Local adaptation in *Drosophila melanogaster*—

Molecular and morphological aspects

Annegret Werzner

München 2011

Aufforderung zur Bescheidenheit

von Erich Kästner

Wie die Dinge nun mal liegen,

und auch wenn es uns missfällt:

Die Menschen sind wie Eintagsfliegen

an den Fenstern dieser Welt.

Unterschiede sind fast keine,

und was wär auch schon dabei!

Nur: die Fliege hat sechs Beine,

und der Mensch hat höchstens zwei.

Erklärung:

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

Ehrenwörtliche Versicherung:

Ich versichere ferner hiermit ehrenwörtlich, dass die vorgelegte Dissertation von mir selbstständig, ohne unerlaubte Hilfe angefertigt wurde.

München, 20.07.2011 …….....

Annegret Werzner

Local adaptation in *Drosophila* melanogaster—

Molecular and morphological aspects

Dissertation

an der Fakultät für Biologie

der Ludwig-Maximilians-Universität

München

vorgelegt von

Annegret Werzner

aus Karl-Marx-Stadt

München 2011

Dekan: Prof. Dr. Benedikt Grothe

Erstgutachter: Prof. Dr. Wolfgang Stephan

Zweitgutachter: Prof. Dr. Dirk Metzler

Tag der Abgabe: 20.07.2011

Tag der mündlichen Prüfung: 20.10.2011

Acknowledgments

I am thankful to my supervisor, Prof. Wolfgang Stephan, whose guidance and support enabled me to conduct my PhD at the LMU in Munich. I also would like to deeply thank Prof. Dirk Metzler. I benefited from his teaching and support in all aspects of statistical analysis and for running the program R. In summary, that means: interesting, small group classes, we know now that Joe Felsenstein is a sly dog, and stochastic can be handled by biologists, at least on average. Also, I would like to deeply thank my committee members: Prof. John Parsch and Prof. Susanne Renner. Their direction was particularly valuable in the early stages of the PhD. They always had time to discuss and advance my thesis work.

Special thanks to Pavlos Pavlidis for fruitful collaborations and many nice dinners with the family. I address special thanks to Anne Steincke for helping me with the experimental setups and practical matters. Moreover, you became a friend. For the helpful and kind work atmosphere I appreciate the institute members Iris Fischer and Rayna Stamboliyska with whom I shared office. Furthermore I want to thank Anja Hörger, Stefan Laurent, Susanne Voigt, Nicolas Svetec, Lisha Naduvilezhath, Sarah Peter, Stephan Hutter, Aurélien Tellier, and Claus Kemkemer. I will keep you in mind.

My time at the LMU Biocenter was made enjoyable in large part due to my friends: Jonathan Jeschke, Alex Loebel, Hannes Lüling, and the other ‚Neuros'. Furthermore, I address a special thank to my close friends Thomas Knoll, Ben Sonneson, Ina Rost, Conny Müller, Marianne Schindler, and Jana Krämer. Ihr seid bergauf mein Antrieb und Schwung.

Lastly, I would like to thank my family for all their love and encouragement. My mother Brigitte Werzner supported me in all my pursuits and my boyfriend Matthias Köhler was steadily providing advice and love. Thank you.

List of Abbreviations

ABC	Approximate Bayesian Computation
Adh	*Alcohol dehydrogenase* gene
Ala	Alanine
ANOVA	Analysis of Variance
Asn	Asparagine
bp	Base pair
CCRT	Chill coma recovery time
cDNA	Complementary DNA
CG	Coding gene
CIM	Composite Interval Mapping
CLR	Composite likelihood ratio
cM	Centimorgan
Cor	Correlation coefficient
Cov	Covariance
CV	Coefficient of variance
D	Tajima's *D*
DNA	Deoxyribonucleic acid
EEFG	Ecological and evolutionary functional genomics
Flo-2	*Flotillin-2* gene
GSI	Genotypeby-sex interactions
GWAS	Genome-wide association study
h	Hour

H^2	Heritability in a broad sense
H^2_X	Sex-specific X-chromosomal heritability in a broad sense
Hsp	Heat shock protein
IM	Interval Mapping
kb	Kilobase pair
LD	Linkage disequilibrium
LOD	Likelihood of Odds
LR	Likelihood ratio
Mb	Megabase pair
mg	Milligram
ml	Milliliter
µm	Micrometer
mRNA	Messenger ribonucleic acid
u	Mutation rate
n	Sample size
NGS	Next generation sequencing
ns.	Non-significant
ω	Omega
ORF	Open reading frame
PCR	Polymerase chain reaction
Pfcrt	*Plasmodium falciparum chloroquine resistance transporter* gene
Pgd	*Phosphogluconate dehydrogenase* gene
ph-d	*Polyhomeotic distal* gene
ph-p	*Polyhomeotic proximal* gene

QTL	Quantitative trait loci
QTN	quantitative trait nucleotide
r	Recombination rate
r_{GS}	Crosssex genetic correlation coefficient
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SCP	Supercooling point
SCS	Single colonization scenario
se	Standard error
Ser	Serine
SFS	Site frequency spectrum
σ^2	Sigma, measure of variance
σ^2_L	Between-line variance
σ^2_E	Within-line variance
SNP	Single nucleotide polymorphism
θ_W	Theta Watterson
TSHR	*Thyroid stimulating hormone receptor* gene
UTR(s)	Untranslated region(s)
v	*Vermilion* gene
wapl	*Wings apart-like* gene
wt	Wild type
XR	X chromosomal recombinant
y-ac	*Yellow-achaete* gene

Summary

The present work is focused on the identification of positively selected genes that are involved in local adaptation in European *Drosophila melanogaster*. This species is globally distributed as human commensal and occupies almost every ecozone. The ancestral range, however, is afrotropical and the questions arise how and when the fruit fly managed to invade new environments that differed in environmental parameters.

One phenotype that might have been of crucial importance for the cosmopolitan distribution of an insect species is cold tolerance. Using fly samples from the ancestral range of tropical Africa and the derived range of temperate Europe we compared their cold tolerance by measuring the time they need to recover from a cold induced chill coma. We picked the most divergent African and European fly lines as parental lines and created a huge population of X chromosomal recombinants. We searched their X chromosome for quantitative trait loci (QTL) that caused phenotypic divergence between the parental lines and identified several loci that were associated with chill coma recovery time. Subsequently, we went back to the original population samples from Africa and Europe and characterized a European selective sweep that was co-localized with one QTL.

We established a novel colonization model to tackle the question when *D. melanogaster* spread from Africa and invaded new environments, such as Europe and Asia. We sequenced ~280 fragments of the X and third chromosome of an Asian population sample and aligned them with the corresponding fragments of the African and European sample that were already sequenced before. By means of Approximate Bayesian Computation (ABC) we found one common ancestor of European and Asian *D. melanogaster* that left Africa around

16,800 years ago. We reject an ancient colonization event from Africa to Asia, which could have led to the strongly divergent Asian phenotype of the 'Far East Race'.

A formerly performed genome scan of X-linked genetic variation of the European and African sample revealed interesting candidates of European-specific adaptation. To analyze one candidate region more closely we conducted a follow-up study and sequenced the entire candidate in both population samples. We found multiple European specific genetic variants one of which was an insertion/deletion polymorphism that generates a new transcript of the *flotillin-2* gene. This transcript (*Flo-2-C*) is unique to *D. melanogaster* and encodes a truncated version of flotillin, a membrane-anchoring scaffolding protein. An expression analysis revealed that the *Flo-2-C* transcript is expressed in most fly lines independent of the gene structure in third instar larvae. Thus, a disordered gene structure does not prevent the process of transcription and might reflect a young gene variant.

Zusammenfassung

Die vorliegende Arbeit untersucht Gene, die in einer Europäischen Population der Taufliege (*Drosophila* melanogaster) unter positiver Selektion stehen. Die Taufliege lebt als menschlicher Kommensale und besiedelt inzwischen fast alle Biogeografischen Zonen. Ihr ursprüngliches Verbreitungsgebiet ist jedoch das tropische Afrika. Die globale Verbreitung in den unterschiedlichsten Ökosystemen wirft die Fragen auf, wie und wann *D. melanogaster* sich an derart unterschiedliche Umweltbedingungen angepasst hat.

Ein Phänotyp, der dabei wahrscheinlich eine besonders wichtige Rolle gespielt hat, ist Kältetoleranz. Um den Zusammenhang zwischen Verbreitungsgebiet und Kältetoleranz zu untersuchen, verwendeten wir Fliegenlinien aus Afrika und Europa und verglichen die Zeit, die die einzelnen Linien benötigten, um sich von einem Kälteschock zu erholen. Wir wählten die zwei unterschiedlichsten Fliegenlinien aus Afrika und Europa als Elterntiere aus und erstellten eine große Population von Linien mit rekombinierten X Chromosomen. Diese verwendeten wir für eine QTL-Studie und fanden mehrere Genorte die Unterschiede zwischen den Elternlinien hervorrufen und den Phänotyp der Kältetoleranz codieren. Anschließend untersuchten wir unsere Fliegenlinien aus der Europäischen Population auf das Signal von positiver Selektion, das mit dem Signal der Kältetoleranz überlappt und identifizierten einen ‘Selective Sweep’.

Um die Frage über die Kolonisierungsgeschichte von *D. melanogaster* zu beantworten verwendeten wir DNA-Sequenzdaten von drei Populationen: der Afrikanischen Stammpopulation und zwei abgeleiteten Populationen aus Europa und Asien. Mit Hilfe der ABC Methode fanden wir heraus, dass *D. melanogaster* Afrika vor rund 16.800 Jahren

verließ und Europa und Asien in einer einzigen Migrationswelle besiedelte. Wir konnten eine einzelne, unabhängige und sehr alte Besiedlung Asiens ausschließen, die als Ursache des morphologisch divergenten asiatischen Phänotyps der 'fernöstlichen Rasse' vorgeschlagen wurde.

Eine Genomuntersuchung des X Chromosoms der Europäischen Fliegenlinien ergab mehrere interessante Kandidatenregionen, die unter positiver Selektion stehen. In einer Anschlussstudie analysierten wir die vollständige Sequenz einer dieser Kandidatenregionen und verglichen die Daten mit den Afrikanischen Fliegenlinien. Der Vergleich zeigte genetische Veränderungen in den Europäischen Fliegen, unter anderem eine völlig neue Variante des *flotillin-2* Gens. Diese Variante (*Flo-2-C*) ist in den Europäischen Fliegen fast vollständig vorhanden und codiert eine verkürzte Version des Flotillins, einem Membran-verankernten Gerüstprotein. Die meisten Afrikanischen Fliegenlinien besitzen eine fehlerhafte Genstruktur des *Flo-2-C*. Expressionsdaten zeigen jedoch, dass *Flo-2-C* im dritten Larvalstadium von *D. melanogaster* trotz fehlerhafter Genstruktur exprimiert wird. Da *Flo-2-C* in anderen Arten nicht vorkommt, handelt es sich unter Umständen um ein sehr junges Gen in seiner Entstehungsphase.

Table of Contents

List of Figures

List of Tables

General Introduction

The process of evolution was first studied by observing phenotypic differences between natural populations of the same species. One of the main results was the high degree to which species suit their environmental conditions. The underlying evolutionary mechanism of this phenomenon is natural selection. Adaptive traits evolve to a given optimum by natural selection and become more frequent in the population. This leads to a local reduction in trait variation and accelerates population differentiation. The main goal in evolutionary biology is to detect adaptive traits and describe their ecological relevance. The first empirical studies predate the era of molecular biology and even *Drosophila* as model organism for evolution; still they provide us with valuable insights.

Mutations as source of variation In the year 1886, Hugo de Vries, a Dutch botanist and geneticist, examined a colony of evening primroses (*Oenothera lamarckiana*) that covered an abandoned potato field nearby the village of Hilversum. De Vries aimed to observe the process of speciation by spontaneous "single variations". He studied each primrose; profoundly convinced that speciation occurs within populations and fits into C. Darwin's concept "*On the Origin of the Species by Natural Selection*" (Darwin 1859). During his weekly visits, de Vries identified two unknown variants of *O. lamarckiana* within his small population: one exhibited a shorter stylus and one had flat leaves. Due to these distinct characters, he named the variants *O. brevistylis* and *O. laevifolia*. Back in his own garden, de Vries was able to consistently reproduce both variants from their seeds. He was, therefore, convinced to have attended the natal hour of two new species of the genus *Oenothera* and described how variation is introduced into a population.

The small and sudden phenotypic changes that appear by speciation were, according to de Vries, mutability-based. He used the term 'mutation' to define sudden and rare events that alter the individual's "anatomy", the unit of traits that defines a species. Interestingly, he even made use of the term substitution while explaining: "To simplify matters, we can compare mutations by chemical substitutions" (de Vries 1901 p. 4/684).

De Vries was surprisingly prescient in his understanding of the divergence between species, long before the discovery of DNA as genetic material (Watson & Crick 1953). The modern term 'mutation' refers to changes in the DNA sequence, such as nucleotide substitutions (point mutations), insertions, deletions, duplications and inversions. Point mutations alter the DNA sequence and create variation in terms of single nucleotide polymorphism (SNP).

Since de Vries' efforts to observe naturally occurring speciation, scientists were interested to understand how SNPs alter a species phenotype, how they are distributed in natural populations, and to which extent species vary genetically. The study of the frequencies of genetic variants in populations as well as their temporal and spatial frequency changes belong to the field of population genetics. It provides a coherent framework for Charles Darwin's principle of natural selection and Mendelian genetics, which were jointly summarized in the book "Evolution - The Modern Synthesis" first published in 1942 by Julian Huxley.

In the case of *O. lamarckiana*, the flood of more and more mutant phenotypes in the gardens and greenhouses of scientists revealed that chromosomal duplications cause the aberrant anatomy in these phenotypes (Emerson 1935; Catcheside 1937). De Vries'

interpretation inflated the magnitude to which a mutation affects a trait; nonetheless it was the impulse and starting point for a new field of biology.

Mutations happen in different frequencies among organisms formalized by the mutation rate *u*. It measures the probability by which an organism gains a new mutation and is either given as probability per site per generation or per genome per generation (Drake *et al.* 1998). It can vary between species and even between genomic regions within a single species (Woodruff *et al.* 1984; Li & Stephan 2006; Haag-Liautard *et al.* 2007). In accordance with de Vries' principle, mutations are sudden and rare events. However, their impact might not be as harmless to an individual. Whenever two genetic variants differ in fitness, natural selection can act on it.

A high fraction of newly arising mutations is immediately removed from the population due to functional constraints (purifying selection) and thus never appears as a polymorphism in natural populations. Studies of *D. melanogaster* protein coding regions show up to 90% of mutations that alter the amino acid sequence of a protein are disadvantageous. Most of the nucleotide polymorphisms that are actually present in natural populations are expected to be almost neutral (Kimura 1968; Kaplan *et al.* 1989). Kimura suggested that most genetic changes in a population can be attributed to genetic drift, i.e. the stochastic process acting randomly on SNP frequencies either fixing them in the population or eliminating them. Both fates –fixation and loss– reduce genetic variation within a population. However, if it comes to fixation, the new mutant will substitute the former population variant leading to divergence between populations.

The rate of evolution in terms of mutant substitutions within populations is accelerated in functionally less important parts of a gene whereas in functionally constraint regions

purifying selection predominates (Kimura 1983; Hughes 2007). Under neutral conditions the average rate of substitution equals the mutation rate, persistently driving the neutral evolution of natural populations (Kimura 1968). In his 'Neutral theory of molecular evolution' (1983) Kimura further stated neutral mutations to be the foundation of genetic variation within species. He thus linked two differing forces that shape genetic variation within a population: neutral mutations that introduce variation and drift that potentially eliminates variation.

Deviation from neutrality Beside serious DNA damages, mutations also provide the raw material for adaptation. The introduction of sequencing techniques (Maxam & Gilbert 1977; Sanger *et al.* 1977) and the availability of DNA sequence data from natural populations allowed population geneticists to resolve the distribution of genetic variation in the genome by direct sequence comparison (*alcohol dehydrogenase* (*Adh*), Kreitman 1983; *vermilion* (*v*), Stephan & Langley 1989; *yellow-achaete* (*y-ac*), *phosphogluconate dehydrogenase* (*Pgd*), Begun & Aquadro 1991). These studies analyzed gene regions (loci). Interestingly, the studied loci either exhibited too high (*Adh*, *Pgd*) or too low (*y-ac*, *v*) levels of genetic variation compared to neutral expectations. The authors suggested positive selection on beneficial variants as cause for the variation pattern at those loci.

In the cases of *Adh* (Kreitman & Aguadé 1986) and *Pgd* genetic variation was maintained at levels above neutral expectations as more than one variant was favorable over time (balancing selection). At loci *y-ac* and *v*, genetic variation is reduced due to positive directional selection of one strongly favored variant. Maynard Smith & Haigh (1974) introduced a model of adaptive evolution that described how genetic variation is reduced because of linkage between a newly arisen beneficial mutation and neutral variants. Due to positive directional selection, the beneficial mutation will rise in frequency until it becomes

fixed in the population. At the same time, the selectively linked neutral variants will jointly increase in frequency and become fixed in the population, unless recombination breaks the linkage apart. This process is known as genetic hitchhiking. Each beneficial mutation that becomes fixed within the population will be surrounded by a 'window' of reduced genetic variation, called a selective sweep. The dimension of such a selective sweep depends on the population size, the strength of selection, and the rate of recombination (Kaplan *et al.* 1989).

Charlesworth and coworkers (1993) found background selection as an alternative explanation for the reduced variability at the *y-ac* locus due to the joint elimination of deleterious mutations and linked neutral variants. In regions of intermediate to high levels of recombination, however, the impact of background selection should be low and patterns of genetic variability should be consistent with the neutral theory (Charlesworth *et al.* 1993).

Within a sweep region, low levels of genetic variation are accompanied by two more features. One is the skew of SNP frequencies within a population. After a beneficial mutation has swept through the population rare and high-frequency derived variants are overrepresented in that genetic region (Braverman *et al.* 1995; Fay & Wu 2000). The frequencies of SNPs are summarized in the site frequency spectrum, which departs from the neutral expectations in the case of hitchhiking. Another feature of a selective sweep is the high level of linked polymorphism on either side of the positively selected variant, creating elevated patterns of linkage disequilibrium (LD) (Kim & Nielsen 2004). This LD pattern does not extend between neutral loci across the selected site due to recombination and secondary hitchhiking on the recombinants that also carry the beneficial allele (Stephan *et al.* 2006).

Selective sweep mapping The characteristics of selective sweeps were used to detect recent positive selection on ecologically favored variants to unravel the molecular basis of

adaptive evolution. Numerous species were scanned for selective sweeps using neutral DNA and microsatellite polymorphism data of multiple loci along recombining chromosomes. Williamson *et al.* (2007) discovered more than 100 regions along the entire human genome that showed evidence for strong and recent selective sweeps, containing genes of great biological interest related to pigmentation, olfactory receptors, nervous system development, and temperature tolerance. Another example of a selective sweeps was found in the human malaria parasite *Plasmodium falciparum* (Wootton *et al.* 2002). Asian and African populations of *P. falciparum* showed a conspicuous decline in genetic variation and high levels of linkage disequilibrium at the genomic region of the gene *pfcrt* on chromosome seven. *Pfcrt* is a key gene for chloroquine-resistance, a drug that was widely used to prevent malaria during the twentieth century (Fidock *et al.* 2000). Selective sweep mapping was also successfully applied to study plant and animal domestication resulting from artificial selection. Tian *et al.* (2009) used candidate loci at chromosome 10 in *Zea mays* that were previously associated with the domestication process. The authors revealed evidence for a selective sweep at these loci, combining morphological and genetic data to unravel changes that promoted the shift from a wild to a domesticated species. In chicken, the thyroid stimulating hormone receptor (*TSHR*) on chromosome five was found to be a domestication locus that swept through all domesticated chicken species (Rubin *et al.* 2010). An amino acid substitution relaxes the strict regulation of seasonal reproduction, which is absent in domesticated animals.

The multilocus approach of sweep mapping is an established tool for localizing the footprints of positive selection and for identifying the genetic basis of adaptive evolution because it harbors several advantages. First, a preliminary selection of candidate genes can be done but is not necessarily required for this method. According to the marker density, the

genome can be scanned for selective sweeps without any prior knowledge of candidate genes or gene regions in a "bottom-up" manner (Wright & Gaut 2005). The corresponding phenotypes are subsequently analyzed for functional innovations that promoted the selective phase. Second, the underlying phenotype is not necessarily visible. Studies of selective sweeps frequently detect candidate loci of less apparent function in physiological pathways such as immune response, neuronal development, and metabolic processes (de Groot *et al.* 2002; Ihle *et al.* 2006; Williamson *et al.* 2007). Third, natural populations can be screened for selective sweeps directly without any experimental crosses. The method thus is straight forward: a natural population is scanned for patterns of DNA polymorphism, loci that deviate from neutrality are detected, and their biological function might be identified (Schlötterer 2003).

However, sweep mapping has some serious difficulties. Natural populations may regularly experience fluctuations in size, passing through bottlenecks or spreading by expansion. Such a demographic background generates two error sources. First, it violates the neutral theory's assumptions of a constant population size in creating the polymorphism patterns seen in selective sweeps, thus leading to false positives (type I error) (JENSEN *et al.* 2007a). Second, it disturbs the signature of a selective sweep in the genome leading to false negatives (type II error) (Pavlidis *et al.* 2010). Such confounding effects of demography and selection make it indispensable to reconstruct the demographic history of the populations under consideration.

When the fixation of a beneficial variant is complete, the signal of the sweep vanishes inexorably. The high-frequency SNP's that hitchhiked along are subsequently fixed in the population and do not contribute to polymorphism anymore (Kim & Stephan 2000). New

mutations enter the population and straighten the selective sweep while ongoing recombination breaks linkage apart. Consequently detecting old selective sweeps from random loci has a low success rate (Przeworski 2002). Finally, if the scan for selective sweeps is accomplished successfully and candidate genes are identified with a "bottom-up" approach, missing data on the biological function of the candidate genes untimely stops the search for adaptation: a situation in which selective sweep mapping at first leads nowhere.

Population geneticists have taken advantage of model organisms to meet this challenging situation. One established model organism in genetics is the fruit fly *D. melanogaster.* It is an African sub-tropical species that extended its geographic range at first to Europe and Asia and subsequently to the rest of the world, becoming a cosmopolitan human commensal (David & Capy 1988; Lachaise & Silvain 2004). Li & Stephan (2006) provide a coherent model for the demographic history of the African and European natural populations based on DNA polymorphism data of the X chromosome. Their findings suggest that the European population has split off from an expanding African ancestor approximately 15,000 years ago. The European founder event was accompanied by a quite severe bottleneck phase with a reduction of the X chromosomal population size down to only 2,200 effective individuals for a time of around 340 years. The date of the bottleneck corresponds well to the beginning of the Holocene after the last glaciation (around 11,500 years ago, Gulliksen *et al.* 1998) when Europe became a suitable environment for *Drosophila*.

The range expansion challenged *D. melanogaster* with new ecological constraints and induced local adaptation which was mapped to the X-chromosome by several authors (Glinka *et al.* 2003; Ometto *et al.* 2005; Li & Stephan 2006; Hutter *et al.* 2007). As population sample, they used 24 highly inbred lines of *D. melanogaster*: 12 lines from Africa (Lake

Kariba, Zimbabwe) and 12 lines from Europe (Leiden, Netherlands). Follow-up studies identified the genes targeted by positive selection with respect to the demographic history. Glinka *et al.* (2006) reported the genes *CG1677* and *CG2059* ("fragment 125" in Glinka *et al.* 2003) to have experienced nucleotide substitutions in the European sample which probably were derived from standing genetic variation in the ancestral population. Beisswanger *et al.* (2006) investigated the *wapl* region ("fragment 10" in Glinka *et al.* 2003) encompassing a large number of genes which could be scaled down in further analysis to the tandemly duplicated genes *ph-d* and *ph-p* (Beisswanger & Stephan 2008). Surprisingly, all follow-up studies traced the signal back to the African population supporting the theory that some selective sweeps may have originated in the ancestral range (Pool *et al.* 2006). In addition, Pool *et al.* (2006) hypothesized that in cosmopolitan populations the window of reduced variation might be enlarged by the bottleneck and thus be easier to detect. Until now, there is no candidate study of a selective sweep that is unique to the European sample of *D. melanogaster* and the "bottom-up" approach of selective sweep mapping so far failed to identify the genetic basis of ecologically relevant phenotypes in European *D. melanogaster*. The Asian population of *D. melanogaster* could be an equally promising candidate for selective sweep mapping (Schlötterer *et al.* 2006), but the demographic model and thus the neutral scenario has been missing up to now. However, it should not be complicated to identify adaptive phenotypes that affect survival in different environments for a species as well studied as *Drosophila* whenever a reliable population colonization model is at hand.

Quantitative traits and population genetics Naturally occurring phenotypic variation within a species indicates positive selection in terms of population differentiation. Going "top-down" from the phenotype to the underlying genetic architecture one can identify loci that account for trait variation (quantitative trait loci, QTL) using QTL mapping (Mackay

2001). In the case of *D. melanogaster*, many interesting traits may contribute to the success of worldwide colonization, especially for the switch from the ancestral tropical to a temperate environment. Several traits follow latitudinal clines in populations on different continents including developmental time (James & Partridge 1995), temperature tolerance (James *et al.* 1997; Hoffmann *et al.* 2002), egg size (Azevedo *et al.* 1996), diapause (Mitrovski & Hoffmann 2001) and body size (David & Capy 1988; Van 't Land *et al.* 1999).

All these traits are distributed continuously in the population and do not follow the discrete pattern of Mendelian inheritance. Ronald A. Fisher (1918) was the first to unravel the apparent contradiction of natural selection causing continuous trait variation. He proved that multiple genes or what he referred to as *allelomorphs*, coding for one trait by multifactorial inheritance can cause that kind of continuous variation as observed in natural populations. Investigations on the petal length of F_2 *Oenothera* hybrids revealed a wide range of variation for flowers of the same individual as well as petals of the same flower, even exceeding the combined parental forms (Gates 1923). These unexpected observations could not be explained by Mendelian inheritance and led to the conclusion that information on size is repeated in the same organism, most likely on different chromosomes.

The possibility of multifactorial inheritance in size motivated Sax (1923) to study seed size of the common bean (*Phaseolus vulgaris*). This pioneering QTL study provided the very simple idea of QTL mapping using F_2 hybrids with continuous trait variation in seed size and phenotypic Mendelian markers with distinct color patterns, such as pigmentation and mottling that were closely linked to the trait of interest. Thus, the markers allowed observing what is otherwise unobservable: tightly linked to the genetic loci that encode seed size they correctly reflect the genotype of these loci. Sax (1923) reported evidence for two facts: *i)* seed size is

encoded at several loci in the genome with different relative effects on trait variation; *ii)* markers are useful to trace the genetic variant of each QTL back to the parental line from which they were inherited. The use of morphological markers such as color patterns were later displaced by genome-wide neutral molecular markers based on DNA polymorphism, such as microsatellites, SNPs, and insertion sites of transposable elements which can be scored for significant trait association. The higher a marker-trait association the lower is the recombination frequency between the QTL and the marker position, which does not necessarily refer to their actual distance in base pairs (bp) along the chromosome. Thus, the recombination frequency is estimated from the data and translated into map units (in centimorgan, cM) either using Haldane's map function (without interference, Haldane 1919) or Kosambi map function (with interference, Kosambi 1943). Sophisticated statistical methods like Interval Mapping (IM, Lander & Botstein 1989) and Composite Interval Mapping (CIM, Zeng 1994) allow precise localization of QTL along chromosomes. As pointed out by Mackay (2001) the pivotal questions are: where in the genome is the trait of interest encoded? How many QTL affect trait variation? Does the single locus have major or minor effects? Do effects orchestrate in an additive or multiplicative manner? What are the frequencies of QTL variants within and between populations?

Genome-wide screens for genetic loci affecting trait variation were carried out in numerous species; first and foremost among these is *D. melanogaster* that was studied for QTL in bristle number (Gurganus *et al.* 1998), longevity (Nuzhdin *et al.* 1997), lipid content (Wang *et al.* 2005b), thermotolerance (Morgan & Mackay 2006; Norry *et al.* 2008), aggressive behavior (Edwards & Mackay 2009), and many more. *Drosophila* turned out to be a profitable model organism for evolutionary quantitative genetics for several reasons: it exhibits a rich repertoire of traits; the large number of offspring and the short generation time

facilitate the construction of (recombinant) inbred lines and artificial selection lines; a large number of individuals with the same genotype can be studied under controlled laboratory conditions; and the construction of chromosome substitution lines is easily realized by means of balancer chromosomes that inhibit recombination potentially for all three major chromosomes.

Unlike selective sweep mapping, where the signature of positive selection is locally quite restricted even in non-equilibrium populations (Pavlidis *et al.* 2010), QTL mapping has the disadvantage of discovering broad candidate regions with an immense number of genes (Mackay 2010). In turn, sweep mapping does not necessarily identify genes with known biological function or a SNP distribution that was shaped in an adaptive manner. Thus, neither approach by itself is satisfactory in terms of detecting ecologically relevant variants of single positively selected genes. To solve this problem, multiple molecular population genetic analyses have to be combined, such that several characteristics of the dataset are covered by the corresponding statistical tools.

Present work The aim of the thesis presented here was the identification of the genetic architecture of adaptive traits in a sample of European *D. melanogaster*. The subtropical origin of this species suggests inevitable positive selection for adapting to the temperate environment and to establish a stable population. Furthermore, we provide a new colonization model of how *D. melanogaster* spread in Europe and Asia, which is the foundation of subsequent scans for selection.

Chapter 1) summarizes the results of a QTL analysis of the X chromosome related to cold tolerance. We were able to show that multiple QTL along the X chromosome affect this trait and applied modern statistical approaches to identify the single allelic effects. Subsequent

selective sweep mapping precisely identified the gene *CG16700*, positively selected only in the European population.

The first part of Chapter 2) provides a new colonization model of the African and European *D. melanogaster* including Asia as a third population. We reject the hypothesis of an ancient Asian colonization and show that European and Asian *D. melanogaster* share a common ancestor. The second part of this chapter comprises a morphological analysis of *D. melanogaster* samples from the above-mentioned populations to investigate whether or not the Asian flies show some distinct characters that were described as Far East Race.

Chapter 3) presents a selective sweep analysis of a genetic region of the European population of *D. melanogaster* identified by a former genome scan (Ometto *et al.* 2005). We analyzed the complete DNA sequence of the candidate region and discovered a novel splicing variant of the *flotillin* gene that is in very high frequency in the European *Drosophila* sample.

Each study of this thesis is based on the same set of highly inbred fly lines that are maintained in the lab under stable monitored conditions. These lines are used for sweep mapping, for modeling the demographic history, as well as parental lines for the QTL analysis. By using multiple statistical approaches in the consistent framework of a common species sample, we entered the area of ecological and evolutionary functional genomics (EEFG, Loeschcke *et al.* 2004) to address the fundamental question: Which genes influence phenotypic traits and affect adaptation?

Chapter 1

Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis

Svetec N [1], Werzner A [1,2], Wilches R, Pavlidis P, Alvarez-Castro JM, Broman KW, Metzler D, Stephan W

([1] shared first authorship, [2] corresponding author)

MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 530–544 doi: 10.1111/j.1365-294X.2010.04951.x

Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis

NICOLAS SVETEC,*†[1] ANNEGRET WERZNER,*[1] RICARDO WILCHES,* PAVLOS PAVLIDIS,* JOSÉ M. ÁLVAREZ-CASTRO,*‡ KARL W. BROMAN,§ DIRK METZLER* and WOLFGANG STEPHAN*

*Section of Evolutionary Biology, Biocenter, University of Munich, 82152 Planegg-Martinsried, Germany, †Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, Ontario, Canada M5S 3B2, ‡Department of Genetics, University of Santiago de Compostela, Avda. Carvalho Calero, ES-27002 Lugo – Galiza, Spain, §Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI 53706, USA

Abstract

***Drosophila melanogaster* is a cosmopolitan species that colonizes a great variety of environments. One trait that shows abundant evidence for naturally segregating genetic variance in different populations of *D. melanogaster* is cold tolerance. Previous work has found quantitative trait loci (QTL) exclusively on the second and the third chromosomes. To gain insight into the genetic architecture of cold tolerance on the X chromosome and to compare the results with our analyses of selective sweeps, a mapping population was derived from a cross between substitution lines that solely differed in the origin of their X chromosome: one originates from a European inbred line and the other one from an African inbred line. We found a total of six QTL for cold tolerance factors on the X chromosome of *D. melanogaster*. Although the composite interval mapping revealed slightly different QTL profiles between sexes, a coherent model suggests that most QTL overlapped between sexes, and each explained around 5–14% of the genetic variance (which may be slightly overestimated). The allelic effects were largely additive, but we also detected two significant interactions. Taken together, this provides evidence for multiple QTL that are spread along the entire X chromosome and whose effects range from low to intermediate. One detected transgressive QTL influences cold tolerance in different ways for the two sexes. While females benefit from the European allele increasing their cold tolerance, males tend to do better with the African allele. Finally, using selective sweep mapping, the candidate gene *CG16700* for cold tolerance colocalizing with a QTL was identified.**

Keywords: cold stress, QTL analysis, selective sweep mapping, thermal adaptation

Received 4 May 2010; revision received 20 October 2010; accepted 1 November 2010

The search for adaptive signatures in genomic data is essential to unravel the evolutionary history of a species (Bersaglieri *et al.* 2004; Palaisa *et al.* 2004; Beisswanger *et al.* 2006; Kane & Rieseberg 2008). Comprehensive studies of genome-wide patterns of variation provided insight into allele distributions and ecologically favoured genotypes in different species (Hinds *et al.* 2005; Nordborg *et al.* 2005; Ometto *et al.* 2005; Borevitz *et al.* 2007; Hutter *et al.* 2007). However, these studies were based on patterns of genetic variation that can be altered by nonselective forces like population demography. Mapping methods for quantitative trait loci (QTL)

Correspondence: Annegret Werzner, Fax: 0049 89 2180 74104; E-mail: werzner@bio.lmu.de

[1]These authors contributed equally to this work.

overcome this problem (Mackay 2001). Once the QTL are discovered, one can quantify their impact by estimating allelic effects and interactions and find plausible candidate genes that control the trait.

