.222	0.639
		1,974,821	0.946	1.242	1.216	0.638
		1,903117	0.942	1.212	1.202	0.635
		1,974,723	0.923	1.081	1.162	0.628
		2,090,045	0.901	0.958	-1.521	0.181
		2,179,117	0.926	1.095	-1.573	0.174
		2,136,315	0.941	1.203	-1.584	0.171
		1,963,657	0.928	1.108	-1.609	0.170
		2,181,270	0.948	1.257	-1.589	0.170
		1,984,277	0.933	1.147	-1.644	0.166
		2,171,163	0.953	1.309	-1.626	0.166
		1,994,497	0.961	1.394	-1.673	0.160
		2,069,894	0.954	1.315	-1.691	0.160
		2,179,478	0.973	1.550	-1.702	0.157
		2,141,928	0.976	1.606	-1.733	0.153
		2,153,801	0.973	1.557	-1.742	0.153
		1,965,347	0.979	1.669	-1.782	0.148
		2,139,505	0.971	1.525	-1.846	0.144

Africa	5	2,054,375	0.998	2.698	1.731	0.674
		1,972,008	0.959	1.365	1.343	0.599
		1,972,114	0.958	1.354	1.335	0.598
		2,054,995	0.955	1.331	1.324	0.595
		2,112,292	0.945	1.238	1.290	0.589
		2,141,480	0.939	1.190	1.296	0.590
		2,053,119	0.899	0.951	1.215	0.573
		1,944,464*	0.880	0.864	1.178	0.566
		2,039,998*	0.865	0.807	1.140	0.558

BayeScan analyses were run on a X-chromosomal 300-kb sequence window around the *ph* locus. Significant outlier SNPs at an FDR of 5% are shown here. For the African-exclusive analysis, also outliers at an FDR of 7% are given (marked by asterisks). Six outlier SNPs are located in the region of significant CLR values. Bold font highlights the five outlier SNPs that are located in *ph-p/CG3835* intergenic region and that are not already significantly segregating when only African populations are considered. The SNP at X-chromosomal position 2,039,998 (FlyBase release 5) is located within in the region of significant CLR values, but not between *ph-p* and *CG3835*.

DECREASED THERMOSENSITIVITY OF *VESTIGIAL* GENE EXPRESSION IN TEMPERATE POPULATIONS OF *DROSOPHILA MELANOGASTER*

Susanne Voigt, Anna Christina Erpf, Wolfgang Stephan

(unpublished manuscript)

ABSTRACT

Drosophila melanogaster originated in tropical Africa and has successfully adapted to temperate environments worldwide. A major limiting factor of temperate climates has probably been its low and varying temperatures. Gene regulation by the Polycomb group (PcG) may be disrupted by ambient temperature resulting in increased expression of PcG-regulated genes when temperature is low. However, there is evidence that this temperature-induced expression plasticity has been reduced during the colonization of temperate environments. In this study, we focus on the PcG target gene *vestigial* (*vg*) and provide evidence that the thermosensitivity of PcG regulation has been buffered in populations from temperate climates. We investigated thermosensitivity of *vg* gene expression in six natural populations including four temperate populations (three from Europe and one from high-altitude Africa), and two tropical populations from the ancestral species range. Temperate populations exhibited a lower degree of temperature-induced expression plasticity than tropical populations. Decreased *vg* expression plasticity likely evolved more than once, since the underlying mechanisms seem to differ between temperate populations.

INTRODUCTION

Species colonizing new environments need to adapt to novel biotic and abiotic conditions. As a cosmopolitan species, the fruit fly *Drosophila melanogaster* successfully adapted to a wide range of new habitats. Its origin is thought to be in tropical southern-central Africa from where it spread around the world (David and Capy 1988; Lachaise and Silvain 2004; Pool *et al.* 2012). After an initial expansion throughout Africa, it finally reached the Eurasian continent after the last glaciation around 10,000 years ago (Li and Stephan 2006; Stephan and Li 2007), and later moved on to colonize Asia and Europe (Laurent *et al.* 2011). A main limiting factor while settling in Europe and also at high altitudes in Africa has probably been the temperate climate with its low and varying temperatures.

Chromatin-based gene regulation in *D. melanogaster* is known to be disrupted by ambient temperatures (Fauvarque and Dura 1993; Gibert *et al.* 2011). In the case of genes regulated by the Polycomb group (PcG) of proteins, transcriptional output increases with decreasing temperature at which flies are reared or held (Fauvarque and Dura 1993; Chan *et al.* 1994; Zink and Paro 1995; Bantignies *et al.* 2003; Gibert *et al.* 2011). This phenomenon prompted the hypothesis that adaptation to temperate environments might have included the buffering of this thermosensitive regulatory process (Levine and Begun 2008). Several studies, indeed, provided evidence for positive directional selection acting on proteins that are involved in PcG regulation (Harr *et al.* 2002; Levine and Begun 2008; Gibert *et al.* 2011). Another study found positive selection acting on *cis*-regulatory sites that led to decreased thermosensitivity of PcG gene regulation in European populations (Voigt *et al*. 2015). The selected sites were highly differentiated between African and European populations, and were located in a Polycomb response element (PRE), a *cis*-regulatory module that recruits PcG proteins to their target genes.

In this study, using the gene *vestigial* (*vg*), we again provide evidence that temperature-induced expression plasticity of PcG-regulated genes has been

buffered in temperate populations from *D. melanogaster*. The PcG target gene v*g* (Schuettengruber *et al.* 2009; Schwartz et *al.* 2010) is well-studied and located on the 2R chromosome arm of the *D. melanogaster* genome. It encodes a transcription factor that plays a central role in the development and patterning of the *Drosophila* wing (Williams *et al.* 1991). It is also known as a wing selector gene: a loss of *vg* results in the failure of wings to develop (Lindsley and Zimm 1992) and the ectopic expression of *vg* leads to the outgrowth of ectopic wing tissue (Kim *et al.* 1996). We investigated thermosensitive expression of *vg* in natural populations of *D. melanogaster* from six different locations. These included three temperate populations from Europe, one temperate population from a high-altitude location in Africa, and two tropical populations from the ancestral species range. The degree of expression plasticity differed between the populations with a higher degree in those from tropical climates than in those from temperate climates. This buffering effect likely evolved more than once, since the underlying mechanisms appear to differ between temperate populations.

MATERIALS & METHODS

Expression analysis

Six population samples, each consisting of eight isofemale lines, from Sweden (Umeå), The Netherlands (Leiden), France (Lyon), Rwanda (Gikongoro), Zimbabwe (Lake Kariba), and Zambia (Siavonga) (Table 1) were selected for expression analysis in adult *D. melanogaster*. Tissue-specific expression analysis in third instar larvae was performed using samples of four of the aforementioned populations from Sweden, The Netherlands, Zimbabwe, and Zambia. Flies were reared on a standard cornmeal-molasses medium with a 14/10h light/dark cycle at 17°C or 28°C. Five males and five females were allowed to mate and oviposit for seven or three days at 17°C and 28°C, respectively. Eight males aged 4-6 days of the resulting progeny, one per line, were pooled for RNA extraction. For tissue-specific expression-analysis, 50 eggs per vial

were allowed to hatch and develop into wandering third instar larvae from which tissue was sampled. Eight brain and sixteen wing imaginal discs of each line were pooled for RNA extraction. After dissection, tissues were immediately stored in RNAlater (Qiagen, Hilden, Germany; http://www.qiagen.com). RNA was extracted using the MasterPure RNA Purification Kit (Epicentre, Madison, WI, USA; http://www.epibio.com). RNA purity was assessed via the ratio of absorbances at 260 and 280 nm (A260/A280 > 1.8). It was then reverse transcribed into cDNA using random primers and SuperScript® III Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA;http://www.lifetechnologies.com). RT-qPCR reactions were performed with iQ™ SYBR® Green Supermix (BioRad, Hercules, CA, USA; http://www.bio-rad.com) on a CFX96™ real-time PCR cycler (BioRad, Hercules, CA, USA; http://www.biorad.com). Primers for target genes and two reference genes for normalization (*RpS20 and RpL32*) were designed applying the QuantPrime software (Arvidsson *et al*. 2008). Three biological replicates per population sample, rearing temperature and tissue were run in triplicates and primer specificity was confirmed by a melting curve analysis. Negative controls included no template controls (NTCs) and no reverse-transcription controls (NRTs) to exclude contamination. Relative expression was calculated using the qBase relative quantification framework (Hellemans *et al.* 2007). Both reference genes were stably expressed across samples. This was assessed by calculating the coefficient of variation and the M stability parameter according to Hellesmans *et al.* (2007). Log-transformed normalized relative quantities were subjected to a Welch Two Sample t-test to test for statistically significant expression differences between the different rearing temperatures and population samples. False discovery rate was controlled using the multiple testing correction method by Benjamini and Hochberg (1995).

Table 1 Population samples

Population sample	Latitude	Longitude	Altitude (m)	Climate
Umeå, Sweden (SU)	63.83	20.26	12	cold - temperate
Leiden, The Netherlands (NL)	52.17	4.48	0	warm - temperate
Lyon, France (FR)	45.76	4.84	175	warm - temperate
Gikongoro, Rwanda (RG)	-2.49	28.92	1927	warm - temperate
Lake Kariba, Zimbabwe (ZK)	-16.52	28.80	619	tropical
Siavonga, Zambia (ZI)	-16.54	28.72	530	tropical

Climate data was taken from *http://en.climate-data.org/*.

Sequence analysis

DNA polymorphisms at the *vg* locus were analyzed in five of the populations that were used in the expression analysis. Sequence analysis was performed for population samples from Sweden (Umeå), The Netherlands (Leiden), France (Lyon), Rwanda (Gikongoro), and Zambia (Siavonga). French (8 lines), Rwandan (24 lines), and Zambian (27 lines) sequences were taken from assembled full-genome data as provided by the Drosophila Population Genomics Project (DPGP) (http://www.dpgp.org). Swedish (14 lines) and Dutch (10 lines) sequences were extracted from full genomes assembled following the approach of Pool *et al.* (2012). Nucleotides with known admixture or identity-by-descent according to Pool *et al.* (2012) were replaced with missing value labels in the analysis. The same was done for sites exhibiting heterozygosity since heterozygotes are not expected in sequence data obtained from haploid embryos (Pool *et al.* 2012). Sites with an excess of missing data that is less than six lines per population not labeled as missing were excluded from the analysis. Nucleotide diversity was estimated in terms of π (Tajima 1989), and divergence was calculated against a *Drosophila simulans* sequence (Hu *et al.* 2013). F_{ST} per SNP was estimated according to Weir and Cockerham (1984). Calculations of summary statistics were performed in the R package diveRsity (Keenan *et al.* 2013).

RESULTS

Expression analysis in adult D. melanogaster

Male adult flies from different natural populations were reared at two different temperatures (17° and 28°C) to examine whether the PcG target gene *vg* (Schuettengruber *et al.* 2009; Schwartz *et al.* 2010) exhibits thermosensitivity in its expression. Population samples were from six different locations with different climates (Table 1). Samples from tropical regions included two from the putative ancestral *D. melanogaster* species range: Zambia and Zimbabwe (Pool *et al.* 2012). The other four samples were derived from populations from temperate regions: three from warm-temperate climates including two European samples (France and The Netherlands) and an African high-elevation population sample from Rwanda, and one from the cold-temperate climate of Sweden.

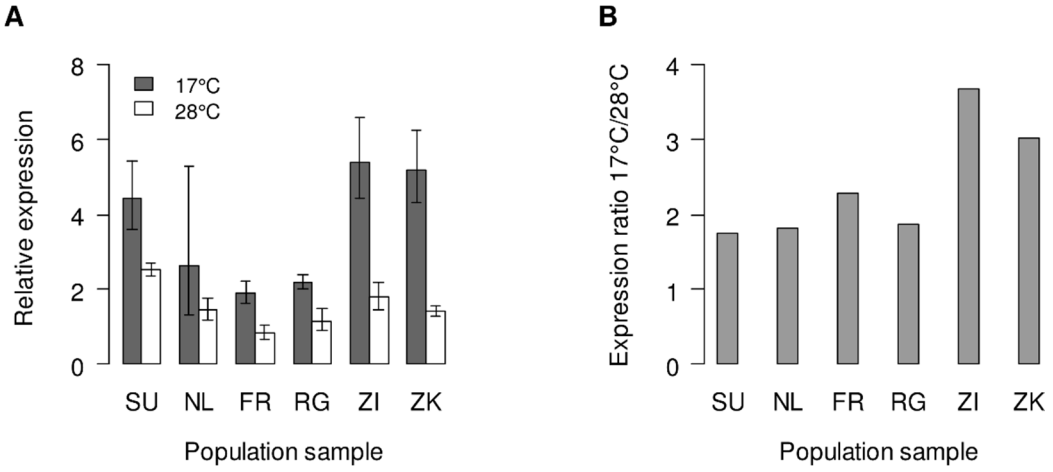


Figure 1 Expression of *vg* in adult *D. melanogaster*. mRNA abundance in adult males was measured via RT-qPCR. (A) Expression in Swedish (SU), Dutch (NL), French (FR), Rwandan (RG), Zambian (ZI), and Zimbabwean (ZK) flies reared at either 17°C (gray bars) or 28°C (white bars). Error bars represent the 95% confidence interval. (B) Ratio of mean *vg* expression between rearing temperatures (17°C to 28°C).

Thermosensitivity of *vg* gene expression due to variation in rearing temperature was observed for all population samples but not for the control gene *Aats-asp* that is adjacent to *vg* and is not known to be regulated by the PcG proteins (Schuettengruber *et al.* 2009; Schwartz *et al.* 2010) (Figure 1, Table 2). Expression of *vg* was consistently higher at 17°C than at 28°C, and this difference was highly significant for all samples except for the Dutch one in which expression at 17°C varied too much between biological replicates (Figure 1A, Table 2). Interestingly, with a ~2-fold higher expression at 17°C, the ratio between *vg* expression at 17°C and at 28°C was lower in all four temperate population samples than in the tropical samples in which *vg* expression was more than 3-fold higher at 17°C (Figure 1B, Table 2). Therefore, temperature-induced expression plasticity of *vg* appears to be buffered in derived temperate genotypes when compared to ancestral tropical genotypes.

Table 2 Fold-changes in gene expression between rearing temperatures in adult *D. melanogaster*

Gene	Gene expression ratio 17°C/28°C					
	Sweden (SU)	The Netherlands (NL)	France (FR)	Rwanda (RG)	Zimbabwe (ZK)	Zambia (ZI)
vg	1.76*	1.82	2.28**	1.88**	3.68**	3.02**
Aats-asp	1.15	1.52	0.90	0.92	1.01	1.17

Statistical testing included t-tests and correction for multiple testing. *P<0.05, **P<0.01 (FDR=0.05).

Although the amount of buffering among temperate samples appears to be the same, the mechanisms of how the lowered expression ratio between rearing temperatures comes about seem to be different. For the cold-temperate sample from Sweden, increased *vg* expression relative to the other population samples was observed at 28°C. The difference was statistically significant for all comparisons between the Swedish and the other samples, except for the one to Zimbabwe, which was of borderline significance (P=0.07) (Table 3). In contrast, for the three warm-temperate population samples, *vg* expression was decreased at 17°C compared to

the other three samples. Statistical significance was found for all comparisons except for those including the Dutch sample (Figure 1A, Table 2). Again, the reason for this is the high variation between biological samples in the Dutch *vg* expression at 17°C which might be eliminated by increasing sample size. At least for the French and Rwandan samples, this decreased *vg* expression is also visible at 28°C, though to a lower extent than at 17°C (Figure 1A, Table 2).

Table 3 Fold-changes in gene expression between population samples in adult *D. melanogaster*

Gene	Gene expression ratio between population samples		
	Population samples	Rearing temperature	
		17°C	28°C
vg	Sweden (SU)/The Netherlands (NL)	1.69	1.75*
	Sweden (SU)/France (FR)	2.35**	3.04**
	Sweden (SU)/Rwanda (RG)	2.03*	2.16**
	Sweden (SU)/Zimbabwe (ZK)	0.82	1.41
	Sweden (SU)/Zambia (ZI)	0.85	1.78***
	The Netherlands (NL)/France (FR)	1.39	1.75*
	The Netherlands (NL)/Rwanda (RG)	1.20	1.24
	The Netherlands (NL)/Zimbabwe (ZK)	0.49	0.81
	The Netherlands (NL)/Zambia (ZI)	0.50	1.02
	France (FR)/Rwanda (RG)	0.86	0.71
	France (FR)/Zimbabwe (ZK)	0.35**	0.65**
	France (FR)/Zambia (ZI)	0.36**	0.59*
	Rwanda (RG)/Zimbabwe (ZK)	0.41**	0.65**
	Rwanda (RG)/Zambia (ZI)	0.42**	0.82
	Zimbabwe (ZK)/Zambia (ZI)	1.04	0.79
Aats-asp	Sweden (SU)/The Netherlands (NL)	0.69*	0.92
	Sweden (SU)/France (FR)	0.96	0.76
	Sweden (SU)/Rwanda (RG)	0.88	0.70
	Sweden (SU)/Zimbabwe (ZK)	0.90	0.91
	Sweden (SU)/Zambia (ZI)	0.87	0.76
	The Netherlands (NL)/France (FR)	1.39	0.83
	The Netherlands (NL)/Rwanda (RG)	1.26	0.77
	The Netherlands (NL)/Zimbabwe (ZK)	1.29	0.99
	The Netherlands (NL)/Zambia (ZI)	1.25	0.83
	France (FR)/Rwanda (RG)	0.91	0.93
	France (FR)/Zimbabwe (ZK)	0.93	1.20
	France (FR)/Zambia (ZI)	0.90	1.00
	Rwanda (RG)/Zimbabwe (ZK)	1.02	1.29
	Rwanda (RG)/Zambia (ZI)	0.99	1.08
	Zimbabwe (ZK)/Zambia (ZI)	1.03	1.20

Statistical testing included t-tests and correction for multiple testing. *P<0.05, **P<0.01, ***P<0.001 (FDR=0.05).

Overall, two main conclusions can be drawn: first, populations from temperate environments show a reduction of temperature-induced plasticity in *vg* expression compared to tropical populations, and second, the mechanisms responsible for this buffering seem to differ between temperate populations. However, with this rather crude experiment, one cannot distinguish whether this is in general found for *vg* expression or whether differences between different tissues or developmental stages can be observed.

Tissue-specific expression analysis in third instar larvae of D. melanogaster

Adult structures are derived from larval structures called imaginal discs. Since the *vg* gene product has its main function in the control of wing formation (Williams *et al.* 1991; Lindsley and Zimm 1992; Kim *et al.* 1996), the *vg* gene shows an enriched expression in the wing imaginal disc (Williams *et al.* 1991). Therefore, this tissue was chosen to further examine *vg* expression under different rearing temperatures. The same was done for larval brains in order to monitor *vg* expression in a tissue in which expression of the gene is known to be low (Chintapalli *et al.* 2007). Gene expression was measured in four samples of the aforementioned populations from Sweden, The Netherlands, Zambia and Zimbabwe (Table 1).

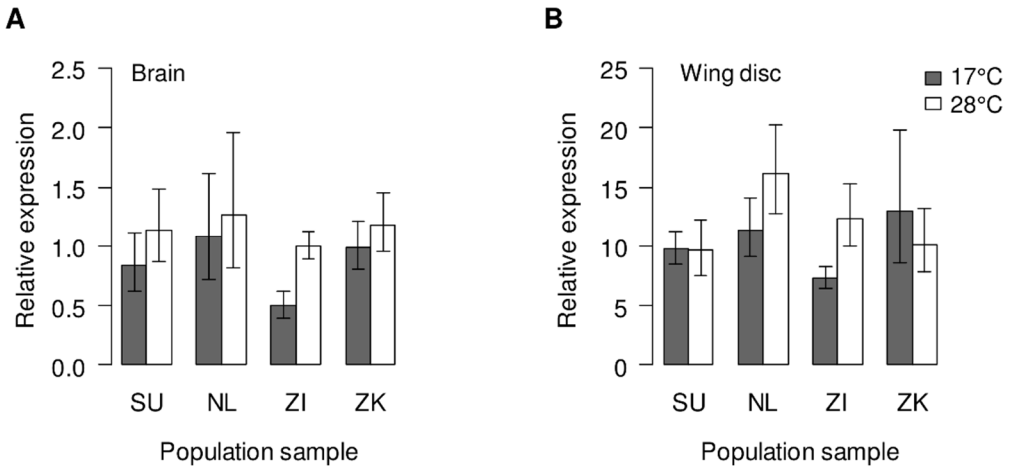


Figure 2 Tissue-specific *vg* expression in third instar larvae of *D. melanogaster*. mRNA abundance was assessed by RT-qPCR in brains (A) and wing discs (B) from wandering third instar larvae reared at either 17°C (gray bars) or 28°C (white bars). Expression was measured in population samples from Sweden (SU), The Netherlands (NL), Zambia (ZI), and Zimbabwe (ZK). Error bars represent the 95% confidence interval.

