Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften an der Fakultät

für Biologie der Ludwig-Maximilians-Universität München

Searching for genes involved in the adaptation of *Drosophila melanogaster* to the European climate

Nicolas Henri Svetec

aus

Colmar, France

2009

Erklärung:

Diese Dissertation wurde im Sinne von §12 der Promotionsordnung von Prof. Dr. Wolfgang 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 selbständig, ohne unerlaubte Hilfe angefertigt wurde.

München, ……

Nicolas Svetec

Dekan: Prof. Dr. Jürgen Soll

1. Gutachter: Prof. Dr. Wolfgang Stephan

2. Gutachter: Prof. Dr. John Parsch

Dissertation eingereicht am: 17.03.2009

Datum der Disputation: 29.04.2009

1. Table of contents

2. Note

In this thesis I present my doctoral research, all of which has been done by me except for the following: Pavlos Pavlidis performed *CLR* and *SweepFinder* tests and ran the simulations for the sweep age estimations. Both Pavlos Pavlidis and Prof. Wolfgang Stephan contributed to the research article cited below that served as a basis for the present thesis manuscript. The generation of the substitution lines for the QTL mapping experiment was a collaborative work with Dr. David de Lorenzo.

The results from my thesis have contributed to the following publication:

Svetec, N., P. Pavlidis, and W. Stephan (2009). "Recent strong positive selection on *Drosophila melanogaster HDAC6*, a gene encoding a stress surveillance factor, as revealed by population genomic analysis." Mol. Biol. Evol. (in press)

3. List of abbreviations

AFP	Anti-freeze protein
bp	Base pair
CCR	Chill coma recovery
CI	Confidence interval
CIM	Composite interval mapping
CLR	Composite likelihood ratio
cM	Centimorgan
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
gen.	Generation
GO	Gene Ontology
GOF	Goodness-of-fit
HDAC	Histone deacetylase
HSP	Heat shock protein
IM	Interval mapping
kb	Kilobase pair
LD	Linkage disequilibrium
LR	Likelihood ratio
Mb	Megabase pair
mRNA	Messenger ribonucleic acid
N_e	Effective population size
PCR	Polymerase chain reaction
QTL	Quantitative trait locus

QTN	Quantitative trait nucleotide
r	Recombination rate
RAPDs	Randomly Amplified Polymorphic DNAs
RFLPs	Restriction Fragment Length Polymorphisms
RIL	Recombinant inbred line
RNAi	Ribonucleic acid interference
S	Number of segregating sites
SFS	Site frequency spectrum
SNP	Single nucleotide polymorphism
THP	Thermal hysteresis protein
tRNA	Transfer ribonucleic acid
UTR	Untranslated region
V_A	Additive variance
V_D	Dominance variance
V_E	Environmental variance
V_G	Genetic variance
V_I	Epistatic variance
V_P	Phenotypic variance
WT	Wild type
θ	Population mutation rate
θ_W	Watterson estimator of θ
π	Nucleotide diversity (average number of pairwise differences)
μ	Mutation rate

4. Zusammenfassung

Ziel der vorliegenden Arbeit ist es, Gene zu identifizieren, die bei der ökologischen Anpassung von *D. melanogaster* eine Rolle gespielt haben. Diese Spezies, die in Afrika beheimatet ist, hat in den letzten 100.000 Jahren sowohl in Afrika als auch nach ihrer Kolonisierung von Habitaten außerhalb Afrikas unter sehr unterschiedlichen klimatischen Bedingungen gelebt. Um besagte Gene zu finden, wurden zwei verschiedene Ansätze gewählt: (1) eine genomische Region, die eine Reihe von ökologisch relevanten Kandidatengenen enthielt und die gemäß vorheriger Untersuchungen unter positiver Selektion stand, wurde mit Hilfe der *selective sweep* Methode feinkartiert, und (2) wurde ausgehend von einem in der Anpassung an temperierte Zonen wichtigen Phänotyp (Kältetoleranz) eine *quantitative trait locus* (QTL) Analyse durchgeführt.