Traditionally, QTL studies start out with parental lines, such as inbred lines of certain plant or animal species, artificially selected for a divergent phenotype (Long *et al.* 1995; Moehring & Mackay 2004; Norry *et al.* 2008). Likewise, natural populations used for QTL analysis should exhibit divergent traits that have been driven apart by natural or sexual selection. *Drosophila melanogaster* is a cosmopolitan species with locally adapted populations, which makes it a suitable model for naturally based QTL studies (David & Capy 1988). The ancestral population in central equatorial Africa split off from the *melanogaster-simulans* ancestor around 2–3 Myr ago (Cariou 1987; Lachaise & Silvain 2004). By contrast, the non-African populations have a by far younger history. The habitat expansion to the Eurasian continent started in sub-Saharan Africa around 15 000 BP (David & Capy 1988; Baudry *et al.* 2004; Stephan & Li 2007). Stable derived populations in major parts of Europe and Asia as well as new populations in North America and Australia suggest recent adaptation to new environmental conditions. The occurrence of geographical patterns and latitudinal clines in populations on different continents for several phenotypic traits, such as developmental time (James & Partridge 1995), temperature tolerance (James *et al.* 1997; Hoffmann *et al.* 2002), egg size (Azevedo *et al.* 1996), diapause (Mitrovski & Hoffmann 2001) and body size (David & Capy 1988; Van 't Land *et al.* 1999), suggest climatic factors as important selective pressures. Thus, one of the first important adaptive steps for a tropical species, like *D. melanogaster*, is an increase in cold tolerance while colonizing temperate regions.

Indeed, numerous studies suggest that *D. melanogaster* steadily adapts to new temperature conditions (Mckenzie 1975; Azevedo *et al.* 1996; James *et al.* 1997; Gibert *et al.* 2001; Hoffmann *et al.* 2003; Anderson *et al.* 2005). It generally maintains a stable population within the thermal limits of 12–32 °C, even though far colder temperatures can temporarily be accepted (David & Clavel 1967; Stanley *et al.* 1980). All developmental stages of *D. melanogaster* have the capacity for cold tolerance adaptation; nonetheless, the adult stage is the strongest candidate to survive drops in temperature on a seasonal scale (Mckenzie 1975; Tucić 1979; Izquierdo 1991). The larval phase apparently runs through critical periods of development, which are highly susceptible to thermal changes and tend to collapse when exposed to cold. After eclosion, the flies overcome these critical periods of development and selection might favour adults for overwintering.

Outside the tropical regions, however, *D. melanogaster* is in any case closely dependent on human activity (Lachaise & Silvain 2004). When temperature drops below 6 °C, they locally hide in human settlements and enter quiescence, a resting stage that might have serious consequences on sex-specific physical constraints (Izquierdo 1991). With rising temperature in springtime, populations might regenerate and the capacity for reconstructing a population is mostly attributable to inseminated females (Izquierdo 1991). Certainly, the most critical factor influencing the survival of *Drosophila* populations in temperate regions is wintertime.

Tucić (1979) first suggested that cold tolerance in *D. melanogaster* is controlled by various groups of genes at particular developmental stages. He proposed that the genetic factors controlling cold tolerance are spread along the entire genome and that the relative importance of genes varies between chromosomes, such that most genes are located on chromosomes 2 and 3, and only relatively few are present on the X chromosome. However, in contrast to the autosomes, QTL on the X chromosome for cold tolerance have not been detected in previous studies (Morgan & Mackay 2006; Norry *et al.* 2008).

Here, we present the results of a QTL analysis of cold tolerance in *D. melanogaster* concentrating on the X chromosome. In this context, cold tolerance is referred to as a short time response to a cold shock that is induced by a temporary and severe cold treatment (0 °C) (for the definition of cold shock, see Loeschcke & Sørensen 2005). The individual recovery time of such a cold shock was shown to be a reliable measure of the organismal defence abilities against low temperatures in *Drosophila* species (Gibert *et al.* 2001). We chose the X chromosome for two reasons: (i) to revisit earlier studies that did not find QTL on the X and (ii) to compare the QTL analysis with our map of selective sweep regions (that is most detailed for the X chromosome; Li & Stephan 2006) to search for candidate genes of cold tolerance. In other words, the selective sweep approach is employed here to fine-map QTL by testing which genes under a QTL were subjected to recent strong positive selection and may therefore be used for further analysis.

For the QTL study, we used two highly inbred parental lines with homogenized genetic background and unique X chromosomes that were each derived from wild populations of two very different thermal environments: sub-Saharan Africa (Lake Kariba, Zimbabwe) and Western Europe (Leiden, The Netherlands). The parental lines exhibit significantly different phenotypes for cold tolerance, such that the European line is much better adapted to cold than the African one. Reciprocal crosses of the parental lines were used to establish a panel of X-chromosomal recombinant (XR) lines. With these XR lines, we addressed the following questions: (i) Do we

find a major QTL with large effect or several QTL with intermediate effects? (ii) Which genomic regions of the X chromosome are associated with cold tolerance? (iii) Are there interactions between QTL? (iv) Do QTL–sex interactions play a role for cold tolerance? For the selective sweep analysis, we used samples that have been collected from the same two local populations from Africa and Europe. In fact, the two lines that led to the construction of the XR lines were derived from these samples.

Materials and methods

Drosophila stocks

Experiments were carried out using highly inbred isofemale lines of *D. melanogaster*. All stocks were kept at 23 °C, 45% humidity and under constant light conditions. Development took place on a high-nutrient killed yeast food medium (12 mL) in glass vials of 200 mL. Two unrelated lines of wild origin served as starting point for the QTL analysis: line A157 was sampled in Africa (Lake Kariba, Zimbabwe; Begun & Aquadro 1993) and line E14 was collected in Europe (Leiden, Netherlands 1999). The demographic and selective history of the populations from which these two lines were sampled has been studied in detail using genome scans (Glinka *et al.* 2003; Ometto *et al.* 2005; Beisswanger *et al.* 2006; Li & Stephan 2006; Hutter *et al.* 2007). The cold tolerance profiles between lines A157 and E14 differ significantly (Welch's *t*-test, $P < 0.001$).

By constructing substitution lines, we assessed the presence of cold tolerance factors on the X chromosome. Using the balancer line 6418 (FM7j balancer X chromosome obtained from Bloomington Stock Center; see legend to Supplementary Fig. S1), we introgressed a wild-type (*wt*) X chromosome into the balancer line background. First, we isolated a single female offspring from a *wt*-balancer cross and then back-crossed it for seven generations to males of the balancer line (Supplementary Figs S1 and S2). The resulting lines carry the genetic background of the balancer line (mitochondria, 2nd, 3rd and 4th chromosome) and the X chromosome of either the A157 or the E14 wild-type. These lines will be referred to as A* and E*, respectively.

We checked subsequently whether these lines carry the same autosomal background by sequencing several short DNA fragments of the autosomes (data not shown). This confirmed that A* and E* were identical except for their X chromosomes. Moreover, both lines still show significantly different cold tolerance profiles (Welch's *t*-test, $P < 0.001$), which made them suitable parental lines to establish an array of XR lines by directed crossings. The offspring of two reciprocal crosses between A* and E* was allowed to mate freely for five generations to accumulate recombination events. During this phase, generations were strictly separated by removing the adults just before their offspring hatched. Male flies of the F5 generation were then crossed and back-crossed to the balancer line to establish a population of 186 XR lines each carrying a unique recombinant X chromosome. Once the XR lines were created, they were maintained in the laboratory by full-sib mating and monthly mass transfer.

Cold tolerance phenotype

As a measurement of cold tolerance, we used a modified protocol for the chill coma recovery test of David *et al.* (1998). At temperatures below 0 °C, *D. melanogaster* adults lose their mobility and feeding capacity experiencing a knockdown that is called chill coma. When they are brought back to a higher temperature, they progressively recover from chill coma and regain normal activity. A fly is considered as having recovered when it is able to stand on its legs (David *et al.* 1998). The time they take for doing so is called the chill coma recovery time (CCRT). The CCRT is tightly linked to cold tolerance as the recovery from a cold treatment is strongly correlated with the duration of cold treatment (David *et al.* 1998). The test is nonlethal, and the time metric corresponds to a quantitative variable. To remove any influence of breeding conditions, all tested flies were maintained at low densities (Peters & Barbosa 1977). As thermotolerance is highly correlated with maturity (David *et al.* 1998), male and female flies were all tested at the age of 5 days.

For the chill coma recovery test, flies were individually placed into empty 8-ml vials without anaesthesia (Milton & Partridge 2008), fixed in racks and placed in an ice-water bath of 0 °C. After 7 h of exposure, flies were brought back to room temperature (23 ± 1 °C) and CCRT was recorded in minutes (min). Dead individuals were not included into the analysis. They represented less than 1% of the total tested flies and were randomly distributed among lines. As controls, flies from the lines A* and E* were tested at the same time as the XR lines.

Marker selection and genotyping

A previously conducted genome scan of the X chromosome revealed that lines A157 and E14 differ by numerous single nucleotide differences (SNPs) (Glinka *et al.* 2003; Ometto *et al.* 2005). These SNPs are thus reliable markers for the African and European genotypes. A total of 23 SNPs from neutrally evolving fragments that were uniformly distributed along the chromosome were selected to serve as genetic markers for the QTL analysis. To maximize the mapping resolution, the marker

density was adjusted to the recombination rate: marker density was thus increased in regions of higher recombination rate (for further information on markers, see Supplementary Table S1).

By extracting genomic DNA from pools of 20 females with the Puregene DNA isolation kit (Gentra System, Minneapolis, USA), we genotyped each XR line. Custom Taqman SNP genotyping assays from Applied Biosystems (ABI, Foster City, USA) were performed on the DNA samples and ran on an ABI 7500 Real Time thermocycler PCR machine. The presence of either genotype at each marker position was determined with the 7500 Fast Software version 2.0 provided by ABI. To infer marker linkage, we used the frequency of recombination events between marker pairs (RecombRate software, Comeron *et al.* 1999) and applied Haldane's map function to obtain a marker map in centimorgan (cM) (Haldane 1919). Markers show significant segregation distortion (SD) with the European alleles being overrepresented. As this phenomenon affects the whole X chromosome, it is not likely to result from either genotyping errors or linked segregation distortion loci. One possibility that would lead to the chromosome-wide overrepresentation of the European genotype would be the reduced inheritance of African alleles in an early step of XR line establishment. The reciprocal crosses of the A* and E* lines (see Material and Methods, *Drosophila* stocks) might be a crucial point at which such a shift could happen by chance. A sensible reduction in offspring number of males carrying the African X chromosome can reduce the amount of African alleles in the XR line population. The SD is, however, not expected to produce a higher false positive rate, nor will it significantly change the position or the effect of QTL, particularly in additive models of a large XR line population (Xu 2008; Zhang *et al.* 2010).

Quantitative genetic analysis

The statistical analysis was performed using the R software (version 2.8.0). To determine the differences in cold resistance between the parental lines A* and E*, we log-transformed the CCRT measurements to improve normality of the residuals (ln(CCRT)). A two-way ANOVA was used to infer the effects of line (A* or E*), sex (male or female) and their interactions.

The data obtained from the XR line population were analysed using different statistical approaches. We used a linear regression model to reveal the genotype-by-sex interactions (GSI). According to the model $y = \mu + a_L + b_S + \gamma(a_L \times b_S) + \varepsilon_{L,S,i}$, we quantified the influence of the line a_L, the sex b_S and their interactions γ on the variance of the ln(CCRT) y. The parameter μ is the population mean; the residuals $\varepsilon_{L,S,i}$ are assumed to follow a normal distribution with constant variance. Each XR line represents a different multi-locus genotype for the X chromosome, and within-line phenotypic variation can exclusively be traced back to trait plasticity as there is no within-line genetic variation. We estimated the sex-specific X-chromosomal heritability in a broad sense: $H_X^2 = \sigma_L^2/(\sigma_L^2 + \sigma_E^2)$, where σ_L^2 is the between-line variance and σ_E^2 is the within-line variance for ln(CCRT). Another measurement of GSI is the cross-sex genetic correlation coefficient of mean ln(CCRT) of males and females between the XR lines $r_{GS} = cov_{MF}/(\sigma_M \times \sigma_F)$.

Interval mapping

To investigate the QTL positions, their contributions to variation among lines and to estimate their effect, we used the Windows QTL cartographer version 2.5 (Wang *et al.* 2007) and R/qtl (Broman *et al.* 2003). The input for the analysis is the line-dependent mean values of ln(CCRT) and the vector of marker genotypes (0 = African genotype, 1 = European genotype). Within an interval between two markers, we used a fixed number of grid positions with a density of 1 cM to estimate the probabilities of the underlying genotypes conditioned on the observed marker data using the Haley–Knott regression (Haley & Knott 1992). The results are expressed as LOD-score (Sen and Churchill 2001). This score is plotted at every 1 cM grid position of the X-chromosome to give a likelihood profile. A QTL is more likely to be located at the region with high LOD values. We performed a hierarchical interval mapping for males and females separately, starting with a one-QTL analysis, which was then expanded to a more advanced two-QTL analysis and the composite interval mapping (CIM) to handle the nature of more complex QTL. The number of marker cofactors for the CIM should not exceed $2\sqrt{n}$ where n is the number of available markers (Jansen & Stam 1994). Thus, we used up to 10 putative cofactors performing a forward and backward selection of the CIM model with a window size of 10 cM. Each significant threshold was established by randomly permuting the trait values among the marker genotypes 1000 times. The empirical critical LOD value corresponding to $\alpha = 0.05$ was determined according to the 5% cut-off of the distribution of maximal LOD values obtained by each permutation. To test whether there is evidence for a QTL and whether the two-QTL analysis is a significant improvement over a one-QTL analysis, the LOD score had to exceed its threshold.

Multiple-QTL model

The multiple-QTL model was used to regress over the main candidate QTL positions to identify the most

important elements that shape the genetic architecture of cold tolerance on the X chromosome of *D. melanogaster*. As there is a continuum of potential covariates, we used the knowledge of the interval mapping to pick candidate QTL regions for model fitting. Again, the likelihood ratio test statistic was used to infer the model fit expressed in LOD scores. Model search was conducted by backward elimination: we started with the largest model comprising all candidate loci that were found in both the male and female data sets and stepwise removed the covariate that had the smallest LOD value. Hence, the starting point of the analysis was identical for males and females. By backward selection, we constructed a nested sequence of models of decreasing size all the way to the null model. For model selection, we chose the one with the largest model LOD score. By including pairwise interactions, we enforced a hierarchy under which the inclusion of the interaction of certain markers requires the inclusion of main effects for each marker individually.

Mapping of QTL–sex interactions

Loci that influence the cold tolerance differently for males and females underlie QTL–sex interactions. These interactions might preserve genetic variation within the population as males and females lack a consistent response to natural selection. To localize such regions, we used the average difference of ln(CCRT) between males and females of each XR line as the phenotype in a QTL analysis. If there is no variation in the differences between males and females, there can be no QTL–sex interactions, but if the differences vary between the XR lines, we might find associations of this variation and a chromosomal region. Again, a significance threshold was derived by 1000 permutations.

Analysis of a selective sweep region

Of the selective sweep map of the X chromosome inferred by Li & Stephan (2006) that consisted of 54 and 55 putative 100-kb long fragments in the African and European populations, respectively, where sweeps were identified, we investigated one region in detail (window 55 of the European population, cytological position 15E3). This region was chosen because it fulfils necessary conditions for the occurrence of a (complete) sweep in a local population influencing a quantitative trait: (i) the sweep colocalizes with a QTL, namely the QTL at position 56 cM (cytological position 13E-20E, Fig. 3), which is significant in both males and females and is lacking QTL–sex interactions; (ii) the sweep is specific to a population. Because the sweep occurred in the European, but not in the African population, allelic differences at the gene(s) affecting the trait may therefore be found. In contrast, if a recent sweep was shared between these populations (such as for the *D. melanogaster polyhomeotic* locus, Beisswanger & Stephan 2008), the gene targeted by positive selection would not contribute to the quantitative trait.

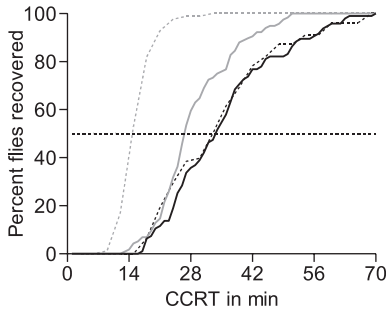


Fig. 1 Chill coma recovery curves of the parental lines (A* in black, E* in grey) separated for males (doted lines) and females (solid lines). Both males and females of the European line recovered significantly earlier from the chill coma than those of the African line (Welch's *t*-test, $P < 0.001$). The black dotted horizontal line shows the time at which 50% of the line-specific flies recovered.

While the analysis by Li & Stephan (2006) was carried out using the site frequency spectrum of SNPs averaged over a limited number of 500-bp long fragments (of at least three in a 100-kb window), we increased the amount of sequence data in window 55 in two steps as described by Svetec *et al.* (2009): first, 12 additional fragments of 500-bp length were PCR-amplified and re-sequenced in both the African and European samples used by Li & Stephan (2006) (11 lines for the African sample and 12 for the European one), and second, a region of 6.4 kb (between absolute positions 16 992 569 and 16 998 925; release 5.29 of Flybase, http://flybase.org) was completely re-sequenced in these 23 lines. This fine-scale analysis allowed us to determine the target of selection very precisely (i.e. down to the level of individual genes). The new sequence data have been deposited in GenBank (accession nos HQ005309 to HQ005331).

This data set was then subjected to an analysis of linkage disequilibrium (LD) using the ω statistic (Kim & Nielsen 2004). Elevated values of ω provide evidence of a selective sweep, and the peak of this statistic (ω_{MAX}) indicates the location of the target of selection in the genome (Pavlidis *et al.* 2010). Positions containing insertions or deletions (indels) were excluded.

The study by Li & Stephan (2006) provided evidence for recent positive selection in this genomic region based on the site frequency spectrum. In this study, we confirm their results using a LD-based statistic. In the statistical hypothesis testing, we did not use the ascertainment bias correction of Thornton & Jensen (2007). Their correction is applied when a genomic region is chosen for selective

sweep analysis based on *a priori* information (e.g. reduced polymorphism levels). In the current analysis, we already know that this genomic region deviates from neutrality (Li & Stephan 2006). A low *P*-value for the ω_{MAX} is therefore expected because there is evidence that the genomic region is an outlier when the null (neutral) hypothesis is assumed to be correct.

To assess the statistical significance of the maximum value ω_{MAX}, we ran 10 000 neutral simulations with the *ms* software (Hudson 2002). The demographic scenario of the African and European populations of *D. melanogaster* (Li & Stephan 2006) represented the null hypothesis. The mutation rate (1.47×10^{-9}) was estimated from the observed number of polymorphisms in the African population using the method of Zivkovic & Wiehe (2008). Thus, the African population was used as a proxy for selective neutrality. The recombination rate (3.6×10^{-8}) was obtained from the *D. melanogaster* recombination rate calculator (Fiston-Lavier *et al.* 2010). Only the European subset of each simulation was used to assess the significance of ω_{MAX}.

Results

Thermotolerance in the parental lines

The CCRT of the original lines A157 and E14 differed significantly; i.e. A157 males: 62.1 ± 1.3 (SE) min, females: 66.4 ± 3.8 (SE) min; E14 males: 28.6 ± 0.5 (SE), females: 23.5 ± 1.3 (SE). The CCRT of the parental lines A* and E* were also significantly different: 39.4 ± 2.3 (SE) and 35.5 ± 1.3 (SE) min for males and females with the African X chromosome, respectively, and 27.5 ± 1.0 (SE) and 29.1 ± 0.8 (SE) min for males and females with the European X chromosome (Fig. 1, Table 1, two-way ANOVA for log-transformed data). However, there was no significant variation in CCRT for sexes among the parental lines as well as no detectable line–sex interaction (Table 1).

Thermotolerance in the XR lines

Mean values ranged from 22.9 ± 1.3 (SE) min to 42.2 ± 3.5 (SE) min for males and from 22.7 ± 1.2 (SE) min to 46.7 ± 4.1 (SE) min for the females. The results of the two-way ANOVA of the XR lines revealed highly significant line-specific variation for ln(CCRT) (Table 2). The main effect of sex is only slightly significant as it is not constant across lines; the line–sex interaction is highly significant (Table 2). Thus, genotype-by-sex interactions are present in our XR line population. The increase in mean ln(CCRT) between the XR lines is smooth, suggesting multiple QTL with small to intermediate effects (Fig. 2a,b). In females, one XR line significantly exceeded the cold tolerance of the parental E* line (Welch's *t*-test and Holm–Bonferroni correction, $P < 0.05$), showing a transgressive phenotype.

Table 1 Results of the two-way ANOVA of log-transformed chill coma recovery time of the parental lines A* and E*

	d.f.	Sum Sq	Mean Sq	*F* value	Pr(>*F*)
Line	1	3.0128	3.0128	31.7842	4.511e–08***
Sex	1	0.0936	0.0936	0.9879	0.3212 ns
Line × sex	1	0.2334	0.2334	2.4626	0.1178 ns
Residuals	258	24.4556	0.0948		

ns, not significant; ***, significant with $P < 0.001$.

Table 2 Results of the ANOVA of log-transformed chill coma recovery time of X-chromosomal recombinant lines

	d.f.	Sum Sq	Mean Sq	*F* value	Pr(>*F*)
Line	189	64.64	0.34200	3.9779	2.2e–16***
Sex	1	0.53	0.53241	6.1925	0.01286*
Line × sex	182	25.32	0.13910	1.6179	4.411e–07***
Residuals	5867	504.42	0.08598		

* significant with $P < 0.05$, *** $P < 0.001$

Sex-specific X-chromosomal heritability in the broad sense, H^2_X, varied between sexes ranging from 0.2 for males to 0.13 for females. The reduced linear regression model among lines within sexes was highly significant for males and females, jointly indicating an X-linked genetic basis for cold tolerance in both sexes ($P < 0.001$). The coefficient of cross-sex genetic correlation r_{GS} was low (0.35) but highly significant (Pearson's product–moment correlation, $P < 0.001$), displaying a weak correlated response of both sexes under selection. These estimates support the hypothesis that X-linked cold tolerance is low but present, and a mixture of commonly inherited and sex-specific loci affects males and females of *D. melanogaster*.

One-QTL and two-QTL analyses

The one-QTL interval mapping revealed one significant QTL position for cold tolerance on the X chromosome of *D. melanogaster* for both males and females. Although the LOD curve of males shows at least three peaks, only one exceeds the threshold of the 5% significance level (Fig. 3a, dotted line). This QTL position is located towards the right tip of the chromosome close to the centromere between the last two markers at 56 cM. Thus, for the one-QTL analysis, this is the strongest candidate in males. The LOD curve of females shows a double peak with the highest LOD value at position 22 cM between markers six and eleven

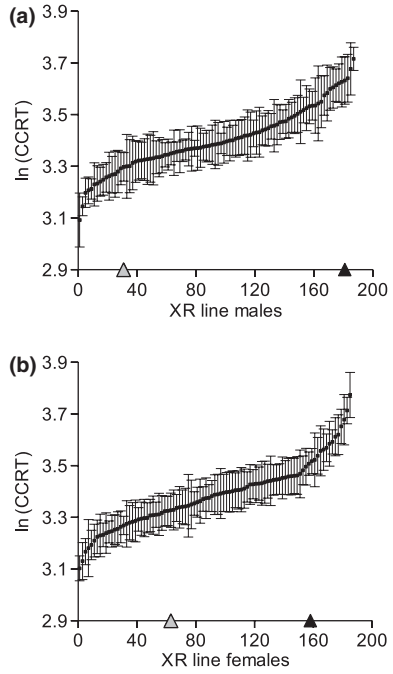


Fig. 2 Mean ln[chill coma recovery time (CCRT)] in males (a) and females (b) of the X-chromosomal recombinant (XR) lines of *Drosophila melanogaster*. The smooth increase indicates multiple quantitative trait loci (QTL) with small to intermediate effects. Paternal cold tolerance (E*, grey triangle on the *x*-axis) as well as cold sensitivity (A*, black triangle) was not significantly higher in the males of the XR lines, whereas the females of one XR line showed a significantly shorter CCRT than the E* line (Welch's *t*-test and Holm–Bonferroni correction, $P < 0.05$).

(Fig. 3b, dotted line). The second peak at around 9 cM also exceeds the threshold but cannot be reliably decoupled from the main candidate at 22 cM. The thresholds of 2.5 LOD were established by 1000 permutations for males and females separately.

The two-QTL analysis revealed insight into a more complex QTL structure. The first model extension comprehended the possibility for a second QTL with additive effects (additive model, H_{1a}), while the second model extension allowed for a second QTL with epistatic effects (full model, H_{1f}). For males of the XR line population, the additive and the full models reach their maxima at the same positions, 0 cM and 56 cM, showing very highly significant LOD scores for both models (additive model: lod.add = 7.45; full model: lod.full = 8.32). By comparing the two-QTL analysis to the one-QTL analysis, we exclusively quantified the support of the second QTL position. Under the additive model as well as under the full model, there is significant evidence for a second QTL in the male data set, concerning the two candidate regions at 0 cM and 56 cM (lod.av1 = 3.09, lod.fv1 = 3.97). There is no clear hint for an interaction between them (lod.int = 0.87).

The females of the XR line population show similar consistent results for the two-QTL genome scan as the males. The additive and the full models maximize at analogous positions (additive model positions at 8 cM and 22 cM, full model positions at 10 cM and 23 cM), thus overlapping with the maxima of the one-QTL analysis. The scores are high, giving strong support for at least one QTL at either of these positions (lod.add = 7.17, lod.full = 8.23). The evidence for the second QTL is congruent in both models. The implementations of a second QTL in the additive model and the full model with the interaction term exceed the thresholds (lod.av1 = 1.85, lod.fv1 = 2.90), giving significant support for the second QTL position at 22–23 cM.

Composite interval mapping

For the males of the XR line population, four cofactors were included in the CIM model (marker 1, 3, 7 and 23, Fig. 3a, black triangles), leading to one more candidate region that exceeded the threshold at position 18 cM (Fig. 3a, solid line). The formerly detected QTL positions at 0 cM and 56 cM remained significant. For females of the XR line population, the QTL positions remained unchanged at around 9 cM and 23 cM although the candidate region slightly increased. Overall, three markers were included as cofactors (Fig. 3b, black triangles, marker 4, 10 and 23).

Multiple-QTL model

The most significant QTL positions defined by the two-QTL analysis and CIM entered the multiple linear regression analysis to estimate the main QTL and their interaction effects. For both males and females, an additive model with interactions was favoured to summarize the main QTL and their effects. After the backward elimination procedure, overall four main effects and two interactions remained important in the multiple-QTL model for the males of the XR line population. In descending order, these are position 56 cM (LOD = 6.64), 0 cM (LOD = 4.82), 18 cM (LOD = 4.08), 24 cM (LOD = 2.66), interaction 24 × 18 cM (LOD = 2.17) and 0 × 56 cM (LOD = 1.16). The general model fit is good (LOD = 12.40) and explains around 26% of the trait variance (Table 3). For the females of the XR line population, overall four main effects and one interaction remained important in the model fit after the elimination procedure. These are position 24 cM (LOD = 6.95), 17 cM (LOD = 4.31), 9 cM (LOD = 2.99), 56 cM (LOD = 2.68) and the interaction 24 × 17 cM (LOD = 4.18). The general model fit is good (LOD = 12.89) and explains around 28% of the trait variance (Table 4).

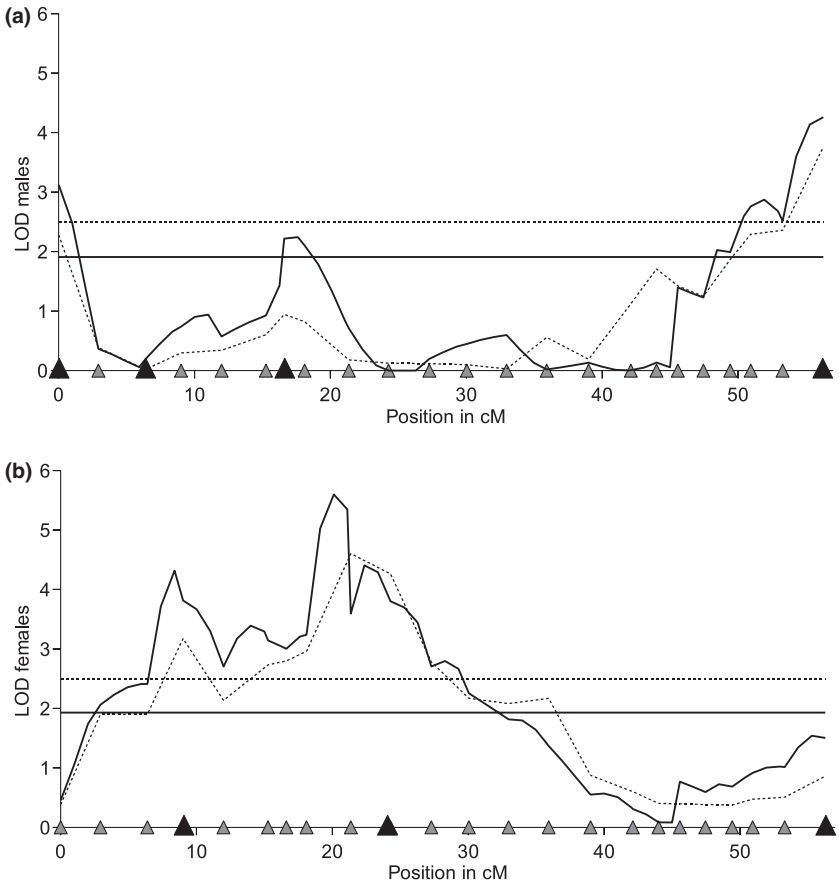


Fig. 3 Results of interval-mapping methods for ln[chill coma recovery time (CCRT)] in males (a) and females (b) of *Drosophila melanogaster*. The one-quantitative trait loci (QTL) analysis (dotted lines) and the composite interval mapping (CIM) (solid lines) for males and females give consistent results, but the CIM better resolved the genetic architecture of multiple QTL for cold tolerance on the X chromosome. The horizontal lines correspond to the threshold values (dotted line for the one-QTL model and the solid one for the CIM). The distribution of the 23 genetic markers used is shown on the *x*-axis by triangles, where larger black triangles indicate the locations of QTL cofactors. Each position at which the LOD score exceeds the threshold value is a genomic region showing significant association with CCRT—i.e. a QTL.

Overall, males and females of *D. melanogaster* show similar genetic architecture for cold tolerance on the X chromosome. They share most of the detected QTL, such as positions 17 cM, 18 cM, 24 cM and 56 cM, although it needed a coherent multiple-QTL model to show this. Two QTL, however, remained unique to one of both data sets: position 0 cM to males and position 9 cM to females. The estimated additive effects have mainly a negative sign, which indicates that the African alleles generally reduce cold tolerance. Three of the studied QTL positions, however, have a positive sign: the QTL at position 24 cM as well as the interaction 0×56 cM in males and the QTL at position 17 cM in females. Whereas the effect of the QTL at position 17 cM in females is low, the effect of the QTL at position 24 cM in males is a transgressive QTL. Males benefit from the African allele to increase their cold tolerance. The interaction effect 0×56 cM is based on the product of the genotypes at position 0 cM and 56 cM. The effect of position 0 cM depends on whether a male fly has the African or European allele at position 56 cM. Thus, the positive sign of the interaction effect indicates an increased cold tolerance when both loci exhibit the African genotype, although the QTL at positions 0 cM and 56 cM individually show negative effects.

QTL–sex interactions

Consistent with the highly significant genotype-by-sex interactions of the XR line population, the differences between males and females show considerable variation. The result of the one-QTL analysis shows a strong association between the differences in cold tolerance between sexes and the position at 24 cM (Fig. 4). The LOD value of 3.83 clearly exceeds the threshold of 1.79. At this position, females substantially increase their cold tolerance by the European allele (Welch's *t*-test,

Table 3 Multiple-QTL model for males of the XR line population of *Drosophila melanogaster*

Model formula: $y = Q1 + Q3 + Q5 + Q2 + Q2{:}Q5 + Q3{:}Q1$

	d.f.	SS	MS	LOD	%Var
Model	6	0.70	0.12	12.40	26.31
Error	180	1.95	0.01		
Total	186	2.65			

Drop one QTL at a time ANOVA table

	Cytologic position	d.f.	Type III SS	LOD	%Var	Estimated allele effect (SE)
Intercept						3.51 ± 0.02
X@56	13E-20E	2	0.35	6.64	13.09	−0.16 ± 0.03
X@0	1A-3F	2	0.25	4.82	9.30	−0.13 ± 0.03
X@18	7B-8E	2	0.21	4.08	7.79	−0.06 ± 0.03
X@24	8E-11D	2	0.13	2.66	4.99	0.06 ± 0.03
X@24:X@18		1	0.11	2.17	4.04	−0.16 ± 0.05
X@0:X@56		1	0.06	1.16	2.13	0.13 ± 0.06

QTL, quantitative trait loci; XR, X-chromosomal recombinant.

$P < 0.001$), but in males, both genotypes perform equally well (Welch's *t*-test ns, Fig. 5). The difference of genotype-based cold tolerance of males and females assumes its maximum at position 24 cM and gives reliable evidence for QTL–sex interactions on the X chromosome.

A selective sweep associated with CCRT colocalizes with a QTL

The profile of the ω statistic in the fully re-sequenced region is shown in Fig. 6. The dashed line depicts the 95th percentile of ω_{MAX} obtained by the neutral simulations (see Materials and Methods). The *P*-value for the observed ω_{MAX} is 0.037. The maximum of the ω distribution is located in the intergenic region between the computer-predicted genes *CG16700* and *CG4991*. In a very recent study, the expression variation observed at *CG16700* has been associated with CCRT in another derived population (from North America; Ayroles *et al.* 2009). This suggests that *CG16700* (but not *CG4991*) is a candidate gene of cold tolerance.

Discussion

Rationale

Numerous QTL studies used artificially selected parental lines (Tucić 1979; Heyen *et al.* 1999; Tuiskula-Haav-

Table 4 Multiple QTL model for females of the XR line population of *Drosophila melanogaster*

Model formula: $y = Q2 + Q6 + Q4 + Q1 + Q2{:}Q6$

	d.f.	SS	MS	LOD	%Var
Model	5	0.77	0.15	12.89	27.58
Error	178	2.02	0.01		
Total	183	2.79			

Drop one QTL at a time ANOVA table

	Cytologic position	d.f.	Type III SS	LOD	%Var	Estimated allele effect (SE)
Intercept						3.52 ± 0.02
X@24	8E-11D	2	0.38	6.95	13.75	−0.05 ± 0.03
X@17	6C-10B	2	0.23	4.31	8.25	0.02 ± 0.03
X@9	4C14-5C7	1	0.16	2.99	5.63	−0.11 ± 0.03
X@56	13E-20E	1	0.14	2.68	5.02	−0.08 ± 0.02
X@24:X@17		1	0.22	4.18	7.96	−0.25 ± 0.06

QTL, quantitative trait loci; XR, X-chromosomal recombinant.

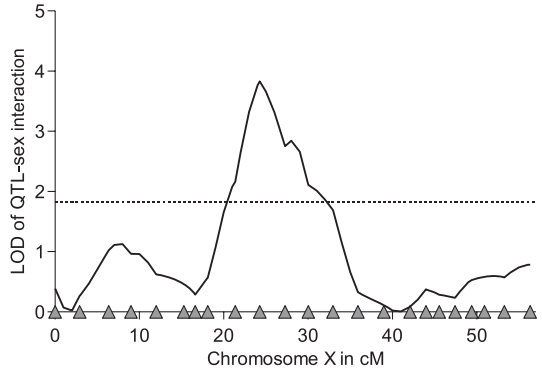


Fig. 4 Localization of quantitative trait loci (QTL)–sex interactions on the X chromosome of *Drosophila melanogaster*. The one-QTL analysis was sufficient to reveal a peak at 24.3 cM, which exceeds the 5% threshold of LOD = 1.79 (dotted line).