As expected *vg* expression was significantly higher in wing discs than in brains at both rearing temperatures and in all population samples (Table S1). A thermosensitive expression pattern as it is often observed for PcG-regulated genes (Fauvarque and Dura 1993; Chan *et al.* 1994; Zink and Paro 1995; Bantignies *et al.* 2003; Gibert *et al.* 2011) and that we found for adults with an increased expression at lower rearing temperature was neither detected in wing discs nor in brains for *vg* or the control gene *Aats-asp* (Figure 2, Table 4). Decreased expression levels in the Zambian sample at 17°C relative to 28°C were the only statistically significant differences observed in *vg* expression (Table 4, Table S2). Thus, neither thermosensitivity in *vg* expression as observed for adults nor a buffering of it in temperate populations appear to play a role in the larval tissues examined here.

Table 4 Fold-changes in tissue-specific gene expression between rearing temperatures in third instar larvae

Gene	Tissue	Gene expression ratio 17°C/28°C			
		Sweden (SU)	The Netherlands (NL)	Zimbabwe (ZK)	Zambia (ZI)
vg	Wing disc	1.02	0.71	1.29	0.59*
	Brain	0.73	0.85	0.84	0.50*
Aats-asp	Wing disc	0.95	0.87	0.99	0.88
	Brain	0.90	0.99	1.07	0.81

Statistical testing included t-tests and correction for multiple testing. *P<0.05 (FDR=0.05).

Sequence analysis

In order to identify putative *cis*-regulatory changes, *i.e.* changes in regulatory sequences at the *vg* locus, between derived and ancestral populations that might contribute to the observed expression differences, F_{ST} per SNP (Weir and Cockerham 1984) was calculated in the *vg* gene region. The *vg* gene region was defined as the interval between the two outer insulators (±500 bp) (Nègre *et al.* 2010) that flank the *vg* locus (Figure 3B). Insulators are DNA sequence elements that can block the interaction between regulatory elements and genes, as well as prevent the spread of regions of modified chromatin (Nègre et al. 2010). Therefore, most of the *cis*-regulatory sequences with an effect on *vg* expression are expected to be located within this region. DNA sequences were extracted from available full genome data of five of the populations that were used in the expression analysis. F_{ST} per SNP (biallelic) was calculated between each of the temperate population samples (Sweden, The Netherlands, France, and Rwanda) and the ancestral tropical sample (Zambia). We arbitrarily chose $F_{ST} > 0.6$ as a cutoff value for candidate SNPs of *cis*-regulatory changes that are highly differentiated between populations (*i.e.*, high allele frequencies in the derived populations and low frequencies in the ancestral population), and therefore have the potential to account for a large proportion of the expression differences observed between derived and ancestral populations. Since the Swedish expression pattern differed from the one in the other derived populations, putative candidate SNPs are not expected to be shared among the Swedish and the other derived populations. Thus, candidate SNPs for the Swedish

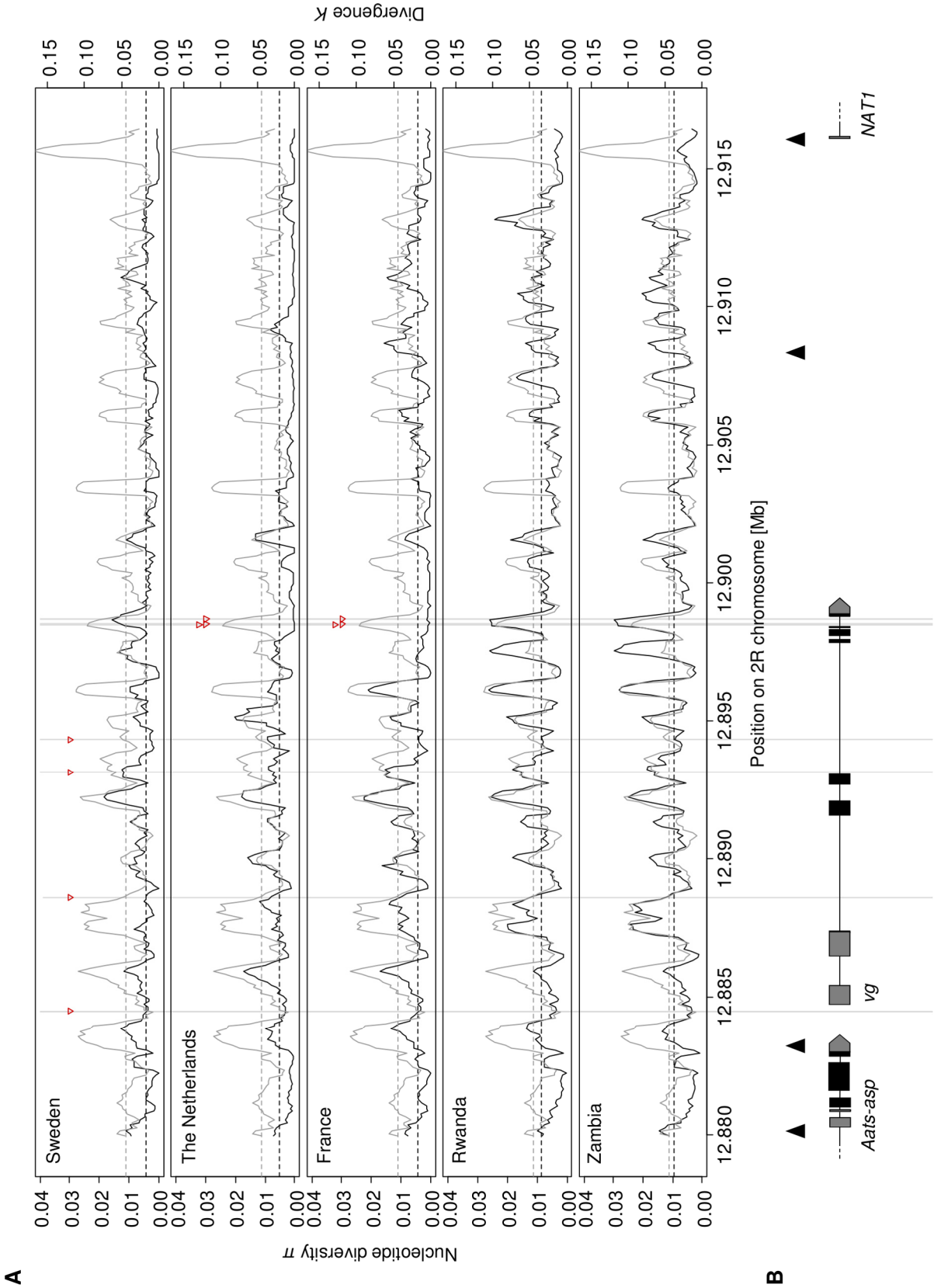
population are not expected to be highly differentiated between the other derived populations and the ancestral population (F_{ST} < 0.4), and vice versa. Such SNPs were also excluded as candidates. The unlikely scenario of two alleles differing between the Swedish and the other derived populations at a multi-allelic site and both alleles being also highly differentiated to the ancestral population was considered as well, yet no such site was observed. Four candidate SNPs putatively responsible for the Swedish expression pattern were detected (Figure 3, Table S3). The three other temperate populations were similar in their expression response to temperature, therefore the simplest scenario would be a common genetic basis causing this response. However, no SNP with an F_{ST} above 0.6 to the ancestral population shared among all three derived population was observed. Thus, it seems less likely, if there is a common genetic basis in the three populations, that it would be largely due to shared *cis*-regulatory changes. However, highly differentiated insertions/deletions not detected by next-generation sequencing and sites excluded due to an excess of missing data could also be responsible (see Materials and Methods). Since the Rwandan and European populations are rather geographically distant populations, convergent evolution between the Rwandan and the two European populations was also considered as an explanation for the similar expression response to temperature. In turn, for the two geographically close European populations, three candidate SNPs with F_{ST} values above the cutoff value of 0.6 that were shared between the two populations could be identified and that were not highly differentiated (F_{ST} < 0.4) to the ancestral population in the other derived populations (Figure 3, Table S3). For the comparison of the Rwandan and the ancestral population samples, no SNP above the cutoff could be detected which renders it less likely that a large proportion of the Rwandan expression response is due to *cis*-regulatory changes. Yet again, highly differentiated undetected insertions/deletions and sites excluded due to missing data cannot be ruled out as being responsible.

The French and Dutch candidate SNPs are clustered within a small region of about 200 bp that shows strongly reduced nucleotide diversity (π) relative to the mean

across a 100-kb window surrounding the vg locus. In contrast, in all the other population samples, derived and ancestral, levels of π are rather high as well as divergence to *D. simulans* (Figure 3A). Thus, it seems unlikely that this reduction in genetic variation is the result of this region being under abnormally high selective constraint or having an unusually low mutation rate, but could indicate the action of positive directional selection.

Except for one, all European candidate SNPs are located in introns of the vg gene (Figure 3, Table S3), which are known to contain many different regulatory elements (Pierre *et al.* 2014). Two of the three French and Dutch candidate SNPs fall into two known transcription factor binding sites (TFBSs) (Table 3). A Swedish candidate SNP in the *vg* promoter region co-localizes with several experimentally validated TFBSs including one for the Trithorax-like (Trl)/GAGA factor (GAF) protein, a PcG recruiter protein (Kassis and Brown 2013, Steffen and Ringrose 2014), as well as a putative PRE which was identified by chromatin immunoprecipitation (ChIP) experiments (Nègre et al. 2011) (Figure 3, Table S3).

Figure 3 Genetic variation in the *vg* gene region and candidate SNPs (next page). (A) Nucleotide diversity (π) in the five population samples from Sweden, The Netherlands, France, Rwanda, and Zambia (black solid lines), as well as divergence to *D. simulans* (gray solid lines) in a sliding window across the *vg* gene region (window size = 500 bp, step size = 100 bp). Means of π and divergence in a 100-kb window surrounding the *vg* locus, of which 36 kb are shown here, are given as black and gray dashed lines, respectively. Positions of candidate SNPs are given by red triangles and gray lines across panels. (B) Genetic map of the *vg* region. Coding regions are represented by black boxes, untranslated regions by gray boxes, and introns by black lines. Arrowheads indicate the direction of transcription and the positions of insulators are given by black triangles.


A
Nucleotide diversity π
Divergence K
Sweden
The Netherlands
France
Rwanda
Zambia
0.04
0.03
0.02
0.01
0.00
0.15
0.10
0.05
0.00
B
Aats-asp
vg
NAT1
12.880
12.885
12.890
12.895
12.900
12.905
12.910
12.915
Position on 2R chromosome [Mb]

DISCUSSION

Here we examined the expression response to temperature of the PcG target gene *vg* in natural populations of *D. melanogaster* from different latitudes and altitudes. The PcG together with another group of proteins, the Trithorax group (TrxG), are important epigenetic regulators of gene expression that act antagonistically to maintain repressed and activated transcription states, respectively. They are recruited to their target genes via PREs that influence the expression of their target genes in combination with other *cis*-regulatory DNA sequences (*i.e.* enhancers) in a cell- or tissue-specific manner. Enhancers initially determine the level of transcriptional output of PcG-regulated genes, and this transcriptional output is then epigenetically maintained by PREs (Schwartz *et al.* 2010; Kassis and Brown 2013; Steffen and Ringrose 2014).

Another feature of PcG-regulation in *Drosophila* is that it is often observed to be sensitive to the temperature at which flies are held or reared with the transcriptional output increasing when temperature decreases (Fauvarque and Dura 1993; Chan *et al.* 1994; Zink and Paro 1995; Bantignies *et al.* 2003; Gibert *et al.* 2011). This phenomenon led to the hypothesis that adaptation of *D. melanogaster* to temperate environments might have included the buffering of this thermosensitive regulatory process (Levine and Begun 2008). Several studies explored the question whether the temperature-induced expression plasticity of PcG-regulated genes has been detrimental while settling in temperate habitats, and whether this plasticity has been buffered by the action of positive selection. Indeed, less expression plasticity due to temperature was observed in populations from temperate environments than in those from tropical environments (Levine *et al.* 2011). Other studies presented evidence for positive selection acting on PcG proteins (Harr *et al.* 2002; Levine and Begun 2008; Gibert *et al.* 2011), and for positive selection acting on *cis*-regulatory sequences that led to decreased thermosensitivity of PcG gene regulation in temperate populations (Voigt *et al*. 2015).

In this study, we observed temperature-sensitive expression of the PcG target gene *vg* in six populations from *D. melanogaster* with a decreased thermosensitivity in the four temperate populations relative to the two tropical populations from the ancestral species range. Temperate populations were derived from a range of different locations including high-latitude Europe and high-altitude Africa. The consistent response to temperature across all temperate populations indicates positive directional selection acting to decrease thermosensitivity of *vg* expression in those populations.

However, the mechanisms underlying this reduced expression plasticity seem to differ between the temperate populations. Reduced *vg* expression at 17°C compared to the ancestral expression level led to a buffering of temperature-induced expression plasticity in the three population samples from warm-temperate climates. In contrast, in the cold-temperate sample from Sweden, increased *vg* expression at 28°C relative to the ancestral and the other population samples resulted in the observed buffering effect. Other ecological constraints due to the colder climate in Sweden could be a possible explanation for the observed difference. Higher overall *vg* expression, for instance, might have been additionally beneficial in the colder climate of Sweden. In contrast, convergent evolution leading to the same degree of buffering seems to be less likely, since Sweden was probably colonized by European flies already exhibiting reduced *vg* expression plasticity like the French and Dutch populations.

We detected putative *cis*-regulatory candidate SNPs responsible for the Swedish expression response that are highly differentiated between Sweden and Zambia but not between the other temperate populations and the ancestral population. Since the three other temperate populations were similar in their expression response to temperature, the simplest scenario would be a common genetic basis causing this response. However, no highly differentiated SNP was detected that is shared among the three populations. Therefore, if there is a common genetic basis between the

three populations, it seems to be less likely that it is to a large extent explained by *cis*-regulatory changes. Given the large geographic distance between the Rwandan and the European populations, convergent evolution might also be a likely explanation for the similar expression response to temperature. In turn, for the two geographically close European populations, we could detect shared candidate SNPs that are clustered within a region that also shows signs indicative of positive selection. No highly differentiated SNP could be detected between the Rwandan and the Zambian populations. The lack of *cis*-regulatory candidates might indicate *trans*-regulatory changes, which alter the abundance or activity of factors controlling *vg* expression (*e.g.*, PcG proteins), to be largely responsible. However, since accurate detection of short insertions and deletions is still difficult using next-generation sequencing, highly differentiated insertions/deletions cannot be excluded as possible explanations for the observed expression differences. The same is true for sites that were excluded from the analysis due to an excess of missing data (see Materials and Methods). Support for *trans*-regulatory changes being important might come from another study that conducted a genome-wide scan to identify genes under positive selection within the Rwandan sample (Pool *et al.* 2012). Genes identified in this scan included two PcG/TrxG genes, *pleiohomeotic like* (*phol*) and *female sterile (1) homeotic* (*fs(1)h*), which are in turn putative candidates for *trans*-regulatory factors responsible for the Rwandan decreased thermosensitivity. One of the four Swedish *cis*-regulatory candidate SNPs is located in the *vg* promoter region and a putative PRE identified by ChIP experiments (Nègre *et al*. 2011). In a reporter gene analysis that linked changes at c*is*-regulatory sites to differences in the thermosensitivity of PcG regulation, the sites were also found in a region harboring a known PRE (Voigt *et al.* 2015). The other six European candidate SNPs are located in intronic regions of the *vg* gene. It was shown before that mutations in *vg* introns can lead to changes in the thermosensitivity of *vg* expression (Silber *et al.* 1997).

Since the main function of *vg* is in wing development (Williams *et al.* 1991; Lindsley and Zimm 1992; Kim *et al.* 1996), we also searched for temperature-induced

expression plasticity and its possible buffering in wing discs of wandering third instar larvae. In this tissue and at this developmental stage, *vg* is in an activated state and highly expressed (Williams *et al.* 1991). As a control tissue, we used the larval brain in which *vg* gene expression is low (Chintapalli *et al.* 2007). Thermosensitive expression as it is often observed for PcG-regulated genes and as we found for adults could not be detected in either of the two larval tissues. Therefore, selective pressure against such a temperature-induced expression plasticity might be much stronger in larval tissues compared to adult tissues and therefore is not observed in any of the populations, or PcG regulation is in itself not affected by temperatures in larval tissues like it is in adult tissues. At least for wing discs, the former explanation might be more likely. Silber *et al.* (1999) found mutations in *vg* introns causing thermosensitive expression of the gene in wing discs of these mutant flies, whereas no thermosensitivity was observed for wild type wing discs.

In which adult tissue(s) the temperature-induced expression plasticity and its buffering in temperate populations is taking place, and whether the detected candidate SNPs are causing the observed buffering are questions still to be answered. However, here we have presented further evidence that temperature-induced expression plasticity of PcG regulated genes has been buffered in temperate populations of *D. melanogaster*.

ACKNOWLEDGEMENTS

We thank Simone Lange, Hilde Lainer, Gisela Brinkmann, and Gaby Kumpfmüller for their excellent technical assistance. We are grateful to Charles Langley (UC Davis) for providing us with the assembled genomes of the Dutch population. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Research Unit grants STE 325/12-1 and 12-2.

REFERENCES

Arvidsson, S., M. Kwasniewski, D. M. Riaño-Pachón, and B. Mueller-Roeber, 2008 QuantPrime–a flexible tool for reliable high-throughput primer design for quantitative PCR. BMC Bioinformatics 9: 465.

Bantignies, F., C. Grimaud, S. Lavrov, M. Gabut, and G. Cavalli, 2003 Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. Genes Dev. 17: 2406–2420.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 57: 289–300.

Chan, C. S., L. Rastelli, and V. Pirrotta, 1994 A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. EMBO J. 13: 2553.

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

Fauvarque, M. O., and J. M. Dura, 1993 *polyhomeotic* regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in *Drosophila*. Genes Dev. 7: 1508–1520.

Harr, B., M. Kauer, and C. Schlötterer, 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U S A 99: 12949–12954.

Hellemans, J., G. Mortier, A. De Paepe, F. Speleman, and J. Vandesompele, 2007 qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8: R19.

Hu, T. T., M. B. Eisen, K. R. Thornton, and P. Andolfatto, 2013 A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. Genome Res. 23: 89–98.

Kassis, J. A., and J. L. Brown, 2013 Polycomb Group Response Elements in *Drosophila* and Vertebrates, pp. 83–118 in *Advances in Genetics*, edited by J. C. D. and S. F. G. Theodore Friedmann. Academic Press.

Keenan, K., P. McGinnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl, 2013 diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol. Evol. 4: 782-788.

Kim, J., A. Sebring, J. J. Esch, M. E. Kraus, K. Vorwerk, J. Magee, and S. B. Carroll, 1996 Integration of positional signals and regulation of wing formation and identity by *Drosophila vestigial* gene. Nature 382: 133-138.

Lachaise, D., and J.-F. Silvain, 2004 How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. Genetica 120: 17-39.

Laurent, S. J., 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.

Levine, M. T., and D. J. Begun, 2008 Evidence of spatially varying selection acting on four chromatin-remodeling loci in *Drosophila melanogaster*. Genetics 179: 475–485.

Levine, M. T., M. L. Eckert, and D. J. Begun, 2011 Whole-genome expression plasticity across tropical and temperate *Drosophila melanogaster* populations from Eastern Australia. Mol. Biol. Evol. 28: 249–256.

Li, H., and W. Stephan, 2006 Inferring the demographic history and rate of adaptive

substitution in *Drosophila*. PLoS Genet. 2: e166.