Mit Hilfe der *selective sweep* Methode, die im Stephan Labor pioniert wurde, konnte ein kleiner Bereich von nur 2.7 kb in der anzestralen afrikanischen Population gefunden werden, die mit dem 3' Ende des *HDAC6* Gens überlappt. Dieses Gen kodiert für einen Streßkontrollfaktor. HDAC6 ist eine ungewöhnliche Histondeacetylase. Sie ist im Cytoplasma lokalisiert und besitzt Ubiquitin-Bindungs- und Tubulin-Deacetylase Aktivitäten. Diese Eigenschaften machen HDAC6 zu einem wichtigen Regulator cytotoxischer Streß-Resistenz.

Die QTL Analyse zeigte, dass afrikanische und europäische Populationen sich stark in ihrer Fähigkeit, Kälte zu ertragen, unterscheiden. Durch Ersetzen aller Chromosomen (außer dem X Chromosom) konnte ich zeigen, dass ein beträchtlicher

Teil dieser phänotypischen Varianz durch die Variablität von Genen auf dem X Chromosom bestimmt ist. Diese Gene wurden schließlich in zwei genomischen Regionen des X Chromosoms lokalisiert. Durch Vergleiche meiner Resultate mit denen von Assoziationstudien und mit Genontologie (GO) Tabellen konnte ich eine Liste von X-gekoppelten Kandidatengenen zusammenstellen, die die Kältetoleranz von *D. melanogaster* bestimmen. Diese Liste ist sehr kurz und wird deshalb für zukünftige Studien der natürlichen Selektion, die an diesen Loci gewirkt hat, von großer Hilfe sein.

5. Abstract

The aim of the present work was to identify the genes that played a role in ecological adaptation in *D. melanogaster*. This species, which originated in Africa, successfully adapted to a broad range of climates during the last 100.000 years. To find the genes involved, I used two different approaches: (1) a genomic region containing several ecologically relevant candidate genes and putatively carrying footprints of selection was investigated using selective sweep mapping, and (2) cold tolerance that might have been an important phenotype for the adaptation to the temperate climates was investigated using a QTL analysis.

Using the technique of selective sweep mapping pioneered in the Stephan's group, I detected evidence that recent strong positive selection has been acting on a small DNA region of 2.7 kb overlapping with the 3' end of the *HDAC6* gene in the ancestral African population. This gene codes for a newly characterized cell stress surveillance factor. HDAC6 is an unusual histone-deacetylase. It is localized in the cytoplasm and has a ubiquitin-binding and a tubulin-deacetylase activity. These properties make HDAC6 a key regulator of cytotoxic stress resistance.

The phenotypic analyses show that the African and the European populations have very strong cold tolerance differences. By removing the effects of the autosomes, I showed that a significant amount of the phenotypic variance is due to genetic factors carried by the X chromosome. These factors were then more precisely mapped to two genomic regions of the X chromosome. By comparing the present results with other association studies and the Gene Ontology database, it was possible to determine a list

of candidate genes influencing cold tolerance in *D. melanogaster*. As this list is limited to a very small number of genes, additional investigations for footprints of selection in these regions may be used to confirm their role in ecological adaptation.

6. General introduction

6.1 The theory of species evolution

The field of modern biology was founded by the French naturalist Jean-Baptiste de Lamarck (1744-1829) at the end of the 18th century. His goal was to unify the study of all living organisms into a single scientific field: biology (in greek *bios* "life", and *logos* "study"). Another contribution of J.-B. Lamarck in his famous "*Philosophie Zoologique*" (1809) was to establish two major concepts in biology. First, he proposed that species could transform over time, and second that some phenotypic characters are heritable. Even if his theory of the inheritance of acquired characters has been disproved later, he highlighted two major properties of living organisms: they reproduce such that offspring inherit some characters from their parents, and species evolve over time.

Some time later, a major contributor to the field of biology was Charles Darwin (1809-1882). His world tour on the *Beagle* and especially the observations he made on birds of the Galapagos Islands, brought C. Darwin to reject Lamarckian concepts that where unable to explain a fundamental idea, namely that species are wonderfully well adapted to their way of life. This forced C. Darwin to build his own theory: because of competition between individuals, only those that are the best adapted survive and give birth to more offspring, therefore increasing their frequency in the population. Then, when the environment changes a new category of traits becomes advantageous and the process restarts. Interestingly, at the same time, Alfred Russel Wallace (1823-1913) developed independently a very similar theory with the same two central ideas of natural selection and species evolution. They exchanged ideas for a long time before C.