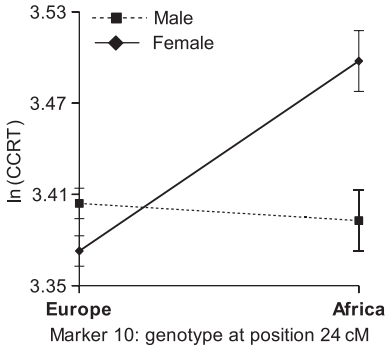


Fig. 5 ln[chill coma recovery time (CCRT)] results in relation to the X-chromosomal recombinant (XR) line genotype at marker position 24 cM for males and females. While females significantly decrease their CCRT when having the European genotype at this marker position (Welch's *t*-test, $P < 0.001$), males tend to increase their CCRT. However, the difference in CCRT between the males with European and African genotype was not significant. Such contrasting patterns for males and females are called quantitative trait loci (QTL)–sex interactions.

isto *et al*. 2002; Norry *et al*. 2004, 2007; Morgan & Mackay 2006) to maximize the initial genetic differences. Strongly divergent genetic effects increase the chances to detect significant QTL. With the recent advances in statistics and molecular biology, it is possible to detect QTL that are based on much lower levels of phenotypic variation. We thus decided to investigate naturally segregating variance for cold tolerance occurring between two wild lines of *D. melanogaster* (A157 and E14) that were sampled in sub-Saharan Africa and Western Europe. These two lines exhibit a divergent cold tolerance phenotype: line A157 is less tolerant to cold than line E14. This is an expected result as previous studies showed that cold tolerance follows a geographical pattern and latitudinal clines (Bubliy *et al*. 2002; Hoffmann *et al*. 2002).

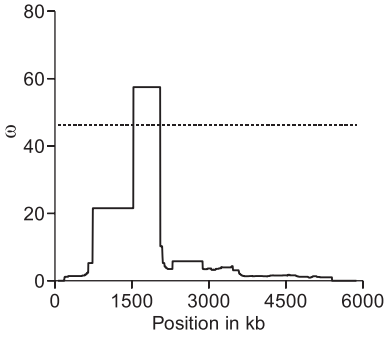


Fig. 6 Profile of the ω statistic across the completely re-sequenced region of the European population. This 6.4-kb region consists of the intergenic fragment between the genes *CG16700* and *CG4991*. The region shown on the *x*-axis is shorter than 6.4 kb because the genomic segment at the right end, between positions 5921 and 6379, was not included in the analysis because of smaller sample size (missing data). These positions correspond to the 3'UTR of *CG4991*. Likewise, positions containing indels were excluded. The dotted line represents the 5% significance level.

The presence of cold tolerance factors on the X chromosome of *D. melanogaster* is demonstrated by the fact that the two substitution lines A* and E* exhibit divergent cold tolerance phenotypes. To map these factors on the X chromosome, an XR line population was constructed and studied. The broad sense heritability of cold tolerance estimated for the XR line population confirmed that 13–20% of the phenotypic variance has a genetic basis on the X chromosome. We thus used a QTL mapping approach to locate the X-linked factors influencing cold tolerance. Furthermore, using selective sweep analysis, we fine-mapped the genomic region colocalizing with one of the QTL.

QTL analysis

As we detected a significant effect of sex as well as line–sex interactions on cold tolerance, we decided to treat the male and female data sets separately to study sex antagonistic effects, which are referred to as QTL–sex interactions. Splitting the data set decreased the sample size and could thus have reduced the power to detect the cold tolerance factors that are shared between sexes. However, our results clearly show that each of the two data sets remained sufficiently large to detect significant QTL effects. Moreover, the high number of scored individuals (about 10 individuals per XR line), the high marker density (23 along the X chromosome) and the use of a homogenized genetic background certainly raised our chance to detect QTL.

The basic one-QTL analysis was not sufficient to explain the trait variance in our XR line population.

Therefore, we extended the one-QTL analysis and considered multiple QTL to distribute residual variance to several loci with putative low to intermediate effects. The two-QTL analysis allows for the possibility of two QTL positions, while the CIM method relaxes the limitation of two QTL using a number of marker cofactors. The selection of cofactors, however, is critical and somehow arbitrary, which can result in conspicuous randomness in the choice of cofactors and misleading conclusions of QTL positions. Using the candidates of the two-QTL analysis and the CIM, comprising QTL positions, cofactors and interactions, we constructed a coherent model for both sexes, based on a backward selection process to quantify the impact of main factors as well as interactions.

Overall, we found six QTL that were significantly associated with cold tolerance on the X chromosome, explaining about 26–28% of the total phenotypic variance observed in males and females. Consistently, the distribution of mean CCRT in the XR line population (Fig. 2) does not show a biphasic response as expected in case of one major QTL. Indeed, each of the detected QTL accounts for 5–14% of the total phenotypic variation. This also explains the failure of the one-QTL analysis as it lacks the possibility to sufficiently distribute the residual variance of the trait and dissolve its complexity. It is important to notice that on average, the QTL effects are likely to be slightly overestimated because the XR lines have a homogenous autosomal background. In other words, in the present study, we focused on phenotypic variation in cold tolerance generated by naturally occurring alleles of the X chromosome only.

It is a common finding that genes contribute differently to complex traits in males and females. Among the six QTL, two were exclusively detected in either males (0 cM) or females (9 cM), two are partially overlapping (17 cM and 18 cM), and two were found at the same position in males and females (24 cM, 56 cM). But sex specificity has to be taken with caution. Detecting a QTL in one sex and not in the other might be attributable to a lack of power in the analysis. Precise investigations into the sex-specific effects of the different alleles must be performed to assess the sex-specific role of any of these QTL. The QTL positions 17 cM and 18 cM do not perfectly colocalize but partially overlap and are both involved in interactions with the QTL position at 24 cM. This could suggest that female position 17 cM and male position 18 cM are confounded. Only further fine-mapping experiments would confirm whether the same sets of genes contribute to cold tolerance in both males and females.

One CIM cofactor was found in the male data set at position 6 cM, but it does not appear as main effect in the multiple-QTL analysis, neither in males nor in females. Thus, it is the sole cofactor whose effect in the CIM could not be verified with the multiple-QTL model. However, one has to keep in mind that a classical drawback of association mapping studies also applies to the present study: we simply might miss some additional QTL because their effects are below the detection level of the interval mapping.

In Fig. 2, the ranges of phenotypic values of the parental lines and XR lines are different for the two sexes. More precisely, the parental values are closer, and the XR line values are more spread out for the females than for the males. In other words, transgressive segregation (i.e. the existence of hybrid offspring with phenotypes exceeding the E* line) is present when looking at females. One female line significantly exceeds the parental cold tolerance. As dominance is excluded (the XR lines are homozygotes), additive effects between loci are the most likely explanation for the female transgressive segregation observed in Fig. 2.

Interactions between QTL

We detected several QTL that showed significant levels of interactions (in males: 0 cM × 56 cM and 18 cM × 24 cM, in females 17 cM × 24 cM). Interactions, also called epistasis, are referred to as departure from additivity, mainly multiplicativity of allelic effects. QTL mapping allows one to detect such interactions, but it does not provide any information about the nature of these interactions: protein–protein (for example, cofactors), DNA–protein (like gene expression regulation) or physiological pathway interactions (for example, enzyme cascades in a metabolic pathway).

Generally, interactions might not necessarily be the consequence of gene action but reflect the way how selection operates at these loci (Crow 2010). Genes with small additive effects might accumulate slightly deleterious mutations until a certain threshold. Individuals that reached the threshold will be subject to truncation selection and are eliminated from the population. This means deleterious mutations in multiple QTL are jointly removed, creating the signal of interactions. Crow (2010) called this 'quasi-epistasis' that is attributable to selection's grouping of alleles with similar effect.

QTL–sex interactions

The QTL at position 24 cM exhibits two interesting features: it interacts with both QTL positions at 17 cM and 18 cM, but it also generates a QTL–sex interaction. We thus included sex differentiation factors as candidate genes for this region only (Supplementary Table S2).

Moreover, for the QTL at position 24 cM (marker 10), the female cold tolerance pattern is consistent with what is observed all along the chromosome: the European allele confers a significantly higher cold tolerance than the African allele. Males, however, exhibit a transgressive QTL and reduce their CCRT by having the African allele. This finding agrees with the highly significant but low cross-sex genetic correlation coefficient of males and females, indicating a weak correlated response of both sexes under selection. Sex antagonistic selection maintains genetic variation within a population and is thus not suitable for a selective sweep analysis in the context of cold tolerance. Although the coherent multiple-QTL model revealed similar architectures for males and females, there are some noticeable differences, one of which can be quantified by means of a one-QTL analysis (Fig. 4).

Several scenarios could explain that cold tolerance genes are present in the African population. First, in the past, the African population might have experienced colder periods and, because of its large population size, some cold tolerance alleles could still segregate. A second explanation would be that some positively selected traits in Africa, such as desiccation resistance or heat tolerance, could confer higher cold tolerance. However, the reason why the African QTL allele at position 24 cM generates such strong divergent effects on males and females requires more investigations.

Candidate genes and selective sweep mapping

Thermal adaptation in *Drosophila* is a complex topic that comprises a variety of physiological processes (whether considering a short-term or long-term treatment, Loeschcke & Sørensen 2005). In the case of chill coma recovery, adaptation might occur by a rapid cold response triggered by multiple genes. Besides the direct temperature resistance, further mechanisms such as energy allocation shifts and metabolic regulation could increase cold tolerance (see Supplementary Data S1).

Thus far, the X chromosome was thought to be rather poor in cold tolerance factors as all major genes (*ppk*, *trap1*, *Fst*, *Sas*, *desat2*, *Catsup*, *Ddc*, *nompA*, *hsp 70*, *Dca* and *hsr-omega*) were mapped on the autosomes (Goto 2000, 2001; Morgan & Mackay 2006; Norry *et al.* 2007, 2008; Sinclair *et al.* 2007). Recent expression analyses, however, produced many candidates on the X chromosome that are associated with the cold tolerance phenotype measured by cold shock treatments. Telonis-Scott *et al.* (2009) created lines of *D. melanogaster* that were artificially selected for cold resistance by breeding individuals with elevated chill coma recovery abilities (protocol described in Anderson *et al.* 2005). By comparing gene expression levels of selected and nonselected lines, they found ten genes on the X chromosome that significantly altered their expression level, eight of which are located within our QTL. Another expression study (Ayroles *et al.* 2009) revealed 144 genes with expression levels correlated with the ability to recover from chill coma. Out of this set, 91 genes are located within our QTL.

Nonetheless, it remains challenging to narrow down the list of candidate genes that are promising for further study. Here, we used selective sweep analysis for fine-mapping the genomic region colocalizing with a QTL to search for suitable candidate genes. Selective sweep mapping under the QTL most proximal to the centromere revealed evidence for recent positive selection in the European population. The target of selection was located in the intergenic region between two adjacent genes, *CG16700* and *CG4991* (Fig. 6). In combination with the recent results of Ayroles *et al.* (2009) that showed an association of *CG16700* with CCRT in another derived population (from North America), our sweep analysis suggests that *CG16700* is a candidate gene for cold tolerance. Functional investigations, however, are needed to validate this result.

This example shows that fine mapping based on selective sweeps may provide a powerful approach to identify candidate genes that colocalize with QTL. A similar strategy can be used in cases in which other fine-scale analyses of QTL, such as deletion mapping, are not applicable (Rubin *et al.* 2010).

Additional genes on the X chromosome that potentially affect CCRT and colocalize with the identified QTL have been obtained in database searches (see Supplementary Table S2).

Acknowledgements

We thank David De Lorenzo for discussion, Aurélie Bonin and her team of reviewers for valuable suggestions on previous versions of the manuscript, and Anne Wilken and Susanne Voigt for excellent technical assistance. This research was funded by grant STE 325/12 from the DFG Research Unit 1078 to WS, National Institutes of Health grant GM074244 to KWB, and doctoral fellowships from the Volkswagen-Foundation to PP and RW. S.

References

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

Ayroles JF, Carbone MA, Stone EA *et al.* (2009) Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics*, **41**, 299–307.

Azevedo RBR, Vernon F, Partridge L (1996) Thermal evolution of egg size in *Drosophila melanogaster*. *Evolution*, **50**, 2338–2345.

Baudry E, Viginier B, Veuille M (2004) Non-African populations of *Drosophila melanogaster* have a unique origin. *Molecular Biology and Evolution*, **21**, 1482–1491.

Begun DJ, Aquadro CF (1993) African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature*, **365**, 548–550.

Beisswanger S, Stephan W (2008) Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated *polyhomeotic* genes in *Drosophila*. *Proceedings of the National Academy of Sciences, USA*, **105**, 5447–5452.

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

Bersaglieri T, Sabeti PC, Patterson N *et al.* (2004) Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, **74**, 1111–1120.

Borevitz JO, Hazen SP, Michael TP *et al.* (2007) Genome-wide patterns of single-feature polymorphism in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA*, **104**, 12057–12062.

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**, 889–890.

Bubliy OA, Riihimaa A, Norry FM, Loeschcke V (2002) Variation in resistance and acclimation to low-temperature stress among three geographical strains of *Drosophila melanogaster*. *Journal of Thermal Biology*, **27**, 337–344.

Cariou ML (1987) Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D. sechellia* and *D. orena*. *Genetical Research*, **50**, 181–185.

Comeron JM, Kreitman M, Aguadé M (1999) Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics*, **151**, 239–249.

Crow JF (2010) On epistasis: why it is unimportant in polygenic directional selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 1241–1244.

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

David J, Clavel MF (1967) Influence of temperature during the course of development on various biometric characteristics of adult *Drosophila melanogaster* Meigen. *Journal of Insect Physiology*, **13**, 717–729.

David JR, Gibert P, Pla E *et al.* (1998) Cold stress tolerance in *Drosophila*: analysis of chill coma recovery in *D. melanogaster*. *Journal of Thermal Biology*, **23**, 291–299.

Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA (2010) *Drosophila melanogaster* recombination rate calculator. *Gene*, **463**, 18–20.

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

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

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

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

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. *Journal of Genetics*, **8**, 299–309.

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315–324.

Heyen DW, Weller JI, Ron M *et al.* (1999) A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiological Genomics*, **1**, 165–175.

Hinds DA, Stuve LL, Nilsen GB *et al.* (2005) Whole-genome patterns of common DNA variation in three human populations. *Science*, **307**, 1072–1079.

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

Hoffmann AA, Sørensen JG, Loeschcke V (2003) Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *Journal of Thermal Biology*, **28**, 175–216.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W (2007) Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. *Genetics*, **177**, 469–480.

Izquierdo JI (1991) How does *Drosophila melanogaster* overwinter? *Entomologia Experimentalis et Applicata*, **59**, 51–58.

James AC, Partridge L (1995) Thermal evolution of rate of larval development in *Drosophila melanogaster* in laboratory and field populations. *Journal of Evolutionary Biology*, **8**, 315–330.

James AC, Azevedo RB, Partridge L (1997) Genetic and environmental responses to temperature of *Drosophila melanogaster* from a latitudinal cline. *Genetics*, **146**, 881–890.

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, **136**, 1447–1455.

Kane NC, Rieseberg LH (2008) Genetics and evolution of weedy *Helianthus annuus* populations: adaptation of an agricultural weed. *Molecular Ecology*, **17**, 384–394.

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

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

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

Loeschcke V, Sørensen JG (2005) Acclimation, heat shock and hardening—a response from evolutionary biology. *Journal of Thermal Biology*, **30**, 255–257.

Long AD, Mullaney SL, Reid LA *et al.* (1995) High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics*, **139**, 1273–1291.

Mackay TFC (2001) Quantitative trait loci in *Drosophila*. *Nature Reviews Genetics*, **2**, 11–20.

Mckenzie J (1975) The influence of low temperature on survival and reproduction in populations of *Drosophila melanogaster*. *Australian Journal of Zoology*, **23**, 237–247.

Milton CC, Partridge L (2008) Brief carbon dioxide exposure blocks heat hardening but not cold acclimation in *Drosophila melanogaster*. *Journal of Insect Physiology*, **54**, 32–40.

Mitrovski P, Hoffmann AA (2001) Postponed reproduction as an adaptation to winter conditions in *Drosophila melanogaster*: evidence for clinal variation under semi-natural conditions. *Proceedings of the Royal Society in London. Series B, Biological Sciences*, **268**, 2163–2168.

Moehring AJ, Mackay TFC (2004) The quantitative genetic basis of male mating behavior in *Drosophila melanogaster*. *Genetics*, **167**, 1249–1263.

Morgan TJ, Mackay TFC (2006) Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity*, **96**, 232–242.

Nordborg M, Hu TT, Ishino Y *et al.* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology*, **3**, e196.

Norry FM, Dahlgaard J, Loeschcke V (2004) Quantitative trait loci affecting knockdown resistance to high temperature in *Drosophila melanogaster*. *Molecular Ecology*, **13**, 3585–3594.

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

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

Ometto L, Glinka S, De Lorenzo D, Stephan W (2005) Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Molecular Biology and Evolution*, **22**, 2119–2130.

Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. *Proceedings of the National Academy of Sciences, USA*, **101**, 9885–9890.

Pavlidis P, Jensen JD, Stephan W (2010) Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*, **185**, 907–922.

Peters TM, Barbosa P (1977) Influence of population density on size, fecundity, and developmental rate of insects in culture. *Annual Review of Entomology*, **22**, 431–450.

Rubin CJ, Zody MC, Eriksson J *et al.* (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, **464**, 587–591.

Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. *Genetics*, **159**, 371–387.

Sinclair BJ, Gibbs AG, Roberts SP (2007) Gene transcription during exposure to, and recovery from, cold and desiccation stress in *Drosophila melanogaster*. *Insect Molecular Biology*, **16**, 435–443.

Stanley S, Parsons P, Spence G, Weber L (1980) Resistance of species of the *Drosophila melanogaster* subgroup to environmental extremes. *Australian Journal of Zoology*, **28**, 413–421.

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

Svetec N, Pavlidis P, Stephan W (2009) Recent strong positive selection on *Drosophila melanogaster HDAC6*, a gene encoding a stress surveillance factor, as revealed by population genomic analysis. *Molecular Biology and Evolution*, **26**, 1549–1556.

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

Thornton KR, Jensen JD (2007) Controlling the false-positive rate in multilocus genome scans for selection. *Genetics*, **175**, 737–750.

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

Tuiskula-Haavisto M, Honkatukia M, Vilkki J *et al.* (2002) Mapping of quantitative trait loci affecting quality and production traits in egg layers. *Poultry Science*, **81**, 919–927.

Van 't Land J, Putten PV, Zwaan B, Kamping A, Delden WV (1999) Latitudinal variation in wild populations of *Drosophila melanogaster*: heritabilities and reaction norms. *Journal of Evolutionary Biology*, **12**, 222–232.

Wang S, Basten CJ, Zeng Z-B (2007) *Windows QTL Cartographer 2.5*. Department of Statistics, North Carolina State University, Raleigh, NC.

Xu S (2008) Quantitative trait locus mapping can benefit from segregation distortion. *Genetics*, **180**, 2201–2208.

Zhang L, Wang S, Li H *et al.* (2010) Effects of missing marker and segregation distortion on QTL mapping in F2 populations. *Theoretical and Applied Genetics*, **121**, 1071–1082.

Zivkovic D, Wiehe T (2008) Second-order moments of segregating sites under variable population size. *Genetics*, **180**, 341–357.

N.S. has special interest in the behavior, the genetics, and the life history of insects. A.W. works on QTL and selective sweep mapping in model organisms. R.W. works on adaptive consequences of standing genetic variation in natural populations, including humans. P.P. develops algorithms to detect selective sweeps in populations that have experienced past demographic changes. J.C. is focusing on theoretical models of genetic effects for improving quantitative analyses in general and QTL analyses in particular. K.B's research interests include the characterization of meiotic recombination and the development of improved methods and software for QTL mapping. The main focus of D.M.'s research is on the development of model-based methods for the analysis of genetic data. W.S. works on various questions of theoretical and empirical population genetics (in particular adaptation).

Data accessibility:

data deposited at Dryad: doi:10.5061/dryad.8083.

Supporting information

Additional supporting information may be found in the online version of this article.

Data S1 Candidate genes.

Fig. S1 Construction of the substitution lines A* and E*.

Fig. S2 Homogenization of maternally inherited factors in A* and E*.

Table S1 Marker information and related accession nos of sequenced fragments of isofemale fly lines E14 and A157 of *D. melanogaster* (Glinka *et al*. 2003; Ometto *et al*. 2005).

Table S2 Candidate gene list for all six QTL positions.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Chapter 2

Part I: Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia

Laurent SJY, Werzner A, Excoffier L, Stephan W

Molecular Biology and Evolution

Approximate Bayesian Analysis of *Drosophila melanogaster* Polymorphism Data Reveals a Recent Colonization of Southeast Asia

Stefan J.Y. Laurent,*[1] Annegret Werzner,[1] Laurent Excoffier,[2] and Wolfgang Stephan[1]

[1]Section of Evolutionary Biology, Department of Biology II, University of Munich, Munich, Germany

[2]Department of Biology, Computational and Molecular Population Genetics Laboratory, Zoological Institute, University of Berne, Berne, Switzerland

***Corresponding author:** E-mail: laurent@bio.lmu.de.

Associate editor: Jody Hey

Abstract

Southeast Asian populations of the fruit fly *Drosophila melanogaster* differ from ancestral African and derived European populations by several morphological characteristics. It has been argued that this morphological differentiation could be the result of an early colonization of Southeast Asia that predated the migration of *D. melanogaster* to Europe after the last glacial period (around 10,000 years ago). To investigate the colonization process of Southeast Asia, we collected nucleotide polymorphism data for more than 200 X-linked fragments and 50 autosomal loci from a population of Malaysia. We analyzed this new single nucleotide polymorphism data set jointly with already existing data from an African and a European population by employing an Approximate Bayesian Computation approach. By contrasting different demographic models of these three populations, we do not find any evidence for an early divergence between the African and the Asian populations. Rather, we show that Asian and European populations of *D. melanogaster* share a non-African most recent common ancestor that existed about 2,500 years ago.

Key words: *Drosophila melanogaster*, demography, approximate Bayesian computation.

Research article

Introduction

The demographic history of wild populations of the fruit fly *Drosophila melanogaster* is a subject of investigation for several decades (David and Capy 1988; Baudry et al. 2004; Stephan and Li 2007). This is mainly due to the fact that this species is used as a model organism in studies of local adaptation (Ometto et al. 2005; Li and Stephan 2006; Pool et al. 2006). These studies attempt to detect evidence for past events of positive selection by scanning the genome for specific patterns of genetic variation. In these analyses, the identification of major demographic events that affected the populations in the recent past plays an important role for at least two reasons. First being able to identify the ancestral and derived populations and obtaining reliable estimates for the times at which the derived populations colonized new habitats is critical in understanding to which environmental conditions these populations had to adapt. Second studies of local adaptation that are trying to detect genes targeted by positive selection are based on the assumption that the adaptive event required the fixation of beneficial alleles. The rapid increase in frequency of a favorable allele may leave a typical signature in DNA polymorphism data (Maynard Smith and Haigh 1974) that is called a selective sweep. Such signatures can be detected by recently developed methods (Kim and Stephan 2002; Kim and Nielsen 2004; Nielsen 2005; Stephan et al. 2006; Jensen et al. 2007; Pavlidis, Jensen, et al. 2010). Some of these methods rely on a demographic model that helps to control for the fact that neutral demographic forces (such as population size bottlenecks in the recent past) may generate similar signatures as selective sweeps (Barton 1998). Estimation of demographic models can therefore help to reduce the rate of false positives and to identify chromosomal regions that are not compatible with neutral demographic scenarios.

Previous studies showed that ancestral populations of *D. melanogaster* live in the African mainland south of the Saharan desert (Tsacas and Lachaise 1974). Furthermore, historical and morphological lines of evidence indicate that derived populations may be categorized into ancient populations that have colonized the Eurasian continent during prehistoric times, and new populations that have colonized the American and Australian continents during historic times along with recent human migrations (David and Capy 1988). More recently, studies based on a rigorous statistical analysis of genome-wide samples of nucleotide polymorphism data could confirm that *D. melanogaster* originated in sub-Saharan Africa (Li and Stephan 2006). These studies could also show that divergence between African and European populations occurred about 16,000 years ago and that this event was associated with a population size bottleneck (Li and Stephan 2006; Thornton and Andolfatto 2006).

In contrast, the timing of the colonization process of the Asian populations has not been identified yet. The first study that analyzed Asian populations of *D. melanogaster* contrasted patterns of morphological variation between derived and ancestral populations (David et al. 1976). This

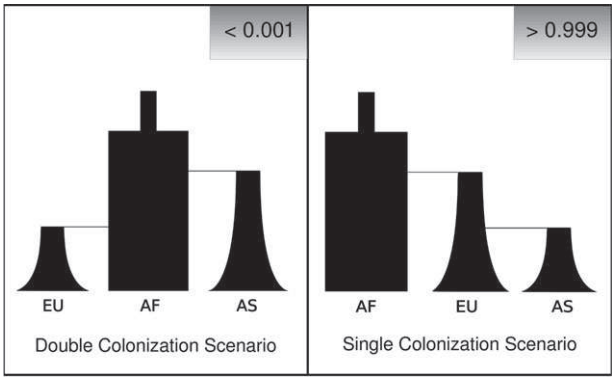


Fig. 1. Demographic models with associated posterior probability. AF: Africa, AS: Asia, and EU: Europe.

study showed that Asian populations are characterized by specific morphological and physiological properties, such as slower development growth, higher fresh weight, and smaller ovariole numbers than African and European populations. Based on these results, the authors proposed the existence of a "Far Eastern Race" of *D. melanogaster*. One of the hypotheses that has been proposed to explain the morphological divergence was the occurrence of an ancestral divergence between African and Asian populations. This early colonization of Asia would have predated the divergence between African and European flies and occurred before or during the last ice age. In contrast to these results, recent population genetics surveys revealed that non-African populations of *D. melanogaster* share a unique origin (Baudry et al. 2004; Schlötterer et al. 2006). However, these latter studies do not provide an estimation of the time at which *D. melanogaster* colonized the Asian continent. Furthermore, they do not explicitly model the colonization process.

In this study, we conducted a population genetic analysis of three populations of *D. melanogaster* from Africa, Europe, and Southeast Asia. We sequenced nearly 280 fragments on the X and third chromosome of the Asian population and analyzed this new data set together with existing data that have been collected for the African and European populations (Glinka et al. 2003; Ometto et al. 2005; Hutter et al. 2007). We employed an Approximate Bayesian Computation (ABC) approach (Beaumont et al. 2002) to investigate the demographic histories of these three populations. We found that a model in which Asian and European populations share a common non-African ancestor is more likely than a scenario with an independent early colonization of the Southeast Asian region. We show that the divergence between Southeast Asian and European *D. melanogaster* populations occurred about 2,500 years ago.

Materials and Methods

Samples

The individuals analyzed come from 36 inbred lines sampled from an African population from Zimbabwe, a European population from The Netherlands, and an Asian population from Kuala Lumpur (Glinka et al. 2003, 2005).

Collection of DNA Sequence Data

We sequenced from the Asian sample a subset of the loci that have also been sequenced in the African and European samples (Glinka et al. 2003; Ometto et al. 2005; Hutter et al. 2007). We reused the same primers to sequence 222 fragments of about 550 bp on the X chromosome and 51 fragments on the third chromosome. Sequences were generated as described in Glinka et al. (2003). We aligned the new Asian sequences to the already existing African and European data sets and retained for the demographic analysis only the fragments for which data was available from at least nine individuals in every population (208 X-linked and 50 autosomal fragments). Sequences have been deposited in GenBank (accession numbers: HQ396779–HQ396790 and HQ860796–HQ863959).

Evolutionary Scenarios

We analyzed two different demographic scenarios to investigate how the Asian population is related to the Africa and European ones. Graphical representations of the two models are given in figure 1, and a description of the prior distributions of the model parameters can be found in table 1. Both models are characterized by the absence of migration between populations and a stepwise expansion in population size of the ancestral African population, as it has been described in Li and Stephan (2006).

The first model that we called the double colonization scenario (DCS) describes a demographic history where the Asian and European populations split off independently from the African population (fig. 1). This model has been designed to test the hypothesis that the migration of *D. melanogaster* from Africa to Asia predated the colonization of the European continent. The second scenario that we called the single colonization scenario (SCS) describes a situation in which *D. melanogaster* has migrated out of Africa through a single colonization route; that is, all non-African (cosmopolitan) populations of *D. melanogaster* share a non-African common ancestor (Baudry et al. 2004). To model this scenario, we relied on previous demographic analyses of the African and European populations. These studies showed that the European population derived from the African one and went through a population size bottleneck (Li and Stephan 2006; Thornton and Andolfatto 2006). We modeled the Asian population assuming that it split off from the European population and underwent a population size bottleneck associated with this founding event (fig. 1). Because this model is making the assumption that the European population underwent one bottleneck against two for the Asian population, we also investigated an additional model that we called the SCS-2 model, where we inverted the situation and applied one bottleneck to the Asian and two to the European population. This approach allowed the European and Asian populations to experience different levels of genetic drift in our models. It is important to note here that our modeling approach

Table 1. Prior Distributions of the Demographic Models.

Parameter	Prior Distribution			Models
	Min	Max	Distribution	
Sizes				
Current African size	10^5	3×10^7	Uniform	All
Current European size	10^4	5×10^6	Uniform	All
Current Asian size	10^4	5×10^6	Uniform	All
Bottleneck size of the Asian population	10	10^5	Uniform	All
Bottleneck size of the European population	10	10^5	Uniform	All
Ancestral African population size	10^5	2×10^7	Uniform	All
Times				
Exit out of Africa	10^2	10^5	Uniform	SCS and SCS-2
Divergence between European and Asian populations	10^2	Exit out of Africa	Uniform	SCS and SCS-2
Exit out of Africa of the Asian population	10^2	10^5	Uniform	DCS
Exit out of Africa of the European population	10^2	10^5	Uniform	DCS
Expansion time of the African population	10^2	4×10^5	Uniform	All

NOTE.—Sizes are given in effective numbers of individuals (N_e), and times are given in years assuming ten generations per year. In the coalescent simulations, times were scaled in units of $4N_e$ generations for the autosomal data set and in units of $3N_e$ generations for the X-linked data set.

doesn't make any assumption concerning the geographic location of the split between European and Asian populations. (The fact that the Asian population splits off from the European one in the SCS model in figure 1 does not mean that this split occurred in Europe).

Approximate Bayesian Inference

To estimate the posterior probabilities of different demographic models and posterior distributions of the parameters of these models, we took an ABC approach. ABC is a computational method in Bayesian statistics that is used in population genetics to perform model-based parameter inference when suitable likelihoods are not available or computationally prohibitive (Pritchard et al. 1999; Beaumont et al. 2002; Excoffier et al. 2005). The method relies on the comparison of a vector of summary statistics computed on the observed data, Δ_{obs}, with those computed on a large number of simulated data sets for which the parameters of interest are known, Δ_{sim}. Here, we implemented our ABC algorithm following Excoffier et al. (2005). The algorithm used to estimate the parameters of a model is composed of three steps: a simulation step, a rejection step, and an estimation step. The simulation step consisted in simulating, for every evolutionary scenario, 1 million data sets that were identical to our observed data set in terms of numbers of loci and sample sizes. Every evolutionary scenario was defined by a set of parameters (population sizes, age of different demographic events), and every parameter was characterized by a prior distribution (table 1). For each evolutionary scenario, we sampled from the prior distribution and used the randomly picked parameter values to perform coalescent-based simulations. The way in which the rejection and the estimation steps have been applied to these simulated data sets differed for the model choice and the parameter estimation procedures and are described later in the corresponding sections.

To simulate these data sets, we incorporated available external information about the local mutation rates (μ) and the local crossing-over rate (r). This allowed us to directly estimate posterior distributions for effective population sizes instead of estimating them for the compound population parameters $\theta = 4N_e\mu$ and $\rho = 4N_e r$ where θ is the coalescent mutation parameter, ρ the coalescent recombination rate, and N_e the effective population size. Mutation rates for every fragment were calculated based on genetic divergence to the sister species *D. simulans* (Kimura 1980) following Li et al. (1999); that is, assuming a divergence time between *D. melanogaster* and *D. simulans* of 2.3 My and correcting for prespeciation divergence. However, divergence-based estimates of mutation rates can potentially be biased by the long-term action of purifying selection on noncoding regions. We therefore took this uncertainty into account by putting a uniform prior distribution on the mutation rate of every simulated fragment centered around the local divergence-based estimate μ_{est} with lower and upper boundaries $\mu_{est}/2$ and $2\mu_{est}$.

Similarly, external information about the local recombination rates was used to generate our simulated data sets. Recombination rates are given as the local rates of crossing-over per site per generation and were calculated using the "Recombination Rate Estimator" web-based program (Singh et al. 2005; see also Hutter and Stephan 2009) available at http://petrov.stanford.edu/software.html (accessed February 24, 2011). However, these estimates have been obtained using a North American population, and it is reasonable to think that the real recombination rates in our populations may deviate from these values. To show that the results of our study are robust to this potential deviations, we put a prior distribution on the recombination rate at each locus centered on the Singh et al. (2005) estimate r_{est}, with lower and upper boundaries set at $r_{est}/2$ and $2r_{est}$.

The coalescent simulations were performed using a slightly modified version of the program ms that allows to simulate data sets with unequal sample sizes across loci (Hudson 2002; Ross-Ibarra et al. 2008). However, to sample parameter values from prior distributions and to compute summary statistics on the simulated data efficiently, we developed our own code. For this, we used the Gnu Scientific Library C++ library and the libsequence C++ library (Thornton 2003). Simulations were launched on a 64-bit

Linux cluster with 510 nodes. We checked for errors in our code by comparing simulation results with similar computations performed by msABC, a coalescent simulator for ABC simulations (Pavlidis, Laurent, et al. 2010).

The Model Choice Procedure

To summarize our data sets for the model choice procedure, we computed the following statistics in the three populations: the average and the variance across all fragments of the number of polymorphic sites (S), the average and the variance of Tajima's D (Tajima 1989), and the average Z_{nS} (Kelly 1997). Additionally, we computed for all pairs of populations the average distance of Nei, D_A (Nei and Li 1979), and the average proportion of shared polymorphisms between populations, S_S. Monomorphic fragments were removed from the analysis. These statistics have been chosen based on their correlation with important demographic parameters. For example, the number of polymorphic sites is expected to increase with the effective population size, and variations of Tajima's D statistic can reflect past fluctuations in population size. The Z_{nS} statistic (Kelly 1997) is a measure of linkage disequilibrium (LD) defined as the average r^2 over all pairwise comparisons of polymorphic sites in a sample of sequences, where r^2 is the squared correlation of allelic identity between two loci (Hartl and Clark 1989, p. 53–54). This statistic is expected to be sensitive to variations in the length of the oldest branches of gene genealogies and can therefore carry information about ancestral population sizes and the severity of population size bottlenecks. The distance of Nei and the proportion of shared polymorphisms are measures of genetic differentiation between our populations and are expected, in a model without migration like ours, to correlate with times of divergence of two populations. All summary statistics presented in this study have been computed using the routines of the C++ library "libsequence" (Thornton 2003).