Lindsley, D. L., and G. G. Zimm, 1992 *The Genome of* Drosophila melanogaster. San Diego, New York, Boston, London, Sydney, Tokyo,Toronto: Academic Press Inc.

Nègre, N., C. D. Brown, L. Ma, C. A. Bristow, S. W. Miller *et al.*, 2011 A *cis*-regulatory map of the *Drosophila* genome. Nature 471: 527-531.

Nègre, N., C. D. Brown, P. K. Shah, P. Kheradpour, C. A. Morrison, *et al.*, 2011 A comprehensive map of insulator elements for the *Drosophila* genome. PLoS Genet. 6: e1000814.

Pierre, S. E. S., L. Ponting, R. Stefancsik, P. McQuilton, and others, 2014 FlyBase 102—advanced approaches to interrogating FlyBase. Nucleic Acids Res. 42: D780–D788.

Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno *et al.*, 2012 Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet. 8: e1003080.

Schuettengruber, B., M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek *et al.*, 2009 Functional anatomy of polycomb and trithorax chromatin landscapes in *Drosophila* embryos. PLoS Biol 7: e1000013.

Schwartz, Y. B., T. G. Kahn, P. Stenberg, K. Ohno, R. Bourgon *et al.*, 2010 Alternative epigenetic chromatin states of polycomb target genes. PLoS Genet. 6: e1000805.

Steffen, P. A., and L. Ringrose, 2014 What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat Rev Mol Cell Biol 15: 340–356.

Stephan, W., and H. Li, 2007 The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity 98: 65–68.

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

Voigt, S., S. Laurent, M. Litovchenko, and W. Stephan, 2015 Positive selection at the *polyhomeotic* locus led to decreased thermosensitivity of gene expression in temperate *Drosophila melanogaster*. Genetics 200

Weir, B. S., and C. C. Cockerham, 1984 Estimating *F*-statistics for the analysis of population structure. Evolution 38: 1358–1370.

Williams, J. A., J. B. Bell, and S. B. Carroll, 1991 Control of *Drosophila* wing and haltere development by the nuclear vestigial gene product. Genes Dev. 5: 2481–2495.

Zink, D., and R. Paro, 1995 Drosophila Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target DNA. EMBO J. 14: 5660.

SUPPORTING INFORMATION

Table S1 Fold-changes in gene expression at different rearing temperatures between tissues in third instar larvae

Gene	Rearing temperature	Gene expression ratio wing disc/brain			
		Sweden (SU)	The Netherlands (NL)	Zimbabwe (ZK)	Zambia (ZI)
vg	17°C	11.77***	10.53**	13.19***	14.64**
	28°C	8.45***	12.71**	8.59***	12.35***
Aats-asp	17°C	1.10	0.99	1.10	1.32
	28°C	1.04	1.13	1.20	1.23

Statistical testing included t-tests and correction for multiple testing. ***P*<0.01, ****P*<0.001 (FDR=0.05).

Table S2 Fold-changes in tissue-specific gene expression between population samples in third instar larvae

Gene	Tissue	Gene expression ratio between population samples		
		Population samples	Rearing temperature	
			17°C	28°C
vg	Wing disc	Sweden (SU)/The Netherlands (NL)	0.86	0.60
		Sweden (SU)/Zimbabwe (ZK)	0.75	0.74
		Sweden (SU)/Zambia (ZI)	1.35	0.78
		The Netherlands (NL)/Zimbabwe (ZK)	0.87	1.59
		The Netherlands (NL)/Zambia (ZI)	1.56	1.30
		Zimbabwe (ZK)/Zambia (ZI)	1.79*	0.82
	Brain	Sweden (SU)/The Netherlands (NL)	0.77	1.14
		Sweden (SU)/Zimbabwe (ZK)	0.84	0.90
		Sweden (SU)/Zambia (ZI)	1.67	0.97
		The Netherlands (NL)/Zimbabwe (ZK)	1.09	1.07
		The Netherlands (NL)/Zambia (ZI)	2.17*	1.27
		Zimbabwe (ZK)/Zambia (ZI)	1.99	1.18
Aats-asp	Wing disc	Sweden (SU)/The Netherlands (NL)	0.94	0.86
		Sweden (SU)/Zimbabwe (ZK)	0.84	0.87
		Sweden (SU)/Zambia (ZI)	1.06	0.97
		The Netherlands (NL)/Zimbabwe (ZK)	0.90	1.01
		The Netherlands (NL)/Zambia (ZI)	1.13	1.13
		Zimbabwe (ZK)/Zambia (ZI)	1.26	1.12
	Brain	Sweden (SU)/The Netherlands (NL)	0.85	0.93
		Sweden (SU)/Zimbabwe (ZK)	0.84	1.00
		Sweden (SU)/Zambia (ZI)	1.27	1.14
		The Netherlands (NL)/Zimbabwe (ZK)	1.00	1.08
		The Netherlands (NL)/Zambia (ZI)	1.50	1.23
		Zimbabwe (ZK)/Zambia (ZI)	1.51*	1.15

Statistical testing included t-tests and correction for multiple testing. **P*<0.05 (FDR=0.05).

Table S3 Candidate SNPs

Population samples	Genomic position	F_{ST}	Gene region	Regulatory region	TFBS
Sweden (SU)	2R:12,884,497	0.62	promoter	PRE	*dl, Med, D Trl, bab1, twi, da, prd, sens, hb*
	2R:12,888,606	0.73	intronic	-	
	2R:12,893,153	0.65	intronic	-	
	2R:12,894,323	0.74	intronic	-	
The Netherlands (NL) /France (FR)	2R:12,898,469	0.63/0.61	intronic	-	*chinmo, cad*
	2R:12,898,497	0.63/0.61	intronic	-	*chinmo, cad*
	2R:12,898,678	0.73/0.71	intronic	-	

Genomic position is given according to FlyBase release 6. F_{ST} (Weir and Cockerham 1984) was calculated between the respective derived and the ancestral Zambian population samples. Information about regulatory regions and transcription factor binding sites (TFBS) that were identified by immunoprecipitation (ChIP) experiments were also derived from FlyBase release 6 (Pierre *et al.* 2014).

Fine-Mapping and Selective Sweep Analysis of QTL for Cold Tolerance in *Drosophila melanogaster*

Ricardo Wilches,[1] Susanne Voigt, Pablo Duchen, Stefan Laurent, and Wolfgang Stephan
Section of Evolutionary Biology, Department of Biology II, Ludwig-Maximilian University of Munich, 82152 Planegg-Martinsried, Germany
ORCID ID: 0002-7862-1507 (R.W.)

ABSTRACT There is a growing interest in investigating the relationship between genes with signatures of natural selection and genes identified in QTL mapping studies using combined population and quantitative genetics approaches. We dissected an X-linked interval of 6.2 Mb, which contains two QTL underlying variation in chill coma recovery time (CCRT) in *Drosophila melanogaster* from temperate (European) and tropical (African) regions. This resulted in two relatively small regions of 131 kb and 124 kb. The latter one co-localizes with a very strong selective sweep in the European population. We examined the genes within and near the sweep region individually using gene expression analysis and *P*-element insertion lines. Of the genes overlapping with the sweep, none appears to be related to CCRT. However, we have identified a new candidate gene of CCRT, *brinker*, which is located just outside the sweep region and is inducible by cold stress. We discuss these results in light of recent population genetics theories on quantitative traits.

KEYWORDS
cold tolerance
QTL
fine-mapping
selective sweep
brinker

Quantitative genetics assumes that selection on an adaptive trait involves simultaneous selection at multiple loci contributing to this trait. This causes small to moderate allele frequency shifts at these loci (Barton and Keightley 2002). Therefore, adaptation does not require new mutations in the short-term. Instead, selection may use alleles that are found in the standing genetic variation (Pritchard and Di Rienzo 2010). Genome-wide data suggest that this quantitative genetic view is relevant (Mackay *et al.* 2012). In particular, association studies confirm that quantitative traits are typically polygenic.

However, there is the view that the molecular population genetics scenario of selective sweeps is also important in describing selection on quantitative traits. These sweeps may be caused by new mutations or low-frequency alleles from the standing variation. Empirical evidence for the occurrence of sweeps at QTL has been reported in studies of artificial selection, including the domestication of chickens (Rubin *et al.* 2010), dogs (Axelsson *et al.* 2013), pigs (Rubin *et al.* 2012), and cattle (Qanbari *et al.* 2014). Furthermore, there is growing evidence of sweeps associated with positive directional selection on quantitative traits in natural populations. Linnen *et al.* (2013) studied coat color adaptation in deer mice controlled by a single large gene that shows multiple signatures of sweeps. Incomplete sweeps in the enhancer region of the gene *ebony* have contributed to the darker color of the abdominal segments of high-altitude *Drosophila melanogaster* from Uganda (Pool and Aquadro 2007; Rebeiz *et al.* 2009). Sweeps have also been observed at the EDA locus in threespine sticklebacks associated with the reduction of lateral plate armor in fresh water environments (Cano *et al.* 2006). In the common sunflower, selective sweeps have revealed candidate genes for adaptation to drought and salt tolerance (Kane and Rieseberg 2007).

Sophisticated methods have been developed to detect selective sweeps in a genome (Thornton *et al.* 2007; Pavlidis *et al.* 2008; Stephan 2010). In this study, we utilized these methods in combination with quantitative genetics tools to dissect a QTL interval for cold tolerance in *D. melanogaster*. Cold tolerance has been shown to be driven by environmental selection (Hoffmann *et al.* 2002; Schmidt *et al.* 2005) and to have a highly polygenic basis (Morgan and Mackay 2006; Norry *et al.* 2008; Svetec *et al.* 2011; Mackay *et al.* 2012). In a previous analysis, using strains from African (tropical) and European (temperate) regions, we have identified X-linked QTL controlling chill coma recovery time (CCRT), a proxy for cold tolerance (Svetec *et al.* 2011). Here, we chose one interval of size 6.2 Mb, which contains two QTL shared between both sexes, and fine-mapped it using quantitative

doi: 10.1534/g3.114.012757
Manuscript received March 25, 2014; accepted for publication June 20, 2014; published Early Online June 26, 2014.

Supporting information is available online at http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.114.012757/-/DC1
[1]Corresponding author: Department of Biology II, Ludwig-Maximilian University of Munich, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany. E-mail: wilches@bio.lmu.de

complementation tests. This resulted in a 131-kb region and a 124-kb region. In the European population, a strong selective sweep co-localized with the 124-kb region. We then analyzed this region in detail using population genetics and gene expression analyses. We found that the genes within the selective sweep region are probably not related to the trait, but a gene (*brinker*) just outside the sweep is induced by cold stress.

MATERIALS AND METHODS

Fly lines

To conduct quantitative complementation tests on chromosomal deletions, a set of available deficiency lines were ordered at the Bloomington stock center (http://flystocks.bio.indiana.edu). Although the QTL interval was defined by Svetec *et al.* (2011) to be 6.2 Mb long (6C to 11D), the set of available deletions spans 5.8 Mb of its total length between coordinates 6,642,419 and 12,461,494 that correspond to cytological bands 6C to 11B (Figure 1). This 5.8-Mb-long interval is covered by a total of 24 overlapping deletions with known breakpoints at the sequence level in 92% of the cases. Additional deficiencies were tested if necessary.

The African and European versions of the X-chromosome used in the complementation tests are contained in fly lines A* and E*, created by introgressing one X-chromosome from a population of Zimbabwe and one from the Netherlands into a common laboratory strain (Svetec *et al.* 2011). Hence, these two lines bear different X-linked alleles while the rest of the major nuclear chromosomes and mitochondrial DNA are the same. These two lines are the parents of an X-recombinant inbred population used to localize the QTL interval that concerns us in this project.

Prior to CCRT scoring experiments, virgin female flies bearing the deficiency chromosome and the respective balancer were mated with males of the A* and E* lines, respectively. Eggs were allowed to develop in the same medium in which they were laid at 23°. Female F1 were sorted on hatching by phenotype as balancer or deletion bearer. Because all balancer types used to maintain the deletions have a dominant mutation for eye shape at the *Bar* (B^1) locus, F1 flies exhibiting the B^1 mutant phenotype were considered as balancer bearers, whereas wild-type appearance was indicative of bearing the deletion. Sorted flies were kept at room temperature until CCRT scoring on their fourth to sixth day of life.

Assessments of expression levels of candidate genes were conducted using 4- to 6-day-old female flies belonging to the Netherlands population (isofemale lines: NL01, NL12, NL14, NL15, NL16, NL18, NL19, NL20) and the Zimbabwean population (isofemale lines: ZK84, ZK131, ZK145, ZK157, ZK186, ZK229, ZK377, ZK398). Flies were reared at 23° and subjected to cold stress in the same manner as reported for CCRT scoring. Three flies per line were used as controls (not exposed to cold). Three flies of each line were snap-frozen in liquid nitrogen at 10 min after being brought to room temperature, whereas three remaining flies per line were scored for their CCRT and frozen 15 min after the minute in which they were reported as recovered. Control flies, which remained at 23° in glass vials during the 7 hr of treatment, were also snap-frozen at the end of this time period. Frozen material was stored at −80° until RNA extractions were performed. Population pools per line/treatment were made prior to RNA extraction. Each population pool per treatment consisted of eight flies of the same population. Three population pools per treatment were made for both the Netherlands and Zimbabwe.

CCRT scoring

Once flies reached 4 to 6 days of age, they were scored for CCRT following the protocol of Svetec *et al.* (2011). Briefly, flies were transferred to glass vials without anesthesia and placed in an ice-water bath of 0° for 7 hr. At the end of this time period, flies were brought back to room temperature (23°) and observed in time intervals of 1 min. The minute in which a fly was standing on its feet was recorded as its CCRT.

Quantitative complementation tests on deficiencies

On average, 35 female flies per each of the four resulting genotypes *E*/def*, *A*/def*, *E*/bal*, and *A*/bal* were scored. For ANOVA analysis, log-transformed CCRT scores per genotype, line (L), and genomic background (G) were kept as fixed effects. We focused on the significance of the interactions of these two factors ($L \times G$) as well as on the following two conditions to call the procedure quantitative failure to

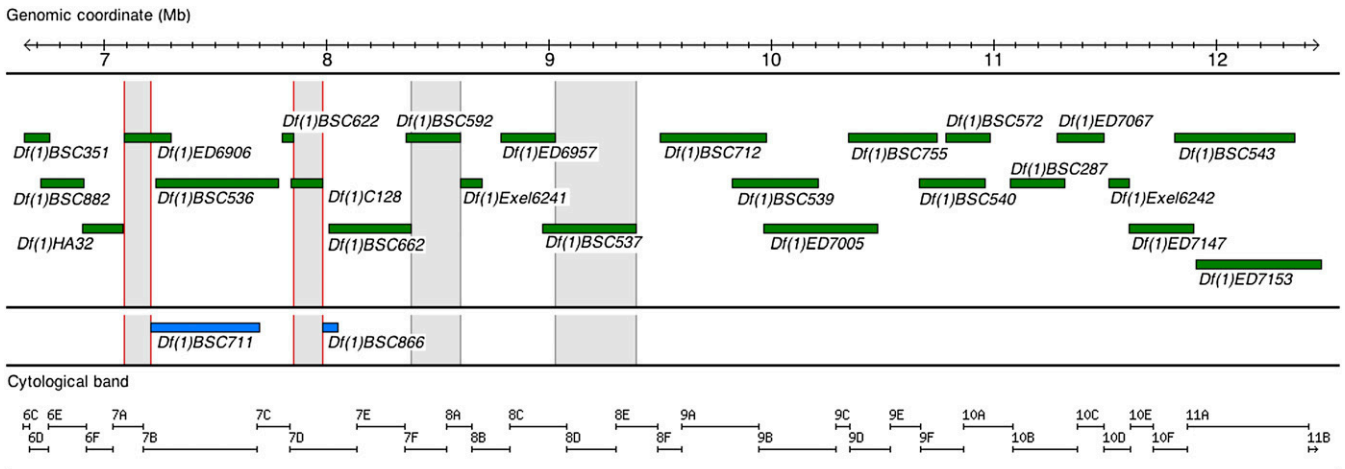


Figure 1 Map of tested deletions within the QTL interval undergoing study. All deletions, represented by green or blue bars, span a 5.8-Mb fraction of the 6.2-Mb-long interval of interest on the X-chromosome. Deletion breakpoints at the base pair level are known for all deletions except *Df(1)HA32* and *Df(1)C128*, for which only cytological bands are reported. Both physical and cytological coordinates are provided. The 24 deletions represented in green represent the minimum available set spanning the 5.8-Mb QTL interval; deletions in blue were tested on failure to complement of one of the overlapping deletions in green. Fractions of the QTL interval with light gray shading indicate regions of interest under deletions that show failure to complement. Red borders of this gray background indicate highly significant failure to complement ($P < 0.01$).

complement: the differences in CCRT for the genotypes bearing the balancer had to be small compared with that of the genotypes bearing the deletion; and in the latter case, the E^*/def flies should show reduced CCRT with respect to the A^*/def genotypes. It is expected that if these conditions are satisfied, failures to complement due to the presence of other QTL outside the region in question can be ruled out. However, a failure to complement detected with a given deficiency either can be caused by its interaction with QTL alleles in the region uncovered by a deficiency (allelism) or arise by interaction between this deficiency and QTL alleles elsewhere in the genome (epistasis) (Pasyukova *et al.* 2000; Service 2004). Because we are interested in the allelic cases of failure to complement by using the E^* and A^* lines (as well as its inbred wild-type progenitor lines) in the tests, we control for most of the epistatic effects that can be caused by loci residing on chromosomes 2 and 3. Bonferroni correction was applied to control for multiple testing.

RNA extraction and cDNA synthesis

RNA was extracted from pools using the MasterPure RNA Purification Kit (Epicentre Biotechnologies, Madison, WI), followed by DNase treatment. Purified RNA was quantified with a nanodrop apparatus and tested for genomic DNA contamination based on a PCR (Phusion) protocol using a primer pair binding in nontranscribed regions of the X-chromosome (Primer code: X-1435; sequence available on request). Samples tested positive for genomic DNA were excluded from downstream protocols. cDNA synthesis was performed with SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, CA) on 1400 ng of RNA per reaction.

RT-qPCR assays

RT-qPCR assays for candidate genes *CG1958*, *CG1677*, *CG2059*, *unc-119*, *brk*, and *Atg5* were performed with primers designed using the online tool QuantPrime (www.quantprime.de) to match all possible transcript types per candidate gene. The ribosomal genes *RpS20* and *RpL32* were taken as reference genes, against which relative gene expression levels of our genes of interest were normalized. RT-qPCR assays consisted of a total of three biological replicates each run in triplicate and were conducted on a Real-Time thermal cycler CFX96 platform (BioRad, Hercules, CA). Each reaction was taken to a final volume of 10 μl using iQ SYBR Green Supermix (BioRad, Hercules, CA). Further details of the experimental setup, such as amplification efficiency assessments with dilution series, can be provided on request.

Obtained Cq values per replicate within line (or pool) and treatment were transformed to calibrated normalized relative quantities (CNRQ) following Hellemans *et al.* (2007). Log-transformed CNRQs were then used to test the hypothesis of expression differences between pairs of lines (or pool) within and between treatments. For this purpose, Welch two-sample *t*-tests were performed on comparisons with fold differences above a threshold (defined by the variance between technical replicates). The Benjamini and Hochberg (1995) *P*-value correction was applied to control for false positives, due to the high number of simultaneous tests performed.

Basic population genetics analysis

Molecular variation was characterized in the genomic fragment uncovered by the deletion *Df(1)ED9606*; *i.e.*, a total of 124 kb between coordinates 7,089,000 and 7,212,999. Publically available whole-genome sequences generated by Illumina NGS technology for four *D. melanogaster* populations were retrieved from the Drosophila Population Genomics Project (DPGP) at http://www.dpgp.org. The populations include the Netherlands (NL) with 11 lines (NL01, NL02, NL11, NL12, NL13, NL14, NL15, NL16, NL17, NL18, NL19), France (FR) with 8 lines (FR14, FR151, FR180, FR207, FR217, FR229, FR310, FR361), Rwanda (RG) with 25 lines (RG2, RG3, RG4N, RG5, RG6N, RG7, RG8, RG9, RG10, RG11N, RG13N, RG15, RG18N, RG19, RG21N, RG22, RG24, RG25, RG28, RG2, RG32N, RG33, RG36, RG37N, RG38N), and Southeast Africa consisting of a pool of 18 lines from Zambia (ZI91, ZI261, ZI268, ZI468, ZO12, ZO65), Zimbabwe (ZK84, ZK131, ZK186, ZS5, ZS11, ZS56), and Malawi (MW6, MW11, MW28, MW38, MW46, MW63). The corresponding *D. simulans* sequence (Hu *et al.* 2013) was used as outgroup.