Darwin finally published the theory in the famous book: "*The Origin of Species*". C. Darwin's book convinced the scientific community that species do evolve but the fact that natural selection is the driving force of evolution was still controversial. Indeed, he was missing the basic explanations of the mechanism of heredity when he formulated his theory of natural selection. Because of this lack, his theory remained controversial for a long time.

Nearly at the same time, Gregor Mendel (1822-1884), a monk and Austrian botanist, was working on the precise elements C. Darwin was missing: the mechanisms of heredity - that later became the basis of modern genetics. However, it is only in the first half on the 20th century that the two theories were combined into what is now called neo-Darwinism. This work has mainly been carried out by R.A. Fisher (1890-1962), J.B.S. Haldane (1892-1964) and S. Wright (1889-1988). T. Dobzhansky (1900-1975) who was working on *Drosophila* also largely contributed to the popularization of the Darwinian theory of evolution through the several successive editions of his book the "*Genetics and the Origin of species*."

Nowadays, the theory of biological evolution is probably one of the most important theories in science. Within biology, the most interesting aspect of the theory is its unifying role as it has implications across all scales at which biology can be studied: from the microscopic world of molecules that an organism is made of, to the nature of species communities that constitute ecosystems. During the last century, evolutionary theory thus contributed greatly by bringing new research perspectives into a broad range of biological fields including systematics, genetics, embryology, paleontology, biometrics, ethology, medicine, psychology, and also others sciences such as cosmology and linguistics (reviewed in Derry 2009). This can be illustrated by

a famous quotation from T. Dobzhansky: “Nothing in biology makes sense except in the light of evolution”.

6.2 The population genetics approach of evolution

History of the main theories: One of the main goals of evolutionary biologists is the study of how species adapt to their environment and the detection of examples of adaptation. In this research area evolutionary biology benefited from the input of population geneticists whose main objective is the study of genetic variation in natural populations. As genetic variability is the substrate of natural selection, a key question is to understand how these genetic variations (such as mutations) appear and spread in natural populations. In 1968, Kimura proposed his theory of neutral evolution, which posits that at the molecular level most of the nucleotide changes have negligible impact on fitness. This idea suggests that most of the DNA (deoxyribonucleic acid) and protein evolution is driven by random processes like genetic drift (leading to random fixation or loss of an allele) rather than by natural selection.

In the 1980s-90s, molecular population geneticists, such as Stephan and Langley (1989), Aguadé *et al.* (1989) and Begun and Aquadro (1992), made some observations that were inconsistent with the neutral expectations while studying levels of DNA polymorphism in fruit flies. They observed that the level of nucleotide polymorphism is positively correlated with the recombination rate. Following the neutral model this would mean that the mutation rate must be correlated with the recombination rate. But this was not the case, as divergence (which depends on the mutation rate and the time since two species split) did not follow such pattern. They thus rejected the neutral hypothesis, and tried to come up with an explanation.