The posterior probabilities of different demographic models can be estimated on the basis of the Euclidean distance δ between the observed summarized data set and the simulated summarized data sets of all models. The inference procedure consists in retaining only simulations for which the Euclidean distance is smallest. Pritchard et al. (1999) proposed that the posterior probability of a model can be approximated by the proportion of retained simulations under that model, relative to the number of retained simulations under all models. Beaumont (2008) proposed an improvement of the method that corrects for the fact that retained simulations never exactly match the observed data. The method is based on a weighted multinomial logistic regression procedure, where the response variable is the indicator of the model and the predictor variables are the summary statistics (Fagundes et al. 2007; Beaumont 2008). We computed posterior probabilities for every demographic model, based on the 500 simulations associated with smallest Euclidean distance following the method of Beaumont (2008) and applied this procedure to the X-linked and the autosomal data set independently. Because the same number of data sets have been simulated under each model, we computed Bayes factors as the ratio of the posterior probabilities. To investigate if the results of our estimations were stable with regard to the proportion of retained simulations, $P_{\delta,}$ we computed posterior probabilities for our three models using several values of P_δ ranging from 0.025% to 1%.

We also investigated the accuracy of our model choice procedure, following Peter et al. (2010). We therefore simulated 1,000 pseudo-observed data sets, with the same number of fragments and sample sizes as our autosomal data set, under each demographic model (DCS and SCS) and computed for each one of them the posterior probability of having been generated under the DCS model and the posterior probability of having been generated under the SCS model. To decide whether a pseudo-observed data set should be assigned to the DCS or the SCS model, we used as an arbitrary threshold value, a ratio of posterior probabilities of 10 in favor of one of both models. Precision was measured as the proportion of correctly identified data sets under each model.

Parameter Estimation

In ABC estimations, the quality of the analysis generally relies on well-chosen summary statistics (Joyce and Marjoram 2008). However, at least two problems arise when full genetic polymorphism data sets have to be summarized into vectors of summary statistics. On the one hand, using a small number of summary statistics may lead to a substantial loss of information compared with the information carried by the full data set, but on the other hand, increasing the number of summary statistics can cause two problems. First, summary statistics that are not related to the parameters of the model or that correlate with other summary statistics will be uninformative and will only add noise to the Euclidean distance. Second, correlations between summary statistics will violate the assumption of singularity, which is required when performing the locally weighted linear regression for estimating the parameters (Beaumont et al. 2002). To overcome this problem, Wegmann et al. (2009) proposed to reduce the dimensionality of the summarized data set by performing a partial least square (PLS) transformation. The advantage of the PLS transformation is 2-fold. First, similar to a principal component analysis, it allows extracting a small number of orthogonal components from a matrix composed of a larger number of summary statistics of our data set. This is leading to a reduction of the uninformative signals of the Euclidean distance and ensures the singularity of the final matrix of summary statistics. Second, in a PLS transformation, the dimensionality reduction is coupled with a regression model, and the latent components (i.e., the transformed summary statistics) are constructed to maximize the prediction of the response variable of the regression model (i.e., the parameters of our demographic model) (Boulesteix et al. 2007). Although this approach can be applied when estimating the posterior distributions of the parameters of a given demographic

model, it cannot be applied for model choice. The reason is that PLS components are constructed independently for every single demographic model, whereas our model choice procedure requires that the set of summary statistics remains identical for all compared models.

We summarized our nucleotide polymorphism data sets into 12 summary statistics that carry information about the level of polymorphism, the site frequency spectrum, LD, and the amount of differentiation between all three populations. All statistics have been computed for every population separately and for the pooled data set. Summary statistics describing differentiation between populations have been computed for all possible pairwise comparisons. The list of all summary statistics that we used is given in supplementary table S1, Supplementary Material online. We used this set of summary statistics to summarize our observed and our simulated data. We constructed the PLS latent components using 10,000 simulated data sets under the best demographic model for the chromosomes X and 3. To do this, we employed code available in the ABCtoolbox package (Wegmann et al. 2010). Choosing the best number of PLS components for parameter estimations has been done by investigating the decrease of the root mean square error (RMSE) for every parameter as a function of the number of PLS components. The RMSE indicates the percentage of variation unexplained by the PLS components and is constructed by comparing the simulated parameter values with the ones predicted using a given number of PLS components. We chose the number of components to be used in the parameter estimation procedure such that additional components do not decrease the RMSE of any parameter of the model. The retained PLS components were used to transform the observed and the simulated data sets. The rejection step consisted in computing the Euclidean distance δ between simulated and observed sets of summary statistics and to retain the 5,000 simulations closest to the observed data based on their value of δ. Finally, we estimated posterior distributions of the parameters of interest by applying the locally weighted multivariate regression method of Beaumont et al. (2002) implemented in the abcEst program (Excoffier et al. 2005). We estimated the marginal posterior probability distribution of each demographic parameter using the kernel density estimation method implemented in the R core package and reported the mode and the 95 credibility intervals of these distributions. To investigate if the results of our estimations were stable with regard to the proportion of retained simulations, P_δ, we reestimated the marginal posterior distributions of all parameters of the best model using several values of P_δ ranging from 0.01% to 5%.

Before the regression, we transformed the parameter values as

$$y = -\ln\left(\tan\left(\frac{x-\min}{\max-\min}\frac{\pi}{2}\right)^{-1}\right),$$

where min and max are the lower and upper bounds of the prior, respectively. This transformation guarantees that the posterior distribution obtained after the regression step remains within the range of the prior distribution (Hamilton et al. 2005).

Table 2. Polymorphism Patterns in the Asian Population.

	X Chromosome	Autosome
No. of loci	222	51
Average sample size	11.64	11.57
Average length of alignment	567.16	551.04
No. of invariant loci	39	4
Average (SE) θ_w in %	0.35 (0.02)	0.39 (0.0004)
Average (SE) π in %	0.37 (0.03)	0.38 (0.0005)
Average (SE) divergence (K) in %	6.8 (0.25)	5.5 (0.54)
Average (SE) θ_w /K	0.05 (0.004)	0.086 (0.013)
Average (SE) Tajima D	0.17 (0.08)	0.19 (0.12)
Average (SE) Z_{nS}	0.5 (0.02)	0.53 (0.04)

Posterior Predictive Simulations

To see how well our best model is able to reproduce the observed data, we performed posterior predictive simulations (Gelman et al. 2003; Thornton and Andolfatto 2006). This has been done by sampling parameter values from the probability density functions of the marginal posterior distributions of our best demographic model and by using them to simulate multilocus summaries of the data. This procedure allowed us to check which aspect of the data could be explained by our model and which aspects might indicate some limitations of our model. We generated 1,000 simulated data sets that had the same number of fragments and sample sizes as our X-linked and autosomal data sets and summarized them into the mean and variance of a large number of summary statistics. For every summary statistic, we computed the probability that the simulated mean and variance are smaller than the observed ones.

Results

X-Linked and Autosomal Polymorphism Patterns in the Asian Population

For the X chromosome, we gathered polymorphism data from a total of 222 fragments, spanning 126,154 nt sites (gaps were excluded). 1,309 of these sites are polymorphic. Information about this new genome scan is summarized in table 2. On average data could be obtained from 11.6 (of 12) lines. The sequenced fragments show recombination rates ranging between 2.1 and 4.3×10^{-8} per bp per generation. The means (standard error [SE]) of π (Tajima 1983) and θ_w (Watterson 1975) across the X chromosome are 0.0037 (0.0003) and 0.0035 (0.0002), respectively. We found 39 loci with no polymorphism. The average and variance of Tajima's D along the X chromosome are 0.17 and 1.28, respectively (table 2). In order to contrast X-linked with autosomal genetic variation, we sequenced 51 autosomal loci spanning 28,441 sites from which 323 are polymorphic. Glinka et al. (2005) identified four autosomal inversions in the Asian population segregating at frequencies ranging from 0.04 to 0.24. For this study, only lines harboring no inversions were used. The fragments are randomly spread along the third

chromosome. On average data could be obtained for 11.6 lines and sequenced fragments show recombination rates between 1.22 and 3.33 $\times$ 10^{-8} per bp per generation. The means (SE) of π and θ_w for the autosomal fragments are 0.0039 (0.0005) and 0.0038 (0.0004), respectively, which is slightly higher than for X-linked data. Only four fragments are monomorphic. The average and the variance of Tajima's *D* along chromosome 3 are 0.19 and 0.81, respectively. Tables with the number of segregating sites, diversity indices, and basic statistics for all X-linked and autosomal loci will be provided upon request.

Comparison of the African, European, and Asian Populations

In order to compare our new Asian data with the data that have previously been collected for the African and European populations (Glinka et al. 2003; Ometto et al. 2005; Hutter et al. 2007), we aligned data from the three populations and analyzed the data jointly. Data for all three populations were available for 208 X-linked and 50 autosomal fragments. We computed the number of SNPs *S*, Tajima's *D*, Z_{nS}, and F_{ST} values (Hudson et al. 1992) for every population and every fragment and generated distributions of these statistics across X and chromosome 3. The results of these analyses are summarized in figure 2. Compared with the African and European populations, the Asian population shows the lowest amount of nucleotide diversity (fig. 2*A*) both for X-linked and autosomal data. The distribution of Asian Tajima's *D* values is similar to those observed in the European population although the average *D* is slightly more positive on the Asian X chromosome and the variance slightly smaller on the Asian third chromosome (fig. 2*B*). The Asian population also harbors the highest levels of LD among all three populations both on the X and autosome (fig. 2*C*). The distributions of F_{ST} values between all three pairs of populations indicate that the African and Asian populations are the most differentiated populations and that the Asian and European populations are the most closely related ones (fig. 2*D*). These analyses show that the Asian population shares several characteristics with the European population: low diversity, large variance in Tajima's *D*, and high LD. This suggests that, similar to the European population (Stephan and Li 2007), the Asian population is also derived, characterized by past population size fluctuations, and might share part of its recent demographic history with the European population. This hypothesis was evaluated by the following ABC analysis.

We also compared the three populations on the basis of their ratios of X-linked with autosomal diversity (X/A diversity ratios). Recent studies showed that European populations of *D. melanogaster* have a smaller X/A diversity ratio compared with African populations (Hutter et al. 2007). A possible explanation for this observation may lie in the colonization process of *D. melanogaster* in Europe. Pool and Nielsen (2008) showed that in a multiply mating species, like *D. melanogaster*, serial founder events reduce levels of genetic variability more severely on the X chromosome than on the autosomes. We computed the X/A diversity ratios for the three populations, following Pool and Nielsen (2008), as:

$$\frac{X}{A}\text{diversity ratio} = \frac{\pi_X/K_X}{\pi_A/K_A},$$

where π_x and π_A are the nucleotide diversities, and K_x and K_A the number of substitutions per site between *D. melanogaster* and *D. simulans* for chromosome X and 3, respectively. Dividing nucleotide diversity by between-species divergence corrects for mutational biases. The values of the X/A diversity ratios for the African, European, and Asian populations were 0.87, 0.52, and 0.63, respectively. These results show that the Asian population has lost more X-linked than autosomal genetic variability during its range expansion, as is expected for cosmopolitan populations (Pool and Nielsen 2008) and that this effect is smaller in Asia than in Europe.

Statistical Evidence for a Single Colonization Event out of Africa

The results of the model choice procedure showed strong evidence in favor of the SCS model. The Bayes factor (Kass and Raftery 1995) was larger than 100 for both X-linked and autosomal data sets (fig. 1). The method also revealed that the SCS model where the Asian population undergoes two successive bottlenecks is associated with higher posterior probabilities for both X-linked and autosomal data, when compared with the SCS-2 model in which the Asian population is bottlenecked only once. The posterior probabilities were 0.99 versus 0.01 for the X chromosome and 0.9996 versus 0.0004 for chromosome 3. When letting P_δ, the proportion of retained simulations in the model choice procedure, vary from 0.025% to 1%, we observed that the strength of evidence in favor of the SCS model remained decisive for all values of P_δ and both chromosomes (supplementary table S2, Supplementary Material online). For the comparison between model SCS and SCS-2, the results also remained stable for both chromosomes. Finally, the results of our analysis of the accuracy of our model choice procedure showed that both the SCS and the DCS model could be correctly identified by the method in 97.9% and 96.8% of the cases, respectively. We thus identified the SCS model as our best demographic model and estimated the marginal posterior distribution of each parameter.

Estimation of the Demographic Parameters

We summarized the X-linked and the autosomal data sets using eight PLSs components (table 3). Using the larger X-linked data set, we find that the effective population size of the ancestral African population is about 1,837,000 individuals (932,000, 2,531,000) and that the out-of-Africa migration occurred approximately 16,800 years ago (9,400, 33,500). The colonization of the Southeast Asian

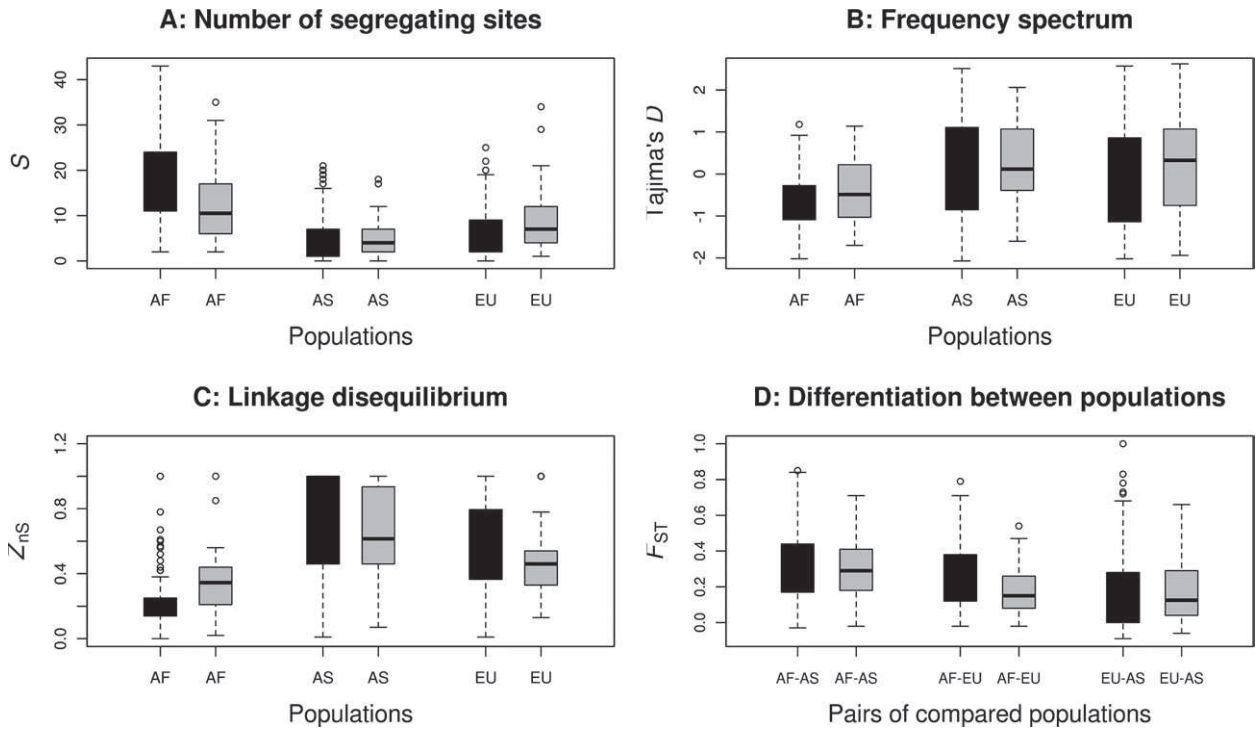


Fig. 2. Boxplots representing the distributions of four summary statistics for all three populations. (*A*) Number of segregating sites, (*B*) site frequency spectrum, (*C*) linkage disequilibrium, and (*D*) differentiation between populations. Results from the analysis of X-linked data are shown in black, whereas those of autosomal data are shown in gray.

continent occurred about 2,500 years ago (700, 5,200), and the effective size of the Asian founding population was approximately 8,300 individuals (3000, 77,000). The corresponding estimates based on the autosomal data are comparable except for the estimate of the time of colonization of the Southeast Asian continent that is larger for chromosome 3: 5,000 years (1,000, 11,000). Estimates of the demographic parameters for the African and European populations can also be found in table 3, and graphical representations of all marginal posterior distributions for chromosome X and 3 are in figure 3. These estimates do roughly agree with earlier results that were obtained with a maximum likelihood analysis of the joint mutation frequency spectrum of the African and European population (Li and Stephan 2006). When letting P_{δ}, the proportion of retained simulations in the parameter estimation procedure, vary from 0.01% to 5%, we observed that the mode and the credibility intervals of our posteriors distributions remained stable (supplementary fig. S1, Supplementary Material online).

Table 3. Estimation of the Demographic Parameters.

Parameters	Priors	X Chromosome	R^2	Autosome	R^2
Current African population size	(100,000, 30,000,000)	4,786,360 (2,040,701, 29,208,295)	0.62	3,134,891 (1,371,066, 28,013,950)	0.55
Current European population size	(10,000, 5,000,000)	1,632,505 (780,907, 4,870,580)	0.40	878,506 (383,361, 4,775,964)	0.30
Current Asian population size	(10,000, 5,000,000)	414,113 (151,168, 4,567,219)	0.18	512,748 (143,082, 4,542,090)	0.12
European bottleneck population size	(10, 100,000)	22,066 (14,338, 81,102)	0.47	32,128 (15,968, 95,162)	0.48
Asian bottleneck population size	(10, 100,000)	8,279 (2,971, 77,482)	0.36	11,862 (4,255, 81,044)	0.35
Exit out of Africa	(100, 100,000)	16,849 (9,392, 33,452)	0.63	12,843 (7,095, 31,773)	0.62
Southeast Asia colonization time	(100, Exit out of Africa)	2,467 (711, 5,195)	0.72	5,012 (992, 11,084)	0.68
African Expansion time	(100, 400,000)	25,553 (1,698, 376,730)	0.02	37,323 (3,636, 379,212)	0.01
Ancestral African population size	(10,000, 20,000,000)	1,837,229 (931,637, 2,530,609)	0.83	1,705,328 (609,393, 2,458,653)	0.82

Note.—The time estimations (i.e., modes and credibility intervals) are provided in years assuming ten generations per year. Population sizes are given in effective numbers of individuals. R^2 is the coefficient of determination.

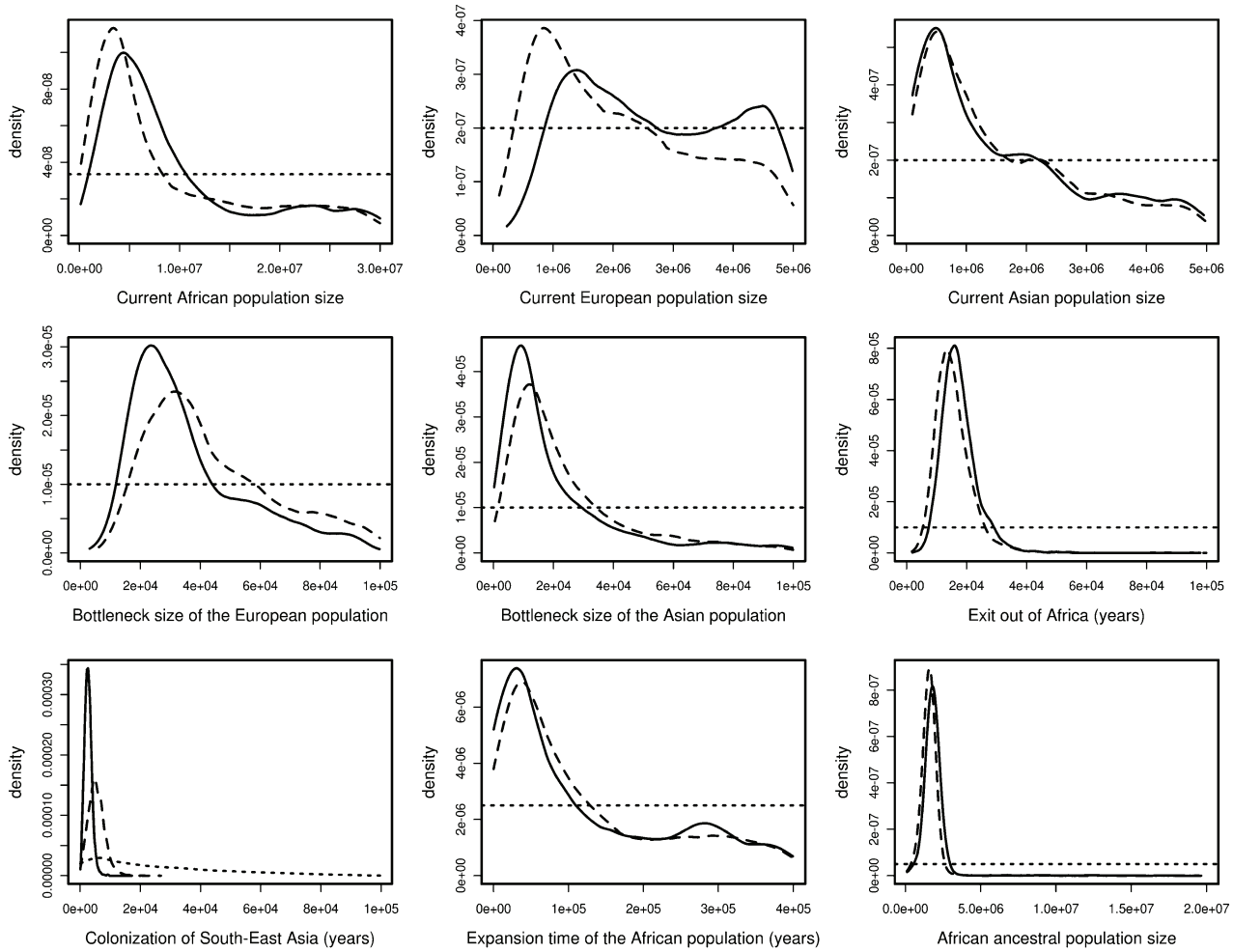


Fig. 3. Posterior distribution of the parameters of the SCS model inferred by ABC estimation. Dotted lines: Priors; Solid lines: Posteriors for the X-linked data set; and Dashed lines: Posteriors for the autosomal data set.

Discussion

The results of the performance analysis we carried out on the model choice procedure shows that, for a data set similar to the one we analyse in this study, the method is able to distinguish between the two demographic scenarios we investigated in this study: the SCS, where African and non-African populations have a non-African most recent common ancestor (MRCA), and the DCS, where these populations have an African MRCA (fig. 1). We also showed that the results of the model choice procedure were stable with regard to variations of the proportion of retained simulations $P_{\delta,}$ These results give credibility to our finding that the SCS provides a better fit to our observed data than the DCS.

The quality of the estimation of each parameter of the SCS can be assessed by the coefficient of determination (R^2). Previous studies have shown that parameters for which R^2 values are smaller than 5–10% are usually difficult to estimate (Neuenschwander et al. 2008). For the parameters of our best model, R^2 values are higher than 10% (table 3) except for the parameter representing the time of the African expansion (1% for chromosome X and 2% for chromosome 3). These low values for this parameter could indicate that a simple expansion model is not the best demographic model to explain the patterns of genetic diversity observed in the African population or that the statistics we used to summarize the full data set (supplementary table S1, Supplementary Material online) did not capture enough information on this parameter. As for the model choice procedure, we showed that our parameter estimates are robust to the proportion of retained simulations P_{δ} (supplementary fig. S1, Supplementary Material online).

Migration of human populations toward Southeast Asia started during the last ice age about 55–65 ky ago (Fagundes et al. 2007). Due to the current commensal relation between *D. melanogaster* and humans, it could be envisioned that this species had colonized Southeast Asia together with the first human migrants. Such an early colonization scenario has once been used to explain the strong morphological differentiation of the Asian populations (David et al. 1976). However, our present demographic estimations do not support such a scenario. Rather, our results indicate that the Southeast Asian population was founded by flies that diverged from the first non-African populations at a later time. We estimated that the divergence between the African and the derived

population occurred about 16,800 ago (9,400, 33,500) for the chromosome X and about 12,800 years ago (7,100, 31,800) for chromosome 3. These results are in line with a previous maximum likelihood analysis of the African and European populations that suggested a date of divergence of 15,800 years (12,000, 19,000) (Li and Stephan 2006). At this time, rising temperatures following the end of the last ice age could have favored the colonization of northern territories. Early human sedentary settlements during the natufian period (starting 13,000 years ago) in the Fertile Crescent (Bar-Yosef 1998) could already have helped to maintain a relatively large population of *D. melanogaster* in this region. Our results reject the hypothesis of an ancient migration of *D. melanogaster* to Southeast Asia, but they do not provide information about the geographic location of the split between the European and the Southeast Asian populations. Even if it seems likely that migrating *D. melanogaster* populations have entered the region of the Middle East by crossing the Sinai peninsula, additional information, such as genetic polymorphism data from a Middle East population, would be necessary to learn more about the geographic location of the divergence between European and Southeast Asian populations.

Furthermore, we found that the divergence between Asian and European populations occurred about 2,500 years ago (700, 5,200) for the chromosome X and 5,000 years ago (990, 11,000) for chromosome 3. This late colonization of Asia and Europe, although favorable climatic conditions were already present before that time, suggests that *D. melanogaster* depended on the establishment of human populations (and perhaps agriculture) to colonize these new areas. Therefore, a long divergence time associated with a neutral process may not explain the observed morphological characteristics of the Asian populations (i.e., smaller number of ovarioles, smaller eggs, and higher fresh weight than the ancestral population) as suggested by David et al. (1976).

We also estimated the effective size of the founding population that diverged from the ancestral gene pool (22,100 and 32,100 individuals for chromosome X and 3, respectively; table 3), and the size of the founder population that colonized the Southeast Asian region (8,300 and 11,900 individuals for chromosome X and 3, respectively; table 3). These estimates have to be taken with caution since the impact of genetic drift on the history of the derived populations may be much more complex than that produced by a simple population size bottleneck. Recurrent fluctuations in population size (Pool and Nielsen 2008) as well as population substructure might also have influenced the way genetic drift shaped the patterns of polymorphism we observe in present populations. Therefore, these estimates should be seen as a simple way of summarizing the amount of genetic drift needed to reproduce the skew in the site frequency spectrum we observe in these populations. This model can then be used to correct the rate of false positives of selection detection methods that are based on the site frequency spectrum.

The results of the predictive simulations (supplementary table S3, Supplementary Material online) show how well our model is able to reproduce the observed data sets. Unability to reproduce certain summary statistics of the data set indicates that some aspects of the demographic model are not optimal. The nature of the summary statistics that are not well predicted by our model can also, in some cases, indicate how the model might be improved. Although most summary statistics are correctly predicted by our best model, it fails to account for some aspects of the observed data sets (supplementary table S3, Supplementary Material online). Our model predicts higher values for the average and lower values for the variance of Tajima's *D* compared with observed data on the X-chromosome in the European population. Therefore, it seems that patterns of genetic variation found on the X chromosome of European *D. melanogaster* cannot be fully explained by a single population size bottleneck alone. Discrepancies between patterns of genetic variation found at X-linked and autosomal genes in the European population could be due to biased sex ratios and/or different intensities of selective pressures on the X chromosome (Hutter et al. 2007). Also, predicted values of Z_{nS} for the African populations are higher than observed values. This could indicate that this population experienced more complicated size fluctuations than predicted by the expansion model alone or that recombination rates in the African population are higher than what we assume in this study. Interestingly, most aspects of the Asian data set can be correctly predicted by our model, with the exception of D_A on the third chromosome where predictions are smaller than the observed value.

X/A diversity ratios in the three populations reflect an aspect of the data set that cannot be explained by our model. Pool and Nielsen (2008) showed that the small values of the X/A diversity ratio in the European population may be explained by serial founder events and multiple mating. Because the aim of this study was to understand how Asian populations relate to the African and European ones, we decided to analyse the X-linked and the autosomal data sets separately instead of increasing the complexity of our models. We note, however, that we do not observe large differences between the posterior distributions estimated from X-linked and autosomal data sets, which suggests that departure from expected X/A diversity ratios does not affect much our demographic estimates. Interestingly, the X/A diversity ratio is higher in the Asian population than in the European population (0.63 vs. 0.52, respectively), which could indicate that sex ratios and/or the number of founder events have been different in these two populations (Pool and Nielsen 2008). However, additional analyses, based on more complex demographic models including serial founder events, multiple mating, and varying sex ratios, would need to be performed to further investigate this phenomenon.

In conclusion, our study generated a new data set of X-linked and autosomal nucleotide polymorphism in a Malaysian population of *D. melanogaster*. In addition, we performed a demographic analysis of this data, jointly with one African and one European data set, by employing an ABC approach. Our results reject the hypothesis that an

early migration event could have led to the colonization of Southeast Asia and account for the observed morphological differentiation between African, European, and Asian *D. melanogaster*. The statistical model of demographic evolution that we inferred suggests that *D. melanogaster* colonized Eurasia after the Neolithic period, when the rise of agriculture turned small communities of hunter gatherers into larger sedentary settlements. Because *D. melanogaster* is now a human commensal, it is possible that the adaptive history of this species is mostly characterized by adaptations to post-neolithic human societies rather than by adaptations to the new climatic conditions that the non-African populations encountered. Local adaptation studies that may be based on our demographic analyses will help to answer this question by identifying the genes that were targeted by positive selection during this colonization process.

Supplementary Material

Supplementary tables S1–S3 and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank the Munich Drosophila group and Matthieu Foll for discussions, two anonymous reviewers for helpful comments on the manuscript, Daniel Wegmann for sharing code, and the Deutsche Forschungsgemeinschaft for funding (grants STE 325/7 and STE 325/12 of the Research Unit 1078). L.E. was partially supported by a Swiss National Science Foundation grant No 3100,126074.

References

Barton NH. 1998. The effect of hitch-hiking on neutral genealogies. *Genet Res*. 72:123–133.

Bar-Yosef O. 1998. The natufian culture in the levant, threshold to the origins of agriculture. *Evol Anthropol*. 6:159–177.

Baudry E, Viginier B, Veuille M. 2004. Non-African populations of *Drosophila melanogaster* have a unique origin. *Mol Biol Evol*. 21:1482–1491.

Beaumont MA. 2008. Joint determination of topology, divergence time, and immigration in population trees. In: Matsumura S, Forster P, Renfrew C, editors. Simulations, genetics and human prehistory. Cambridge (UK): McDonald Institute Monographs. p. 135–154.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.

Boulesteix AL, Strimmer K. 2007. Partial least squares: a versatile tool for the analysis of high dimensional genomic data. *Brief Bioinform*. 8:32–44.

David J, Bocquet C, Pla E. 1976. New results on the genetic characteristics of the Far East race of *Drosophila melanogaster*. *Genet Res*. 28:253–260.

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

Excoffier L, Estoup A, Cornuet JM. 2005. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* 169:1727–1738.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci U S A*. 104:17614–17619.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2003. Bayesian data analysis, 2nd ed. London: Chapman and Hall/CRC Press.

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

Glinka S, Stephan W, Das A. 2005. Homogeneity of common cosmopolitan inversion frequencies in Southeast Asian *Drosophila melanogaster*. *J Genet*. 84:173–178.

Hamilton G, Stoneking M, Excoffier L. 2005. Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. *Proc Natl Acad Sci U S A*. 102:7476–7480.

Hartl DL, Clarck AG. 1989. Principles of population genetics. Sunderland (MA): Sinauer Associates.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583–589.

Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W. 2007. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. *Genetics* 177:469–480.

Hutter S, Stephan W. 2009. Recombination rates may affect the ratio of X to autosomal noncoding polymorphism in African populations of *Drosophila melanogaster*—reply. *Genetics* 181:1703.

Jensen JD, Wong A, Aquadro CF. 2007. Approaches for identifying targets of positive selection. *Trends Genet*. 23:568–577.

Joyce P, Marjoram P. 2008. Approximately sufficient statistics and Bayesian computation. *Stat Appl Genet Mol Biol*. 7:1–16.

Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc*. 90:773–795.

Kelly JK. 1997. A test of neutrality based on interlocus associations. *Genetics* 146:1197–1206.

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

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

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 16:111–120.

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

Li YJ, Satta Y, Takahata N. 1999. Paleo-demography of the *Drosophila melanogaster* subgroup: application of the maximum likelihood method. *Genes Genet Syst*. 74:117–127.

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

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A*. 76:5269–5273.

Neuenschwander S, Largiadèr CR, Ray N, Currat M, Vonlanthen P, Excoffier L. 2008. Colonization history of the Swiss Rhine basin by the bullhead (*Cottus gobio*): inference under a Bayesian spatially explicit framework. *Mol Ecol*. 17:757–772.

Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet*. 39:197–218.

Ometto L, Glinka S, De Lorenzo D, Stephan W. 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, Jensen JD, Stephan W. 2010. Searching for footprints of positive selection in whole-genome SNP data from non-equilibrium populations. *Genetics* 185:907–922.
Pavlidis P, Laurent S, Stephan W. 2010. msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis. *Mol Ecol Resour*. 10:723–727.
Peter BM, Wegmann D, Excoffier L. 2010. Distinguishing between population bottleneck and population subdivision by a Bayesian model choice procedure. *Mol Ecol*. 19:4648–4660.
Pool JE, Aquadro CF. 2006. History and structure of sub-Saharan populations of*Drosophila melanogaster*. *Genetics* 174:915–929.
Pool JE, Nielsen R. 2008. The impact of founder events on chromosomal variability in multiply mating species. *Mol Biol Evol*. 25:1728–1736.
Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol*. 16:1791–1798.
Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut B. 2008. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS One*. 3:e2411.
Schlötterer C, Neumeier H, Sousa C, Nolte V. 2006. Highly structured Asian *Drosophila melanogaster* populations: a new tool for hitchhiking mapping? *Genetics* 172:287–292.
Singh ND, Arndt PF, Petrov DA. 2005. Effect of recombination on patterns of substitution in*Drosophila*. *Genetics* 169:709–722.
Stephan W, Li H. 2007. The recent demographic and adaptive history of *Drosophila melanogaster*. *Heredity* 98:65–68.
Stephan W, Song YS, Langley CH. 2006. The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* 172:2647–2663.
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. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* 19:2325–2327.
Thornton K, Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* 172:1607–1619.
Tsacas L, Lachaise D. 1974. Quatre nouvelles espèces de la Côte-d'Ivoire du genre *Drosophila*, groupe *melanogaster*, et discussion de l'origine du sous-groupe *melanogaster*. *Ann Univ Abidjan E Ecol*. 7:193–211.
Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7: 256–276.
Wegmann D, Leuenberger C, Excoffier L. 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics* 182:1207–1218.
Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*. 11:116.