The following quality-control steps during the initial handling of the sequence data were used. First, nucleotides with a PHRED score lower than 21 were set to N. Unless otherwise stated, this quality criterion was applied to all analyses in which DPGP sequence data were used. Second, if a given polymorphic site in the alignment showed a frequency of N higher than 10%, then it was excluded from the analysis. The following summary statistics were then computed on 2-kb-long nonoverlapping windows: θ_π (Tajima 1983), θ_W (Watterson 1975), and divergence (D_{xy}) to the outgroup (Nei 1987). In addition, pairwise F_{ST} was calculated as normalized distance of Nei (Nei and Li 1979). Neutrality tests based on the site frequency spectrum (SFS) using the Tajima (1989) *D* statistic were also calculated.

Composite likelihood ratio test of positive selection

To investigate whether the observed SFS in the region of interest is compatible with the one expected after a selective sweep, we calculated the composite likelihood ratio (CLR) statistic (Kim and Stephan 2002; Nielsen *et al.* 2005; Pavlidis *et al.* 2010) as it is implemented in the software SweeD (Pavlidis *et al.* 2013). This likelihood ratio test statistic compares a selective sweep model and a neutral model that is calibrated with the genomic background frequency spectrum. We used the parallel version of the software (SweeD-P) to calculate the CLR statistic along the X-chromosome in our European sample (19 lines from the Netherlands and France). In addition to the classes of the SFS (*i.e.*, 1 to *n*-1, where *n* is the sample size), we considered two additional site classes consisting of sites that are monomorphic in the European sample and polymorphic in the Rwandan sample. Extending the SFS in this way was shown to improve the power of the method to detect selective sweeps (Nielsen *et al.* 2005). SweeD was run on a 16-core CPU using the command line option "- -monomorphic" with 500,000 grid points. The background SFS was taken from the complete X-chromosome. However, following Pool *et al.* (2012), we excluded from the analysis telomere and centromere regions of the X-chromosome due to their very low recombination rate. The coordinates of the excluded regions range from the origin until position 2,222,391 for the telomere and from position 20,054,556 to the end for the centromere region. Finally we compared the CLR profile of our region of interest to the profile calculated for the complete chromosome.

The significance level of the CLR-test statistic was calculated by simulating large genomic regions with the coalescent simulator fastsimcoal2 (Excoffier and Foll 2011) under a neutral model that takes into account our current knowledge of the demography of European populations of *D. melanogaster* (Laurent *et al.* 2011). For every one of the simulated datasets, we computed the CLR-test statistic in the same way as we did for the observed dataset and recorded the maximum CLR value. We used the 95th quantile of the distribution of top CLR values as our significance threshold. Because this analysis becomes computationally intensive as the size of the simulated genomic region increases, we investigated the relation between the threshold value and the size of the simulated region. We simulated

batches of 100 datasets of increasing size from 50 to 5000 kb in length and took the asymptotic value reached as the chromosomal threshold.

F_{ST}-based scan for positive selection

For the set of F_{ST} analyses performed with BayeScan (Foll and Gaggiotti 2008) (http://cmpg.unibe.ch/software/bayescan/), input files were prepared following the authors' instructions. The different runs were performed using default parameters with the sequence data of the same DPGP samples from the Netherlands, France, Rwanda, and Southeast Africa. In addition, lines from the following African populations were included: Ethiopia (ED2, ED3, ED5N, ED6N, ED10N, EZ2, EZ5N, EZ9N, EZ25), Cameroon (CO1, CO2, CO4N, CO8N, CO9N, CO10N, CO13N, CO14, CO15N, CO16), and South Africa (SP80, SP173, SP188, SP221, SP235, SP241, SP254). SNP exclusion criteria were as follows: positions showing more than two segregating alleles as well as sites with less than 50% base calls in one population were excluded from the analysis.

RESULTS

We first describe the results of the deficiency mapping and then those of the population genetics analysis and the gene expression studies.

Quantitative deficiency mapping

To dissect the QTL interval reported by Svetec *et al.* (2011) corresponding to the interval at 6C-11D of approximately 6.2 Mb in length, we conducted quantitative complementation tests for 24 overlapping deletions spanning 94% of this interval. The chromosome fractions comprising the remaining 6% of the interval were left untested due to lack of suitable deletions. With the set of 24 deficiencies (Figure 1), we could potentially uncover the effect of line-specific alleles (line-specific refers to the type of X-chromosome involved in the test, which is either African or European) at 588 (95%) of the 622 annotated and computationally predicted genes within the interval. Overall, 14 of 24 of the tested deficiencies showed significant line effects at the 5% level, whereas 9 of 24 showed a significant effect of the genomic background on CCRT scores (with the term "genomic background" we refer to the involved deletion and balancer chromosomes; see *Materials and Methods*). We observed failure to complement in 4 of the 24 tested deletions (Table 1). Failure to complement implied both a significant effect of line (L) and a significant line by genomic background interaction ($L \times G$) as long as the differences in CCRT followed the expected direction. That is, there were shorter CCRT times for flies bearing the E* X-chromosome in the presence of the deletion compared with the corresponding flies bearing the A* X-chromosome in the presence of the same deletion, but no difference was shown between the CCRT of the flies bearing the E* and A* X-chromosomes in the presence of the balancer chromosome.

Deletions *Df(1)ED6906* and *Df(1)C128* were the only ones that revealed a highly significant failure to complement. In the case of *Df(1)ED6906*, the difference between the means of the CCRT scores for the flies bearing this deletion is 9.18 min, whereas that of the flies harboring the balancer chromosome is 1.82 min (Table 1). This means that deletion *Df(1)ED6906* has potentially uncovered E* X-chromosome alleles and/or alleles with the opposite effect residing on the A* X-chromosome. A similar interpretation can be given for the results obtained for *Df(1)C128*, *Df(1)BSC592* and *Df(1)BSC537*, which also failed to complement, as evidenced by the significant L and $L \times G$ effects and by the higher CCRT differences in the presence of the deficiency than in the presence of the balancer. However, for the two last deletions, these effects were only marginally significant (Table 1). Thus, they are not considered for further study.

Although our results with *Df(1)ED6906* and *Df(1)C128* meet the requirements to be considered allelic failures to complement, these can also be interpreted as an epistatic failure to complement due to interactions of these deficiencies with other loci that affect CCRT residing elsewhere on the X or in the other two major chromosomes. We are aware of the limitation of quantitative deficiency mapping to tell these two causes apart (Pasyukova *et al.* 2000). This is also a problem in similar studies (Fanara *et al.* 2002; De Luca *et al.* 2003; Geiger-Thornsberry and Mackay 2004; Harbison *et al.* 2004). However, because we used the E* and A* lines that share the same genetic background for their respective wild-type–derived X-chromosomes and not the wild-type inbred lines NL14 and ZK157, we can exclude interactions with factors located outside the X-chromosome.

The fact that we used a set of overlapping deficiencies allowed us to better define the stretch that revealed quantitative failure to complement. With respect to the 210-kb-long deletion *Df(1)ED6906*, the 67.15 kb overlapping with deletion *Df(1)BSC536* were subtracted from the stretch of interest (Figure 1). Furthermore, the results of the complementation tests with yet another overlapping deficiency at the same end (*Df(1)BSC711*) allowed us to subtract an additional 19.64 kb from the 210 kb encompassing *Df(1)ED6906* (Figure 1). At the other end of deletion *Df(1)ED6906*, its overlap with deletion *Df(1)HA32* is not known at the base pair level. Thus, the redefined fraction of interest under deletion *Df(1)ED6906* encompasses 124 kb between coordinates 7.09 and 7.21 Mb. Similarly, for the other highly significant deletion *Df(1)C128*, the redefined region of interest has a of length 131 kb (between coordinates 7.85 and 7.98 Mb).

This quantitative complementation mapping approach based on overlapping deletions has allowed us to reduce the number of initial candidate genes within the QTL undergoing study from 622 to a subset of 21. A total of 7 genes are located within the 124 kb uncovered by deletion *Df(1)ED6906*, and 14 genes were uncovered by deletion *Df(1)C128*. This is remarkable given the substantial fraction of uncharacterized genes in the 6.2 Mb of the QTL defined by Svetec *et al.* (2011) and the absence of known *a priori* candidate genes for CCRT in this chromosomal region.

In the next section we show that of the two regions identified by the complementation tests, the 124-kb region uncovered by deletion *Df(1)ED6906* overlaps with a selective sweep identified in several previous analyses (Glinka *et al.* 2006; Boitard *et al.* 2012; Langley *et al.* 2012), whereas we did not detect evidence for positive directional selection in deletion *Df(1)C128* (see below). Therefore, in the following section we focus on deletion *Df(1)ED6906*.

Molecular population genetics analysis

We characterized molecular variation in the genomic region of 124 kb uncovered by deletion *Df(1)ED9606* in natural populations of *D. melanogaster*. First, we calculated a set of summary statistics on a 2-kb nonoverlapping window basis using next-generation sequence data from two European (the Netherlands and France) and two African (Rwanda and Southeast Africa) populations. The Netherlands population and a set of Southeast African lines represent the gene pools from which the E* and A* lines were derived. The additional two populations consisted of French and Rwanda sequence data from the DPGP (Pool *et al.* 2012). These four populations allowed us to draw conclusions on patterns of variation in temperate and tropical populations.

For each population we obtained nucleotide diversity estimates measured by the average number of pairwise differences (θ_π) and Watterson's estimator (θ_W). The European populations showed a three-fold to four-fold reduction in nucleotide diversity when compared

Table 1 Deficiency analysis of X-linked QTL affecting CCRT in female flies

		Mean CCRT (SD), min									
Deletion	Balancer	E*/Deletion	E*/Balancer	A*/Deletion	A*/Balancer	Δdef	Δbal	P L	P G	P L × G	Stock No.
Df(1)BSC351	*FM7h*	31.70 (7.99)	30.91 (7.88)	31.46 (9.42)	32.73 (8.72)	0.24	−1.82	0.070057	0.838897	0.343401	24375
Df(1)BSC882	*FM7h*	29.39 (9.94)	30.91 (7.88)	32.69 (12.41)	32.73 (8.72)	−3.3	−1.82	0.018773	0.149609	0.574953	30587
Df(1)HA32	*FM7c*	37.75 (8.24)	32.16 (9.01)	41.61 (10.42)	33.25 (9.28)	−3.86	−1.09	0.177930	0.000001	0.454433	947
Df(1)ED6906	*FM7h*	26.93 (5.66)	30.91 (7.88)	36.28 (8.52)	32.73 (8.72)	−9.35	−1.82	0.000103	0.779708	0.000289	8955
Df(1)BSC711	*FM7h*	35.73 (6.52)	30.91 (7.88)	34.03 (8.45)	32.73 (8.72)	1.7	−1.82	0.944120	0.011307	0.121501	26563
Df(1)BSC536	*FM7h*	36.27 (9.81)	30.91 (7.88)	36.43 (11.34)	32.73 (8.72)	−0.16	−1.82	0.055280	0.002486	0.473532	25064
Df(1)BSC622	*Binsinscy*	33.37 (9.22)	35.59 (8.83)	34.50 (8.65)	36.60 (9.59)	−1.13	−1.01	0.386305	0.121365	0.820123	25697
Df(1)C128	*FM6*	26.97 (6.16)	30.11 (8.63)	37.44 (8.11)	32.63 (7.74)	−10.48	−2.52	0.000012	0.606252	0.000885	949
Df(1)BSC866	*Binsinscy*	36.26 (8.16)	35.59 (8.83)	37.46 (11.06)	36.60 (9.59)	−1.21	−1.01	0.480645	0.501206	0.850165	29989
Df(1)BSC662	*Binsinscy*	40.29 (8.98)	35.59 (8.83)	39.89 (9.70)	36.60 (9.59)	0.39	−1.01	0.330040	0.002076	0.615274	26514
Df(1)BSC592	*Binsinscy*	31.42 (7.47)	35.59 (8.83)	37.92 (9.16)	36.60 (9.59)	−6.51	−1.01	0.063548	0.639100	0.031094	25426
Df(1)Exel6241	*FM7c*	31.57 (7.29)	32.16 (9.01)	30.29 (8.76)	33.25 (9.28)	1.29	−1.09	0.685768	0.224516	0.294193	7715
Df(1)ED6957	*FM7j*	27.90 (7.68)	32.16 (9.01)	29.21 (6.77)	33.25 (9.28)	−1.31	−1.09	0.261153	0.001693	0.795931	8033
Df(1)BSC537	*FM7h*	29.64 (7.36)	30.91 (7.88)	35.36 (8.17)	32.73 (8.72)	−5.71	−1.82	0.004959	0.528960	0.099098	25065
Df(1)BSC712	*FM7j*	40.21 (8.39)	35.59 (8.83)	39.66 (9.36)	36.60 (9.59)	0.55	−1.01	0.623921	0.005147	0.565518	26564
Df(1)BSC539	*Binsinscy*	31.50 (7.07)	35.59 (8.83)	34.30 (7.52)	36.60 (9.59)	−2.8	−1.01	0.239470	0.026941	0.414266	25067
Df(1)ED7005	*FM7h*	29.63 (6.45)	30.91 (7.88)	29.21 (6.46)	32.73 (8.72)	0.43	−1.82	0.060255	0.075504	0.350039	9153
Df(1)BSC755	*Binsinscy*	30.36 (7.49)	30.11 (8.63)	33.33 (6.92)	32.63 (7.74)	−2.98	−2.52	0.009057	0.547656	0.88738	26853
Df(1)BSC540	*FM7h*	32.20 (7.51)	30.91 (7.88)	34.73 (9.31)	32.73 (8.72)	−2.53	−1.82	0.022620	0.141156	0.845784	25068
Df(1)BSC572	*FM7h*	31.11 (6.04)	30.91 (7.88)	37.05 (7.54)	32.73 (8.72)	−5.94	−1.82	0.009129	0.070999	0.152617	25391
Df(1)BSC287	*Binsinscy*	36.11 (9.57)	35.59 (8.83)	35.23 (8.06)	36.60 (9.59)	0.87	−1.01	0.596120	0.852371	0.643564	23672
Df(1)ED7067	*FM7h*	28.89 (7.99)	30.91 (7.88)	29.58 (7.80)	32.73 (8.72)	−0.68	−1.82	0.039538	0.029350	0.702833	9154
Df(1)Exel6242	*FM7c*	33.96 (7.48)	32.16 (9.01)	32.93 (7.68)	33.25 (9.28)	1.03	−1.09	0.585559	0.397537	0.406384	7716
Df(1)ED7147	*FM7h*	31.43 (7.44)	30.91 (7.88)	34.96 (8.42)	32.73 (8.72)	−3.52	−1.82	0.010630	0.272976	0.497331	9171
Df(1)BSC543	*FM7h*	30.88 (6.43)	30.91 (7.88)	33.75 (6.67)	32.73 (8.72)	−2.88	−1.82	0.012423	0.689596	0.674376	25071
Df(1)ED7153	*FM7h*	29.80 (6.30)	30.91 (7.88)	30.47 (6.82)	32.73 (8.72)	−0.67	−1.82	0.028176	0.134114	0.603622	8953

Summary of quantitative deficiency tests performed with the listed deletions. Δ*def* is the difference between the average CCRT of flies bearing E* and A* chromosomes in the presence of a given deletion. Negative differences suggested the presence of CCRT reducing alleles at sites potentially uncovered by the deletion. Δ*bal* is the difference between the average CCRT of flies bearing E* and A* chromosomes in the presence of a given balancer chromosome. Note that deletions held with the same balancer show the same the Δ*bal* values. *P* L is the value for the line effect (E* or A*) from two-way ANOVA analysis. *P* G is the value for the genomic background effect (deletion or balancer) effect from two-way ANOVA analysis. *P* L × G P is the value for the interaction between the two aforementioned variables. Stock no. is the code number under which the fly line bearing the deletion can be ordered at the Bloomington Stock Center in Indiana.

with the African populations (Table 2). Supporting Information, Figure S1 depicts the values of the 2-kb windows along the entire region of 124 kb in the four populations. The θ_π values in the Netherlands and French populations are along a 40-kb fragment (positions 65,000 to 105,000) 1 SD lower than the average over the 124-kb region. This pattern is in contrast to that observed in the two African populations for the same coordinates, for which nucleotide diversity values tend to be above their respective population averages. F_{ST} estimates along the 124-kb region between each of the European populations and the Southeast African pool parallel the diversity estimates such that F_{ST} vales are higher where diversity is low in the two European populations (Figure S1). Regarding divergence of each population to *D. simulans*, the averages of the two European populations are approximately 10% whereas those of the African populations have values between 12% and 13% (Table 2).

We also analyzed possible deviations from the standard neutral SFS using Tajima's *D* statistic (Figure S2). The African samples from Rwanda and Southeast Africa show generally negative *D* values that are typical for these populations (Glinka *et al.* 2003; Pool *et al.* 2012). The European samples have average *D* values near zero and a larger variance, as was also found previously (Glinka *et al.* 2003). Most interestingly, however, is the 40-kb window from position 65,000 to 105,000, in which Tajima's *D* has lower than average values (except for a peak around coordinate 87,000).

The patterns of polymorphism observed in the region of interest in the European and African populations revealed a conspicuous reduction of genetic variability and a negative Tajima's *D* in both European populations that extends for approximately 40 kb between relative position 65,000 and 105,000 in the 124-kb region. This reduction already has been identified as a footprint of positive selection in non-African populations (Glinka *et al.* 2006; Boitard *et al.* 2012; Langley *et al.* 2012). In this work, motivated by the link to the QTL, we revisited these analyses.

Because demographic scenarios (particularly bottlenecks) can create similar genomic patterns as positive directional selection (Stephan and Li 2007), we subjected the data of the available European samples

Table 2 Summary statistics average for the QTL undergoing study in four *D. melanogaster* populations

Population	θ_π, Mean (SD)	θ_W, Mean (SD)	D_{xy}, Mean (SD)	*Tajima's D*, Mean (SD)
The Netherlands	0.0010 (0.0007)	0.0010 (0.0006)	0.0953 (0.0352)	−0.4995 (1.0403)
France	0.0008 (0.0007)	0.0008 (0.0006)	0.1042 (0.0382)	−0.1010 (0.869)
Rwanda	0.0031 (0.0011)	0.0041 (0.0013)	0.1286 (0.0422)	−0.9671 (0.4024)
Southeast Africa	0.0034 (0.0011)	0.0042 (0.0012)	0.1248 (0.0482)	−0.7834 (0.3875)

(pooling the Netherlands and French lines) to the most advanced composite likelihood ratio test (Pavlidis *et al.* 2013). This likelihood ratio was computed between a selective sweep model and a neutral model that is calibrated with the genomic background frequency spectrum. The background SFS was obtained from 20 Mb of the X-chromosome, excluding the telomere and centromere regions (see *Materials and Methods*). In our region of interest, the fragment between relative positions 65,000 and 105,000 exhibits a SFS that is in contrast to that of the genomic background and is better described by a selective sweep model (Figure 2A). The CLR values obtained for this interval (Λ_{CLR} >300) are within the top 1% of CLR values along the entire region of the X-chromosome analyzed (Figure S3) and are above the significance threshold of 72 that corresponds to the 95th quantile of the top CLR values of 100 simulated subgenomic regions of 5 Mb. This value did not increase when larger genomic regions were simulated (Figure S4). Simulations were based on our current understanding of the demographic history of European populations (Laurent *et al.* 2011; Duchen *et al.* 2013).

Using the same CRL test, we also analyzed in detail *Df(1)C128*, the second highly significant deletion uncovered by the quantitative complementation test (between coordinates 7.85 and 7.98 Mb). However, we could not find evidence for selective sweeps (see Figure S3).