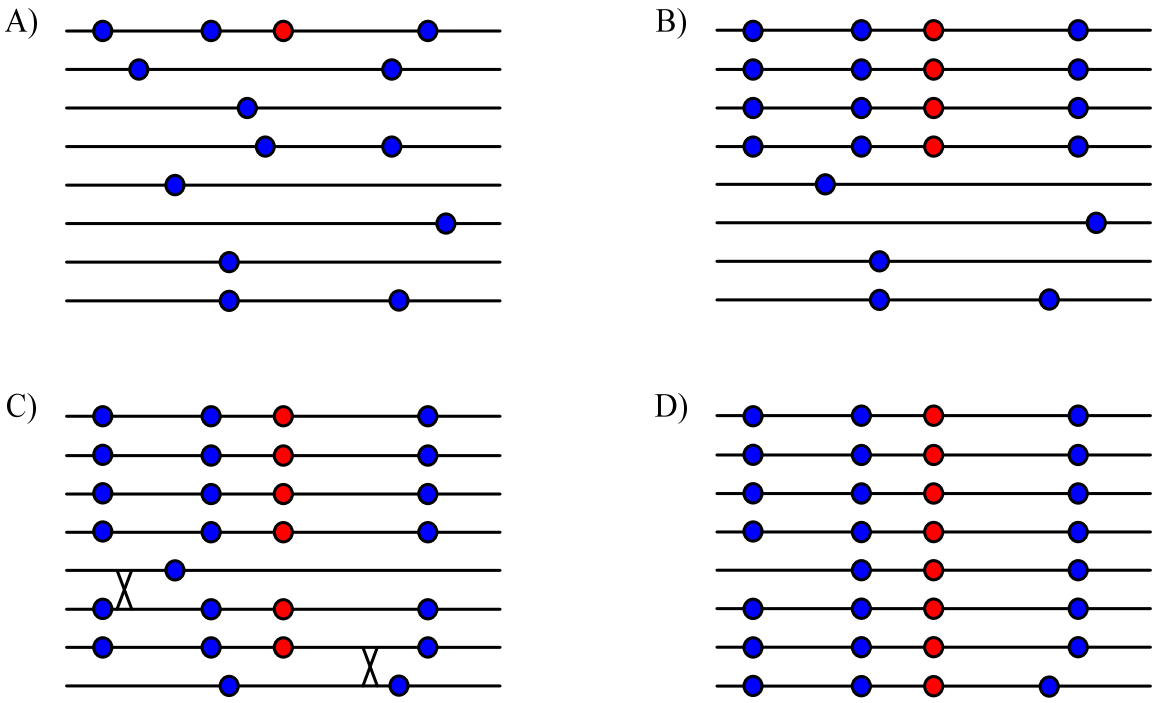


Figure 1: Genetic hitchhiking. A beneficial mutation (red dot) occurs into a population (A). Because of the effect of natural selection, its frequency rises in the population (B). Because of its physical linkage to the surrounding DNA adjacent neutral polymorphisms (blue dots) also increase in frequency. Quickly the level of variability around the beneficial mutation decreases and only the regions further away maintain variability because of recombination (C). Finally the beneficial mutation reaches fixation (D).

An alternative explanation for the variation-recombination correlation is provided by the genetic hitchhiking model proposed by Maynard Smith and Haigh (1974). This model assumes that some DNA regions evolve under strong positive selection. When a beneficial mutation occurs, it confers to its bearer a selective advantage. Therefore, the mutation is positively selected and increases in frequency in the population, until it eventually gets fixed. However, because of its physical linkage to the surrounding DNA, adjacent neutral polymorphisms may also rise in frequencies - in other words, they hitchhike along with the beneficial mutation (see Figure 1). In such a case, the theory predicts the emergence of a specific polymorphism pattern: (1) diversity vanishes around the site of selection, (2) the site frequency spectrum (SFS) of polymorphisms is shifted toward low- and high-frequency derived variants (Braverman *et al.* 1995; Fay and Wu 2000), and (3) linkage disequilibrium (LD) is elevated in the early phase of the fixation process (Kim and Nielsen 2004; Stephan *et al.* 2006). Importantly, the width of the valley of reduced variation is mainly determined by the ratio of the rate of recombination around the site of selection and the strength of selection. This is a plausible explanation of the variation-recombination correlation (Kaplan *et al.* 1989).

Population genetics tests for selection: As the theory clearly predicts the signature of a selective sweep, there has been an increasing desire of mapping such selective event. It, thus, motivated the development of a large number of population genetics statistical tools that can be used to screen genomes for footprints of selection. The simplest indicator one can look at is the level of nucleotide polymorphism. It can be estimated from sequence data either by using the average number of segregating