Part II: Comparative study of morphological traits in *D. melanogaster* from Africa, Europe, and Asia (unpublished work)

Natural populations of *D. melanogaster* are differentiated spatially according to genetic and morphological characters (David & Capy 1988; Capy *et al.* 1993; Caracristi & Schlötterer 2003; Schlötterer *et al.* 2006). The Asian *D. melanogaster* show an interesting morphological divergence to flies from other parts of the world. This so called 'Far Eastern Race' is characterized by a greater adult weight, lower ovariole number, larger egg size, and slow growth. Flies showing these characters are known from Japan, Taiwan, Singapore, Indochina, Vietnam, and Sri Lanka (David *et al.* 1976; Lemeunier *et al.* 1986; Capy *et al.* 1993).

The origin of the Far Eastern Race, however, remained unknown. It had been argued that natural populations of *D. melanogaster* in Europe were able to disperse to the Asian continent and establish a stable population (Davis & Tsacas 1981). A second bottleneck accompanied this colonization process and reduced the population size. Thus, the morphological traits might recently drift away from European values (Lemeunier *et al.* 1986). Another explanation for the evolution of the Far Eastern Race is an ancient colonization of Asia and a long divergence time (more than 10,000 years) associated with local adaptation due to different ecological strategies (Capy *et al.* 1993).

Since we characterized the demographic history of European and Asian *D. melanogaster* we can now shed light on the origin of the Far East Race. We scanned our Asian population from Kuala Lumpur for the phenotype of the Far Eastern Race in order to combine the colonization model with the phenotypic characters.

Material and Methods

Morphological analysis We scanned the African, European and Asian populations using 12 isofemale lines per sample. Per line we used 40 individuals, 20 males and 20 females. We generated three replicates per line. The following traits were measured: adult fresh weight of males and females, ovariole number, egg size, and time of development. The fresh weight of each adult fly was measured with an electronic microbalance within 12 hours after emergence, since fluctuations of weight occur quickly (Katz & Young 1975). For egg size determination, newly emerged flies were placed into mating chambers at low density (10 females) for oviposition. The food medium consisted of molasse-agar which provided a dark and firm structure in Petri dishes. After oviposition, the Petri dishes were photographed with a digital camera. Egg size (i.e. egg length and width) was measured using Photoshop CX2/3 (CellP, Hamburg, Germany). The number of ovarioles was analyzed using 20 females per line for dissection, following the protocol of Wong & Schedl (2006). To measure the development time in *D. melanogaster* we used an average number of 20 males and 20 females of all 36 isofemale lines. They were placed in fresh breeding vials allowing them to oviposit within two hours. The time of transfer was taken as starting point of development as egg laying used to take place immediately. The final developmental stage was defined by hatching adults. Thus, all vials were scanned for hatched individuals every hour. The analysis of the experiment was done based on nested ANOVA with the help of the software R (Version 2.8.0), including the replicates, populations, and lines as classes. We estimated the sex specific heritability in a broad sense $H^2 = \sigma_L^2 / \left(\sigma_L^2 + \sigma_E^2\right)$, where σ_L^2 is the between-line variance and σ_E^2 is the within-line variance for the trait under consideration.

Results

Morphological analysis Table 1 presents means, SEs CVs and H^2 of phenotypic variation for all six morphological traits of the three studied populations. The data shows that the average trait values differ between populations, except for development time. This is confirmed by a nested ANOVA. Variation among lines is highly significant, showing that populations are geographically differentiated. The isofemale lines also contribute significantly to the variation. This is to be expected because all traits are highly heritable (Katz & Young 1975; Nunney 1996). This is confirmed by intermediate to high levels of heritability in a broad sense H^2, except for developmental time with low values of heritability for all three populations. Furthermore males and females show a highly significant correlation coefficient for body fresh weight (Pearson correlation coefficient Cor(weight male, female)=0.87, $P<0.001$).

Trait differences are greatest between the ancestral and the Asian population while flies from Europe are middle-sized and show likewise average values for all other studied traits. The Asian flies are on average heavier, have a smaller number of ovarioles, and smaller eggs. Despite the long-time laboratory rearing and standardized growing conditions, our *D. melanogaster* samples differed significantly at the morphological level, in agreement with previously reported geographic patterns (David *et al.* 1976). Flies derived from the Asian population exhibit characteristics of the Far East Race of *D. melanogaster*, namely high adult weight, large eggs, and a low ovariole number.

Table 1: Morphological characteristics of the African, European, and Asian sample of *D. melanogaster*. Displayed are the number of scored individuals (*n*), mean values, standard error (SE), coefficient of variance (CV), broad sense heritability (H^2), and the results of the nested ANOVA concerning the F-value of the line effect and the population effect. Not significant ns., $P<0.01$**, $P<0.001$***

	Africa					Europe					Asia					nested ANOVA	
Trait	*n*	mean	se	CV	H^2	*n*	mean	se	CV	H^2	*n*	mean	se	CV	H^2	pop effect *F*	line effect *F*
Male weight [mg]	40	0.68	0.02	10.26	0.66	35	0.69	0.01	7.24	0.53	34	0.72	0.01	5.12	0.44	55.000 ***	28.438 ***
Female weight [mg]	38	0.97	0.04	11.91	0.54	35	1	0.02	5.88	0.30	32	1.04	0.02	7.35	0.40	43.119 ***	23.294 ***
Ovariole number	60	14.71	0.32	6.82	0.36	57	14.84	0.24	5.49	0.22	59	13.94	0.23	5.79	0.28	83.8670 ***	21.5503 ***
Egg length [µm]	47	499.73	4.14	2.62	0.30	54	513.59	6.91	4.66	0.53	53	522.72	4.19	2.77	0.27	120.817 ***	22.637 ***
Egg width [µm]	47	178.05	1.5	2.67	0.31	54	184.12	1.15	2.16	0.25	53	188.87	1.81	3.32	0.31	134.244 ***	11.620 ***
Development [h]	50	319.9	3.36	3.32	0.22	60	321.76	3.53	3.8	0.27	60	325.72	10.15	3.12	0.22	2.6080 ns.	2.1782 **

Discussion

The studied *D. melanogaster* population samples from Africa, Europe, and Asia significantly differ morphologically. Flies from the Asian population exhibit the characteristics of the Far East Race and are most diverged from the ancestral population in Africa.

The genetic analysis by means of Approximate Bayesian computation (ABC) clearly favored the Single Colonization Scenario (SCS) where all non-African populations share a non-African common ancestor (Laurent *et al.* 2011). Although we find the phenotypic characteristics of the Far East Race there is no genetic evidence for a double colonization event and an ancient Asian population. This study strengthens the argument for recent genetic drift as main force driving divergence between European and Asian *D. melanogaster*. It is the first genetic characterization of the Far Eastern Race, as a combined study of the genotype and the phenotype has so far not been accomplished.

When observing the phenotypic values that match the colonization model with European flies showing an intermediate phenotype between African and Asian flies it gives the idea of correlated trait evolution. However, living species generally lack biological and statistical independence due to a common ancestry.

Chapter 3

Positive selection in the *Flotillin-2* region of European *D. melanogaster* (Unpublished work)

Annegret Werzner, Pavlos Pavlidis, Lino Ometto and Wolfgang Stephan

Localizing genes that are subject to recent positive selection by means of genome-wide analysis is a major goal of population geneticists. Within a population an ecologically favored allele might rise in frequency and even become universal to the population. This process can be caused by environmental changes or colonization events. In case the favored allele already exists in the population selection operates on standing genetic variation, otherwise a favored mutation must arise.

If selection is strong the fixation of a beneficial mutation can have a joint effect on linked neutral sites that are expected to comparably increase their frequencies. This process, known as genetic hitchhiking (Maynard Smith & Haigh 1974), generates a signal of 1) reduced genetic variation at the target of selection (Kaplan *et al.* 1989), 2) a skewed site frequency spectrum (SFS) due to an excess of rare and high-frequency derived alleles (Braverman *et al.* 1995; Fay & Wu 2000), and 3) increased LD on both sides of the target of selection but reduced LD between them (Kim & Nielsen 2004; Stephan *et al.* 2006; Jensen *et al.* 2007b).

The search for adaptive signals is usually carried out using a large number of loci on a genome-wide scale to identify regions that show considerable deviation from neutral expectations. As is generally known, population structure and demographic events such as

population bottlenecks or population size expansion can mimic the genetic footprint of selection and possibly lead to false positives (Fay & Wu 2000; Przeworski 2002; Jensen *et al.* 2005). Thus the best neutral expectation is provided by a demographic null model that is fitted to a species demographic history (Li & Stephan 2006).

Glinka *et al.* (2003) and Ometto *et al.* (2005) performed a chromosome-wide scan of DNA variation in the derived European population of *Drosophila melanogaster* and compared it to a putative ancestral population from East Africa. As a population sample they used 12 highly inbred isofemale fly lines from the Netherlands (Europe) and from Zimbabwe (Africa), respectively. More than 250 loci on the X chromosome were used to evaluate the contribution of adaptive evolution in the European population. The demographic null model of the European population was defined as a bottleneck due to the post-glacial colonization of Eurasia around 10,000-15,000 years ago (David & Capy 1988). The analysis revealed several candidate loci whose level of genetic variation could not be explained by demography alone. Genes within such a candidate region are expected to either show the spatial structure of standing genetic variation or a newly acquired allele due to genetic hitchhiking. Identifying the selected site can give a great insight into the process of functional adaptation.

In this follow-up study, we concentrate on one of these candidate regions, which is centered around the cytological position 13A1 between landmarks X: 14807688-X: 14829464, i.e. about 22 kb. Overall five genes are located within this position: *CG9504, CG9503, CG32591* (unknown function), *CG9009* and eight transcripts of *flotillin-2* (*Flo-2*). *CG9504* and *CG9503* are two out of 12 genes that were recently identified by amino acid sequence characteristics to belong to the GMC oxidoreductase family, clustered in a conserved order and orientation on the X chromosome of *D. melanogaster* (Iida *et al.* 2007).

The functions of these GMC homologues are still unknown except for *CG9504*, which was described as ecdyson oxidase (*Eo*, Takeuchi *et al.* 2005). An expression analysis showed an up-regulation of both genes, *CG9504* and *CG9503*, during embryonic and metamorphic development. This might indicate a joint function in the ecdysteroid metabolism (Iida *et al.* 2007). Remarkably, the GMC cluster is located within the second intron of the *Flo-2* gene with opposing transcription orientations. This arrangement is conserved throughout four distantly related insect species: *D. melanogaster*, *Anopheles gambiae*, *Apis melifera*, and *Tribolium castaneum* (Iida *et al.* 2007) suggesting strong purifying natural selection.

The *CG9009* encodes a long chain fatty acid CoA ligase and is one of seven starvation induced genes that are up-regulated in the case of food deprivation (Grönke *et al.* 2005). Generally, nutritionally regulated genes are expected to orchestrate the energy homeostasis control, including the ability to mobilize stored energy resources.

The *Flo-2* gene is evolutionary highly conserved among a wide range of species (orthologous to vertebrates *reggie 1*) and encodes for flotillin 2, a scaffolding protein that is involved in the formation of non-caveolar lipid rafts (Stuermer *et al.* 2001). It is a single transmembrane protein that is characterized by a short membrane-anchoring segment at the N terminal part of the protein and a large cytoplasmic C terminal domain (Neumann-Giesen *et al.* 2004). The exact mode of membrane association is dependent on co-translational protein modifications such as myristoylation and palmitoylation at highly conserved N terminal regions. If these regions are altered by mutations or the protein is truncated the ability of membrane anchoring is lost.

Flo-2 is expressed predominantly in neuronal structures such as the optic lobes and the central brain during all developmental stages of *D. melanogaster* (Hoehne *et al.* 2005).

Overexpression of *Flo-2* leads to detrimental effects during the development of eyes, ocelli, bristles, and wings while knockout mutants surprisingly show no noticeable phenotypic abnormalities. As *Flo-2* is the most prominent gene within this genomic section we will refer to the candidate region as the '*Flo-2* region'.

Assuming recent positive selection occurred, we intent to characterize the source of selection in the *Flo-2* region and narrow the region down to a single gene or gene variant. Using the full-length sequence of both population samples we applied two statistical tests for selection: 1) SweepFinder based on information of the SFS (Nielsen *et al.* 2005), and 2) the ω statistic based on measures of LD (Kim & Nielsen 2004). The demographic null-model was inferred by simulations from the colonization model that was suggested by Li and Stephan (2006).

Material and Methods

Fly strains and conditions of culture Intraspecific data were collected from highly inbred *D. melanogaster* lines, 12 derived from a European population (Leiden, The Netherlands) and 12 from an African population (Lake Kariba, Zimbabwe). All stocks were kept at 23°C, 45% humidity, and under constant light conditions. Development took place on a high-nutrient killed yeast food medium (12 ml) in glass vials of 200 ml. For interspecific comparison, we used the annotated sequence of *Drosophila sechellia* (http://flybase.org, Release 5.31; Tweedie *et al.* 2009).

Sequence data collection and analysis Primers were designed based on the *D. melanogaster* genome (FlyBase, Release 5.1). We amplified and sequenced the complete genomic region X:14810552-X:14829908 using 36 interdigitated fragments (Primers are given in Table (Appendix C) S 1). This fine-scale analysis allowed us to determine the target of selection down to the level of single genes. DNA sequences were obtained from individual male flies and generated as described in Glinka *et al.* (2003). Sequences were associated and assembled into contigs using the program Seqman of the DNAstar package (DNAstar, Madison, WI, USA). For aligning the entire sequence for all fly lines we used the multiple sequence comparison by the log-expectation alignment algorithm (muscle) kindly provided by http://mobyle.pasteur.fr/cgi-bin/portal.py (Néron *et al.* 2009).

Basic population genetic parameters were estimated using the program DnaSP 5.0 (Librado & Rozas 2009). We estimated nucleotide diversity using π (Tajima 1983) and θ (Watterson 1975). Based on the total numbers of segregating sites we estimated Tajima's *D* (Tajima 1989). The interspecific divergence to *D. sechellia* was determined for all 24 inbred lines of *D. melanogaster*. Furthermore we used *D. sechellia* to polarize the state of the

segregating sites in our population sample. A variant was considered ancestral if it was shared between both species and derived if it was only present in *D. melanogaster*. The intraspecific analysis revealed segregating sites that originated after the bottleneck (Ometto *et al.* 2005). These sites are private to the European population sample and are either fixed or polymorphic.

A phylogenetic analysis was used to study the ancestry of the European *Flo-2* alleles. In case of selection on standing genetic variation the close relation between one African allele and its European descendants would lead to a monophyletic group of one African line and the European lines. In case of selection on a newly emerged European allele the European lines would constitute a monophylum. The phylogeny was constructed using the program MEGA 4.1 (Tamura *et al.* 2007). The Neighbor-Joining option was used to construct the trees under the model: Nucleotide Number of Differences. Bootstrap tests with 500 replicates were performed. *D. sechellia* served as outgroup.

Demographic modeling and selective sweep analysis First, the European dataset was analyzed using the SweepFinder algorithm (Nielsen *et al.* 2005). Traditionally, SweepFinder compares the SFS of a small region of the genome (a ‘‘window’’) to the SFS of the rest of the chromosome which is considered as neutral. In this study, the neutral SFS has been estimated using simulations of the demographic model for the European population inferred by Li and Stephan (2006). For data analysis, we defined 1000 overlapping windows of fluctuating sizes according to the strength of selection and recombination rate.

Using a composite likelihood ratio (CLR) test we calculated the likelihood ratio (LR) for each of these 1000 windows along our sequence alignment for two models: a model without selection based on the neutral SFS vs. a model of a recent selective sweep (Williamson *et al.* 2007). Thus, we consider the spatial pattern of allele frequencies along the

studied genomic sequence, as predicted by a selective sweep model given the background pattern of a bottleneck. Incorporating the demographic history of the European population in the estimation of the null distribution controls the false-positive rate (Pavlidis *et al.* 2010). As the European sample lacks genetic variation for a large part of the sequence we included monomorphic sites in our analysis. We used the combined dataset of the European and African samples to determine the state of monomorphic European sites that were polymorphic in Africa (Williamson *et al.* 2007). This might increase the power for detecting the signature of a selective sweep.

Second, the European dataset was subjected to an analysis of LD using the ω statistic, developed by Kim and Nielsen (2004). The selective sweep model predicts elevated levels of LD within the two flanking regions of the selected site, but does not extend across the two regions. Thus, the sequence alignment was scanned for high values of ω indicating the signature of a selective sweep. Again, the dataset was split into 1000 overlapping windows that spanned a genomic region between 2000 and 10,000 bp. The borders of the two flanking regions are allowed to vary and the window size was picked according to the size for which ω assumes the maximum value (Pavlidis *et al.* 2010).

Statistical significance of both statistics, CLR_{max} and ω_{max}, was inferred from 9.000 neutral coalescent simulations using the demographic model of Li and Stephan (2006). For simulating population parameters we used the observed number of segregating sites of the combined dataset. Thus, the African population was used as a proxy for selective neutrality. The recombination rate (3.64×10^{-8}) was obtained from the *D. melanogaster* recombination rate calculator (Fiston-Lavier *et al.* 2010). Only the European subset of each simulation was used to assess the 95th percentile of the null distribution.

The ascertainment bias correction followed the approach of Thornton and Jensen (2007). In short: the genome scan of the X chromosome identified the *Flo-2* region as candidate for a selective sweep (Ometto *et al.* 2005). This result was based upon a fragment that completely lacked genetic variability over a range of 515 bp in the European dataset. This previous knowledge was used to create the null distribution for both statistics. All simulated European datasets harbored a region that lacked genetic variability for at least 500 bp.

RT-PCR for expression analysis Total RNA from third instar larvae and adults of *D. melanogaster* was isolated using the MasterPure^{TM} RNA Purification Kit (Epicentre® Biotechnologies, Madison, Wisconsin) according to the manufacturer's instructions. Residues of genomic DNA were removed from each RNA sample using DNase I, Amplification Grade (Invitrogen, Karlsruhe, Germany). Reverse transcription was carried out using the SuperScript III reverse transcriptase kit and Oligo$(dT)_{20}$ (Invitrogen, Karlsruhe, Germany). First strand cDNA was used as a template in PCR reactions. The following primer pairs were used to specifically amplify three different *Flo-2* transcripts: reg1 forward 5' AACCGATGTCCAGCGACTGT 3' and reg1 reverse 5' GGAATTGCTCGCTGGCAGTG 3' for amplifying transcript *Flo-2-A/E* and *Flo-2-B/F* (Hoehne *et al.* 2005); 1426 forward 5' ATCAGGCCGACGAGCTTC 3' and 1426 reverse 5' CGATGCCCATACGACCCA 3' for amplifying transcript *Flo-2- A/E*, *Flo-2-C*, and *Flo-2- B/F*; 1428 forward 5' GACACTCACACAACGAAACCACAC 3' and 1428 reverse 5' CAATATGGCTCTCAAATGTCCTTC 3' for exclusively amplifying *Flo-2-C*. PCR conditions were: 94°C for 4 min followed by 40 cycles at 94°C for 1 min, annealing for 1 min and 72°C for 1 min and finally 72°C for 4 min. Annealing temperatures were 55°C for reg1forward/reverse, 57°C for 1426 forward/reverse and 55°C for 1428 forward/reverse.

Results

Sequence data collection and analysis across the *Flo-2* region To investigate the signal of positive selection in the European sample of *D. melanogaster* we sequenced the entire candidate region (discovered by Ometto *et al.* 2005) in two population samples: 12 inbred lines from the Netherlands (Leiden) and 12 inbred lines from the ancestral African range (Zimbabwe). The sequence comprises large parts of the *flotillin-2* gene; *i.e.* 20,011 bp in total (Figure 1).

A sliding window analysis (window size of 1000 bp, 500 bp overlap) showed a conspicuous decline of genetic variability for the European sample, leading to a valley of around 10 kb in size (6.8 kb-16.5 kb, line with black square, Figure 1). This region is below the mean value of genetic variability of the entire X chromosome ($\theta_{X\text{-}Europe}$=0.0044, fine dashed horizontal line in Figure 1; Glinka *et al.* 2003) while the flanking parts steadily increase to levels of variation common for European *D. melanogaster*. The corresponding genetic region in the African sample matches the mean value of the X chromosome without any appreciable decline ($\theta_{X\text{-}Africa}$=0.012, wide dashed horizontal line, Figure 1). Our fully sequenced segment covers several genes including *Eo*, *CG9503 CG32591*, *CG9009* and eight different transcripts of *Flo-2* (Figure 1, Gene map, from left to right). Nonetheless, functional constraint as the sole cause of this pattern of genetic variability in the European sample can be ruled out as the level of genetic variation of the African sample is intermediate to high and divergence to the sibling species *D. sechellia* remains constantly high at around 6%. Tajima's *D* statistic is strongly negative for the European sample and values directly flanking the valley of reduced variation were significantly different from zero (Figure 1, D Europe, solid line). The African values fluctuate around zero with no significant departure (Figure 1, D Africa,

dashed line). The difference in the distribution of variation in the European and African sample suggests that this phenomenon is unique to the European sample.

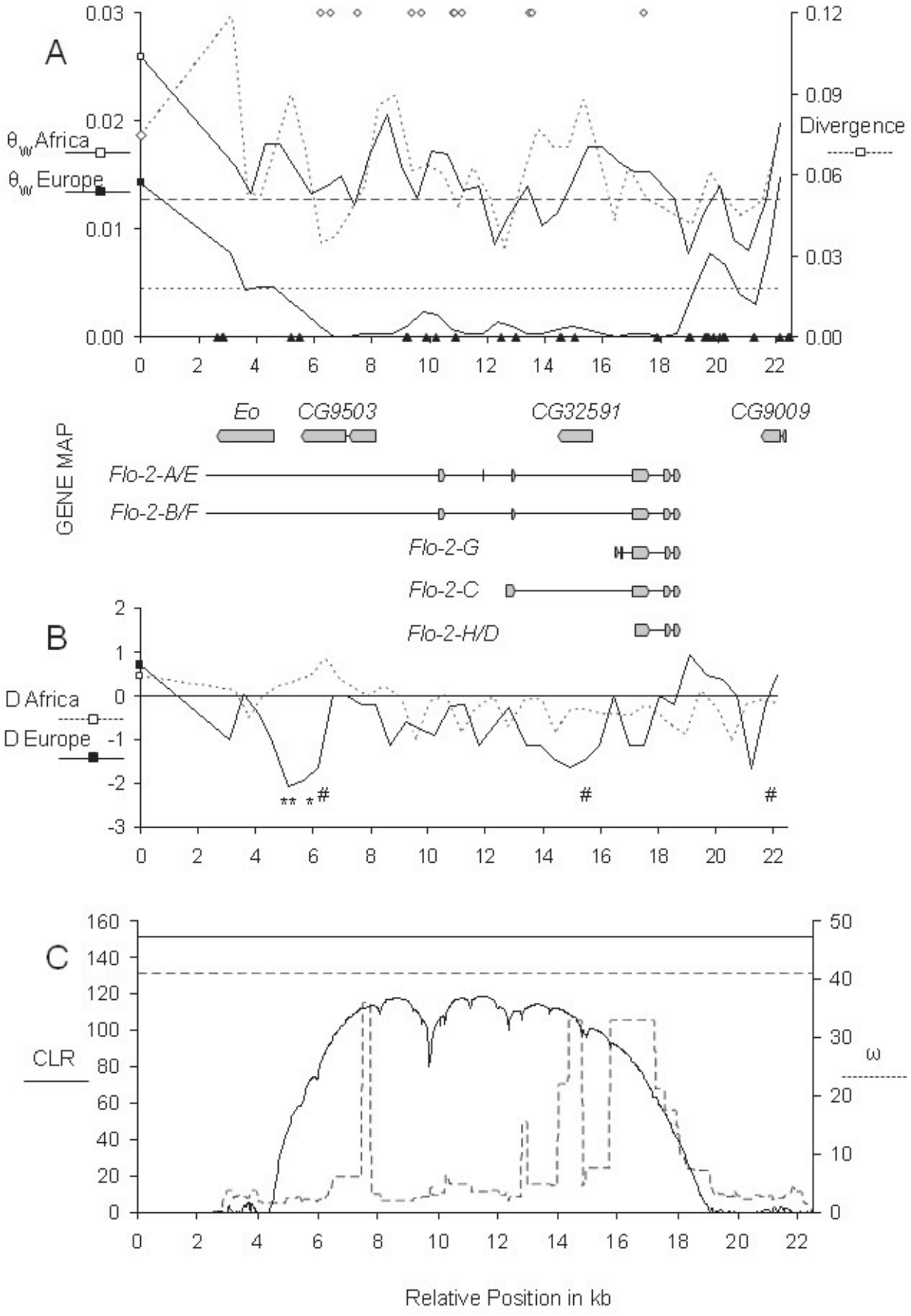

A
θw Africa
θw Europe
Divergence
0.03
0.02
0.01
0.00
0.12
0.09
0.06
0.03
0.00
0 2 4 6 8 10 12 14 16 18 20 22
GENE MAP
Eo
CG9503
CG32591
CG9009
Flo-2-A/E
Flo-2-B/F
Flo-2-G
Flo-2-C
Flo-2-H/D
B
D Africa
D Europe
2
1
0
-1
-2
-3
** * #
#
#
0 2 4 6 8 10 12 14 16 18 20 22
C
CLR
ω
160
140
120
100
80
60
40
20
0
50
40
30
20
10
0
0 2 4 6 8 10 12 14 16 18 20 22
Relative Position in kb

Figure 1: The valley of reduced genetic variation in a European sample of *D. melanogaster*

A) The valley of reduced genetic variation in a European sample of *D. melanogaster* compared to an ancestral population sample from Africa (left y-axis). To the left the fully sequenced 20011 bp are connected with an adjacent fragment of Ometto *et al.* (2005), indicated by squares. The x-axis shows the relative position on the X-chromosome. Black triangles show the position of indels, white diamonds indicate fixed nucleotide differences between both samples. The second y-axis corresponds to divergence of *D. melanogaster* to *D. sechellia* (dashed line). The gene map gives the name and the position of the coding sequence of all genes located within the valley of reduced variation. These genes are: *Eo*, *CG9503* *CG32591*, *CG9009* and eight different transcripts of *Flo-2*.

B) Tajima's *D* statistic is shown for the European (solid line) and African sample (dashed line) of *D. melanogaster*. Significance values correspond to the following signs: # $P<0.10$; * $P<0.05$; ** $P<0.01$.

C) The sweep analysis reveals two profiles: the solid line shows the result of SweepFinder, the dashed line shows the result of the ω statistic. After ascertainment bias correction both statistics are not significant (SweepFinder $P=0.096$, ω $P=0.064$).

The result of the phylogenetic analysis shows that the European fly lines belong to one monophyletic group, which is highly supported by a bootstrap value of 100 (black arrow, Figure 2). Thus, the European clade clearly separates from the paraphyletic African group.

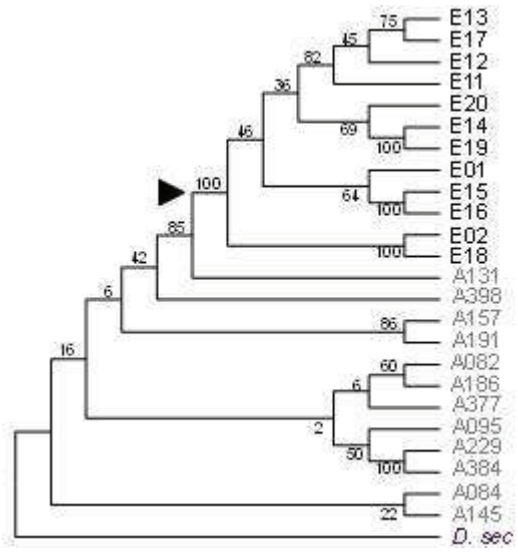


Figure 2: Neighbor-joining tree of European (black) and African (grey) *D. melanogaster*
For visualizing the data we present a neighbor-joining tree of European (black) and African (grey) *D. melanogaster* based on number of nucleotide differences. This tree is an original tree that only displays topology; numbers are bootstrap values for 500 replicates; seed = 64238. *D. sechellia* was used as outgroup. The European fly lines form a highly supported monophylum (black arrow) and the African fly lines are paraphyletic.

Selective sweep analysis The CLR statistic for the entire *Flo-2* region falls below the 5% significance threshold obtained by neutral simulations (see Materials and Methods) (Figure 1, CLR, threshold solid horizontal line, P=0.096). The vast elevated values of the CLR statistic for the valley of reduced variation clearly deviate from the flanking parts of the region, but the population bottleneck inferred by Li & Stephan (2006) regularly creates similar patterns of SNP distributions.

The profile of the ω statistic is shown in Figure 1. In total the ω statistic shows three peaks: the first is associated with the first exon of *CG9503* at relative position 7600, a gene of the GMC oxidoreductase cluster of *D. melanogaster*, the second peak is located at relative position 14,600 and is associated with the coding gene *CG32591*, the third peak is centered around relative position 16,500 and spans the intronic region of *Flo-2* including the first and

second exon of splicing form *Flo-2-G*. None of these peaks exceeds the 5% significance threshold.

Private European alleles Geographically restricted genetic hitchhiking suggests that the target of selection should be fixed in the European population sample and absent in the African population sample. Overall there are 11 fixed different nucleotide substitutions between the African and the European sample within the region of strongly reduced variation (1-11, Table 1, white diamonds Figure 1). Four of them are located within the exon of a gene while the remaining seven are located within introns. In eight cases the European population differs from the ancestral state of the African population (Africa and *D. sechellia* share the same state). These sites form private European alleles.

Table 2: Fixed differences between the European and African sample of *D. melanogaster* for the *Flo-2* region. The derived state is highlighted in grey.

nr.	abs. position	rel. position	location	Outgroup	Africa	Europe	consequence
1	14814106	6218	*CG9503*	G	G	A	Ser → Asn
2	14814438	6550	*CG9503*	T	C	T	syn. (Phe)
3	14815391	7504	*CG9503*	C	T	C	syn. (Asn)
4	14817203	9356	Intron	G	G	A	
5	14817533	9713	Intron	A	A	G	
6	14818608	10798	Intron	A	A	C	
7	14818649	10839	Intron	G	G	T	
8	14818898	11107	Intron	G	G	A	
9	14821087	13429	Intron	C	T	C	
10	14821196	13539	Intron	C	C	T	
11	14824970	17399	*Flo-2*	C	C	A	syn. (Ala)
12	14825835	12986	*Flo-2*	-	-	Ins.	Additional AS: THTHTNT

CG9503 Three fixed differences are located within the gene *CG9503*: one within the first exon, two within the second exon. Within the second exon the substitution that is private to the European sample leads to an alteration of the amino acid sequence. A serin (Ser = ancestral state) is substituted by an asparagin (Asn) at the amino acid position 473. This conservative substitution is not expected to alter the gene function as both amino acids have equivalent characteristics (hydrophil, aliphatic, polar, neutral). The two other fixed differences are synonymous substitutions in the African sample while the European alleles are of ancestral state.

Flotillin-2 One fixed difference is located at a synonymous site of the last but two exon of *Flo-2*. We also found a fixed difference between the European and African samples that is an insertion polymorphism due to a dinucleotide microsatellite within the first exon of the transcript C of the *Flo-2* gene (Figure 1, gene map, Flo-2-C). In total there are 38 indels spread along the Flo-2 region (Figure 1, black triangles on the x-axis) but only one is fixed between the population samples (Table 1, 12). In the European sample the transcript is extended by 20 nucleotides.

The annotated version of transcript form Flo-2-C at Flybase (version FB2010_05, released May 28th, 2010) is strongly supported by a full-length cDNA clone generated and sequenced by Rubin et al. (2000). The sequence had been subjected to integrity checks for accuracy, the presence of an open reading frame (ORF) as well as the start and a stop codon (ATG, TGA). The prediction of the gene structure, particularly of the 5' and 3' untranslated regions (UTRs), were based on the full-length cDNA and are thus highly reliable. Interestingly not all of our studied *D. melanogaster* lineages have an ORF

Table (Appendix C) S 2, ■, Flo-2-C alignment starts at transcription site +46,). Most African lines either show premature stop codons (*) or frame shifts (X) due to microsatellite length polymorphism (9/12). Nonetheless three African lines have a gene version that, at least predictably, forms the Flo-2-C splicing form in the correct way (Table (Appendix C) S 1, ►). In the European sample a probably functional Flo-2-C is even more frequent as the ORF is undamaged (10/12). However, two European lines do not have a functional *Flo-2-C* as their ORF was destroyed by another insertion or recombination event (line 11 and 13, Figure 3, Table (Appendix C) S 1). *D. sechellia* likewise has a microsatellite at the corresponding position. It lacks an ORF for transcript *Flo-2-C* as well as the insertion private to the European *D. melanogaster* sample (Table (Appendix C) S 1).

Expression analysis Expression pattern was studied using third instar larvae and adults of all 24 *D. melanogaster* lines. Using transcript specific primers we can show that the transcripts *Flo-2-A/E* and *Flo-2-B/F* are continuously expressed in larvae and adult flies in both population samples (Figure 3). We have verified the results of Hoehne *et al.* (2005) for all our lines concluding that the short (*reggie1a/Flo-2B/Fa*) and the long splicing form (*reggie1b/Flo-2A/Eb*) are present in our samples. Comparison between expression levels of both population samples reveals that in the larval stage population differences are marginal whereas in the adult stage some European lines show elevated expression levels particularly for the short *Flo-2B/Fa*. In adult flies *Flo-2B/Fa* generally seems to be predominant, whereas in larvae both splicing forms are equally expressed.