Because a large fraction of the region of low variation in Europe (particularly the coding regions of genes *CG1958*, *CG1677*, *CG5059*, and *unc-119*; see the gene model below) (Figure 2B) contains no or very little variation, the CLR tests cannot be used to identify the targets of selection. Instead, we utilized genetic differentiation between African and European populations to obtain model-based F_{ST} coefficients (Foll and Gaggiotti 2008; Riebler *et al.* 2008) for each SNP within the 124-kb region of interest (Figure 2B). We considered SNP data from seven populations along a gradient across Africa and Europe: South Africa, Southeast Africa, Rwanda, Cameroon, Ethiopia, France, and the Netherlands. Using BayeScan (Foll and Gaggiotti 2008), we obtained a pattern of F_{ST} values from a dataset of 7316 SNPs with an average F_{ST} of 0.2621 and revealed four outlier SNPs that show the highest differentiation across populations at an FDR of 5% (Figure 2B). These four SNPs are located within the 40-kb-long fragment enriched for SNPs showing significant CLR values between positions 65,000 and 105,000 (Figure 2A). The 65-kb-long and 19-kb-long flanking regions to the left and to the right of the 40-kb fragment, respectively, are enriched for SNPs showing below-average F_{ST} values (Figure 2B). However, none of these SNPs with low differentiation across populations is significant at the 5% FDR.

The exclusion of European populations from the analysis did not change the pattern of high-differentiation outlier SNPs (results not shown). This suggests that allele frequency differentiation at outlier SNPs had already started within the African continent. Furthermore, because the European populations probably have experienced more severe bottlenecks than the African populations (Li and Stephan 2006; Pool *et al.* 2012), we may conclude that the BayeScan results are relatively robust against demographic changes and are not due to the strong bottleneck in the European population.

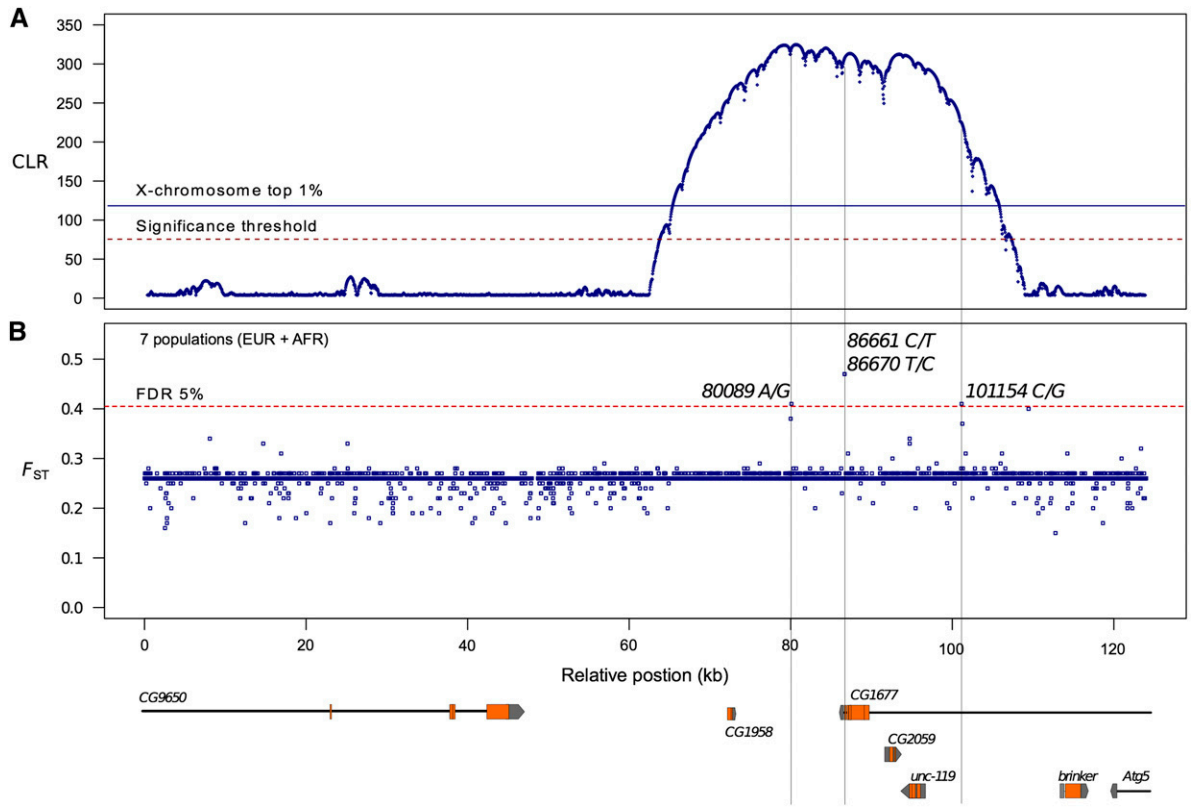


Figure 2 Evidence of positive selection and candidate SNPs in the 124-kb region under deletion *Df(1)ED6906*. (A) Likelihood (CLR) profile along the 124-kb on the X-chromosome using SNP data of two pooled European *D. melanogaster* from the Netherlands and France. Two significance thresholds are depicted. The solid line corresponds to the average of the top 1% CLR values for the X-chromosome in Europe and the dashed red line represents the significance threshold from simulations of equivalent subgenomic regions. (B) Model-based F_{ST} values for 7316 SNPs from a dataset including two European and five African samples (see *Materials and Methods*). Top SNPs (above the false discovery rate of 5%) are marked with position and a thin continuous line across panels.

Among the outlier SNPs that show high differentiation across the entire intercontinental dataset, the top ones are *86,661C/T* (F_{ST}=0.4697, α=2.02, q-value=0.0024) and *86,670T/C* (F_{ST}=0.4654, α=1.98, q-value=0.0042). These two nonsynonymous SNPs are located in exon 5 of the computationally predicted gene *CG1677* and show alleles in perfect linkage disequilibrium (LD) (Figure 3). The TT haplotype (*86,661T–86,670T*) is in high frequency in the Southeast African samples and is intermediate in Rwanda; its frequency decreases with increasing latitude to be replaced in the European populations by the CC haplotype. Both SNPs predict changes in the amino acid sequence of the protein. The common Southeast African form of the protein codes for a threonine (Thr) and an asparagine (Asn) at residues 936 and 939, respectively, whereas the cosmopolitan form has an alanine (Ala) and aspartic acid (Asp) at these two positions. The third highly significant SNP is *80,089A/G* (F_{ST}=0.4146, α=1.54, q-value=0.0313) located between genes *CG1958* and *CG1677* (Figure 2). Its allele frequency distribution across populations is also shown in Figure 3. Finally, SNP *101,154C/G* (F_{ST} = 0.4068, α=1.4804, q-value = 0.0481) is located 5 kb upstream of gene *unc-119* within the large intron of gene *CG1677* (Figure 2).

Candidate gene expression analyses and complementation tests with *P*-element insertion lines

We observed that the CLR profile of the selective sweep does not overlap with *brinker* and *Atg5*, but spans four of the seven candidate genes in the 124-kb region (see gene model below) (Figure 2B). To analyze whether these four genes in the sweep region are related to cold tolerance, we conducted expression analyses; *brinker* and *Atg5* were also included (Figure 4). qPCR assays were performed on cDNA prepared from pools of female flies from the Netherlands and Zimbabwe (see *Materials and Methods*). Expression of candidate genes was measured at two time points after cold stress exposure as well as under control conditions. The two post-cold stress time points were 10 min after the end of cold stress and 15 min after flies recovered from chill coma. Controls consisted of flies of the same lines that were not subjected to cold stress.

Of the six genes, *CG1958* and *brinker* showed significant differences in constitutive expression levels between the Netherlands pool and the Zimbabwean pool ($P < 0.01$) (Figure 4). This difference in expression levels between these populations also has been previously observed (Hutter *et al.* 2008). Furthermore, average expression level appeared to be unaffected by cold stress within pools at 10 min during recovery from chill coma for five of the genes. At this time point, the only highly significant difference between pools was observed for *brinker* ($P < 0.001$). Expression levels measured at 15 min after recovery from chill coma revealed one significant difference within pools: *brinker* was significantly overexpressed with respect to controls in the Netherlands pool ($P < 0.05$). Between-pool contrasts at 15 min after recovery from chill coma revealed only a significant difference for *brinker* ($P < 0.01$).

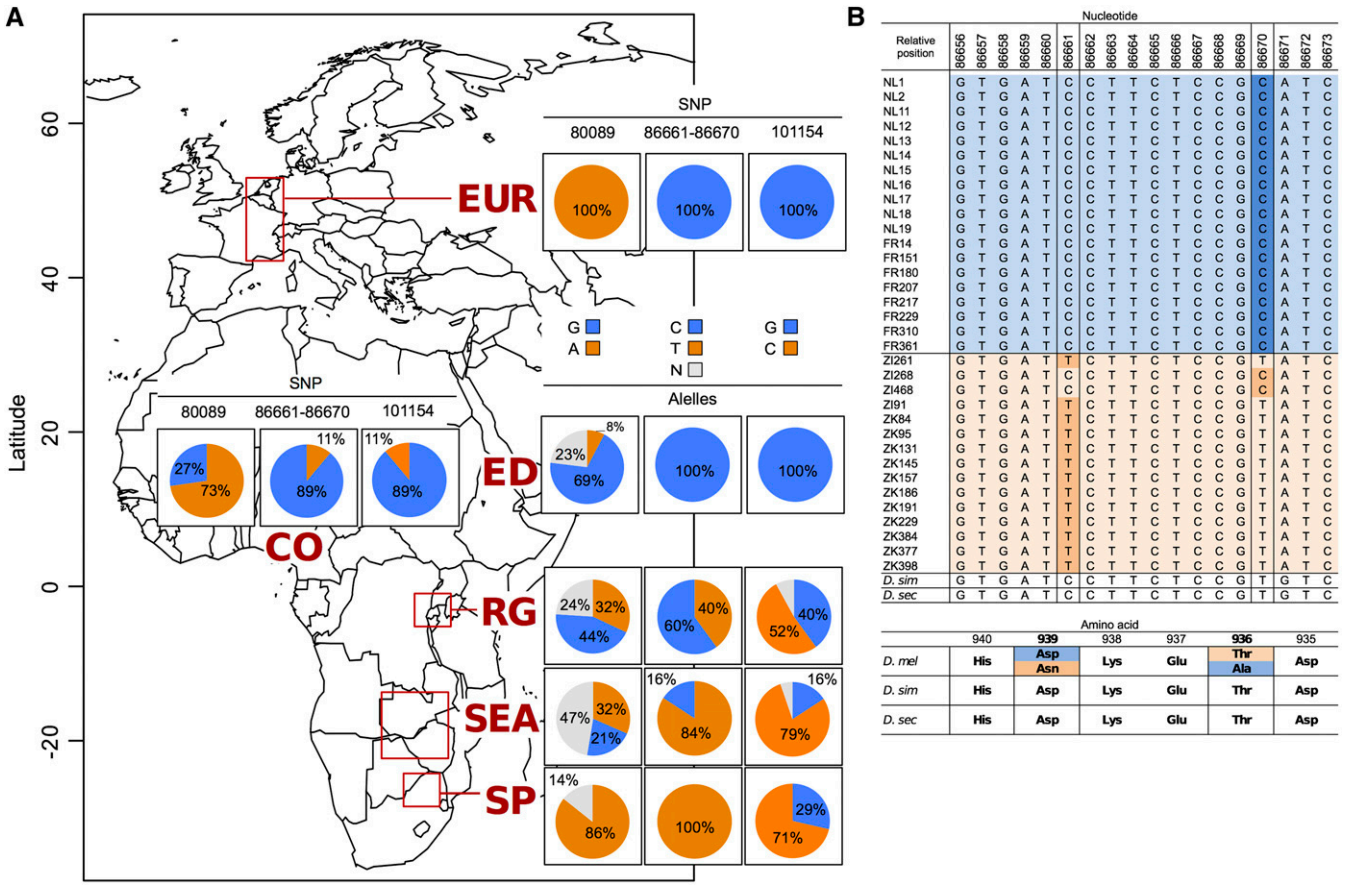


Figure 3 Allele frequency change at highly differentiated SNPs at the QTL of interest. (A) Allele frequencies of the top four highly differentiated SNPs across seven different populations along a latitudinal gradient. Populations are as follows: the Netherlands and France (EUR), Ethiopia (ED), Cameroon (CO), Rwanda (RG), Southeast Africa (SEA), and South Africa (SP) (see *Materials and Methods*). (B) European and Southeast African *D. melanogaster* haplotypes for the two nonsynonymous SNPs (86,661–86,670) in intron 5 of gene *CG1677*. These two SNPs correspond to amino acid positions 939 and 936.

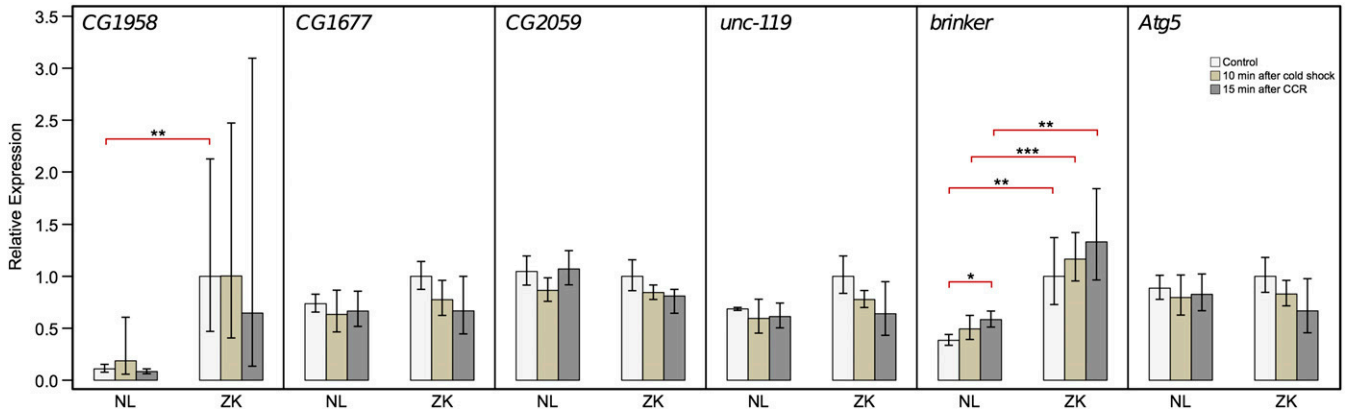


Figure 4 Expression of genes located in the region under deletion *Df(1)ED6906* that was affected by positive selection (see Figure 2A). Relative expression was measured under two cold stress and control conditions in pools of flies from a temperate population [the Netherlands (NL)] and a tropical population [Zimbabwe (ZK)]. Expression levels of these candidate genes were normalized with respect to that of the ribosomal genes *RpS20* and *RpL32*. The height of the bars represents the mean of three calibrated normalized relative quantities (CNRQ) per pool per gene rescaled to that of the corresponding ZK control. Error bars also represent rescaled confidence intervals. Levels of significance for tests of differences in expression levels among treatments within and between populations are indicated with asterisks: $^{*}P < 0.05$; $^{**}P < 0.01$; and $^{***}P < 0.001$.

This suggests that—of the six candidate genes—only *brinker* is induced by cold stress and may contribute to CCRT variation between temperate and tropical populations.

We also investigated *brinker* induction on the A* and E* lines individually. Again, we found that *brinker* is significantly induced 15 min after recovery from chill coma only in the E* line ($P < 0.01$). Furthermore, we noted that the constitutive expression differences between A* and E* disappeared (in contrast to the experiment with the African and the Netherlands pools described above), presumably because the A* and E* lines share the same autosomal background (data not shown).

Finally, we performed quantitative complementation tests on two of the four genes under the sweeps (*CG1677* and *unc-119*) for which lines with *P*-element insertions were available. None of these tests (performed in the same way as with the deletions) revealed quantitative failure to complement (Table 3). This further supports our conclusion that it is unlikely that the genes under the sweep (at least *CG1677* and *unc-119*) affect cold tolerance.

DISCUSSION

Overview

First, we dissected a QTL interval for CCRT (a proxy for cold tolerance) using quantitative complementation tests. This approach revealed two deletions that failed to complement. Second, we used population genetic methods to narrow the number of genes in these two deletions. This approach led to the precise demarcation of a strong selective sweep in deletion *Df(1)ED6906*. Third, we investigated the genes within and near the sweep region by gene expression analysis. We found no evidence that the four genes within the sweep region (*CG1958*, *CG1677*, *CG2059*, and *unc-119*) are related to cold tolerance. However, this analysis also revealed a new candidate gene related to CCRT: *brinker*, a gene located just outside the sweep region that was induced by cold stress. In the following, we discuss these results, including the methods used.

Quantitative complementation tests on deficiencies and gene expression assays

Using a set of deficiencies in the framework of the quantitative complementation test allowed us to narrow the QTL interval to two highly significant deficiencies, one of which contains a selective sweep. Furthermore, the list of genes under the QTL (encompassing the sweep) could be reduced to seven candidate genes. However, there is a caveat, because for both deletions significant $L \times G$ interactions were found in the presence of significant L effects. This is not uncommon in *D. melanogaster* (Fanara *et al.* 2002; De Luca *et al.* 2003; Geiger-Thornsberry and Mackay 2004; Harbison *et al.* 2004), but this means that it is difficult to distinguish between allelic failure to complement at the deficiency and an epistatic interaction between the deficiency and variation elsewhere. However, for the context of this article, this is not important because we have not relied exclusively on quantitative complementation tests to show relatedness to CCRT. In the case of *brinker*, our evidence of an association with CCRT is confirmed by gene expression analysis.

The genes under the selective sweep and the putative targets of positive selection

Although we have not estimated the selection coefficient, the large value of the CLR statistic indicates that the selective sweep at deletion *Df(1)ED6906* in the European population is very strong, the strongest on the entire X-chromosome analyzed (see Figure 2 and Figure S3). It encompasses approximately 40 kb (with boundaries that are sharply defined). Based on the demarcation of the sweep, we observe that *brinker*, the only candidate gene that was induced by cold stress, is located outside the sweep. The four genes under the sweep are not induced in both African and European populations. *CG1958* is differentially expressed at the constitutive level (Figure 4), and it has been reported that *CG1677* expression is increased relative to constitutive levels during sustained cold stress (Graveley *et al.* 2011). Therefore, the expression of these genes may be temperature-dependent. However, it seems unlikely that regulatory elements determining their expression are the target of selection related to cold tolerance. Furthermore, we have performed quantitative complementation tests on two of the four genes under the sweep (*CG1677* and *unc-119*) for which lines with *P*-element insertions were available. None of these tests (performed in the same way as with the deletions) revealed quantitative failure to complement (Table 3). This supports our conclusion that it is unlikely that the genes under the sweep affect cold tolerance.

This leaves us to search for fixed differences in coding sequences on which selection for cold tolerance may have operated. To identify

Table 3 P-element analysis of X-linked candidate genes affecting CCRT in female flies

P-element	Balancer	Mean CCRT (SD), min				Δmut	Δbal	P L	P G	P L × G	Stock No.
		E/Mutation*	*E*/Balancer*	*A*/Mutation*	*A*/Balancer*						
P{EPgy2}[52]CG1677EY06475	*FM7a*	29.28 (8.92)	31.32 (9.48)	29.96 (8.73)	32.44 (10.16)	−0.67	−1.12	0.38070	0.02130	0.83404	17545
P{EPgy2}unc-119EY20221	*FM7a**	27.99 (7.83)	31.32 (9.48)	31.78 (9.28)	32.44 (10.16)	−3.79	−1.12	0.01166	0.02542	0.15217	22375

Summary of quantitative deficiency tests performed with the listed P-element insertions. Δmut is the difference between the average CCRT of flies bearing E* and A* chromosomes in the presence of a given P-element. Negative differences suggest the presence of CCRT reducing alleles at the gene affected by the tested P-element. The other symbols are defined in Table 1.

strong selective fixations (leading to sweeps) in coding regions, we need to analyze the sweep profile in more detail. Yet, because variation is almost completely depleted in this genomic region, we cannot use the CLR approach even if we include LD (Pavlidis *et al.* 2010). Instead, we used an F_{ST}-based method (Foll and Gaggiotti 2008) to identify the target(s) of positive selection. The results are shown in Figure 2B.