sites θ (Watterson 1975) or the average number of pairwise differences π between two sequences (also called nucleotide heterozygosity; Tajima 1983). Under neutral conditions, the level of nucleotide polymorphism is supposed to be equal to the population mutation rate, which is $\theta=4Ne\mu$ for diploid species. However, under the hitchhiking model, one expects a drastic local reduction of nucleotide polymorphism around the selected site. The Tajima's *D* statistic (Tajima 1989) – which is based on the comparison of π to θ – measures deviations of the site frequency spectrum (SFS) from the neutral expectation. Under standard neutral conditions both π and θ are supposed to be equal (see formula above). But under hitchhiking conditions, the fixation of the beneficial allele removed most of the nucleotide polymorphism. The only variants left are the point mutations that occurred after the fixation of the beneficial allele. All these newly arisen mutations are thus in very low frequency making π smaller than θ and driving the Tajima's *D* statistic toward negative values. Fu and Li's (1993) as well as Fay and Wu's (2000) tests are following a similar rationale but use polarized data as selective sweeps are supposed to generate a lack of middle-frequency derived variants.

Although the described methods are relatively straightforward and easy to use, they have the disadvantage to be quite sensitive to some other confounding evolutionary forces that produce SNP (single nucleotide polymorphism) patterns similar to those of a selective sweep. Charlesworth *et al.* (1993, 1995), for example, proposed that some genomic regions may exhibit a lack of nucleotide polymorphism because of background selection. Under this model, the neutral variants linked to deleterious mutations are removed from the population by natural selection. Another important confounding factor is demography, notably population structure, population bottlenecks (reductions of effective population size) or population expansions that may stochastically generate large departures from neutral expectations.

More advanced ways of testing have thus been developed using computer simulations. Basically, the observed data are compared to some artificial data set generated under a special hypothetical scenario. Then, statistical tests confirm or reject if the simulated scenario fits the observed data. Most of the computer simulations are based on the coalescent theory (Kingman 1982). One of these tests, the Composite Likelihood Ratio (*CLR*) test, was proposed by Kim and Stephan (2002). It compares the likelihood of observing a given number of derived variants at a polymorphic site under a selective sweep model to that expected under the standard neutral model. Both likelihoods are based on the site frequency spectrum (SFS) and the spatial distribution of polymorphic sites. Nevertheless, as the *CLR* test uses the standard neutral model as null hypothesis, the violations of the model's assumptions might affect the results of the test (Thornton and Jensen 2007). Thus, Jensen *et al.* (2005) proposed a complementary method to the *CLR* test. Their *GOF* (goodness-of-fit) test allows one to examine whether the rejection of the null hypothesis in the *CLR* test is either due to departure from the neutral expectations or due to a better fit to the selective sweep model. Other improvements of these methods are reviewed in Pavlidis *et al.* (2008). The logic behind the *SweepFinder* test (Nielsen *et al.* 2005) is similar to that of the *CLR* test, but the null hypothesis is estimated from the empirical background distribution of the SFS. The genome wide effects of demography on the SFS are thus taken into account increasing the robustness of the tests for selection (Jensen *et al.* 2005; Nielsen *et al.* 2005; Thornton and Jensen 2007).

In the search for footprints of selection: With these statistical tools available, a common approach has been to screen genomes for footprints of selection. As such an approach requires large amounts of DNA sequence data, most of the studied species are

model organisms like *D. melanogaster* (e.g., Harr *et al.* 2002; Glinka *et al.* 2003; Orengo and Aguadé 2004) mice (Ihle *et al.* 2006) and humans (e.g. Akey *et al.* 2004; International HapMap Consortium 2007). Glinka *et al.* (2003) performed the first scan of the *Drosophila melanogaster* genome for selective sweeps. The study was subsequently expanded by Ometto *et al.* (2005) and Hutter *et al.* (2007). As the species range of *D. melanogaster* is supposed to have expanded out of Africa after the last glaciations, they compared two populations, one originating from Africa (Zimbabwe) and the other from Europe (The Netherlands). The general approach was to sequence short DNA fragments every 50 kb along the X and the third chromosomes. By searching for outlier fragments, Ometto *et al.* (2005) defined a first set of candidate loci for positive selection. Later, Li and Stephan (2006) refined the Ometto *et al.* (2005) analysis and redefined the most likely candidate regions for positive selection. This constituted the subsequent starting point of two studies that confirmed the existence of positive selection on two loci: *brk* and *ph-p* (Glinka *et al.* 2006; Beisswanger *et al.* 2006; Beisswanger and Stephan 2008). Other studies also reported evidence for the occurrence of positive selection in the *D. melanogaster* genome (Orengo and Aguadé 2004; Bauer DuMont and Aquadro 2005; Pool *et al.* 2006; Jensen *et al.* 2007).