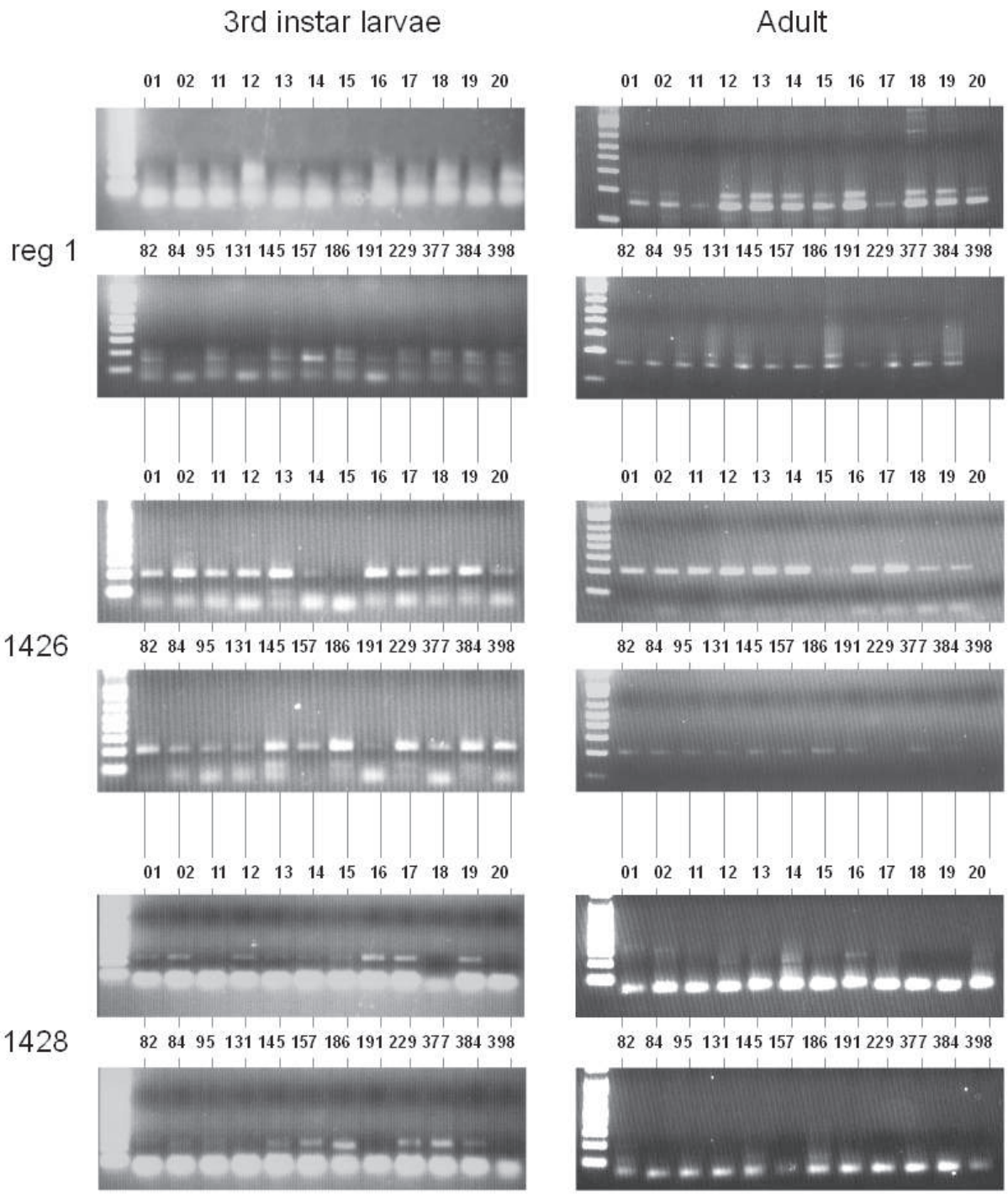
Fragment 1426 reveals expression patterns of five splicing forms: *Flo-2-A/E*, *Flo-2-C*, and *Flo-2-B/F*. Again, all lines show sustained expression in both studied developmental

stages. In larvae there are no detectable population sample differences. In adults, however, the expression level is overall higher in the European sample.

The RT-PCR product of primer 1428 presents the expression pattern of the transcript *Flo-2-C*. In the larval stage transcript *Flo-2-C* was detected in most fly lines of both population samples, except for European lines 11, 18, 20 and African lines 131, 191, 398. Despite a disordered gene structure the gene version *Flo-2-C* is transcribed. In adult flies the transcript is below the detection threshold. Thus, *Flo-2-C* was not proven to be expressed in this developmental stage.

Figure 3: Expression pattern of *Flo-2*

Expression pattern of all studied *Flo-2* transcripts in third instar larvae (left column) and adult (right column) *D. melanogaster*. Primer pairs correspond to the following transcripts: reg 1- *Flo-2-A/E* and *Flo-2-B/F*, 1426- *Flo-2-A/E*, *Flo-2-B/F,* and *Flo-2-C*, 1428- *Flo-2-C*. European lines 01 to 20 (upper picture) and African lines 82 to 398 (lower picture). *Flo-2-A/E* and *Flo-2-B/F* are continuously expressed in larvae and adult flies in both population samples whereas *Flo-2-C* was exclusively detected in third instar larvae (except for European lines 11, 18, 20 and African lines 131, 191, 398).


3rd instar larvae
Adult
reg 1
1426
1428
01 02 11 12 13 14 15 16 17 18 19 20
82 84 95 131 145 157 186 191 229 377 384 398

Discussion

Sequence data analysis across the *Flo-2* region Ometto *et al.* 2005 analyzed X-linked chromosome-wide patterns of DNA variation in the European population of *D. melanogaster* finding candidate loci that deviated from the demographic null model. In a follow-up study, we used one of these candidates to characterize the detailed pattern of nucleotide diversity and to identify the genetic target of positive selection.

Nucleotide diversity of the fully sequenced *Flo-2* region in the European sample shows a sweep-like pattern: a valley of reduced genetic variation of around 10 kb in size, and flanking parts that steadily increase to neutral levels of European variation. Tajima's *D* statistic is strongly negative for the European and some values were significantly different from zero. The *Flo-2* region is characterized by a high density of genes that are functionally and structurally well-conserved, suggesting persistent evolutionary constraint. Divergence to *D. sechellia*, however, remains constantly high at around 6% and excludes purifying selection as major cause of low genetic diversity. Furthermore, the corresponding region in the African sample matches the mean values for X-linked nucleotide diversity without any appreciable decline and Tajima's *D* is not significantly different from zero. The African population as source for the European genetic depletion can thus be ruled out.

Evidence for an African selective sweep in *D. melanogaster* was recently found by Svetec *et al.* (2009). The ecological relevant allele encoding a *D. melanogaster* specific version of the gene *HDAC6* had been fixed in East Africa and was than exported to Europe during colonization. The bottleneck experienced by the European population enlarged the dimension of the ancestral sweep on either side and generated a signal of strong positive selection (window 47, Li & Stephan 2006). This, however, seems not to be the case for the

Flo-2 region. A phylogenetic analysis revealed the European lines to constitute a monophylum although the genealogy of the data may not be tree-like due to recombination. Evidence for the derivation of the beneficial European allele from an African line is very weak as all European lines can be traced back to one common ancestor. Within the *Flo-2* region there are eleven fixed nucleotide differences between the population samples, eight of which are of the derived state in Europe. According to this result, the European population evolved more or less independently after the colonization and locally adapted to the new environment by European genetic novelties. This agrees well with previous studies of the colonization history of *D. melanogaster*. Using a broad population sample Baudry *et al.* (2004) suggested a single founder event for non-African populations and identified several private derived variants, called 'neomutations', which potentially drive local adaptation. These mutations are likely to have appeared in the derived populations after the bottleneck and should be associated with haplotypes of intermediate to high frequency.

Selective sweep analysis We used two statistical tests for the sweep analysis of the *Flo-2* region: ω and SweepFinder. Both statistics capture different aspects of the data: ω is based on LD, while the CLR of SweepFinder is dependent on the SFS. Statistical significance was inferred using neutral coalescent simulations that were based on the demographic model of Li & Stephan (2006) and corrected for an ascertainment bias (Thornton & Jensen 2007). Neither SweepFinder nor ω revealed significant results.

In non-equilibrium populations Pavlidis *et al.* (2010) showed that the combined action of selective sweeps and bottlenecks result in the loss of high-frequency derived alleles, greatly affecting the SFS. We compensate this by including the monomorphic European sites that were polymorphic in Africa, extending the SFS for two allelic classes: 0/12 (polymorphic site

in African sample is missing in European sample) and 12/12 (polymorphic site in African sample is fixed in European sample). Thus, the skew in the unfolded SFS is restored, which increases the power for detecting the signature of a selective sweep (Williamson *et al.* 2007). Given the rough ascertainment bias correction it is, however, likely to lose the power to distinguish the sweep pattern from such severe demographic background.

Thornton & Andolfatto (2006) estimated the age and the size of the bottleneck using X-linked data of the Netherlands population. The analysis indicated that the bottleneck started approximately 16,000 years ago with an effective population size of 38,000 individuals. Laurent *et al.* (2011) showed that the Asian and European populations of *D. melanogaster* share a most recent common ancestor that split off from the African population around 16,800 years ago with effectively 22,000 individuals without recurrent migration from Africa. Thus, both studies suggest a by far higher number of founder individuals than the model of Li & Stephan (2006) and it is questionable whether this model is the best choice for the demographic history of in the Netherlands population.

The profile of the CLR of SweepFinder maximizes at regions of low genetic diversity spanning 14 kb of the entire *Flo-2* region. By comparison, the ω profile is more restricted: one peak is associated with the gene *CG9503*, and two close peaks are associated with the gene *CG32591* and the intron as well as coding parts of *Flo-2*. In non-equilibrium populations SweepFinder is more accurate in discriminating selection from neutrality and thus more reliable (Pavlidis *et al.* 2010). We further discovered that the first peak of the ω statistic is based on singletons and low frequency derived variants that are private to the European population. The source of these variants is speculative. They can either be younger than the selective sweep or they represent ancestral polymorphism that is absent in the African sample

due to restricted sample size. The first case would violate the assumptions of the ω-statistic and screw up the results. When all European polymorphic sites of unpolarized state are removed from the dataset the first peak is lost whereas the second and the third peak merge and have the exact same ω value as before (Figure (Appendix C) S 1). Therefore the second and the third peak of the ω-statistic are more robust to any kind of scenario that can lead to private European polymorphism.

Both statistics are expected to maximize at the target of selection. In non-equilibrium models this was shown to be a difficult task since the predictions of the target mostly were close to random expectations (Pavlidis *et al.* 2010). Therefore we concentrate on fixed differences between the two population samples.

European genetic novelties The European population sample has eight fixed differences between populations that differ from the ancestral state, either from the African sample or from *D. sechellia*, and are therefore classified as private European alleles. Their majority (six out of eight) is located in intronic regions of the *Flo-2* gene. Intron mediated gene differences are usually apparent at the level of expression. Positive and negative regulation has been observed in several species, such as *Drosophila* (Meredith & Storti 1993), mice (Choi *et al.* 1991), and maize (Callis *et al.* 1987). A gene expression analysis of our European and African population samples revealed no significant differences between population samples for males (Hutter *et al.* 2008) and a marginal overexpression of *Eo* in females (Müller *et al.* 2011). Furthermore, intron related expression alteration can be caused by splice site mutations. None of the observed substitutions hit the splicing relevant intron/exon boundary known as splice donor and splice acceptor site or the branch point.

GMC oxidoreductase cluster: The second intron of the *Flo-2* gene harbors the nested gene architecture of GMC oxidoreductases, which is remarkably well-conserved in insects (Iida *et al.* 2007). This gene cluster comprises 12 GMC genes, including the ecdysone oxidase *Eo* and the gene *CG9503*, that all might have an orchestrated function in ecdysteroid metabolism. Expression analysis of the GMC cluster in different developmental stages of *D. melanogaster* revealed varying patterns of temporal expression: the two genes located within the *Flo-2* region, *Eo* and *CG9503*, are highly expressed during embryonic and metamorphic development (Iida *et al.* 2007). Analogous results were found for the expression pattern of *Flo-2* in zebrafish (von Philipsborn *et al.* 2005). In the early fish embryo the *Flo-2* mRNA and protein are abundant in proliferating tissue in all growing structures whereas in late larval stages flotillin becomes a predominantly neuronal protein. This seems to be applicable to *Drosophila* since *Flo-2* is expressed in all developmental stages but experiences a restricted tissue specific expression to optic lobes, central brain, and neuronal structures in later developmental stages. Tissue specificity was also shown for the GMC genes: *Eo* is mainly expressed in the midgut of third instar larvae of *D. melanogaster* whereas *CG9503* is expressed in the wing disc. Interestingly the overexpression of *Flo-2* leads to detrimental effects during the development of wings (Hoehne *et al.* 2005). The evolutionary conservation of the clustered gene architecture and the very similar catalytic activities of the GMC oxidoreductases suggest a joint transcriptional regulation of the GMC complex and *Flo-2*. It has been suggested that *cis*-acting control elements that might be located within or near by this cluster coordinately regulate the GMC oxidoreductases and *Flo-2* in support of a common function. A closer examination of intergenic and intronic regions around the GMC oxidoreductase cluster searching for regulatory elements that influence gene expression will help clarify this issue.

Flo-2-C *Flo-2* is one of two *flotillin* genes in *D. melanogaster* that probably evolved by gene or genome duplication of *Proto-Reggie1/2* in a common ancestor of protostomes and deuterostomes (Malaga-Trillo *et al.* 2002). In *Drosophila* the *Flo-2* gene was subject to intron losses and exon deletion that severely altered the gene structure. Nonetheless the amino acid sequence still has around 60% similarities between *D. melanogaster* and vertebrates.

During the last 2-3 Ma of *D. melanogaster* divergence the *Flo-2* gene experienced a comprehensive transcript diversification. One private *D. melanogaster* transcript is strongly supported by a full-length cDNA clone, referred to as *Flo-2-C* (Rubin *et al.* 2000). This transcript is not ubiquitous in all of our studied *D. melanogaster* lineages. Most African lines either show premature stop codons or frame shifts due to microsatellite length polymorphism (9/12) and two European lines lack a functional *Flo-2-C* as well (line 11 and 13, Figure 3).

Nonetheless three African lines have an allele that, at least predictably, forms the *Flo-2-C* splicing form in the correct way. In the European sample a probably functional *Flo-2-C* is even more frequent as the ORF is undamaged (10/12). We could show that the high frequency of intact European *Flo-2-C* transcripts is due to a microsatellite that is extended by 20 bp. While the transcripts *Flo-2-A/E* and *Flo-2-B/F* are continuously expressed in larvae and adult flies in both population samples we found the *Flo-2-C* to be temporarily expressed in larvae. This is in accordance with the results of Rubin *et al.* (2000) who reported the mRNA to be present in larvae and pupae of *D. melanogaster*. Expression pattern of *Flo-2-C* does not follow the pattern of structural gene integrity which leads to the conclusion that it can be transcribed although it has a disordered gene structure. Such damaged mRNAs are efficiently recognized by mRNA surveillance (also known as nonsense-mediated mRNA decay) that causes their accelerated degradation (reviewed in Wilusz *et al.* 2001).

The function of *Flo-2-C* remains a matter of speculation. Neumann-Giesen *et al.* (2004) showed that a truncated flotillin-2, such as *Flo-2-C*, can not anchor in membrane rafts but still has the ability to form homo-oligomers that in turn seem to enhance membrane association with full-length flotillin-2 variants. Given the fact that *Flo-2-C* is a potentially functional transcript that is private to the *D. melanogaster* lineage but is still not structurally fixed within the species it could be a case of gene diversification. *Flo-2-C* represents a young gene variant that has arisen de novo on the *D. melanogaster* lineage.

General Discussion

The aim of this thesis was to detect positive selection in a European population of *D. melanogaster* by going down to the genetic level and up again to ecologically important traits. We combined QTL and selective sweep mapping in respect to the demographic model for our population sample. Furthermore we extended the demographic model of *D. melanogaster* by an Asian population from Malaysia and clarified the origin of the morphologically distinct Far East Race. A candidate selective sweep in the European population was shown to point at a new gene transcript that is private to the D. melanogaster lineage. Taken together, this thesis provides new information on cold adaptation, demography, morphological differentiation, and early gene development in a model insect.

Combined QTL and selective sweep mapping Many ecologically interesting traits show continuous variation within natural populations due to the assembly of segregating alleles at several loci as well as allelic effects across loci. Chapter one summarizes a QTL study that identified multiple cold tolerance factors encoded on the X chromosome; the first verification since Tucić (1979) associated the major chromosomes with this stress-resistance trait. To achieve the aim of combining QTL and selective sweep mapping, we studied two natural populations of *D. melanogaster*: one from Africa and one from Europe. We crossed one European fly line with an African fly line and established a population of X-chromosomal recombinant (XR) lines for performing the QTL mapping. In a subsequent step we compared our QTL map with a previously generated selective sweep map from the European population (Li & Stephan 2006) searching for sweep-QTL colocalization. We identified the gene *CG16700* that fits all criteria: *i)* it is located within a QTL in both males and females; *ii)* it is

associated with cold tolerance (Ayroles *et al.* 2009), and iii) it is shaped by positive directional selection exclusively in the European population (window 55, Li & Stephan 2006). *CG16700* belongs to the proton-coupled amino acid permease family SLC36 (solute-linked carrier 36, Romero-Calderón *et al.* 2007). Based on sequence similarity, it is predicted to be involved in amino acid transmembrane transport and its expression in *D. melanogaster* is restricted to fly heads.

QTL and selective sweep mapping are two methods that have been used cooperatively in several studies (Wang *et al.* 2005a; Tian *et al.* 2009). To our knowledge, however, the analysis of the X chromosome for cold tolerance is the very first study that successfully reconciles both approaches within the framework of two natural population samples and corresponding XR lines as well as the first European selective sweep in *D. melanogaster* that does not originate from Africa.

Cold tolerance Cold tolerance can be measured in two ways, either as survival rate or recovery time from a cold shock (David *et al.* 1998). In our case the CCRT was a convenient measure to provide a quantitative variable that is strongly correlated with cold adaptation in *Drosophila* and to capture marginal differences between XR lines (Gibert *et al.* 2001). 'Chilling' is defined as cooling state that is rapidly induced by exposure to temperatures between 5°C and 0°C in *Drosophila* (Czajka & Lee 1990). It involves a cascade of molecular mechanisms that are either activated during exposure to cold or during the recovery phase (reviewed in Clark & Worland 2008).

Some species delay chilling by producing heat through shivering with their flight muscles. For example winter moths are able to heat their flight muscles from 0°C to 36°C, the temperature that is required for flying (Heinrich 1987). The queens of the bee genus *Bombus*

even exceed this range: they are capable of heating their muscles from an ambient temperature of -3.6°C to the desired 32-36°C body temperature (Bruggemann, 1958, cited in Heinrich, 1979). The inability to move, however, is the result of flight muscle inactivation. When the electrical activity drops to 0 the insect enters chilling.

Ecological relevance of cold tolerance in *Drosophila* In Europe winter time demonstrably challenges *Drosophila* and adaptation to coldness often becomes a matter of survival. Flies are exposed to a wide range of low temperatures that besides chilling also initiate freezing and supercooling. Freezing is the state when the body water crystallizes. In *Drosophila* it generally occurs at temperatures below -20°C and mostly has fatal consequences (Czajka & Lee 1990). Supercooling delays freezing and is omnipresent in insects (Salt 1961). It ranges from temperatures below 0°C to the whole body supercooling point (SCP), the temperature right before freezing. For insects that do not tolerate freezing the SCP defines the lethal limit. Supercooling and chilling can be summarized as effects to non-freezing temperatures.

The defense against a rapid cold shock is immediately available to the fly. It is generally known that insects have the ability to considerably increase their cold tolerance by cold acclimation or seasonal cold hardening (Salt 1961; Czajka & Lee 1990). Furthermore the rearing temperature of flies significantly influences parameters including wing shape (Cavicchi *et al.* 1978; Frazier *et al.* 2008), body size (David *et al.* 1994; Frazier *et al.* 2008), and pigmentation (Gibert *et al.* 2000). However, this is not the case in our study since fly lines were not acclimated or hardened; in fact all lines experienced constant standardized conditions.

In any case severity of the cold treatment is dependent on the temperature as well as the duration of exposure. Recovery time from chill coma significantly increases with the time of exposure to chilling temperatures (David *et al.* 1998; Colinet *et al.* 2010; Gibert *et al.* 2001). It therefore is crucial to determine a species-specific chilling duration which brings it to the range of high individual discrimination, yet without accelerated mortality. For our population samples we exposed the flies to 0°C for 7 hours (h).

Gibert *et al.* (2001) studied different *Drosophila* species with temperate, tropical, and cosmopolitan distribution. Temperate species like *Drosophila subobscura* were treated for 24 h at 0°C and easily recovered; the tropical species *Drosophila ananassae* in turn was pushed to its limit after 6 h of exposure. The recovery time of our *D. melanogaster* population sample from Europe exceeds the reported values of a 7 h treatment of French flies by 10 minutes and our African sample agrees with the tropical *D. ananassae*. The comparative analysis of Gibert *et al.* (2001) showed that climatic adaptation evolved independently in the genus *Drosophila* and species with similar environmental conditions have the same tolerance levels to temperature extremes. Particularly for cosmopolitan species it is an essential ability to evolve sophisticated mechanisms to cope with environmental stress. *D. melanogaster*, however, has a particular willingness to enter human houses to escape inhospitable conditions. The sheltering nature of human commensalism is a reasonable explanation why this species still exhibit large parts of its tropical phenotypes.

Colonization model The worldwide distribution of *D. melanogaster* is the consequence of long distance dispersal and successful colonization. Chapter two provides a new colonization model including African, European, and Asian population samples. Using the ABC approach we established a single colonization scenario (SCS) where *D.*

melanogaster left the ancestral African range around 16,800 years ago (X chromosomal based estimation) and gave rise to a single bottlenecked Eurasian population. The European and Asian populations diverged approximately 2,500 to 5000 years ago leading to a recurrent bottleneck in the Asian population. There is agreement among scientists that *D. melanogaster* left Africa around 16,000 years ago, in the course of which they colonized Eurasia (Laurent *et al.* 2011; Li & Stephan 2006; Thornton & Andolfatto 2006). The colonization time seems reasonable as it correlates with the end of the glaciation that until then occupied Europe. But maybe the European climate change only is half the story. Jeschke & Strayer (2006) found out that the crucial feature of a successful invader is the strong association with humans. Till these days *D. melanogaster* benefits from constantly available food resources unintentionally put at disposal by humans but it has not always been a commensal. Lachaise & Silvain (2004) suggested the time during the 'post-Aterian hyperarid' phase 18,000 and 12,000 years ago when *D. melanogaster* shifted from a forest-dwelling species to human commensalism. At the time when Europe entered the recent interglacial period of the Holocene *D. melanogaster* already possessed an affinity for human settlements, which set the stage for the colonization of Europe.

A similar founder event was detected in *D. subobscura*, a species that started to disperse from its European home range very recently (Brncic *et al.* 1981). Pascual *et al* (2007) applied the ABC approach to microsatellites and inferred a serial colonization model comprising long-distance dispersal from Europe to Chile and further to North America with two subsequent bottlenecks. Genetic variability was significantly reduced in both non-European populations resulting from a severe bottleneck during colonization of the American continent. The second bottleneck experienced by the North American population was much lower and let a high fraction of South American alleles passing through. *D. subobscura* is a

typically palearctic species and well adapted to the temperate environments of Chile and North America. Nevertheless its successful and rapid invasion was also due to its casual association to humans, if not during its entire life then eventually at crucial moments.

Application of the newly established colonization model The Asian population of *D. melanogaster* is genetically highly structured and less variable than other non-African populations such as Europe and America (Schlötterer *et al.* 2006); and there is evidence for considerable morphological divergence between Asian and non-Asian populations (David & Capy 1988; Capy *et al.* 1993). It has formerly been suggested that the Asian population was founded by an ancient colonization from Africa that pre-dated the European colonization. Over time genetic drift and adaptive pressures could have developed the flies into a morphologically unique 'Race' from the Far East (Capy *et al.* 1993). The second part of chapter two summarizes our study of the Far East Race in the light of our new colonization model.

The population sample from Kuala Lumpur exhibits most of the characters from the 'Far East', except for a longer developmental time. The morphological analysis placed European flies between the African and Asian ones. Using our newly established colonization model we rule out that an ancient colonization event caused the distinct phenotype of flies from Kuala Lumpur. Both, the molecular and morphological data define the European population to be the intermediate state and connecting piece for African and Asian flies. Given the serial colonization from Africa to Europe and further to Asia, the small Asian population size as well as low levels of migration (Schlötterer *et al.* 2006) promote genetic drift leading to divergence between Asian and European *D. melanogaster*, and furthermore to divergence between Asian subpopulations.

None of our population samples from Europe and Asia show recent admixture of native and ancestral alleles, yet admixture is present in other populations of *D. melanogaster*. Caracristi & Schlötterer (2003) analyzed the American population of *D. melanogaster* which was founded in the middle of the 19th century from migrants of the European continent. Recent molecular data, however, places America between Europe and Africa and reflects a more complex demographic history. The North American population is genetically more diverse than the European one, has more private alleles, and therefore does not fit a simple serial founder event like the Asian population. Despite the European origin it might be fed by populations from the Caribbean, which transfer African alleles to the American East Coast on a regular basis.

Recent positive selection at a neuronal locus in European *D. melanogaster* The pattern of genetic variability created by a selective sweep is complex and ephemeral. In chapter three we present a recent selective sweep that is unique to the European population of *D. melanogaster*, pointing to the *flotillin-2* gene. Our study clearly demonstrates the importance of an appropriate demographic null-model for detecting selective sweeps and verifying them statistically. As shown by Pavlidis *et al.* (2010) the demographic effect of a bottleneck confounds the signatures of a selective sweep, in particular when it is as rapid and severe as the one inferred by Li & Stephan (2006). In cases of very strong selection the sweep discovery rate was generally high but for weak selection it considerably decreases. Furthermore it proved difficult to estimate the position of the selected site. In cases of strong selection estimates largely deviated from the true target of selection due to the lack of informative polymorphism. For a shallower bottleneck the discovery rate is slightly higher, but the target of selection is still estimated poorly.

Since we established a new demographic model for the European and Asian population samples we do not find evidence for parameters such extreme as inferred by Li & Stephan (2006). Our model predicts the Eurasian founder population to be one order of magnitude larger that that of Li & Stephan (2006) and is thus very similar to the model of Thornton & Andolfatto (2006). If the analysis is repeated using the model of Laurent *et al.* (2011), the results of the *Flo-2* region may improve significantly, providing another example of an environmental specific selective sweep.

Complex traits in natural populations Detecting the signature of positive selection is a challenging task since most ecologically interesting traits are complex and encoded by multiple genes. Recently improved mapping methods, such as QTL and selective sweep mapping are the key to understand the adaptive history of a population. On the one hand QTL mapping is the method of choice for clarifying the genetic topology of complex traits. It allows estimating the effect of alleles and clarifies the structure of how alleles contribute to complex traits. Selective sweep mapping, on the other hand, identifies adaptive alleles and points to genetic changes that provide a benefit to the carrier.

Within the last century the pioneer work of C. Darwin, H. De Vries, K. Sax, and others who studied patterns of morphological aspects developed into a complex research field with a broad methodological spectrum that combines morphology and genetics. Many complex traits including human diseases, such as Alzheimer, type 2 diabetes, and retinitis pigmentosa can now be associated to certain genes or gene regions (Daw *et al.* 2000; Mori *et al.* 2002; Guyon *et al.* 2007). However, the study of complex traits is strongly supported by the use of model organism's e.g. *D. melanogaster* and *Mus musculus*. The construction of inbred lines, balancer lines, and recombinant lines considerably eases the genetics and simplifies the

analysis. Furthermore, a large sample size that is easily achieved with this species, increases the level to which a complex trait is resolved (Flint & Mackay 2009). It could be shown empirically that enlarging the sample size in QTL studies results in the identification of a much higher number of loci that can be associated to the trait of interest (Turri *et al.* 2001a; Turri *et al.* 2001b). Thus the "axiom of QTL mapping" is validated: "(…) the harder you look, the more QTL you find" (Mackay 2001).

New technologies like next generation sequencing (NGS) open up the opportunity to study the entire genome of an organism and to find associations between the trait of interest and the quantitative trait nucleotide (QTN). Hence, the task of unraveling the genetic architecture of complex traits is still far from being completed. Genome-wide associations studies (GWAS) link certain phenotypes (e.g. type 2 diabetes, Scott *et al.* 2007) to QTNs and contribute to the identification of disease susceptibility genes as well as QTNs that underlie a certain phenotype. Furthermore, NGS provides the ability of molecular diagnostic screens for complex human diseases such as retinitis pigmentosa (Simpson *et al.* 2011). In conclusion, the next step in studying complex traits will be the application of NGS, either in a genome-wide approach or in deep sequencing of candidate regions.

Conclusion One of the main interests of this thesis is to distinguish genetic variation which evolves neutrally from variation that is shaped by natural selection, in particular positive directional selection. We intend to contribute to the broad understanding of local adaptation by choosing different up-to-date statistical approaches coherently applied to the same population samples. Using the model organism *D. melanogaster* we give quantitative insights into selected traits and demographic parameters of populations from Africa, Europe, and Asia.

References

Ayroles JF, Carbone MA, Stone EA*, et al.* (2009) Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics*, **41**, 299-307.

Azevedo RBR, Vernon F, Partridge L (1996) Thermal evolution of egg size in *Drosophila melanogaster*. *Evolution*, **50**, 2338-2345.

Begun DJ, Aquadro CF (1991) Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the yellow-achaete region. *Genetics*, **129**, 1147-1158. Ayroles JF, Carbone MA, Stone EA*, et al.* (2009) Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics*, **41**, 299-307.

Baudry E, Viginier B, Veuille M (2004) Non-African populations of *Drosophila melanogaster* have a unique origin. *Molecular Biology and Evolution*, **21**, 1482-1491.

Begun DJ, Aquadro CF (1991) Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the yellow-achaete region. *Genetics*, **129**, 1147-1158.

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

Beisswanger S, Stephan W (2008) Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated *polyhomeotic* genes in *Drosophila*. *Proceedings of the National Academy of Sciences, USA*, **105**, 5447-5452.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W (1995) The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics*, **140**, 783-796.

Brncic D, Prevosti A, Budnik M, Monclus M, Ocaña J (1981) Colonization of *Drosophila subobscura* in Chile I. First population and cytogenetic studies. *Genetica*, **56**, 3-9.

Bruggemann, P. F. (1958). Insect environment of the high Arctic. *Proc. Int. Congr. Entomol.X,* **1**, 695-702.

Callis J, Fromm M, Walbot V (1987) Introns increase gene expression in cultured maize cells. *Genes Dev*, **1**, 1183-1200.

Capy P, Pla E, David JR (1993) Phenotypic and genetic variability of morphometrical traits in natural populations of *Drosophila melanogaster* and *D simulans*. I. Geographic variations. *Genetics Selection Evolution*, **25**, 1-20.

Caracristi G, Schlötterer C (2003) Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol Biol Evol*, **20**, 792-799.

Catcheside DG (1937) The extra chromosome of *Oenthera lamarckiana lata*. *Genetics*, **22**, 564-576.

Cavicchi S, Giorgi G, Mochi M (1978) Investigation on early divergence between populations of *Drosophila melanogaster* kept at different temperatures. *Genetica*, **48**, 81-87.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289-1303.

Choi T, Huang M, Gorman C, Jaenisch R (1991) A generic intron increases gene expression in transgenic mice. *Mol. Cell. Biol.*, **11**, 3070-3074.

Clark MS, Worland MR (2008) How insects survive the cold: molecular mechanisms — a review. *J Comp Physiol B*, **178**, 917-933.

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

Czajka MC, Lee RE, Jr. (1990) A rapid cold-hardening response protecting against cold shock injury in *Drosophila melanogaster*. *J Exp Biol*, **148**, 245-254.

Darwin C (1859) *On the Origin of the Species by Natural Selection* Murray.

David J, Bocquet C, Pla E (1976) New results on the genetic characteristics of the Far East race of *Drosophila melanogaster*. *Genet Res*, **28**, 253-260.

David JR, Capy P (1988) Genetic variation of *Drosophila melanogaster* natural populations. *Trends in Genetics*, **4**, 106-111.

David J, Moreteau B, Gauthier J*, et al.* (1994) Reaction norms of size characters in relation to growth temperature in *Drosophila melanogaster*: an isofemale lines analysis. *Genetics Selection Evolution*, **26**, 229 - 251.

David JR, Gibert P, Pla E*, et al.* (1998) Cold stress tolerance in *Drosophila*: analysis of chill coma recovery in *D. melanogaster*. *Journal of Thermal Biology*, **23**, 291-299.

Daw EW, Payami H, Nemens EJ*, et al.* (2000) The number of trait loci in late-onset Alzheimer disease. *Am J Hum Genet*, **66**, 196-204.

de Groot NG, Otting N, Doxiadis GGM*, et al.* (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proceedings of the National Academy of Sciences*, **99**, 11748-11753.

de Vries H (1901) Die Mutationstheorie. Versuche und Beobachtungen über die Entstehung von Arten im Pflanzenreich. *Verlag von Veit & Comp. in Leipzig*, **Erste Ausgabe**.

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. *Genetics*, **148**, 1667-1686.

Edwards AC, Mackay TF (2009) Quantitative trait loci for aggressive behavior in *Drosophila melanogaster*. *Genetics*, **182**, 889-897.

Emerson S (1935) The Genetic Nature of de Vries's Mutations in *Oenothera Lamarckiana*. *The American Naturalist*, **69**, 545-559.

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405-1413.

Fidock DA, Nomura T, Talley AK*, et al.* (2000) Mutations in the *P. falciparum digestive vacuole transmembrane protein PfCRT* and evidence for their role in chloroquine resistance. *Mol Cell*, **6**, 861-871.

Fisher RA (1918) The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Philosophical Transactions of the Royal Society of Edinburgh*, **52**, 399-433.

Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA (2010) *Drosophila melanogaster* recombination rate calculator. *Gene*, **463**, 18-20.

Flint J, Mackay TF (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res*, **19**, 723-733.

Frazier MR, Harrison JF, Kirkton SD, Roberts SP (2008) Cold rearing improves cold-flight performance in *Drosophila* via changes in wing morphology. *J Exp Biol*, **211**, 2116-2122.

Gates R (1923) A peculiar type of variability in plants. *Journal of Genetics*, **13**, 13-45.

Gibert P, Moreteau B, David JR (2000) Developmental constraints on an adaptive plasticity: reaction norms of pigmentation in adult segments of *Drosophila melanogaster*. *Evol Dev*, **2**, 249-260.

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

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

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

Grönke S, Mildner A, Fellert S*, et al.* (2005) Brummer lipase is an evolutionary conserved fat storage regulator in Drosophila. *Cell Metabolism*, **1**, 323-330.

Gulliksen S, Birks HH, Possnert G, Mangerud J (1998) A calendar age estimate of the Younger Dryas-Holocene boundary at Kråkenes, western Norway. *The Holocene*, **8**, 249-259.

Gurganus MC, Fry JD, Nuzhdin SV*, et al.* (1998) Genotype-environment interaction at quantitative trait loci affecting sensory bristle number in *Drosophila melanogaster*. *Genetics*, **149**, 1883-1898.

Guyon R, Pearce-Kelling SE, Zeiss CJ, Acland GM, Aguirre GD (2007) Analysis of six candidate genes as potential modifiers of disease expression in canine XLPRA1, a model for human X-linked retinitis pigmentosa 3. *Mol Vis*, **13**, 1094-1105.

Haag-Liautard C, Dorris M, Maside X*, et al.* (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. *Nature*, **445**, 82-85.

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. *Journal of Genetics*, **8**, 299-309.

Heinrich, B. (1979). *Bumblebee Economics.* Cambridge: Harvard University Press. 245.

Heinrich B (1987) Thermoregulation by Winter-Flying Endothermic Moths. *Journal of Experimental Biology*, **127**, 313-332.