We found four significantly differentiated polymorphisms under the selective sweep. The two SNPs that code for amino acid differences in the gene *CG1677* are most interesting. In the Southeast African sample, both combinations, Thr-Asn and Ala-Asp, are present at positions 936 and 939, where the former is more common. No other combinations exist. In Europe, however, Ala-Asp is fixed (Figure 3B). Subjecting the primary protein sequence encoded by this gene to a structure prediction program (Kelley and Sternberg 2009) reveals that both amino acid positions are part of the α-helix, *i.e.*, they are located on neighboring helix turns and can therefore interact. Interestingly, Thr and Asn can form one hydrogen bond between their side-chains more than Ala-Asp. The combination Thr-Asn may therefore make the protein more heat-stable than Ala-Asp (Perl and Schmid 2002), which appears to be an advantage in tropical Africa, given that ambient temperature is an important variant affecting life history traits in fruit flies. Conversely, the combination Ala-Asp may lead to a less rigid structure and thus possibly a more efficient protein, which may be an advantage in the temperate climate of Europe. Ancestral state reconstruction (Lewis 2001) shows that the Thr-Asn combination represents the ancestral state with high probability and that Ala-Asp arose through two point mutations. Because the intermediate states are not observed in the European and African population samples, the transition from Thr-Asn to Ala-Asp probably follows a compensatory evolution model (Kimura 1985; Innan and Stephan 2001) in which the intermediates are assumed to be strongly deleterious.

Do these adaptive fixations have anything to do with cold adaptation? The protein encoded by *CG1677* is part of the spliceosome (Herold *et al.* 2009) whose function may depend on temperature. However, there is no evidence known to us that splicing has a specific function in the protection of flies against cold. The other two significantly differentiated SNPs occur in noncoding regions between genes *CG1958* and *CG1677* and within the huge intron of *CG1677* (see gene model below) (Figure 2B). There is no evidence that they are involved in the regulation of cold tolerance. This leads us to conclude that strong positive selection causing the observed sweep has probably operated on traits (or molecular variants) other than cold tolerance.

Comparison with population genetics theory

One of our salient observations is that the genes within the selective sweep region do not affect CCRT, whereas *brinker* located just outside the sweep is related to this trait. A similar observation was made previously for another QTL of cold tolerance in *D. melanogaster* (Svetec *et al.* 2011). Recent theoretical work has addressed the question of whether we should expect to find selective sweeps at genes controlling a quantitative trait. Chevin and Hospital (2008) presented a model for the footprint of selection at an adaptive QTL in the presence of background variation due to other loci. This analysis is based on the Lande (1983) model that consists of a locus of major effect on the trait and treats the remaining loci of minor effects as genetic background (such that background variation is maintained at a constant amount). This model predicts that adaptive traits that are under stabilizing selection and show the molecular signature of a selective sweep are only a very small subclass of quantitative traits. Pavlidis *et al.* (2012) analyzed a model with *n* loci controlling a trait

under stabilizing selection. In their model, sweeps are more common than in the scenario presented by Chevin and Hospital (2008). They find that a multi-locus response to selection may in some cases prevent selective sweeps from being completed, but that conditions causing this to happen strongly depend on the genetic architecture of the trait. For instance, the probability of fixation of selected mutations decreases with the number n of loci involved and also depends on their effect sizes. Fixations are more common when the effects are approximately equal (in absolute size). This raises the question of to what extent CCRT is under stabilizing selection and to what extent CCRT is under directional selection. Although there is evidence that cold tolerance may have experienced positive directional selection from one optimum in Africa to another optimum in Europe, it is currently unclear whether this optimum shift is sufficiently large to overcome stabilizing selection that is expected to be widespread (*e.g.*, in the form of apparent stabilizing selection due to pleiotropic deleterious effects of mutations).

brinker—a new candidate gene of CCRT

Based on our gene expression study (Figure 4), *brinker* is a candidate gene affecting variation in cold tolerance. However, it is important to note that *brinker* is located outside the large selective sweep described above and thus is not affected by the strong selection generating this sweep. This is consistent with current theory that sweeps at genes controlling phenotypic traits under stabilizing selection are expected to be rare (see above).

Theoretical models of weak selection (particularly for highly polygenic traits) predict the occurrence of allele frequency shifts between populations as a hallmark of polygenic selection (Hancock *et al.* 2010). For this reason, we searched the region upstream of *brinker* and found one conspicuous indel polymorphism (Figure S5A) at relative positions 109,442 to 109,976, *i.e.*, approximately 3 kb upstream of *brinker* and thus also outside the sweep region. This indel is located close to a polymorphic marker (Figure S5A) that is significantly associated with CCRT in a Raleigh population (Mackay *et al.* 2012). Using an extended sample of populations from the DGPG project (Pool *et al.* 2012), we investigated the frequencies of this indel polymorphism in these populations. We classified the indel polymorphism into nondeletion haplotypes and three classes of deletions (see Figure S5B). Based on linear regression analysis of the frequencies of the nondeletion haplotypes, we detected two antiparallel latitudinal clines where one spans from the populations near the equator (Rwanda, Gabon, Cameroon, Ethiopia, and Nigeria) to the north (France and the Netherlands) and another one from the equator to the south (Southeast Africa and South Africa) ($P < 0.05$ in both cases). This agrees with models of weak selection on highly polygenic traits. However, to what extent the observed frequency differences from the equator to the north and to the south explain the expression differences of *brinker* between tropical and temperate populations (see Figure 4) is currently an open question and beyond the scope of this article.

ACKNOWLEDGMENTS

We acknowledge A. Steincke, H. Lainer, and S. Lange for providing technical assistance, and F. Fadel and M. Arias for fly work and CCRT scoring. Furthermore, we thank the editors Steve Chenoweth and Mark Johnston and two anonymous reviewers for their comments that improved the presentation of this paper. This work was supported by grant I/848313 from the Volkswagen Foundation (to R.W. and W.S.), and by the DFG Research Unit grant STE 325/12-2 (to W.S.).

LITERATURE CITED

Axelsson, E., A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster *et al.*, 2013 The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495: 360–364.

Barton, N. H., and P. D. Keightley, 2002 Understanding quantitative genetic variation. Nat. Rev. Genet. 3: 11–21.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc., B 57: 289–300.

Boitard, S., C. Schlötterer, V. Nolte, R. V. Pandey, and A. Futschik, 2012 Detecting selective sweeps from pooled next-generation sequencing samples. Mol. Biol. Evol. 29: 2177–2186.

Cano, J. M., C. Matsuba, H. Mäkinen, and J. Merilä, 2006 The utility of QTL-linked markers to detect selective sweeps in natural populations—a case study of the EDA gene and a linked marker in threespine stickleback. Mol. Ecol. 15: 4613–4621.

Chevin, L.-M., and F. Hospital, 2008 Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180: 1645–1660.

De Luca, M., N. V. Roshina, G. L. Geiger-Thornsberry, R. F. Lyman, E. G. Pasyukova *et al.*, 2003 Dopa decarboxylase (*Ddc*) affects variation in *Drosophila* longevity. Nat. Genet. 34: 429–433.

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.

Excoffier, L., and M. Foll, 2011 fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics 27: 1332–1334.

Fanara, J. J., K. O. Robinson, S. M. Rollmann, R. R. H. Anholt, and T. F. C. Mackay, 2002 *Vanaso* is a candidate quantitative trait gene for *Drosophila* olfactory behavior. Genetics 162: 1321–1328.

Foll, M., and O. Gaggiotti, 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Geiger-Thornsberry, G. L., and T. F. C. Mackay, 2004 Quantitative trait loci affecting natural variation in *Drosophila* longevity. Mech. Ageing Dev. 125: 179–189.

Glinka, S., L. Ometto, S. Mousset, W. Stephan, and D. De Lorenzo, 2003 Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics 165: 1269–1278.

Glinka, S., D. De Lorenzo, and W. Stephan, 2006 Evidence of gene conversion associated with a selective sweep in *Drosophila melanogaster*. Mol. Biol. Evol. 23: 1869–1878.

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

Hancock, A. M., G. Alkorta-Aranburu, D. B. Witonsky, and A. Di Rienzo, 2010 Adaptations to new environments in humans: the role of subtle allele frequency shifts. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 2459–2468.

Harbison, S. T., A. H. Yamamoto, J. J. Fanara, K. K. Norga, and T. F. C. Mackay, 2004 Quantitative trait loci affecting starvation resistance in *Drosophila melanogaster*. Genetics 166: 1807–1823.

Hellemans, J., G. Mortier, A. De Paepe, F. Speleman, and J. Vandesompele, 2007 qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 8: R19.

Herold, N., C. L. Will, E. Wolf, B. Kastner, H. Urlaub *et al.*, 2009 Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. Mol. Cell. Biol. 29: 281–301.

Hoffmann, A. A., A. Anderson, and R. Hallas, 2002 Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. Ecol. Lett. 5: 614–618.

Hu, T. T., M. B. Eisen, K. R. Thornton, and P. Andolfatto, 2013 A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. Genome Res. 23: 89–98.

Hutter, S., S. Saminadin-Peter, W. Stephan, and J. Parsch, 2008 Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biol. 9: R12.

Innan, H., and W. Stephan, 2001 Selection intensity against deleterious mutations in RNA secondary structures and rate of compensatory nucleotide substitutions. Genetics 159: 389–399.

Kane, N. C., and L. H. Rieseberg, 2007 Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. Genetics 175: 1823–1834.

Kelley, L. A., and M. J. E. Sternberg, 2009 Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4: 363–371.

Kim, Y., and W. Stephan, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160: 765–777.

Kimura, M., 1985 The role of compensatory neutral mutations in molecular evolution. J. Genet. 64: 7–19.

Lande, R., 1983 The response to selection on major and minor mutations affecting a metrical trait. Heredity 50: 47–65.

Langley, C. H., K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider *et al.*, 2012 Genomic variation in natural populations of *Drosophila melanogaster*. Genetics 192: 533–598.

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.

Lewis, P. O., 2001 A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50: 913–925.

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

Linnen, C. R., Y.-P. Poh, B. K. Peterson, R. D. H. Barrett, J. G. Larson *et al.*, 2013 Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339: 1312–1316.

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.

Morgan, T. J., and T. F. C. Mackay, 2006 Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. Heredity 96: 232–242.

Nei, M., 1987 *Molecular Evolutionary Genetics*, Columbia University Press, New York.

Nei, M., and W. H. Li, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76: 5269–5273.

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for selective sweeps using SNP data. Genome Res. 15: 1566–1575.

Norry, F. M., A. C. Scannapieco, P. Sambucetti, C. I. Bertoli, and V. Loeschcke, 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*. Mol. Ecol. 17: 4570–4581.

Pasyukova, E. G., C. Vieira, and T. F. C. Mackay, 2000 Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. Genetics 156: 1129–1146.

Pavlidis, P., S. Hutter, and W. Stephan, 2008 A population genomic approach to map recent positive selection in model species. Mol. Ecol. 17: 3585–3598.

Pavlidis, P., J. D. Jensen, and W. Stephan, 2010 Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. Genetics 185: 907–922.

Pavlidis, P., D. Metzler, and W. Stephan, 2012 Selective sweeps in multilocus models of quantitative traits. Genetics 192: 225–239.

Pavlidis, P., D. Živković, A. Stamatakis, and N. Alachiotis, 2013 SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. 30: 2224–2234.

Perl, D., and F. X. Schmid, 2002 Some like it hot: the molecular determinants of protein thermostability. European Journal of Chemical Biology 3: 39–44.

Pool, J. E., and C. F. Aquadro, 2007 The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. Mol. Ecol. 16: 2844–2851.

Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno *et al.*, 2012 Population genomics of Sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet. 8: e1003080.

Pritchard, J. K., and A. Di Rienzo, 2010 Adaptation—not by sweeps alone. Nat. Rev. Genet. 11: 665–667.

Qanbari, S., H. Pausch, S. Jansen, M. Somel, T. M. Strom *et al.*, 2014 Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet. 10: e1004148.

Rebeiz, M., J. E. Pool, V. A. Kassner, C. F. Aquadro, and S. B. Carroll, 2009 Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. Science 326: 1663–1667.

Riebler, A., L. Held, and W. Stephan, 2008 Bayesian variable selection for detecting adaptive genomic differences among populations. Genetics 178: 1817–1829.

Rubin, C.-J., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood *et al.*, 2010 Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464: 587–591.

Rubin, C.-J., H.-J. Megens, A. Martinez Barrio, K. Maqbool, S. Sayyab *et al.*, 2012 Strong signatures of selection in the domestic pig genome. Proc. Natl. Acad. Sci. USA 109: 19529–19536.

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.

Service, P. M., 2004 How good are quantitative complementation tests? Sci. SAGE KE 2004: pe13.

Stephan, W., 2010 Genetic hitchhiking *vs.* background selection: the controversy and its implications. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365: 1245–1253.

Stephan, W., and H. Li, 2007 The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity 98: 65–68.

Svetec, N., A. Werzner, R. Wilches, P. Pavlidis, J. M. Álvarez-Castro *et al.*, 2011 Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis. Mol. Ecol. 20: 530–544.

Tajima, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics 105: 437–460.

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

Thornton, K. R., J. D. Jensen, C. Becquet, and P. Andolfatto, 2007 Progress and prospects in mapping recent selection in the genome. Heredity 98: 340–348.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7: 256–276.

Communicating editor: B. J. Andrews

G3 · Genes | Genomes | Genetics

Fine-mapping and selective sweep analysis of QTL for cold tolerance in *Drosophila melanogaster*

Ricardo Wilches*[1], Susanne Voigt*, Pablo Duchen*, Stefan Laurent* and Wolfgang Stephan*

*Section of Evolutionary Biology. Department of Biology II, Ludwig-Maximilian University of Munich. 82152 Planegg-Martinsried, Germany.

[1]Corresponding author: wilches@bio.lmu.de

DOI: 10.1534/g3.114.012757

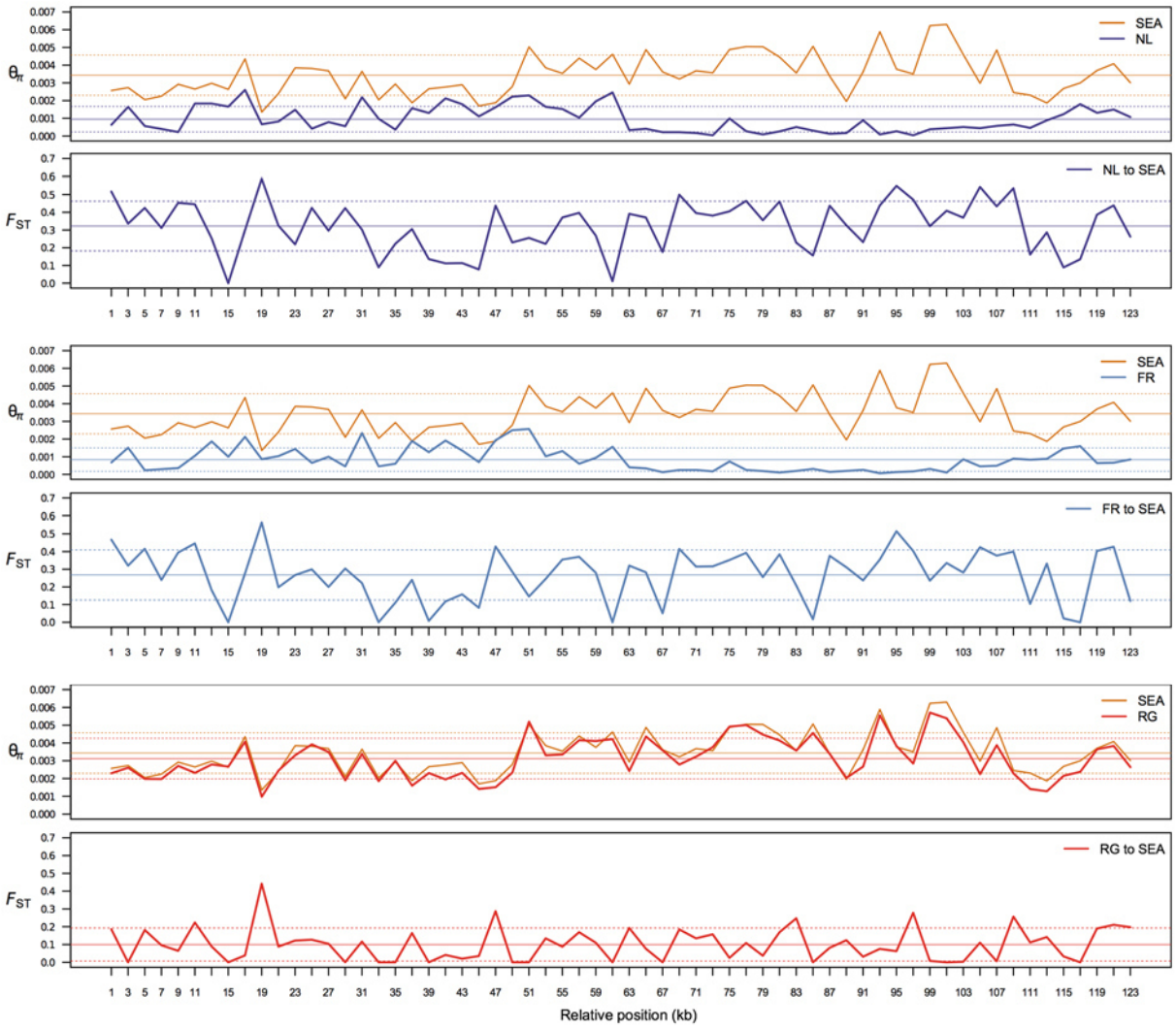


Figure S1 Polymorphism and between-population differentiation along the 124 kb of interest. Nucleotide diversity (θ_π) obtained for consecutive 2-kb long windows in four different populations: the Netherlands (NL), France (FR), Rwanda (RG) and a pool of Southeast African (SEA) lines sampled around Lake Kariba in Zimbawe and Zambia. This pool also includes lines from Malawi. The SEA profile is shown in all three θ_π panels for sake of comparison. Below each θ_π panel, inter-population differentiation profiles are plotted. Differentiation (F_{ST}) was calculated as normalized distance of Nei. Thin continous lines represent the average value for each summary statistic across the 62 windows, dashed lines represent 1 SD above and below the corresponding summary statistic mean.

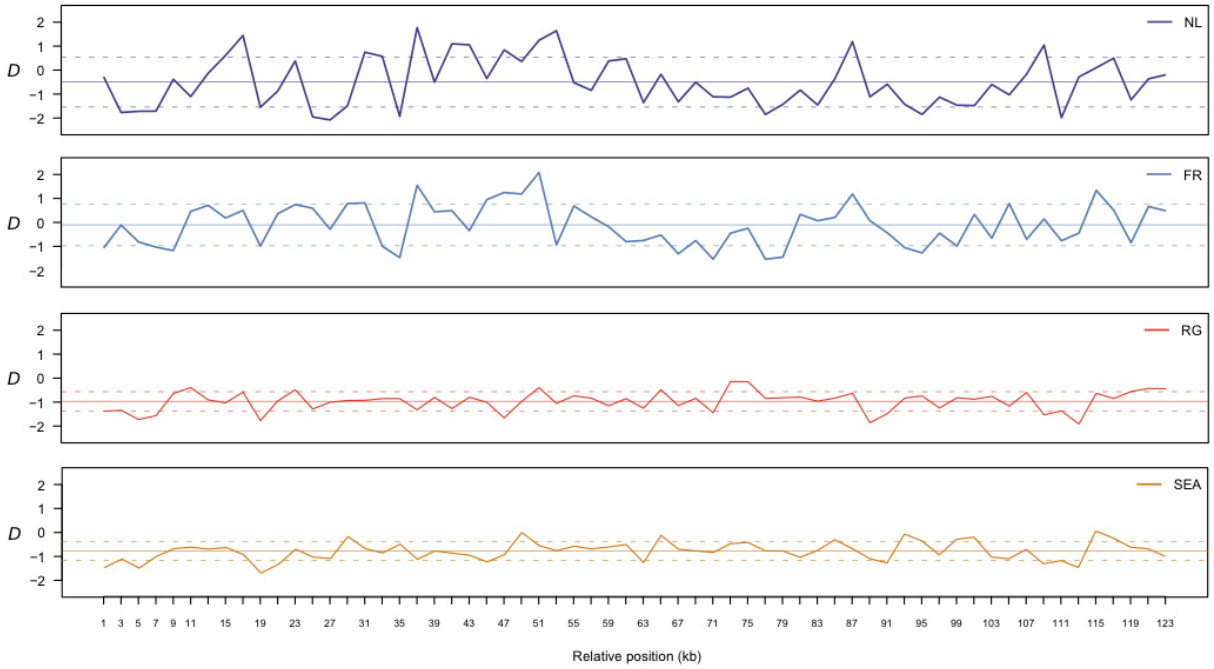


Figure S2 Tajima's *D* statistics. Tajima's *D* profiles along the 124 kb of interest are shown for the following populations: the Netherlands (NL), France (FR), Rwanda (RG), and Southeast Africa (SEA). The solid thin lines represent the corresponding mean value across the entire region, while dashed lines mark 1 SD above and below the corresponding mean.