Although these methods of investigation are very powerful, they have some limitations. Indeed, they rely on the ability of selection to generate very specific polymorphism patterns along the genome. Moreover, demographic factors such as population size bottlenecks may stochastically produce patterns of nucleotide diversity across the genome that resemble those of selective sweeps. Therefore, a major challenge of these analyses has been (and still is) to distinguish the effects of selection from those of demography. Another limitation is the resolution of selection mapping: the candidate regions for selection identified by these tests, were generally very large

(often ~100 kb) and contained many genes (~10). This is particularly the case for humans (e.g. Williamson *et al.* 2007). Although some progress has been made in *Drosophila* (Pool *et al.* 2006; Jensen *et al.* 2007; Orengo and Aguadé 2007; Beisswanger and Stephan 2008), a major challenge ahead is to develop strategies that help to narrow down the target regions of selection such that it is possible to map the site of selection to individual genes or gene regions. This is essential for ultimately understanding adaptation at the functional level. Finally, since detecting footprints of selection does not prove the adaptive role of a molecular variant, the candidate alleles must be tested functionally. These limitations suggest that adaptation cannot be studied from a single perspective only.

6.3 Quantitative genetics for studying evolution

C. Darwin constructed his theory without knowing about the precise mechanisms of genetic inheritance. His whole approach was based on the observation of phenotypic differences between individuals: all these little differences between individuals, which "afford material for selection to act on in the same manner than man accumulates in any given direction individual differences in his domestic productions" ("*The Origin of Species*" C. Darwin). On the functional side, the study of adaptation can receive significant input from quantitative genetics. Indeed, this field investigates the mechanisms of inheritance of complex phenotypic traits - i.e. all these characters that are a matter of degree rather than of kind (Falconer and Mackay 1996). In other words, the goal of quantitative geneticists is to identify which gene codes for which trait. Interestingly, the key founders of quantitative genetics are the same as those of population genetics (R.A. Fisher, J.B.S. Haldane, and S. Wright). This certainly illustrates the particular need to understand how both genetic and phenotypic variability are linked and how they evolve in natural populations.

The basics of quantitative genetics: The basics of quantitative genetics rely on the assumption that the expression of any phenotypic trait is influenced by both the genes and the environment. Thus, in a population of individuals, the phenotypic variance (V_P) of a quantitative trait is given by $V_P = V_E + V_G$, where V_E the variance attributable to environmental effects and V_G is the amount of variance induced by the genotypic variability between individuals.

As a part of the genotypic variability, V_G, is transmitted to the offspring, it is the object of a particular research interest. V_G can be partitioned as follows:

$$V_G = V_A + V_D + V_I,$$

where V_A is the additive, V_D the dominance and V_I the epistatic variance. Although all these effects are genetic only the additive variance V_A is heritable. This variance component is thus the one that will respond to natural and artificial selection. Therefore, one can predict how well a population will respond to selection by estimating the heritability of the trait. The heritability is given by the ratio V_A / V_P. It corresponds to the percentage of phenotypic variance that is transmitted to the next generation. Nevertheless for practical reasons, people usually rather estimate heritability in its broad sense: V_G / V_P (the various heritability calculations are reviewed in Visscher *et al.* 2008). Until recently, most of the quantitative trait studies were restricted to quite descriptive analyses such as parent-offspring trait correlations, estimations of the different variance components and heritability, or analysis of the response to artificial selection. A plausible reason for this is the incapacity of Mendelian genetics to deal with the genetic complexity of quantitative traits. Indeed, the general assumption is that a continuously distributed trait is polygenic (i.e. controlled by a large number of genes of small effect) making the Mendelian approach inappropriate.