Hoehne M, de Couet HG, Stuermer CA, Fischbach KF (2005) Loss- and gain-of-function analysis of the lipid raft proteins Reggie/Flotillin in Drosophila: they are posttranslationally regulated, and misexpression interferes with wing and eye development. *Mol Cell Neurosci*, **30**, 326-338.

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

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity*, **99**, 364-373.

Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W (2007) Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. *Genetics*, **177**, 469-480.

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

Ihle S, Ravaoarimanana I, Thomas M, Tautz D (2006) An Analysis of Signatures of Selective Sweeps in Natural Populations of the House Mouse. *Molecular Biology and Evolution*, **23**, 790-797.

Iida K, Cox-Foster D, Yang X, Ko W-Y, Cavener D (2007) Expansion and evolution of insect GMC oxidoreductases. *BMC Evolutionary Biology*, **7**, 75.

James AC, Partridge L (1995) Thermal evolution of rate of larval development in *Drosophila melanogaster* in laboratory and field populations. *Journal of Evolutionary Biology*, **8**, 315-330.

James AC, Azevedo RB, Partridge L (1997) Genetic and environmental responses to temperature of *Drosophila melanogaster* from a latitudinal cline. *Genetics*, **146**, 881-890.

Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*, **170**, 1401-1410.

Jensen JD, Bauer DuMont VL, Ashmore AB, Gutierrez A, Aquadro CF (2007a) Patterns of sequence variability and divergence at the *diminutive* gene region of *Drosophila melanogaster*: complex patterns suggest an ancestral selective sweep. *Genetics*, **177**, 1071-1085.

Jensen JD, Thornton KR, Bustamante CD, Aquadro CF (2007b) On the utility of linkage disequilibrium as a statistic for identifying targets of positive selection in nonequilibrium populations. *Genetics*, **176**, 2371-2379.

Jeschke JM, Strayer DL (2006) Determinants of vertebrate invasion success in Europe and North America. *Global Change Biology*, **12**, 1608-1619.

Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking effect" revisited. *Genetics*, **123**, 887-899.

Katz AJ, Young SS (1975) Selection for high adult body weight in *Drosophila* populations with different structures. *Genetics*, **81**, 163-175.

Kim Y, Stephan W (2000) Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics*, **155**, 1415-1427.

Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics*, **167**, 1513-1524.

Kimura M (1968) Evolutionary rate at the molecular level. *Nature*, **217**, 624-626.

Kimura M (1983) The Neutral Theory of Molecular Evolution. *Cambridge University Press*, ISBN 0-521-23109-23104.

Kosambi DD (1943) The estimation of map distances from recombination values. *Annals of Human Genetics*, **12**, 172-175.

Kreitman M (1983) Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature*, **304**, 412-417.

Kreitman ME, Aguadé M (1986) Excess polymorphism at the Adh locus in *Drosophila melanogaster*. *Genetics*, **114**, 93-110.

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

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185-199.

Laurent SJY, Werzner A, Excoffier L, Stephan W (2011) Approximate Bayesian analysis of Drosophila melanogaster polymorphism data reveals a recent colonization of Southeast Asia. *Molecular Biology and Evolution*, **28**, 2041-2051.

Lemeunier F, David J, Tsacas L, Ashburner M (1986) in Genetics and Biology of Drosophila (Vol. 3e) (Ashburner M, Carson H L and Thompson J N, eds). *Academic Press*, pp. 147-256.

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

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

Loeschcke V, Sorensen JG, Kristensen TN (2004) Ecologically relevant stress resistance: from microarrays and quantitative trait loci to candidate genes – a research plan and preliminary results using *Drosophila* as a model organism and climatic and genetic stress as model stresses. *J Biosci*, **29**, 503-511.

Mackay TFC (2001) Quantitative trait loci in *Drosophila*. *Nature Reviews Genetics*, **2**, 11-20.

Mackay TF (2010) Mutations and quantitative genetic variation: lessons from *Drosophila*. *Philos Trans R Soc Lond B Biol Sci*, **365**, 1229-1239.

Malaga-Trillo E, Laessing U, Lang DM, Meyer A, Stuermer CA (2002) Evolution of duplicated reggie genes in zebrafish and goldfish. *J Mol Evol*, **54**, 235-245.

Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proc Natl Acad Sci U S A*, **74**, 560-564.

Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res*, **23**, 23-35.

Meredith J, Storti RV (1993) Developmental regulation of the Drosophila tropomyosin II gene in different muscles is controlled by muscle-type-specific intron enhancer elements and distal and proximal promoter control elements. *Dev Biol*, **159**, 500-512.

Mitrovski P, Hoffmann AA (2001) Postponed reproduction as an adaptation to winter conditions in *Drosophila melanogaster*: evidence for clinal variation under semi-natural conditions. *Proceedings of the Royal Society in London. Series B, Biological Sciences*, **268**, 2163-2168.

Morgan TJ, Mackay TFC (2006) Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity*, **96**, 232-242.

Mori Y, Otabe S, Dina C*, et al.* (2002) Genome-Wide Search for Type 2 Diabetes in Japanese Affected Sib-Pairs Confirms Susceptibility Genes on 3q, 15q, and 20q and Identifies Two New Candidate Loci on 7p and 11p. *Diabetes*, **51**, 1247-1255.

Müller L, Hutter S, Stamboliyska R*, et al.* (2011) Population transcriptomics of *Drosophila melanogaster* females. *BMC Genomics*, **12**, 81.

Néron B, Ménager H, Maufrais C*, et al.* (2009) Mobyle: a new full web bioinformatics framework. *Bioinformatics*, **25**, 3005-3011.

Neumann-Giesen C, Falkenbach B, Beicht P*, et al.* (2004) Membrane and raft association of reggie-1/flotillin-2: role of myristoylation, palmitoylation and oligomerization and induction of filopodia by overexpression. *Biochem J*, **378**, 509-518.

Nielsen R, Williamson S, Kim Y*, et al.* (2005) Genomic scans for selective sweeps using SNP data. *Genome Res*, **15**, 1566-1575.

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

Nunney L (1996) The influence of variation in female fecundity on effective population size. *Biological Journal of the Linnean Society*, **59**, 411-425.

Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng ZB, Mackay TF (1997) Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*, **94**, 9734-9739.

Ometto L, Glinka S, De Lorenzo D, Stephan W (2005) Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Molecular Biology and Evolution*, **22**, 2119-2130.

Pascual M, Chapuis MP, Mestres F*, et al.* (2007) Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Molecular Ecology*, **16**, 3069-3083.

Pavlidis P, Jensen JD, Stephan W (2010) Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*, **185**, 907-922.

Romero-Calderón R, Shome RM, Simon AF*, et al.* (2007) A screen for neurotransmitter transporters expressed in the visual system of *Drosophila melanogaster* identifies three novel genes. *Developmental Neurobiology*, **67**, 550-569.

Pool JE, Bauer DuMont V, Mueller JL, Aquadro CF (2006) A scan of molecular variation leads to the narrow localization of a selective sweep affecting both Afrotropical and cosmopolitan populations of Drosophila melanogaster. *Genetics*, **172**, 1093-1105.

Przeworski M (2002) The signature of positive selection at randomly chosen loci. *Genetics*, **160**, 1179-1189.

Rubin GM, Hong L, Brokstein P*, et al.* (2000) A Drosophila Complementary DNA Resource. *Science*, **287**, 2222-2224.

Rubin C-J, Zody MC, Eriksson J*, et al.* (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, **464**, 587-591.

Salt RW (1961) Principles of insect cold-hardiness. *Annual Review of Entomology*, **6**, 55-74.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, **74**, 5463-5467.

Sax K (1923) The Association of Size Differences with Seed-Coat Pattern and Pigmentation in *PHASEOLUS VULGARIS*. *Genetics*, **8**, 552-560.

Schlötterer C (2003) Hitchhiking mapping – functional genomics from the population genetics perspective. *Trends Genet*, **19**, 32-38.

Schlötterer C, Neumeier H, Sousa C, Nolte V (2006) Highly Structured Asian *Drosophila melanogaster* Populations: A New Tool for Hitchhiking Mapping? *Genetics*, **172**, 287-292.

Scott LJ, Mohlke KL, Bonnycastle LL*, et al.* (2007) A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. *Science*, **316**, 1341-1345.

Simpson DA, Clark GR, Alexander S, Silvestri G, Willoughby CE (2011) Molecular diagnosis for heterogeneous genetic diseases with targeted high-throughput DNA sequencing applied to retinitis pigmentosa. *Journal of Medical Genetics*, **48**, 145-151.

Stephan W, Langley CH (1989) Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the vermilion and forked loci. *Genetics*, **121**, 89-99.

Stephan W, Song YS, Langley CH (2006) The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics*, **172**, 2647-2663.

Stuermer CA, Lang DM, Kirsch F*, et al.* (2001) Glycosylphosphatidyl inositol-anchored proteins and fyn kinase assemble in noncaveolar plasma membrane microdomains defined by reggie-1 and -2. *Mol Biol Cell*, **12**, 3031-3045.

Svetec N, Pavlidis P, Stephan W (2009) Recent strong positive selection on *Drosophila melanogaster HDAC6*, a gene encoding a stress surveillance factor, as revealed by population genomic analysis. *Molecular Biology and Evolution*, **26**, 1549-1556.

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.

Takeuchi H, Rigden DJ, Ebrahimi B, Turner PC, Rees HH (2005) Regulation of ecdysteroid signalling during Drosophila development: identification, characterization and modelling of ecdysone oxidase, an enzyme involved in control of ligand concentration. *Biochem J*, **389**, 637-645.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.

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

Thornton KR, Jensen JD (2007) Controlling the false-positive rate in multilocus genome scans for selection. *Genetics*, **175**, 737-750.

Tian F, Stevens NM, Buckler ESt (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proc Natl Acad Sci U S A*, **106 Suppl 1**, 9979-9986.

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

Turri MG, Datta SR, DeFries J, Henderson ND, Flint J (2001a) QTL analysis identifies multiple behavioral dimensions in ethological tests of anxiety in laboratory mice. *Curr Biol*, **11**, 725-734.

Turri MG, Henderson ND, DeFries JC, Flint J (2001b) Quantitative trait locus mapping in laboratory mice derived from a replicated selection experiment for open-field activity. *Genetics*, **158**, 1217-1226.

Tweedie S, Ashburner M, Falls K*, et al.* (2009) FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Research*, **37**, D555-D559.

Van 't Land J, Putten PV, Zwaan, Kamping, Delden WV (1999) Latitudinal variation in wild populations of *Drosophila melanogaster*: heritabilities and reaction norms. *Journal of Evolutionary Biology*, **12**, 222-232.

von Philipsborn AC, Ferrer-Vaquer A, Rivera-Milla E, Stuermer CA, Malaga-Trillo E (2005) Restricted expression of reggie genes and proteins during early zebrafish development. *J Comp Neurol*, **482**, 257-272.

Wang H, Nussbaum-Wagler T, Li B*, et al.* (2005a) The origin of the naked grains of maize. *Nature*, **436**, 714-719.

Wang MH, Harshman LG, Nuzhdin SV (2005b) Quantitative trait loci for lipid content in *Drosophila melanogaster*. *Obes Res*, **13**, 1891-1897.

Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*, **7**, 256-276.

Williamson SH, Hubisz MJ, Clark AG*, et al.* (2007) Localizing Recent Adaptive Evolution in the Human Genome. *PLoS Genet*, **3**, e90.

Wilusz CJ, Wang W, Peltz SW (2001) Curbing the nonsense: the activation and regulation of mRNA surveillance. *Genes Dev*, **15**, 2781-2785.

Wong LC, Schedl P (2006) Dissection of *Drosophila* Ovaries. *JoVE. 1. **http://www.jove.com/index/Details.stp?ID=52,***, doi: 10.3791/3752.

Woodruff RC, Thompson JN, Seeger MA, Spivey WE (1984) Variation in spontaneous mutation and repair in natural population lines of Drosophila melanogaster. *Heredity*, **53**, 223-234.

Wootton JC, Feng X, Ferdig MT*, et al.* (2002) Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature*, **418**, 320-323.

Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. *Mol Biol Evol*, **22**, 506-519.

Zeng ZB (1994) Precision mapping of quantitative trait loci. *Genetics*, **136**, 1457-1468.

Appendix A

Data deposited at Dryad: doi:10.5061/dryad.8083.

Original supporting information of the online version of the published Manuscript

Supplementary Online Material

Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine-mapping by selective sweep analysis

Nicolas Svetec[*,†], Annegret Werzner[*], Ricardo Wilches[*], Pavlos Pavlidis[*], José M. Álvarez-Castro[*,Δ], Karl W. Broman[‡], Dirk Metzler[*], and Wolfgang Stephan[*]

[*]Section of Evolutionary Biology, Biocenter, University of Munich, 82152 Planegg-Martinsried, Germany

[†]Department of Ecology and Evolutionary Biology, University of Toronto, 25 Willcocks St., Toronto, Ontario, M5S 3B2, Canada

[Δ]Department of Genetics, University of Santiago de Compostela, Avda. Carvalho Calero, ES-27002 Lugo – Galiza, Spain

[‡]Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin 53706, USA

[‡]Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin 53706, USA

Table (Appendix A) S 1: Marker information and related accession numbers of sequenced fragments of isofemale fly lines E14 and A157 of *D. melanogaster* (Glinka *et al.* 2003; Ometto *et al.* 2005).

Number	Marker ID	Position in cM	Interval in cM	Cytologic position	Sequence position (release 5.29 of Flybase)	E14 accession number	A157 accession number	Polymorphism [A157/E14]
01	025	0.0	0.0	3A2	X:2285143	AM002070	AM000232	[T/C]
02	056	2.9	2.9	3F1	X:3678740	AJ570920	AJ570929	[G/A]
03	067	6.4	3.5	4C14	X:4596926	AM003704	AM001789	[T/G]
04	094	9.0	2.6	4F4	X:5175417	AM002870	AM001029	[C/A]
05	109	12.0	3.0	5C7	X:5816285	AM002906	AM001048	[G/A]
06	114	15.3	3.3	6C12	X:6654497	AJ571168	AJ571177	[A/T]
07	126	16.6	1.4	7B2	X:7234566	AM002918	AM001067	[A/T]
08	143	18.1	1.5	7E1	X:8167788	AM002942	AM001084	[C/T]
09	167	21.3	3.3	8E1	X:9329835	AM002965	AM001107	[A/G]

10	464	24.3	2.9	9B5	X:10152910	AM002554	AM000717	[T/C]
11	210	27.3	3.0	10B8	X:11292860	AM003060	AM001192	[C/T]
12	493	30.0	2.8	11A6	X:12075922	AM002685	AM000825	[TC/GG]
13	250	32.9	2.9	11D1	X:12761349	AJ569928	AJ569937	[C/T]
14	278	35.9	3.0	12A1	X:13378559	AJ570092	AJ570102	[ATGCCG/TTGCCC]
15	326	39.0	3.1	12E2	X:13995409	AJ570395	AJ570405	[T/G]
16	375	42.1	3.1	12F3	X:14621373	AJ570694	AJ570704	[TGG/GCT]
17	386	44.0	1.9	13A12	X:15008141	AM003501	AM001602	[G/C]
18	282	45.5	1.6	13C3	X:15342614	AJ570147	AJ570157	[C/T]
19	296	47.4	1.9	13E15	X:15659592	AM003264	AM001377	[A/C]
20	726	49.4	2.0	14B6	X:16156731	AM002767	AM000911	[A/T]
21	364	50.9	1.5	15A1	X:16590177	AJ570558	AJ570567	[C/T]
22	426	53.3	2.4	16B1	X:17414064	AM002403	AM000572	[T/C]
23	444	56.3	3.1	17D3	X:18639229	AM002496	AM000670	[GCT/TCG]

Candidate genes

Here we provide a summary of the candidate genes that have been found using a data base search (FlyBase, FB2010_01, Supplementary Table 2). As our QTL are quite large we decided to restrict the candidate gene list to the genes annotated with "response to temperature stimulus" gene ontologies in order to produce a short list of candidate genes. All genes are listed according to the QTL they are located in but will be discussed referring to joint biological functions and characteristics. The term "temperature stimulus" includes both the association with cold and heat stress related phenotypes. QTL influencing cold and heat tolerance rarely co-localize (Morgan & Mackay 2006; Norry *et al.* 2008) although several studies prove that heat shock proteins are also induced by cold treatments (Burton *et al.* 1988; Goto & Kimura 1998; Sejerkilde *et al.* 2003). Thus cross-protection effects between cold and heat stress responses might exist in *D. melanogaster*.

The *Adar* gene has been associated with sensitivity to heat shock (Ma *et al.* 2001) and the genes *Parg* (Ji & Tulin 2009), *hang* (Scholz *et al.* 2005), and *iav* (Chentsova *et al.* 2002) were characterized for heat stress related phenotypes. *Hsc70-3* and *Hsp60* have been associated with thermotolerance phenotypes in numerous studies (Bettencourt *et al.* 2008; Norry *et al.* 2009; Norry *et al.* 2007; Norry *et al.* 2008; Ornelles & Penman 1990) whereas the thermotolerance function of *CG2887, CG32727* and *HspB8* was computationally determined based on sequence similarities. Furthermore, there is a QTL for knockdown resistance to high temperature in the middle of the X chromosome between cytological positions 10A and 12D (Norry *et al.* 2008) that shows some overlap with ours. Only functional analysis of candidate genes could confirm whether the same sets of genes contribute to both cold and heat tolerance phenotypes.

Besides direct temperature resistance there are numerous other temperature induced processes. *per* is an important gene regulating circadian rhythms and diapause (Glaser & Stanewsky 2005). Diapause and quiescence are phenomena that can be described as an energy allocation shift from reproduction to survival that is triggered by decreasing temperature and day length (Saunders *et al.* 1989). The energy re-allocation seems to trigger the expression of numerous protective proteins like heat shock proteins, one of which is located within the QTL at position 24 cM. Another gene that is involved in the synchronization of the circadian clock with temperature fluctuations is *nocte* (Glaser & Stanewsky 2005). It interacts with two of our candidate genes: *tilB* and *per* (Sehadova *et al.* 2009). The *rdgA* gene was studied for thermotaxis behavior (Kwon *et al.* 2008) and the *shi* gene has been extensively used for studying fly neurobiology (Ikeda *et al.* 1976; Kelly & Suzuki 1974). It was shown that mutations in the *shi* gene generate a thermosensitive alteration of synaptic transduction of nervous signals.

Genes related to energy metabolism and cytoskeleton could favor a quicker recovery from cold shock (Clark & Worland 2008). If the cold shock response in *Drosophila* has some similarities with bacteria, cold shock exposure produces a drastic re-programming of gene expression to allow for survival under the new unfavorable conditions (Giuliodori *et al.* 2007). Negative regulation of biosynthetic processes, the initiation of resting stages, and regulation of the biorhythm may play a major role to come up against a temperature stimulus. This assumption is confirmed by a GO analysis (Eden *et al.* 2009) of all genes within the detected QTL positions. Compared to the background region (outside the QTL), the most common gene function is negative regulation of metabolic processes ($P<0.01$).

The QTL at position 9 cM does not contain a gene annotated with a temperature-related function. This is due to the stringency with which we defined our candidates. But it is

likely that some of the genes that create the present cold protective effects are not yet described as temperature related.

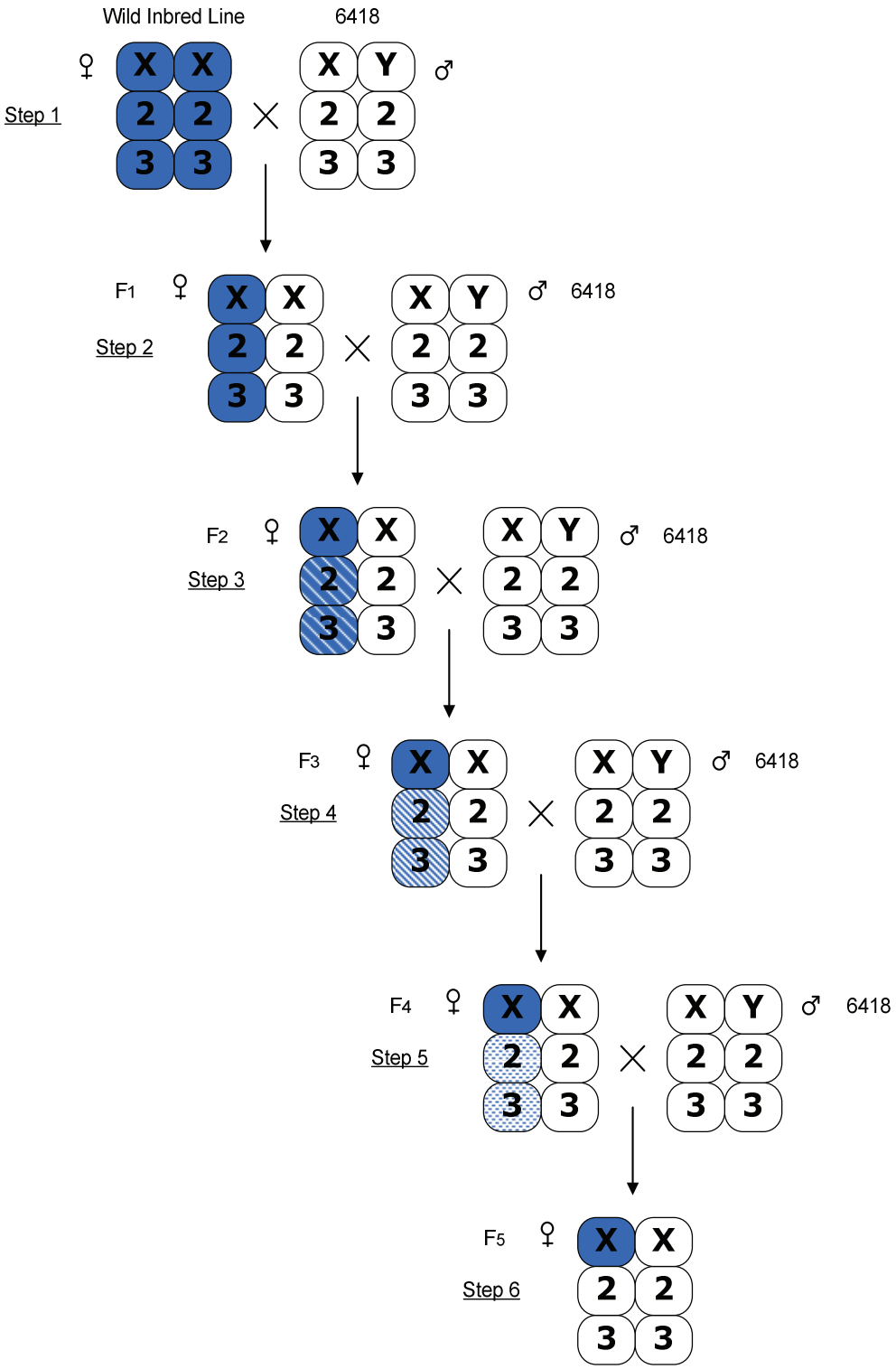
To date, there is no proof that *D. melanogaster* uses proper cold resistance mechanisms like controlled freezing, supercooling, or antifreeze proteins (Izquierdo 1991) even if the *D. melanogaster* genome contains genes annotated as putative antifreeze protein (like *CG6227* that is on the X chromosome but outside the QTL). Nonetheless, there are several other physiological pathways that could increase the cold tolerance of a fly. Any factor implied in desiccation resistance, linked to heat shock proteins or lipids constituting membranes, and has the potential to improve cold tolerance.

Table (Appendix A) S 2: Candidate gene list for all six QTL positions

Sex	QTL Position	Cytologic position	Candidate gene name	Temperature related biological function
Males	0 cM	1A-3F	*Adar*	Adult locomotory behavior, response to heat
			per	Circadian rhythm, locomotor rhythm regulation of circadian sleep/wake cycle, sleep, response to temperature stimulus, negative regulation of transcription from RNA polymerase II promoter
			Parg	Carbohydrate metabolic process, response to heat, regulation of RNA splicing, phagocytosis, engulfment, regulation of histone acetylation
Female	9cM	4C-5C	-	-
Female	17cM	6C-7B	*iav*	Flight behavior, response to heat, ion transport, transmembrane transport
Males	18cM	7B-8E	*CG32727*	Response to heat
			rdgA	Rhodopsine mediated signaling pathway, phototransduction, photoinositide biosynthetic process, thermotaxis, intracellular signaling cascade, phosphatidic acid biosynthetic process

Male / Female	24cM	8E-11D	for thermotolerance:	
			nocte	Entrainment of circadian clock, temperature compensation of circadian clock
			CG2887	Response to heat, protein folding
			Hsp60	Response to heat, protein folding, response to stress, protein targeting to mitochondrion
			Hsc70-3	RNA interference
			For sex differentiation:	
Male / Female	56cM	13E-20E	*SisA, Yp1, Yp2, hop shi*	Anatomical structure development, reproductive process in a multicellular organism, membrane organization, regulation of biological process, learning or memory, mating; memory, associative learning, reproductive cellular process, learning, regulation of actin filament-based process, open tracheal system
			hang	Response to heat, response to stress, behavioral response to ethanol
			HspB8	Response to heat
			tilB	Male courtship behavior, veined wing generated song production; detection of mechanical stimulus involved in sensory perception of sound; sensory perception of sound; temperature compensation of the circadian clock

Supplementary Figures



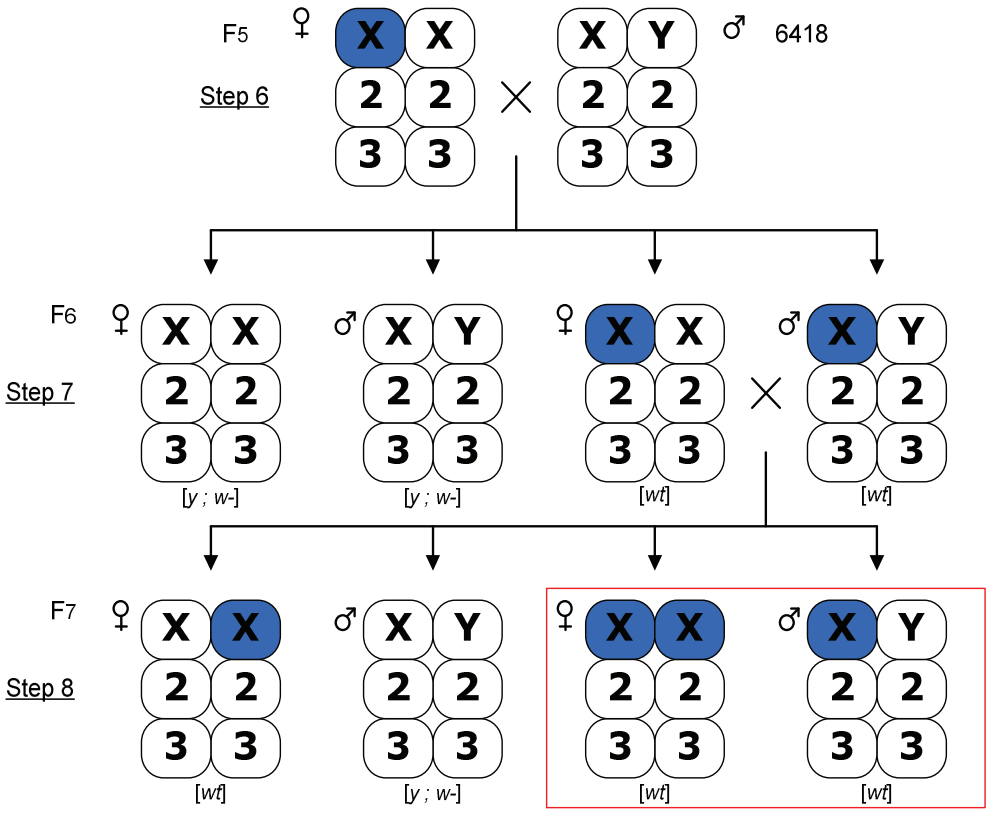

F5
Step 6
6418
F6
Step 7
[y ; w-]
[y ; w-]
[wt]
[wt]
F7
Step 8
[wt]
[y ; w-]
[wt]
[wt]

Figure (Appendix A) S 1: Construction of the substitution lines A* and E*.

Both lines were constructed according to the same crossing scheme: the X chromosome of E14 and A157, respectively, was introgressed into the genetic background of the balancer line 6418. The figure shows chromosomes of wild origin (*wt*) in blue and balancer chromosomes in white. The balancer line 6418 carries an FM7j balancer X chromosome, which has two important properties: first, it is a very efficient suppressor of recombination and second, the homozygous females as well as hemizygous males are viable and fertile. These balancer chromosomes can be traced back by two recessive mutations: w^{1118} and y^{93j} (producing white eyes and yellowish body color, respectively; this phenotype is shown as [*y;w*-] on the figure). For steps 2 to 6, the *wt* heterozygous X_{6418}/X_{wt} female offspring was selected and used for the next generation cross. The crosses from step 1 to 6: the heterozygous X_{6418}/X_{wt} female offspring can be recognized as they are wild-type. They are then backcrossed to the balancer line 6418. Each backcross divides the amount of *wt* genetic material by two. Statistically, at step 7 only 1.06% of the *wt* genetic material is left on the autosomes. Step 7 and 8 are brother-sister crosses of *wt* offspring. At step 8, a single brother-sister pair produced homogenous *wt* offspring that was used to establish a stabilized X chromosomal recombinant (XR) line.

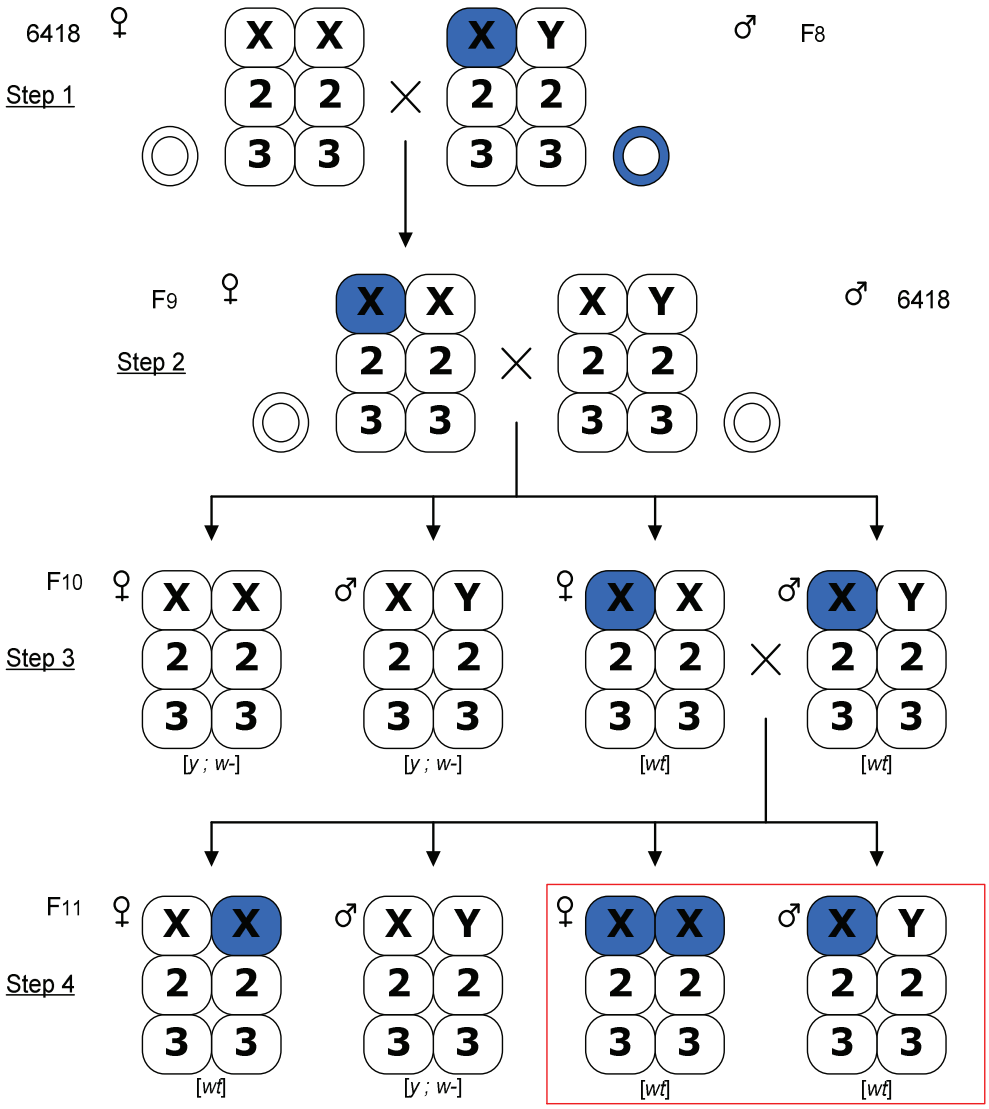

6418 ♀
Step 1
♂ F8
F9 ♀
Step 2
♂ 6418
F10
Step 3
[y ; w-]
[y ; w-]
[wf]
[wf]
F11
Step 4
[wf]
[y ; w-]
[wf]
[wf]

Figure (Appendix A) S 2: Homogenization of maternally inherited factors in A* and E*.

The stabilized lines were previously derived from a wild-type female. The maternally inherited genetic material (like mitochondria) are thus different between A* and E*. The above crossing scheme was used to homogenize the maternally inherited factors represented by a ring (blue for *wt*, white for the balancer line 6418). Step 2 produces individuals that carry maternally inherited factors of the balancer line 6418. Steps 3 and 4 are carried out to establish an XR line. It is important to notice that this step added two more backcrosses to the balancer line 6418, thus lowering the amount of *wt* genetic material left on the autosomes to 0.26%.

Appendix A: Reference

Bettencourt B, Hogan C, Nimali M, Drohan B (2008) Inducible and constitutive heat shock gene expression responds to modification of *Hsp70* copy number in *Drosophila melanogaster* but does not compensate for loss of thermotolerance in *Hsp70* null flies. *BMC Biology*, **6**, 5.

Burton V, Mitchell HK, Young P, Petersen NS (1988) Heat shock protection against cold stress of *Drosophila melanogaster*. *Molecular and Cellular Biology*, **8**, 3550-3552.

Chentsova NA, Gruntenko NE, Bogomolova EV*, et al.* (2002) Stress response in *Drosophila melanogaster* strain inactive with decreased tyramine and octopamine contents. *Journal of Comparative Physiology B*, **172**, 643-650.

Clark MS, Worland MR (2008) How insects survive the cold: molecular mechanisms—a review. *Journal of Comparative Physiology B*, **178**, 917-933.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, **10**, 48.

Giuliodori AM, Gualerzi CO, Soto S, Vila J, TavÍO MM (2007) Review on Bacterial Stress Topics. *Annals of the New York Academy of Sciences*, **1113**, 95-104.

Glaser FT, Stanewsky R (2005) Temperature synchronization of the *Drosophila* circadian clock. *Current Biology*, **15**, 1352-1363.