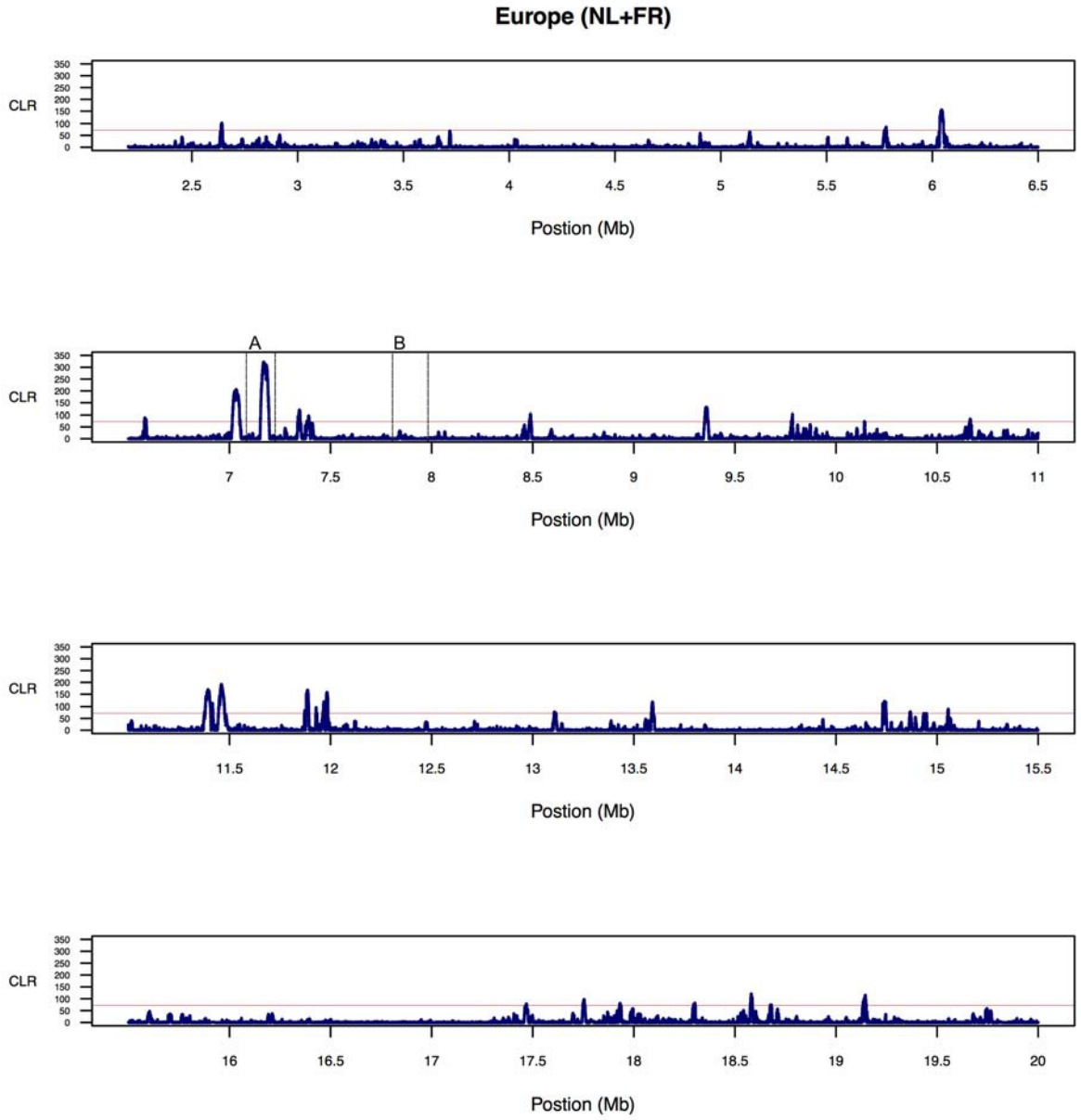


Figure S3 X-chromosome CLR profile for Europe. Composite likelihood ratio (CLR) test results for 18 Mb of a sample of 19 European (the Netherlands and French) *D. melanogaster* X-chromosomes. For this chromosome-wide test all categories (0 to *n*) of the SFS were included. The significance threshold at CLR=72 was obtained from simulated subgenomic datasets (see text and Figure S4). (A) indicates the interval with a CLR peak above 300 corresponding to that under deletion *Df(1)ED6906* (124 Kb long), also depicted in Figure 2A. (B) shows the Interval corresponding to that revealed by deletion *Df(1)C128* (131 Kb long). Note that this interval does not exhibit CLR peaks above the significance threshold.

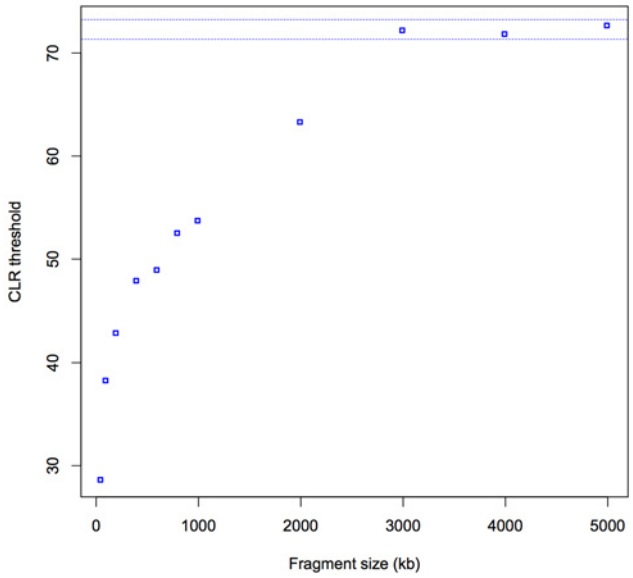


Figure S4 CLR thresholds *vs.* simulated fragment size. CLR thresholds, *i.e.* the top 5% CLR values of 100 simulated fragments of lengths from 5 to 5000 kb reach an asymptotic value around 72 at fragment size ≥3000 kb.

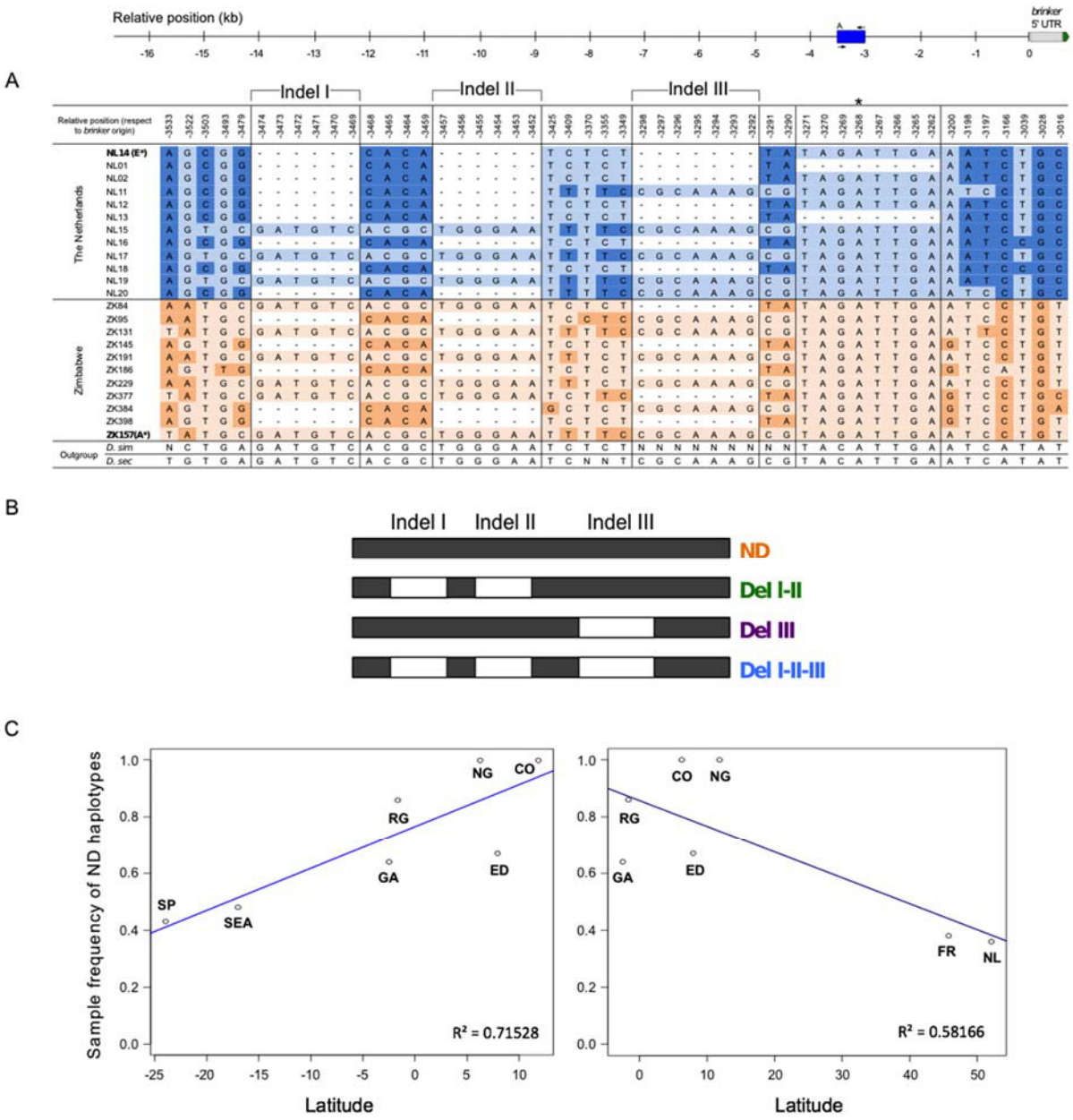


Figure S5 Putative *cis*-regulatory element upstream of *brinker*. A) Polymorphism table of a 534-bp fragment between relative positions -3,000 to -3,553 upstream of *brinker*. The figure depicts SNPs and structural variants (indels) of two *D. melanogaster* population samples from the Netherlands (NL) and Zimbabwe (ZK), including E* (top line) and A* (bottom line) plus two outgroups (*D. simulans* and *D. sechellia*). Light blue and orange indicate the inferred ancestral state of the SNP considering the two outgroups in NL and ZK, respectively, whereas darker tones of the same color represent the derived allele. Deletions are indicated in white background. Relative position -3,268 marked with an asterisk is highly associated with CCRT in the Raleigh population. B) Four haplotypes defined by the presence/absence of deletions and their numbers in the fragment. C) Frequency clines of the non-deletion haplotypes along a latitudinal gradient of *D. melanogaster* populations: the Netherlands (NL), France (FR), Nigeria (NG), Ethiopia (ED), Cameroon (CO), Gabon (GA), Rwanda (RG), Zambia + Zimbabwe + Malawi (SEA), and South Africa (SP).

GENERAL DISCUSSION

Temperature is a crucial factor in determining the distribution and abundance of species. Ectotherms, like most insects, have evolved various physiological and molecular adaptations to survive and prosper in a wide range of thermal environments (Clarke 1996; Doucet *et al.* 2009). *D. melanogaster* as a cosmopolitan species has successfully adapted to diverse thermal environments, and is therefore a useful model system to investigate the genetic basis of thermal adaptation at the intraspecific level. This thesis has tried to shed further light onto genes and genetic changes involved in the adaptation of *D. melanogaster* to temperate environments. Different natural populations from temperate and tropical origins were considered ranging from ancestral populations from tropical southern-central Africa to derived populations from temperate high-altitude Africa, temperate Europe and tropical Asia. Including ancestral and derived populations, in particular derived populations adapted to different temperature regimen, provides the opportunity to compare populations that have undergone adaptations to temperate climates to presumably less adapted populations from tropical regions. In this thesis, two different aspects of temperate climates and its accompanying low temperatures have been considered with which *D. melanogaster* has been confronted while colonizing temperate environments. One is the greater range and variability of ambient temperature and the other are low-temperature extremes such as freezing temperatures. The former is known to disrupt chromatin-based gene regulation (Fauvarque and Dura 1993; Gibert *et al.* 2011) and the latter results in cold stress.

Chromatin-based PcG regulation is known to be thermosensitive. Decreasing the temperature at which flies are reared or held leads to an increased transcriptional output of genes regulated by PcG proteins. This phenomenon was demonstrated in several studies using reporter gene assays with constructs containing specific PREs

(Fauvarque and Dura 1993; Chan *et al.* 1994; Zink and Paro 1995; Bantignies *et al.* 2003; Gibert *et al.* 2011). In chapter 1 and 2, thermosensitivity of PcG-regulated genes was observed in wild-type flies of different natural populations. At both loci, *polyhomeotic* (*ph*) and *vestigial* (*vg*), a lower degree of temperature-induced expression plasticity was consistently detected in populations adapted to temperate environments compared to populations from tropical environments. This supports the idea that temperature-induced expression plasticity of PcG-regulated genes has been detrimental in temperate climates, and that adaptation to such climates required the buffering of this thermosensitive regulatory process (Begun and Levine 2008).

Patterns of DNA sequence polymorphism at the *ph* locus suggest that it has been the target of recent positive selection within European and African populations from *D. melanogaster* (Beisswanger *et al.* 2006; Beisswanger and Stephan 2008). Chapter 1 of this thesis provides evidence that the selective sweep observed in Europe is independent of the ancestral African sweep, and resulted from a selective event that led to reduced thermosensitivity of PcG regulation. Population genetic analysis pinpointed the likely targets of selection in Europe to derived SNPs in the intergenic region of the two PcG target genes *ph-p* and *CG3835*, and subsequent reporter gene assays linked the derived sequence variants to stable levels of gene expression across temperatures. Ancestral sequence variants, in contrast, led to thermosensitive expression with a higher expression at lower temperatures which is often observed for PcG regulation (Fauvarque and Dura 1993; Chan *et al.* 1994; Zink and Paro 1995; Bantignies *et al.* 2003; Gibert *et al.* 2011). Derived variants that conferred stable expression across temperatures were found nearly fixed in temperate populations from Europe, but not in tropical populations from Africa and Asia. These results are consistent with thermosensitivity of PcG regulation being detrimental in temperate environments, and selection acting to buffer this thermosensitivity in populations adapting to such environments.

Thermosensitivity of PcG regulation seems to be restricted to certain tissues. For the *ph* locus, temperature had no effect on gene expression in larval brains, whereas thermosensitive expression was observed in larval midguts and whole adults. Although thermosensitive *vg* expression could be detected in adult flies, lowering the rearing temperature did not result in higher *vg* expression in larval brains and larval wing discs. Possible explanations for this restriction could be that the selective pressure against thermosensitive expression is higher in some tissues than in others, and thus expression is already less environmentally sensitive in ancestral populations, or there is no thermosensitivity in expression in some tissues. Selection pressure could indeed be higher in developing tissues like the larval brain and wing discs than in adult tissues and those not important in development like the larval midgut. Yet to draw this conclusion, expression should be monitored in additional tissues, in particular adult tissues should be considered separately.

Moreover, it appears that higher transcriptional output of PcG-regulated genes at lower temperatures is not in itself detrimental, but rather gene expression plasticity due to variations in temperature. Reduced expression plasticity in temperate populations either resulted from decreased expression levels of PcG-regulated genes at lower rearing temperatures or increased expression levels at higher temperatures when compared to expression levels in ancestral tropical populations. Expression levels of *vg* were decreased at lower rearing temperature in three temperate population samples (France, The Netherlands, and Rwanda) compared to ancestral samples leading to a buffering of thermosensitive expression in those populations. In contrast, the same effect resulted from increased expression levels at higher temperature in Swedish *vg* expression and reporter gene expression from the construct carrying the European sequence variants and the *ph-p* promoter. The higher degree in expression plasticity of PcG target genes therefore seems to have been disadvantageous and needed to be reduced during the colonization of temperate environments.

The difference of how the lower degree of *vg* expression plasticity comes about indicates that underlying regulatory changes also differ between the Swedish population and the other temperate populations. The reduction in *vg* expression plasticity is similar in all temperate population samples, which could be indicative of convergent evolution. However, convergent evolution leading to the same degree of buffering becomes less likely when one takes into account that Sweden was probably colonized by European flies already exhibiting reduced *vg* expression plasticity like the French and Dutch population samples. Other ecological constraints in the cold-temperate climate of Sweden compared to the warm-temperate climate in France, The Netherlands and Rwanda could be a possible explanation for the observed differences. A higher overall expression level of *vg* might have been additionally beneficial in the colder Swedish climate. In contrast, by decreasing *vg* expression with decreasing temperature, the French, Dutch and Rwandan population samples were similar in buffering temperature-induced expression plasticity. The simplest scenario would be a common genetic basis for reduced thermosensitivity in all three populations. If there is a common genetic basis, it seems less likely that it is largely caused by *cis*-regulatory changes in regulatory elements at the *vg* locus, since no highly differentiated SNPs between the derived and the ancestral populations were detected in the gene region that are shared among all three populations. Given the large geographic distance between the Rwandan and the two European populations, convergent evolution might also be a likely explanation for the similar expression response to temperature. However, further studies are required to clarify whether a similar response is due to a common genetic basis or convergent evolution.

While there are candidate SNPs for *cis*-regulatory changes at the *vg* locus in the Dutch and French populations that are shared between both populations and are highly differentiated between the two derived and the ancestral populations, none could be identified for the Rwandan population. The absence of such candidate SNPs might be indicative of *trans*-regulatory changes, which alter the abundance or activity of factors regulating the transcriptional output of *vg* (e.g., PcG proteins),

playing an important role in the response of *vg* expression to temperature in the Rwandan population. Support for this might come from another recent study in which a genome-wide scan was conducted to identify genes under positive selection in the Rwandan population sample (Pool *et al.* 2012). Genes identified in this scan included two PcG/TrxG genes, *pleiohomeotic like* (*phol*) and *female sterile (1) homeotic* (*fs(1)h*), which are in turn putative candidates for *trans*-regulatory factors responsible for the Rwandan decreased thermosensitivity. A change in a *trans*-regulatory factor may simultaneously affect many of its target genes. Thus such a *trans*-regulatory change in the Rwandan population could also contribute to reduced thermosensitivity of the expression of other PcG target genes. At the *ph* locus, thermosensitivity was shown to be decreased by *cis*-regulatory derived sequence variants that are nearly fixed in Europe but rare in sub-Saharan Africa. The derived variants were not observed in the Rwandan population sample, although based on the observations at the *vg* locus one would expect decreased thermosensitivity of PcG regulation also at the *ph* locus in the temperate population from high-altitude Africa. A *trans*-regulatory change that confers decreased thermosensitivity to multiple PcG target genes could be a possible explanation. However, other *cis*-regulatory changes private to the Rwandan population could also be responsible for a decreased thermosensitivity at the *ph* locus in Rwanda, if there is one.

The transcriptional output of PcG target genes is controlled and modulated by a complex interplay of PREs and other *cis*-regulatory DNA sequences (*i.e.*, enhancers) in cell- or tissue-specific manner. Enhancers initially determine the level of transcription which is then epigenetically maintained by PREs (Schwartz *et al.* 2010; Kassis and Brown 2013; Steffen and Ringrose 2014). As a consequence, buffering of thermosensitive expression could result from changes that alter the activity or strength of either of the two types of regulatory elements. This could be mediated, for example, either by an increased activity or abundance of *trans*-regulatory factors binding to the regulatory elements, or directly in *cis* through changes in the DNA sequence of the elements as it was observed at the *ph* locus. It is known that small

changes in the sequence of enhancers or PREs can have large effects on the expressed phenotype, and that both *cis*-regulatory elements evolve rapidly (Hauenschild *et al.* 2008). Changes in only a few base pairs of PREs, for instance, were shown to have profound effects on their function (Okulski *et al.* 2011).