Research on the mapping of genes involved in complex traits has been largely motivated by the direct applications in agriculture (to improve cattle and crop quality) as well as in medicine (to map disease genes). With the advance of genetics and statistics the first Quantitative Trait Loci (QTL) analysis was performed at the beginning of the 20th century. Using morphological markers, Sax (1923) showed an association between seed weight and seed coat color in beans suggesting that the two

traits are coded by factors closely linked on the DNA, such loci is called a QTL. The general logic behind QTL mapping is very simple: it consists in measuring the degree of association between a trait and a genetic marker in a population of recombinant individuals. The higher the association, the closer is the marker to the QTL. Thus, the success of QTL analysis relies on three essential aspects: the generation of the recombinant individuals, the abundance of markers and the genotype-trait association measurement.

QTL mapping designs: Recombinant individuals can be obtained from different crossing designs. The initial step is the same for most of them: two inbred lines (PA and PB) differing for the trait of interest are crossed to produce an F1 generation which is heterozygote (A//B) but genetically completely homogeneous. In the Backcross design, the F1 individuals are crossed to one of the parental lines. If recombination occurred during the formation of gametes at meiosis in the F1, the F2 offspring of this cross will thus receive a parental chromosome and a chromosome supposed to carry a mixture of PA and PB genetic material. The inconvenience of such a design is that a large number of F2 individuals need to be screened, which is tedious particularly in species that have relatively low recombination rates. An alternative design is the construction of Recombinant Inbred Lines (RILs) (Burr *et al.* 1988). In this design the F1 individuals are allowed to mate freely with their siblings for several generations. Strains are then established by inbreeding the offspring of brother-sister pairs. Compared to the backcross design it has the advantage to increase the amount of recombination and thus the precision of the mapping. Nonetheless, for practical or ethical reasons, the construction of RILs is not possible in all species. The two experimental designs described above are the most commonly used in *Drosophila*. For

other organisms such as plants other crossing schemes exist which can be based on selfing for example (Lander *et al.* 1987). Methods for QTL mapping also exist for outbred populations (Lynch and Walsh 1998).

Markers for QTL analysis: Once the recombinant individuals have been generated, they must be genotyped. For that, Sax, who performed the first QTL mapping experiment, used morphological markers. Such markers generally consist of special allelic variants producing visible phenotypes. But the general availability of such morphological markers is limited in most species. This is problematic because the power of QTL mapping directly depends on the number of genetic markers available. However, special efforts have been made for some model organisms like *Drosophila* and a sufficient number of visible markers were identified. This made *Drosophila* a model organism for QTL mapping. Nevertheless, with the increasing number of studies, a second limitation to their use has been discovered: some morphological markers might not be completely neutral toward the traits measured. With the progress in sequencing/genotyping techniques and the availability of fully sequenced genomes, molecular markers turned out to be best adapted for such analysis. Thus, allozymes, RFLPs (Restriction Fragment Length Polymorphisms), or RAPDs (Randomly Amplified Polymorphic DNAs) were the first molecular markers to be used (Botstein *et al.* 1980; Donis-Keller *et al.* 1987; Long *et al.* 1998). Nowadays, because of the much reduced costs of sequencing, SNPs, indels, P-elements insertions, as well as microsatellites provide an unlimited source of very accurate markers that are in general completely neutral toward the trait of interest.

Statistical analysis in QTL mapping: Technical progress allowed the generation of large populations of recombinant individuals and the typing of a large number of markers. This made possible to screen whole genomes for QTL. Nevertheless, the analysis of such large amounts of data required the use of a special statistical framework. One of the simplest designs is the bulked segregant analysis (Arnheim *et al.* 1985; Michelmore *et al.* 1991). In this design, the individuals at each of the two extremes of the trait distribution are pooled. Across the two groups, the markers that are associated with the QTL are expected to have a non-random distribution. The statistical testing is thus simple and can be done by common tests such as the chi-square test. Another approach is to consider each marker independently and look for association with the trait of interest (single marker method; Thoday, 1961). Even though this method is statistically simple, it usually has two major disadvantages: first, the exact position of the QTL is unknown, and second the effect of the QTL is underestimated. This can be overcome by using Interval Mapping (IM; Lander and Botstein 1989). Instead of testing single markers, an interval between two adjacent markers is considered. It has the constraint of requiring a precise marker map that takes in account the possibility of double recombinants within an interval, but it greatly increases the mapping precision. Later Zeng (1994) proposed an improvement to the IM method: the Composite Interval Mapping (CIM). In the IM method, the presence of several QTL in the same linkage group can create bias in the estimated position and the effects of a QTL. Indeed, if two QTL are close, it may violate IM's basic assumption that each test interval is independent. CIM is, thus, an extension of IM that uses the special mathematical properties of multiple regression analysis to control for the effects of other possible QTL.