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

Goto SG, Kimura MT (1998) Heat- and cold-shock responses and temperature adaptations in subtropical and temperate species of *Drosophila*. *Journal of Insect Physiology*, **44**, 1233-1239.

Ikeda K, Ozawa S, Hagiwara S (1976) Synaptic transmission reversibly conditioned by single-gene mutation in *Drosophila melanogaster*. *Nature*, **259**, 489-491.

Izquierdo JI (1991) How does *Drosophila melanogaster* overwinter? *Entomologia Experimentalis et Applicata*, **59**, 51-58.

Ji Y, Tulin AV (2009) Poly(ADP-ribosyl)ation of heterogeneous nuclear ribonucleoproteins modulates splicing. *Nucleic Acid Research*, **37**, 3501-3513.

Kelly LE, Suzuki DT (1974) The effects of increased temperature on electroretinograms of temperature-sensitive paralysis mutants of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, USA*, **71**, 4906-4909.

Kwon Y, Shim HS, Wang X, Montell C (2008) Control of thermotactic behavior via coupling of a TRP channel to a phospholipase C signaling cascade. *Nature Neuroscience*, **11**, 871-873.

Ma E, Gu XQ, Wu X, Xu T, Haddad GG (2001) Mutation in pre-mRNA adenosine deaminase markedly attenuates neuronal tolerance to O_2 deprivation in *Drosophila melanogaster*. *Journal of Clinical Investigation*, **107**, 685-693.

Morgan TJ, Mackay TFC (2006) Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity*, **96**, 232-242.

Norry FM, Larsen PF, Liu Y, Loeschcke V (2009) Combined expression patterns of QTL-linked candidate genes best predict thermotolerance in *Drosophila melanogaster*. *Journal of Insect Physiology*, **55**, 1050-1057.

Norry FM, Sambucetti P, Scannapieco AC, Gomez FH, Loeschcke V (2007) X-linked QTL for knockdown resistance to high temperature in *Drosophila melanogaster*. *Insect Molecular Biology*, **16**, 509-513.

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

Ometto L, Glinka S, De Lorenzo D, Stephan W (2005) Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Molecular Biology and Evolution*, **22**, 2119-2130.

Ornelles DA, Penman S (1990) Prompt heat-shock and heat-shifted proteins associated with the nuclear matrix-intermediate filament scaffold in *Drosophila melanogaster* cells. *Journal of Cell Science*, **95 (Pt 3)**, 393-404.

Saunders DS, Henrich VC, Gilbert LI (1989) Induction of diapause in *Drosophila melanogaster*: photoperiodic regulation and the impact of arrhythmic clock mutations on time measurement. *Proceedings of the National Academy of Sciences, USA*, **86**, 3748-3752.

Scholz H, Franz M, Heberlein U (2005) The *hangover* gene defines a stress pathway required for ethanol tolerance development. *Nature*, **436**, 845-847.

Sehadova H, Glaser FT, Gentile C*, et al.* (2009) Temperature entrainment of *Drosophila's* circadian clock involves the gene *nocte* and signaling from peripheral sensory tissues to the brain. *Neuron*, **64**, 251-266.

Sejerkilde M, Sorensen JG, Loeschcke V (2003) Effects of cold- and heat hardening on thermal resistance in *Drosophila melanogaster*. *Journal of Insect Physiology*, **49**, 719-726.

Appendix B

Original supporting information of the online version of the published Manuscript

Supplementary Online Material

Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia

Stefan J.Y. Laurent*, Annegret Werzner*, Laurent Excoffier[§] and Wolfgang Stephan*

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

[§]Computational and Molecular Population Genetics Laboratory, Zoological Institute, University of Berne, 3012 Berne, Switzerland

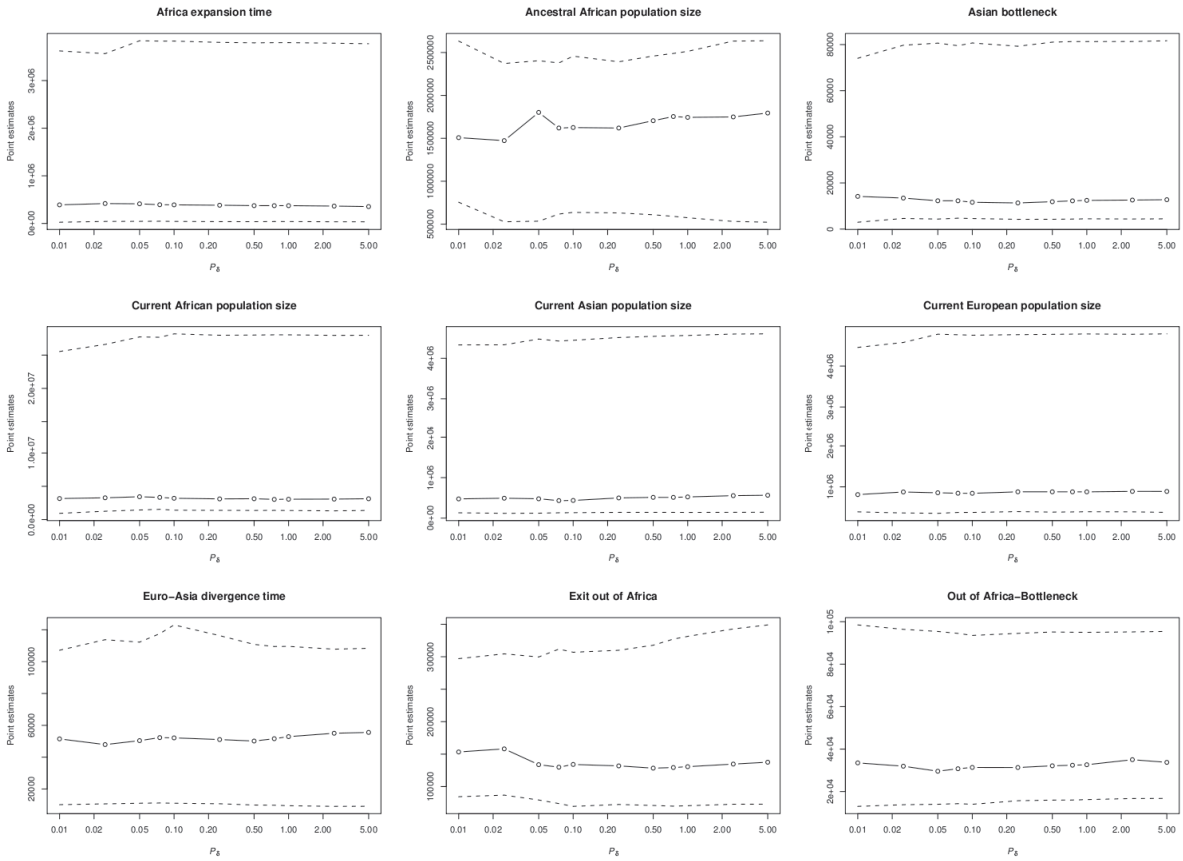


Figure (Appendix B) S 1: Stability of the mode and credibility intervals of our posterior distributions as a function of the proportion of retained simulations (Solid lines: Mode; Dashed lines: 95% credibility intervals)

Table (Appendix B) S 1a: Vectors of observed summary statistics for the ABC analysis

	X				3			
Statistics	Asia	Europe	Africa	All Populations	Asia	Europe	Africa	All Populations
Mean(π)	1.67	2.09	5.07	4	1.78	3.08	4.59	3.83
Var(π)	3.39	4.5	8.44	6.52	2.31	6.97	11.43	7.66
Mean(S)	4.71	6.11	17.42	21.08	5.1	8.84	12.68	17.96
Var(S)	24.72	27.2	77.37	99.08	17.85	46.63	77.98	103.59
Mean(D)	0.16	-0.11	-0.66	-0.9	0.24	0.15	-0.42	-0.64
Var(D)	1.29	1.5	0.37	0.51	0.94	1.21	0.53	0.61
Mean(K)	2.8	3.86	9.48	13.97	3.16	4.92	6.22	11.82
Var(K)	2.18	3.64	5.26	18.11	1.61	2.81	2.42	14.68
Mean(Z_{nS})	0.53	0.43	0.14	0.11	0.55	0.37	0.23	0.13
Var(Z_{nS})	0.1	0.01	0.01	0.01	0.1	0.05	0.02	0.004

Table (Appendix B) S 1b: Vectors of observed summary statistics of population pairs for the ABC analysis

	X			3		
	AS-AF	AS-EU	EU-AF	AS-AF	AS-EU	EU-AF
Mean(D_A)	0.95	0.27	0.77	0.65	0.3	0.4
Mean(S_S)	3.14	3.32	3.99	2.84	3.74	5.12

These are the observed summary statistics that were used in the ABC analysis. S_S is the proportion of shared polymorphic sites between two populations.

Table (Appendix B) S 2: Stability of posterior probability of the different demographic models as a function of the proportion of retained simulations in the model choice procedure

	Chromosome X			**Chromosome 3**		
Tolerance	DCS	SCS	SCS-2	DCS	SCS	SCS-2
250	8.86E-026	0.976	0.024	4.49E-012	1	3.3E-005
500	**2.37E-012**	**0.948**	**0.052**	**3.69E-011**	**0.997**	**0.003**
750	8.94E-016	0.998	0.002	4.02E-010	0.986	0.014
1000	5.22E-065	0.994	0.006	1.52E-009	0.991	0.009
2500	1.15E-007	0.963	0.037	7.07E-009	0.978	0.022
5000	2.46E-009	0.953	0.047	2.87E-008	0.961	0.039
7500	3.52E-005	0.931	0.069	4.91E-008	0.962	0.038
10000	3.20E-005	0.929	0.071	9.13E-008	0.953	0.047

The values in bold correspond to the threshold we used in the model choice estimation

Table (Appendix B) S 3: Results of the predictive simulations

Statistics	X	3	Statistics	X	3
Mean(π)_Africa	0.870	0.891	Mean(Z_{nS})_Europe	0.914	0.619
Var(π)_Africa	0.863	0.835	Var(Z_{nS})_Europe	0.008	0.46
Mean(S)_Africa	0.847	0.869	Mean(π)_Asia	0.749	0.687
Var(S)_Africa	0.765	0.78	Var(π)_Asia	0.901	0.514
Mean(D)_Africa	0.213	0.307	Mean(S)_Asia	0.796	0.755
Var(D)_Africa	0.504	0.778	Var(S)_Asia	0.944	0.619
Mean(K)_Africa	0.683	0.683	Mean(D)_Asia	0.142	0.25
Var(K)_Africa	0.156	0.11	Var(D)_Asia	0.806	0.309
Mean(Z_{nS})_Africa	0.922	0.955	Mean(K)_Asia	0.531	0.634
Var(Z_{nS})_Africa	0.929	0.772	Var(K)_Asia	0.569	0.353
Mean(π)_Europe	0.698	0.893	Mean(Z_{nS})_Asia	0.631	0.619
Var(π)_Europe	0.912	0.887	Var(Z_{nS})_Asia	0.784	0.757
Mean(S)_Europe	0.752	0.925	Mean(D_A)_AF-EU	0.804	0.86
Var(S)_Europe	0.892	0.873	Mean(S_S)_AF-EU	0.754	0.909
Mean(D)_Europe	0.099	0.227	Mean(D_A)_AF-AS	0.831	0.97
Var(D)_Europe	0.995	0.893	Mean(S_S)_AF-AS	0.784	0.677
Mean(K)_Europe	0.437	0.811	Mean(D_A)_EU-AS	0.871	0.792
Var(K)_Europe	0.286	0.271	Mean(S_S)_EU-AS	0.668	0.68

Values represent the probability that the simulated data is smaller than the observed value

Protocol of ovaries dissection in Drosophila

Li Chin Wong, Paul Schedl, Princeton University (2006):

Here we demonstrate a method of dissecting ovaries in Drosophila. As the method of culturing the different cell types that exist in the Drosophila ovary has yet been defined, this is a rapid method to obtain the Drosophila ovarian tissue. Essentially, we show that dissecting Drosophila ovaries involves a simple two- or three-step procedure. Subsequent treatment of these dissected ovaries depends on the downstream experiments that will be performed.

1. Feed flies with yeast 1-2 days prior to dissecting them to fatten up the ovaries.
2. Anesthetize flies using carbon dioxide stream.
3. Using a pair of tweezers, select a female fly.
4. Submerge the female fly into 1X PBS.
5. Grab the fly at its lower thorax with a pair of tweezers.
6. Tug gently at the lower abdomen with another pair of tweezers until the internal organs in the abdomen are exposed.
7. Look for the pair of ovaries and detach it from other organs (e.g. the intestines).
8. Tease apart the ovarioles (if ovaries are to be used for immunostaining or in situ hybridization).
9. Keep ovaries in ice-cold 1X PBS while dissecting the next fly.

Appendix C

DNA Extraction

Protocols of Purgene DNA Isolation kit for one fly (Gentra Systems, Minneapolis, MN, USA).

1. Add 1 fly to the 100 µl chilled Cell-lysis-Solution place on ice in 1.5 ml tube and grind the fly completely.
2. Add 0.5 µl Proteinase K (20 mg/ml).
3. Homogenize thoroughly.
4. Incubate 15 minutes at 65°C.
5. Add 0.5 µl RNase solution A (2 mg/ml) to the tube
6. Mix the sample by inverting the tube and incubate 15 minutes at 37°C.
7. Cool sample at room temperature.
8. Add 33µl Protein-Precipitation-Solution and vortex.
9. Keep the tube 5 minutes on ice.
10. Centrifuge at 13200 rpm for 5 minutes.
11. Transfer the supernatant to a clean 1.5 ml tube.
12. Add 100 µl of 100% Isopropanol and mix the sample by inverting gently.
13. Incubate 5 minutes at room temperature.
14. Centrifuge at 13200 rpm for 5 minutes, remove supernatant.
15. Wash the pellet with 100 µl of 70% ethanol.
16. Centrifuge 2 min at 12000 rpm, carefully remove the ethanol.
17. Dry the pellet at room temperature; resuspend the pellet in 50 µl of distilled water.

Polymerase-chain-reaction (PCR)

Protocols of PeqLab PCR kit (PEQLAB Biotechnologie GMBH, Erlangen, Germany)

Indicated are volumes for a volume final of 25 µl and in parentheses, the concentration

Distilled water	16.12	µl
Buffer specific buffers (10X)	2.5	µl
dNTP's (100 mM)	0.5	µl
Mg (50 mM)	1	µl
Forward Primer (10 mM)	2	µl
Reverse Primer (10 mM)	2	µl
Taq polymerase	0.13	µl
DNA template	1	µl

PCR standard run program

Indicated are the temperature and the duration of each step

1.	Initial denaturation	94°C	2 min
2.	30 amplification cycles		
	Denaturation	94°C	45 sec
	Annealing	X°C	45 sec
	Elongation	72°C	50 sec (for 800 pb)
	(X is specific to each pair of primers; between 50°C and 60°C)		
3.	Final elongation	72°C	7 min
4.	Hold	4°C	until storage at -20°C

PCR products are verified on a 1% agarose gel

PCR products purification

PCR product	20 µl
ExoSAP-IT (USB, Cleavland, OH, USA)	1 µl

Thermocycler programs

37°C	30 min
80°C	15 min
4°C	until the storage at -20°C

Sequencing reaction

Protocol according to the ABI sequencer 3730 using DYEnamic ET terminator cycle sequencing kit (Applied Biosystems, Foster City, CA, USA)

General condition for a 10 µl reaction

Indicated are volumes and in parentheses the concentration

Big Dye v1.1 (2.5X)	1.5	µl
Sequencing buffer (5X)	1.25	µl
Primer (10 µM)	2	µl
DNA template	3	µl
Distilled water	2.25	µl

Thermocycler programs

1. Denaturation 96°C 1 min
2. 40 amplification cycles
   - Denaturation 96°C 10 sec
   - Annealing 50°C 15 sec
   - Elongation 60°C 4 min
3. Hold 4°C until storage at -20°C
4. Add to sample 10 µl of distilled water

RNA extraction for quantitative Real-Time PCR

A Tissue and Cell Lysis

1. Collect three sets of flies (15 male flies) and transfer each set to individual 1.5 ml tubes on ice.
2. Dilute 2.5 µl of 20 mg/ml Proteinase K into 300 µl of Tissue and Cell Lysis Solution for each sample.
3. Homogenize flies and transfer to a microcentrifuge tube.
4. Add 300 µl of Tissue and Cell Lysis Solution containing the Proteinase K and mix thoroughly.
5. Incubate at 65°C for 15 minutes; vortex every 5 minutes.
6. Place the samples on ice for 3-5 minutes and then proceed with total nucleic acid precipitation

B Precipitation of Total Nucleic Acids (for all biological samples)

1. Add 150 µl of MPC Protein Precipitation Reagent to 300 µl of lysed sample and vortex vigorously for 10 seconds.
2. Pellet the debris by centrifugation at 4°C for 10 minutes at ≥10,000 x g in a microcentrifuge.
3. Transfer the supernatant to a clean microcentrifuge tube and discard the pellet.
4. Add 500 µl of isopropanol to the recovered supernatant. Invert the tube 30-40 times.
5. Pellet the total nucleic acids by centrifugation at 4°C for 10 minutes in a microcentrifuge.
6. Carefully pour off the isopropanol without dislodging the total nucleic acid pellet.
7. Rinse twice with 70% ethanol, being careful to not dislodge the total nucleic acid pellet. Centrifuge briefly if the pellet is dislodged. Remove all of the residual ethanol with a pipet.
8. Resuspend the total nucleic acids in 35 µl of sterile, distilled water.

C Removal of contaminating DNA from total nucleic acid preparations using DNase I

Preparation of RNA Sample Prior to RT-PCR:

1. Add the following to an RNase-free, 0.5-ml microcentrifuge tube on ice:
   1 µg RNA sample
   1 µl 10X DNase I Reaction Buffer
   1 µl DNase I, Amp Grade, 1 U/µl
   DEPC-treated water to 10 µl
2. Incubate tube(s) for 15 min at room temperature. Inactivate the DNase I by the addition of 1 µl of 25 mM EDTA solution to the reaction mixture. Heat for 10 min at 65°C. The RNA sample is ready to use in reverse transcription, prior to amplification.

First-Strand cDNA Synthesis Using SuperScript III RT

(Invitrogen, Karlsruhe, Germany)

Indicated are volumes for a volume final of 20 µl and in parentheses, the concentration

1. Add the following components to an Rnase-free 200µl tube:

Oligo(dT_{20}) primers (50-250 ng)	1 µl
Total RNA (5 µg)	x µl
dNTP's (10 mM each)	1 µl
Sterile, distilled water to	12 µl

2. Heat mixture to 65°C for 5 min and quick chill on ice. Collect the contents of the tube by brief centrifugation and add:

First-Strand Buffer (5X)	4 µl
DTT (0.1M)	2 µl

3. Mix contents of the tube gently and incubate at 42°C for 2 min.
4. Add 1µl of SuperScript III RT and mix by pipetting gently up and down and add sterile, distilled water to a 20 µl final volume
5. Incubate tube at 25°C for 10 min
6. Incubate 50 min at 42°C
7. Inactivate the reaction by heating at 70°C for 15 min

Table (Appendix C) S 1: Summary of walking primers, which were used for DNA amplification and sequencing of the *Flo-2* region

ID	Forward Primer	Reverse Primer	TA	Comment
X-01228	CGCAAATATTGAATACCTCTT	CCGTGGGCACTTGTCGCATGG	50.9	
X-01229	CCTCCCGATCGTCGAGAATCA	TGGCGAGGCTGTAGAGGATCA	58.5	
X-01230	GCTTCCAAAGTCGTAGACCAC	ATTGTGCCGTCAACAGTACAC	55.4	
X-01231	CTGAGCCTGAATCAACTCTGG	CTCTCGCCTGCAACAACGTCT	56.6	
X-01232	CAGCCGCATTTAGTAGTACCC	CCTGGCCGCTGCAACACGAAC	57	
X-01233	ACAGCTCACTCAATCGACTGG	ATCGCCTAAATGTCACTTATC	52.5	
X-01234	AGCGTCATGAAATGATCTTAC	GGATGCCAAGCTAAGGGTCTA	50.6	
X-01235	CACAGGCAATCATACTCAGTT	GATCGAGGGCGTGAAGATAGC	53.8	
X-01236	ACCGGGTGATAGATGGTAGAC	TTGGGTGGCGTGGAGGGATTG	55.6	
X-01237	AGCATCAGTTAATCCATGAGC	TCGGAGTGGAGTTCGTTAAGG	55	
X-01238	CGCCCAGACCAATGTGATCCT	GGCCAATACACCATATCATGC	56.5	
X-01239	TATGCAAATTTGGACGTAATC	GACGGAACCATCGGGAACATC	54.1	
X-01240	ATAGCCTCCCAGTTGTCGTAG	ATCCTCGACCATTATGACTTT	54.3	
X-01241	ACGATCCCTTAGGCAGACGTA	CTGTTTGCATTACCCATTAGC	55.1	
X-01242	CCGCTTCCCATTTAGTTAGGG	TGACGAAGCGTATCACGATAG	52.2	
X-01243	GTGATCGATAAATCTAAGTGG	ACTGGTTGCAAGTGAGAGGAC	50.8	
X-01244	TGATGGTGCATCTTTGGATAG	AAACGCAACATCAGTAACAGG	51.3	
X-01245	GTTTGCCTAGGTACAGTGGTA	CGGCAATTCTTCGACGCTTAC	49.6	
X-01406	TCGGAACATATGTAATGCCTATT	CCGCCAAGGGAATTGTGACGAAA	52.0	substitutes 1246
X-01407	CCAAAGTGAAATGCGTAACAGTA	TCGATGGAGAGAGGAGCGGTGAA	54.0	substitutes 1246
X-01408	GCGGCATGAATCACATTCGTTGA	ACCCAACTTTTCCTTTACCTACC	54.6	substitutes 1246
X-01409	TGGCACACTTTATTTCGCTCTCT	TTGGGTGACGTCACACGCACAGT	55.8	substitutes 1246
X-01247	CTCTCCATCGAACTCTCATAC	CACATCGGATTCAGGGTCATC	52.1	
X-01248	AAAGCAAAGCAAGTTACTGAC	CGGGTGCAGGCAGGATACCAA	52.9	
X-01249	TCTCCGCTCGTTTGTGGTTAT	TGTTCGATGAGCTAGTTCTAC	51.9	
X-01250	GAGTGCCGAAAGTAGCGAGTT	GAATCAATTGATGTGAGTGAA	51.2	
X-01251	CTGCGACTGAAGACCGACTTT	TGTGTGGGTGGTTAGGTTAGG	51	
X-01252	AATCGCTGCCCAAGGTTAGTG	CTAGGACACCACACAGACGAG	53.3	
X-01333	CTGCCCAAGGTTAGTGCTCGATATG	GCACCTTATATGGCACAATTCCGTA	54.1	substitutes 1253
X-01254	AACCATTTAAGCAGAGTGTTA	TAAATGCATATTCGTAGTTCA	47.6	
X-01255	CAATTGGCGCTGCTTTACACC	TTGCTTTCTCCATCAGACTGC	50	
X-01256	ACAACACATATTGAACTACGA	TGCGTGCACTCTAACCTTTGA	47.5	
X-01257	AACTGGTTCATTTCGTCTTTA	TCAATCAGGAATGGTACACAG	50.9	
X-01258	GTGCACGCATAAAGATTACAC	AAGGAGCTGATTAAGGTGAAG	54.5	
X-01259	CCATTCCTGATTGAGACGAAC	TATGGCATGACAGAGGCTTCT	58.4	
X-01260	AGACCGTCCTCGTCGTAGAAG	CGGAAACCTATATTCGATAGC	54.2	

Table (Appendix C) S 2: DNA sequence alignment of the upper part of the *Flo-2-C* transcript (+46) in *D. melanogaster* and *D. sechellia* (*D. sec*).
African lines (A-number) show high levels of nonsense mutations (■)…

Rel codon pos	12950			12953			12956			12959			12962			12965			12968			12971			12974		
D. sec	G	A	A	-	-	-	-	-	-	-	-	-	-	-	-	C	C	C	A	C	A	C	A	C	A	C	A
■	E															P			T			H			T		
A082	G	A	A	A	C	C	-	-	-	-	-	-	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T									T			H			T			H			T		
A084	G	A	A	A	C	C	-	-	-	-	-	-	-	-	-	-	-	-	A	C	A	C	A	C	A	C	A
■	E			T															T			H			T		
A095	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T			T			H			T			H			T			H			T		
A131	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T			T			H			T			H			T			H			T		
A145	G	A	A	A	C	C	-	-	-	-	-	-	-	-	-	-	-	-	A	C	A	C	A	C	A	C	A
■	E			T															T			H			T		
A157	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T			T			H			T			H			T			H			T		
A186	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
A191	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T			T			H			T			H			T			H			T		
A229	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T			T			H			T			H			T			H			T		
A377	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
A384	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
A398	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T			T			H			T			H			T			H			T		
E01	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E02	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E11	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T			T			H			T			H			T			H			T		
E12	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E13	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
■	E			T			T			H			T			H			T			H			T		
E14	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E15	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E16	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E17	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E18	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E19	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		
E20	G	A	A	A	C	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A
▶	E			T			T			H			T			H			T			H			T		

… due to frame shifts (X) or premature stop codons (*) compared to the European lines (E-number) which mostly show intact transcripts (►). Codons (shaded in grey) specify an amino acid (below each codon, single letter code, not shaded).

12977	12980	12983	12986	12989	12992	12995	12998	13001	13004
A A C A A A -
N K
C A C A A A C A C T -
H K H S
C A C A A A C A C T -
H K H S
C A C A A A C A C T -
H K H S
C - - A A A C A C T -
Q T L
C A C A A A C A C T -
H K H S
C A C A A A C A C T -
H K H S
C A C A A A C -
H K H
C A C A A A C A C T -
H K H S
C A C A A A C A C T -
H K H S
C A C A A A C A C T -
H K H S
C A C A A A C A C T -
H K H S
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
C A C A A A C A C T C A C A C A C A C A C A C A C A C A C A
H K H S H T H T H T T
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
C A C A A A C A C T - - - - - - - C A C A C A C A C A C A C A
H K H S H T H T T
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
C - - A A A C A C T - - - - - - - C A C A C A C A C A C A C A
Q T L T H T H T H T
Insertion:	1	2	3	4	5	6	7	8	9	10	11	12	13	14

	13007						13010						13013			13016			13019			13022			13025			13028			13031			
D. sec	-	-	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	-	-	-	-
									L			T			H			T			H								
A082	-	-	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	A	A	C	A	-	-	-	-	-
											H			T			Q			T							
A084	-	-	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
											H			T			H			T			*			T	
A095	-	-	-	-	-	-	-	-	-	-	-	-	C	A	C	A	C	A	C	A	C	A	C	A	A	A	C
															H			T			H			K			H
A131	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
								S			H			T			H			T			*			T	
A145	-	-	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
											H			T			H			T			*			T	
A157	-	-	-	-	-	-	-	-	-	-	-	-	C	A	C	A	C	A	C	A	C	A	C	A	A	A	C
															H			T			H			K			H
A186	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A191	-	-	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
											H			T			H			T			*			T	
A229	-	-	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
											H			T			H			T			*			T	
A377	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
									H			T			H			T			H			K			H
A384	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
									H			T			H			T			H			K			H
A398	-	-	-	-	-	-	-	-	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
											H			T			H			T			*			T	
E01	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E02	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E11	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
	Q			T			L			S			H			T			H			I			N		
E12	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E13	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
	Q			T			L			S			H			T			H			I			N		
E14	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E15	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E16	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E17	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E18	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E19	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
E20	C	A	A	A	C	A	C	T	C	T	C	A	C	A	C	A	C	A	C	A	C	A	T	A	A	A	C
			N			T			L			T			H			T			H			K			H
	15	16	17	18	19	20																												

13034			13037			13040			13043			13046			13049				
-	-	-	-	-	-	A	C	G	T	A	C	A	G	A	T	G	A	T	G	D. sec
						T			Y			R			*					
-	-	-	-	-	-	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A082
						T			Y			R			*					
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A084
	Q			T			R			T			D			D			X	
A	C	A	C	-	-	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A095
		T					R			T			D			D			X	
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A131
	Q			T			R			T			D			D			X	
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A145
	Q			T			R			T			D			D			X	
A	C	A	C	-	-	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A157
		T					R			T			D			D			X	
-	-	-	-	-	-	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A186
								V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A191
	Q			T			R			T			D			D			X	
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A229
	Q			T			R			T			D			D			X	
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A377
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A384
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	A398
	Q			T			R			T			D			D			X	
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E01
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E02
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E11
T			N			T			Y			R			*					
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E12
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E13
T			N			T			Y			R			*					
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E14
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E15
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E16
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E17
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E18
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E19
		K			H			V			Q			M			M			
A	C	A	A	A	C	A	C	G	T	A	C	A	G	A	T	G	A	T	G	E20
		K			H			V			Q			M			M			

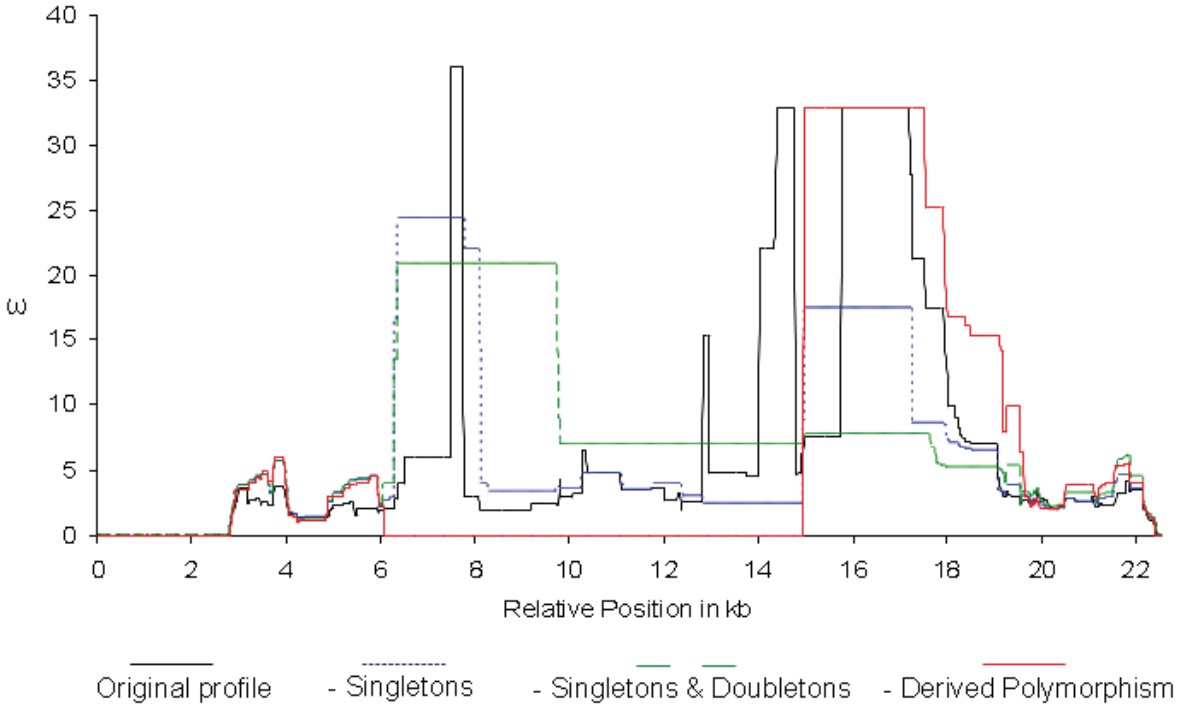


Figure (Appendix C) S 1: ω statistic for the *Flo-2* region.

Given are the profiles for four different datasets: the original dataset (black profile), the dataset without singletons that were private to Europe (- Singletons, dotted blue profile), the dataset without private singletons and doubletons (- Singletons & Doubletons, dashed green profile), and dataset without derived European polymorphism (- Derived Polymorphism, red profile)

Curriculum Vitae

Annegret Werzner (*20.07.1981, Karl-Marx-Stadt, Germany)

E-Mail| Homepage: werzner@bio.lmu.de | http://evol.bio.lmu.de/

Education

2007—2011 Ludwig-Maximilian-University (LMU), Munich, Germany

- **PhD Student**: Population Genetics and Evolutionary Biology

2000—2007 Justus-Liebig-University (JLU), Giessen, Germany

- **Diploma**, **1.1** (M. S.), Molecular Ecology, Department of Animal Ecology
- **Intermediate Diploma** (Bachelor) in Biology

1992—2000 Immanuel-Kant-Gymnasium, Chemnitz, Germany

- General Qualification for University Entrance, **1.9**

Research Experience

2007—2011 **Quantitative Genetics** and Evolution of *Drosophila melanogaster*:

- Selective sweep mapping by means of **sequencing techniques** and **statistical analysis**
- **Association mapping** of quantitative trait loci affecting cold tolerance
- **Morphological analysis** of population structure and local adaptation

2006 Diploma student, Department of Animal Ecology, JLU

- Project SFB 299: "**GIS** based **population genetic analysis** of *Chorthippus parallelus* in two agricultural landscapes"

2005 Research Assistant, Department of Animal Ecology, JLU

- Project: "The effect of **landscape structure** on the symmetry and the mass-size relationship of Carabid beetle species"

Teaching Experience

2009	Laboratory course, **Population Genetics** and **Evolution**, LMU
	Tutorial group, **Statistics for natural sciences**, LMU
2007	Practical course, **Insect Immunology**, Department of Zoology, JLU
2003	Basic class in **Zoology**, Department of Animal Ecology, JLU
2002	Basic class in **Botany**, Department of Plant Ecology, JLU

Selected Presentations

2010 Research Unit **"Next generation sequencing"** (Forschergruppe 1078), LMU

- Oral presentation: "Adaptation to a cold environment- QTL on the X chromosome of *Drosophila melanogaster*"

2009 ESEB 12th congress, Turin, Italy

- Poster presentation: "Recent positive selection in European *Drosophila melanogaster*"

University of Veterinary Medicine Vienna, Austria, **Invited speaker**

- Oral presentation: "How *Drosophila melanogaster* conquer the world- QTL for cold resistance on the X chromosome"

2006 19. Westdt. Entomologentag, Entomologische Gesellschaft Düsseldorf

- Oral presentation: "Population genetic analysis of *Chorthippus parallelus* in two agricultural landscapes"

Publications

Svetec N[1], Werzner A[1,2], Wilches R*, et al.* (2011) Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis. *Molecular Ecology*, **20**, 530-544. **([1]shared first authorship, [2]corresponding author)**

Laurent SJY, Werzner A, Excoffier L, Stephan W (2011) Approximate Bayesian analysis of *Drosophila* melanogaster polymorphism data reveals a recent colonization of Southeast Asia. *Molecular Biology and Evolution*, **28**, 2041-2051.

Skills

Computer/IT: MS-Office; Open Office; Adobe; R; DNA Software such as SeqMan, Oligo Primer, MegAlign, BioEdit and DnaSP; EndNote; Photoshop; CorelDRAW; GIS

Languages: **German** (native), **English** (fluent), **French** (basic)