The question arises whether decreased thermosensitivity of PcG regulation in temperate populations is a global phenomenon detected for many of the several hundred PcG target genes in *Drosophila*, or restricted to a few loci like the two studied in this thesis. The former is supported by a genome-wide study that found greater temperature-induced expression plasticity in tropical than in temperate populations (Levine *et al.* 2011). Furthermore, several studies have provided evidence for positive selection acting on proteins involved in PcG regulation (Harr *et al.* 2002; Begun and Levine 2008; Gibert *et al.* 2011). *Trans*-regulatory changes might also play a dominant role in the reduced thermosensitivity of *vg* expression in the Rwandan population. In contrast to changes in *cis*-regulatory elements controlling expression of specific PcG target genes, changes in *trans* might have an effect on multiple PcG target genes, and could therefore confer a more global buffering effect. Nonetheless, changes in *cis*-regulatory elements of PcG target genes are involved in buffering thermosensitivity of PcG regulation as it was shown for the *ph* locus. Therefore, the relative amount to which either of both, *cis*- and *trans*-regulatory changes contribute to decreased thermosensitivity of PcG regulation in temperate populations is another question interesting to investigate further.

Although the exact mechanisms underlying reduced thermosensitivity of expression of PcG-regulated genes remain unclear, in this thesis further evidence was provided that thermosensitivity of PcG regulation needs to be buffered when *Drosophila* adapts to temperate climates and that this can be facilitated through the action of positive selection.

Increasing cold tolerance is another aspect that probably played a dominant role in adaptation to temperate environments in order to better deal with stress imposed

on *D.melanogaster* by low-temperature extremes. Temperate and tropical populations are known to differ in their cold tolerance and increased tolerance to cold stress is thought to have evolved by natural selection (Hoffmann *et al.* 2002; Schmidt *et al.* 2005). When exposed to extremely low temperatures like those around freezing, *D. melanogaster* as other insects enters a cold-induced coma (David *et al.* 1998; MacMillian and Sinclair 2011). The time required to recover from this chill coma varies between temperate and tropical populations with temperate flies recovering more quickly (Gibert *et al.* 2001; Hoffmann *et al.* 2002; Norry *et al.* 2008; Svetec *et al.* 2011). Chill coma recovery time is often employed as a proxy for cold tolerance to investigate the molecular basis underlying variation in cold tolerance (Hoffmann *et al.* 2002; Norry *et al.* 2008; Svetec *et al.* 2011; Mackay *et al.* 2012). Recovery from chill coma is a quantitative trait that appears to be highly polygenic, however, knowledge about the underlying genes is still limited with only a few known candidate genes (Morgan and Mackay 2006; Colinet et al. 2010; Colinet and Hoffmann 2010; Mackay *et al.* 2012). QTL analyses have been used to uncover genomic regions that harbor such candidate genes (Morgan and Mackay 2006; Norry *et al.* 2008; Svetec *et al.* 2011; Mackay *et al.* 2012). Svetec *et al.* (2011), for instance, localized several QTL regions on the X chromosome that contribute to faster recovery from chill coma in flies from temperate Europe. As it is often the case, the identified QTL regions contained numerous genes, and further effort was necessary to localize candidate genes responsible for observed differences in chill coma recovery between temperate European and tropical African populations. The study presented in chapter 3 dissected a 6.2-Mb QTL interval identified by Svetec *et al.* (2011) trying to single out such candidate genes by combining several fine-mapping approaches. A fine-mapping approach based on quantitative complementation tests yielded two relatively small regions of 131 kb and 124 kb, of which the latter was scrutinized in more depth. Gene expression analysis, population genetic analysis, and additional complementation tests were employed to allow for deeper fine-mapping in the region. The European patterns of DNA sequence polymorphism were

consistent with positive selection and a strong selective sweep which spans about 40 kb of the 124-kb long region. However, none of the four genes within the sweep seemed to be related to cold tolerance as inferred from quantitative complementation tests and gene expression analysis. Expression levels of genes in the 124-kb region were determined at different time points after cold stress exposure. Although the genes within the sweep region were not affected in their expression by cold stress, the gene *brinker* (*brk*) upstream of sweep signal appears to be induced by cold stress. Expression levels of *brk* went up shortly after flies had recovered from chill coma when compared to untreated controls and other time points after cold stress exposure. In two independent experiments, cold stress-induced increase in expression levels of *brk* were detected for temperate genotypes from Europe but not for tropical genotypes from Africa. Therefore, *brk* may contribute to variation in chill coma recovery between temperate and tropical populations, and is a candidate gene involved in the evolution of increased cold tolerance of temperate *D. melanogaster*.

Using only sequence-level information, inference of positive selection might be difficult for highly polygenic traits like cold tolerance. Although a strong selective sweep was detected in the region putatively involved in increased cold tolerance of temperate *D. melanogaster*, genes within the sweep appear not to be related to cold tolerance. Instead a gene upstream of the selective sweep was found to be induced by cold stress and to differ in its expression between temperate and tropical flies. This fits the idea that when positive selection acts on a trait controlled by numerous genes, the effect of any given gene is expected to be rather small, as is the resulting signature of selection (Mackay 2001; Berg and Coop 2014; Wollstein and Stephan 2015).

The gene *brk* encodes for a transcription factor that appears to be important in the triggering of apoptosis (Suissa *et al.* 2011). Apoptosis, or programmed cell death, is associated with cold injury and occurs during cold stress (Yi *et al.* 2007). Programmed

cell death is needed for the removal of damaged cells and often compensated by extra proliferation of neighboring cells, another process whose triggering seems to be dependent on *brk* (Suissa *et al.* 2011). Therefore, *brk* might contribute to increased cold stress tolerance in temperate flies by coping more efficiently with cold injury.

BIBLIOGRAPHY

Bantignies, F., C. Grimaud, S. Lavrov, M. Gabut, and G. Cavalli, 2003 Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. Genes Dev. 17: 2406–2420.

Barrett, R. D. H., and H. E. Hoekstra, 2011 Molecular spandrels: tests of adaptation at the genetic level. Nat. Rev. Genet. 12: 767–780.

Beisel, C., and R. Paro, 2011 Silencing chromatin: comparing modes and mechanisms. Nat. Rev. Genet. 12: 123–135.

Beisswanger, S., and W. Stephan, 2008 Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated *polyhomeotic* genes in *Drosophila*. Proc. Natl. Acad. Sci. U S A 105: 5447–5452.

Beisswanger, S., W. Stephan, and D. De Lorenzo, 2006 Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. Genetics 172: 265–274.

Berg, J. J., and G. Coop, 2014 A Population genetic signal of polygenic adaptation. PLoS Genet. 10: e1004412.

Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner *et al.*, 2004 Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. 74: 1111–1120.

Bonhomme, M., C. Chevalet, B. Servin, S. Boitard, J. Abdallah *et al.*, 2010 Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics 186: 241–262.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan, 1995 The

hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140: 783–796.

Catalán, A., S. Hutter, and J. Parsch, 2012 Population and sex differences in *Drosophila melanogaster* brain gene expression. BMC Genomics 13: 1–12.

Catania, F., M. O. Kauer, P. J. Daborn, J. L. Yen, R. H. Ffrench-Constant *et al.*, 2004 World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster*. Mol. Ecol. 13: 2491–2504.

Chan, C. S., L. Rastelli, and V. Pirrotta, 1994 A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. EMBO J. 13: 2553–2564.

Chung, H., M. R. Bogwitz, C. McCart, A. Andrianopoulos, R. H. ffrench-Constant *et al.*, 2007 *Cis*-regulatory Elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. Genetics 175: 1071–1077.

Clarke, A., 1996 The influence of climate chance of the distribution and evolution of organisms, pp 377–407 in *Animals and Temperature: Phenotypic and Evolutionary Adaptation*, edited by I.A. Johnston and A.F. Bennett. Cambridge University Press, Cambridge.

Colinet, H., and A. Hoffmann, 2010 Gene and protein expression of *Drosophila* Starvin during cold stress and recovery from chill coma. Insect Biochem. Mol. Biol. 40: 425–428.

Colinet, H., S. F. Lee, and A. Hoffmann, 2010 Temporal expression of heat shock genes during cold stress and recovery from chill coma in adult *Drosophila melanogaster*. FEBS J. 277: 174–185.

Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. L. Goff, E. Feil *et al.*, 2002 A single P450 allele

associated with insecticide resistance in *Drosophila*. Science 297: 2253–2256.

David, J. R., F. Lemeunier, L. Tsacas, and A. Yassin, 2007 The historical discovery of the nine species in the *Drosophila melanogaster* species subgroup. Genetics 177: 1969–1973.

Doucet, D., V. K. Walker, and W. Qin, 2009 The bugs that came in from the cold: molecular adaptations to low temperatures in insects. Cell. Mol. Life Sci. 66: 1404–1418.

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.

Fauvarque, M. O., and J. M. Dura, 1993 *polyhomeotic* regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in Drosophila. Genes Dev. 7: 1508–1520.

Fay, J. C., H. L. McCullough, P. D. Sniegowski, and M. B. Eisen, 2004 Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. Genome Biol. 5: 1–14.

Fay, J. C., and C.-I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics 155: 1405–1413.

Foll, M., and O. Gaggiotti, 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Fu, Y.-X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915–925.

Gibert, J.-M., F. Karch, and C. Schlötterer, 2011 Segregating variation in the Polycomb

group gene *cramped* alters the effect of temperature on multiple traits. PLoS Genet. 7: e1001280.

Gibert, P., B. Moreteau, G. Pétavy, D. Karan, and J. R. David, 2001 Chill-Coma Tolerance, a Major Climatic Adaptation Among *Drosophila* Species. Evolution 55: 1063–1068.

Günther, T., and G. Coop, 2013 Robust identification of local adaptation from allele frequencies. Genetics 195: 205–220.

Harr, B., M. Kauer, and C. Schlötterer, 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 99: 12949–12954.

Hauenschild, A., L. Ringrose, C. Altmutter, R. Paro, and M. Rehmsmeier, 2008 Evolutionary plasticity of polycomb/trithorax response elements in *Drosophila* species. PLoS Biol. 6: e261.

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

Hoffmann, A. A., A. Anderson, and R. Hallas, 2002 Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. Ecol. Lett. 5: 614–618.

Hutter, S., S. S. Saminadin-Peter, W. Stephan, and J. Parsch, 2008 Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biol. 9: R12.

Ingram, C. J. E., C. A. Mulcare, Y. Itan, M. G. Thomas, and D. M. Swallow, 2008 Lactose digestion and the evolutionary genetics of lactase persistence. Hum. Genet. 124: 579–591.

Clarke, A., 1996 The influence of climate chance of the distribution and evolution of

organisms, pp 377–407 in *Animals and Temperature: Phenotypic and Evolutionary Adaptation*, edited by I.A. Johnston and A.F. Bennett. Cambridge University Press., Cambridge.

Kaplan, N. L., R. R. Hudson, and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics 123: 887–899.

Kassis, J. A., and J. L. Brown, 2013 Polycomb group response elements in *Drosophila* and vertebrates. Adv. Genet. 81: 83–118.

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

Kim, Y., and R. Nielsen, 2004 Linkage disequilibrium as a signature of selective sweeps. Genetics 167: 1513–1524.

Kim, J., A. Sebring, J. J. Esch, M. E. Kraus, K. Vorwerk *et al.*, 1996 Integration of positional signals and regulation of wing formation and identity by *Drosophila vestigial* gene. Nature 382: 133–138.

Kim, Y., and W. Stephan, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160: 765–777.

Labbé, P., A. Berthomieu, C. Berticat, H. Alout, M. Raymond *et al.*, 2007 Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito *Culex pipiens*. Mol. Biol. Evol. 24: 1056–1067.

Lachaise, D., M.-L. Cariou, J. R. David, F. Lemeunier, L. Tsacas *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup, pp. 159–225 in *Evolutionary Biology*, edited by M. K. Hecht, B. Wallace, and G. T. Prance. Evolutionary Biology 22, Springer US.

Lachaise, D., and J.-F. Silvain, 2004 How two Afrotropical endemics made two

cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle, pp. 17–39 in *Drosophila melanogaster, Drosophila simulans: So Similar, So Different*, edited by P. Capy, P. Gibert, and I. Boussy. Contemporary issues in genetics and evolution 11, Springer Netherlands.

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.

Lenormand, T., D. Bourguet, T. Guillemaud, and M. Raymond, 1999 Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. Nature 400: 861–864.

Lenormand, T., T. Guillemaud, D. Bourguet, and M. Raymond, 1998 Evaluating gene flow using selected markers: a case study. Genetics 149: 1383–1392.

Levine, M. T., and D. J. Begun, 2008 Evidence of spatially varying selection acting on four chromatin-remodeling loci in *Drosophila melanogaster*. Genetics 179: 475–485.

Levine, M. T., M. L. Eckert, and D. J. Begun, 2011 Whole-genome expression plasticity across tropical and temperate *Drosophila melanogaster* populations from Eastern Australia. Mol. Biol. Evol. 28: 249–256.

Lewontin, R. C., and J. Krakauer, 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74: 175–195.

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

Lindsley, D. L., and G. G. Zimm, 1992 GENES, pp. 1–803 in *The Genome of* Drosophila melanogaster, edited by D. L. L. G. Zimm. Academic Press, San Diego.

Linnen, C. R., E. P. Kingsley, J. D. Jensen, and H. E. Hoekstra, 2009 On the origin and spread of an adaptive allele in deer mice. Science 325: 1095–1098.

Linnen, C. R., Y.-P. Poh, B. K. Peterson, R. D. H. Barrett, J. G. Larson *et al.*, 2013 Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339: 1312–1316.

Mackay, T. F. C., 2001 The Genetic architecture of quantitative traits. Annu. Rev. Genet. 35: 303–339.

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.

Mackay, T. F. C., E. A. Stone, and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10: 565–577.

MacMillan, H. A., and B. J. Sinclair, 2011 Mechanisms underlying insect chill-coma. J. Insect Physiol. 57: 12–20.

Morgan, T. J., and T. F. C. Mackay, 2006 Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. Heredity 96: 232–242.

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005 Genomic scans for selective sweeps using SNP data. Genome Res. 15: 1566–1575.

Norry, F. M., A. C. Scannapieco, P. Sambucetti, C. I. Bertoli, and V. Loeschcke, 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*. Mol. Ecol. 17: 4570–4581.

Okulski, H., B. Druck, S. Bhalerao, and L. Ringrose, 2011 Quantitative analysis of polycomb response elements (PREs) at identical genomic locations distinguishes contributions of PRE sequence and genomic environment.

Epigenetics Chromatin 4: 1–16.

Oleksiak, M. F., G. A. Churchill, and D. L. Crawford, 2002 Variation in gene expression within and among natural populations. Nat. Genet. 32: 261–266.

Ometto, L., S. Glinka, D. D. Lorenzo, and W. Stephan, 2005 Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. Mol. Biol. Evol. 22: 2119–2130.

Pavlidis, P., S. Hutter, and W. Stephan, 2008 A population genomic approach to map recent positive selection in model species. Mol. Ecol. 17: 3585–3598.

Pavlidis, P., J. D. Jensen, and W. Stephan, 2010 Searching for footprints of positive selection in whole-genome SNP Data from nonequilibrium populations. Genetics 185: 907–922.

Pavlidis, P., D. Živković, A. Stamatakis, and N. Alachiotis, 2013 SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. 30: 2224–2234..

Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno *et al.*, 2012 Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet. 8: e1003080.

Schmidt, J. M., R. T. Good, B. Appleton, J. Sherrard, G. C. Raymant *et al.*, 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. PLoS Genet. 6: e1000998.

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.

Schuettengruber, B., M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek *et al.*, 2009

Functional anatomy of Polycomb and Trithorax chromatin landscapes in *Drosophila* embryos. PLoS Biol. 7: e1000013.

Schwartz, Y. B., T. G. Kahn, P. Stenberg, K. Ohno, R. Bourgon *et al.*, 2010 Alternative epigenetic chromatin states of Polycomb target genes. PLoS Genet. 6: e1000805.

Simon, J. A., and R. E. Kingston, 2009 Mechanisms of Polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. 10: 697–708.

Smith, J. M., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23–35.

Steffen, P. A., and L. Ringrose, 2014 What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat. Rev. Mol. Cell Biol. 15: 340–356.

Stephan, W., 2010 Detecting strong positive selection in the genome. Mol. Ecol. Resour. 10: 863–872.

Stephan, W., and H. Li, 2006 The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity 98: 65–68.

Stephan, W., Y. S. Song, and C. H. Langley, 2006 The Hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics 172: 2647–2663.

Stephan, W., T. H. E. Wiehe, and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism: Analytical results based on diffusion theory. Theor. Popul. Biol. 41: 237–254.

Storey, J. D., J. Madeoy, J. L. Strout, M. Wurfel, J. Ronald *et al.*, 2007 Gene-expression Variation within and among human populations. Am. J. Hum. Genet. 80: 502–509.

Suissa, Y., O. Ziv, T. Dinur, E. Arama, and O. Gerlitz, 2011 The NAB-Brk signal bifurcates at JNK to independently induce apoptosis and compensatory proliferation. J. Biol. Chem. 286: 15556–15564.

Svetec, N., A. Werzner, R. Wilches, P. Pavlidis, J. M. Álvarez-Castro *et al.*, 2011 Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis. Mol. Ecol. 20: 530–544.

Thornton, K. R., and J. D. Jensen, 2007 Controlling the false-positive rate in multilocus genome scans for selection. Genetics 175: 737–750.

Tishkoff, S. A., F. A. Reed, A. Ranciaro, B. F. Voight, C. C. Babbitt *et al.*, 2007 Convergent adaptation of human lactase persistence in Africa and Europe. Nat. Genet. 39: 31–40.

Townsend, J. P., D. Cavalieri, and D. L. Hartl, 2003 Population genetic variation in genome-wide gene expression. Mol. Biol. Evol. 20: 955–963.

Vignieri, S. N., J. G. Larson, and H. E. Hoekstra, 2010 The selective advantage of crypsis in mice. Evolution 64: 2153–2158.

Voolstra, C., D. Tautz, P. Farbrother, L. Eichinger, and B. Harr, 2007 Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. Genome Res. 17: 42–49.

Whitehead, A., and D. L. Crawford, 2006 Variation within and among species in gene expression: raw material for evolution. Mol. Ecol. 15: 1197–1211.

Williams, J. A., J. B. Bell, and S. B. Carroll, 1991 Control of *Drosophila* wing and haltere development by the nuclear *vestigial* gene product. Genes Dev. 5: 2481–2495.

Wollstein, A., and W. Stephan, 2015 Inferring positive selection in humans from genomic data. Investig. Genet. 6: 5.

Wray, G. A., 2007 The evolutionary significance of *cis*-regulatory mutations. Nat. Rev. Genet. 8: 206–216.

Yi, S.-X., C. W. Moore, and R. E. L. Jr, 2007 Rapid cold-hardening protects *Drosophila melanogaster* from cold-induced apoptosis. Apoptosis 12: 1183–1193.

Zink, D., and R. Paro, 1995 *Drosophila* Polycomb-group regulated chromatin inhibits the accessibility of a *trans*-activator to its target DNA. EMBO J. 14: 5660–5671.

ACKNOWLEDGEMENTS

First, I am very thankful to my supervisor Wolfgang Stephan for giving me the opportunity to conduct my Ph.D. research in his lab, and for the support and advice he provided. I also want to thank Stefan Laurent for his enthusiasm for the projects, we worked on together and all the fruitful discussions. I would like to thank Anne Classen for helpul advice concerning fly work and the Polycomb group. I am very grateful for the amazing support and excellent technical assistance of Simone Lange, as well as for the great help of Hilde Lainer, Gaby Kumpfmüller, Gisela Brinkmann, and Anne Steincke. Ganz, ganz lieben Dank! I would further like to thank all the members of the evolutionary biology group. It was great working with you all. Many thanks to my office mates Pablo and Ricardo, it was a lot of fun sharing an office with you. Special thanks to Claus Kemkemer, Iris Fischer, Stephan Hutter, Dirk Metzler and Katharina Böndel for valuable help and support. I am also grateful to Ingrid Kroiß for all her help with administrative things. Finally, I would like to thank my family and friends for their constant support and encouragement throughout these years.