The *Drosophila* species are perfect model organisms for the QTL approach. The generation of RILs is possible, its genome is not too large and molecular markers are available. A large number of traits have thus been analyzed using this method (reviewed in Mackay 2001, 2004). And, what is of particular importance for this thesis: some of these QTL studies have been developed in an evolutionary context, for example, the mapping of sexual isolation factors between *D. simulans* and *D. mauritiana* (Moehring *et al.* 2004) or the dissection of the genetic basis of maize domestication (reviewed in Doebley 1992). However, QTL analysis also has limitations. In general, the mapping only allows for the identification of broad DNA regions associated with the trait variation. These regions usually contain a large number of genes and further testing is required to narrow down the genes. For this, genetic tools have been developed in *Drosophila*. A standard approach for increasing the mapping precision is deficiency/complementation mapping (reviewed in Mackay 2001). The identification of the precise nucleotide changes (called Quantitative Trait Nucleotides or QTNs) that are responsible for the trait variation is even more difficult. Only few studies reached that level of investigation (e.g., for *Drosophila* De Luca *et al.* 2003, and for tomatoes Frary *et al.* 2000).

6.4 Cold tolerance, an ecologically relevant adaptive trait in *D. melanogaster* cosmopolitan populations

The historical biogeography of *D. melanogaster* suggests an interesting case for the study of climatic adaptation. The species originated in sub-Saharan Africa and about 15,000 years ago, after the end of the last glaciations, it expanded from its native range to Eurasia and then spread across all continents on earth (David and Capy 1988; Lachaise *et al.* 1988; Lachaise and Silvain 2004; Li and Stephan 2006). This means that the species successfully colonized many different types of environments, thereby adapting to a broad range of new climatic conditions. This, among other reasons, makes *D. melanogaster* a perfect model for investigations of climate adaptation. An additional interest may be that *D. melanogaster* demography has been largely influenced by human activities (Keller 2007).

The occurrence of latitudinal clines on different continents for several phenotypic traits such as developmental time (James *et al.* 1995), temperature tolerance (James *et al.*1997), egg size (Azevedo *et al.* 1997) diapause (Mitrovski and Hoffmann 2001) and body size (David and Capy 1988; Huey *et al.* 2000; Arthur *et al.* 2008) strongly suggests that climate has a selective role on *Drosophila* species. Among the climatic factors, temperature has a particular influence on the insects' spatial distribution (Izquierdo 1991). Because insects are ectotherms, their activity is highly dependent on the ambient temperature. The ability to tolerate colder temperature might have been a particular selective constrain during the colonization of higher latitudes by *D. melanogaster*. Natural selection on temperature-related phenotypes should therefore have produced enough phenotypic variation between the African and the European strains to be analyzed, for example, by a QTL analysis.

Phenotypic variation for cold tolerance can be estimated in many different ways (Sinclair and Roberts 2005). Indeed, the protocols that are used in different studies vary greatly with regard to duration of the exposure, age of the tested flies and temperature range. In addition, the metrics used are also variable between different studies. Survival, Chill Coma Recovery (CCR) time, diapause ability, knockdown temperature and fecundity after a cold shock may all be relevant traits for such measurements.

6.5 The present work

The present work attempts to dissect the genetic basis of adaptation to new climatic conditions in *D. melanogaster*. As discussed above, this species offers a case of climate adaptation as well as unique opportunities for performing selection mapping and QTL mapping. As reviewed in Stinchcombe and Hoekstra (2007), in the near future, the combination of population genomics and quantitative genetics might be a powerful approach for the study of evolution. We used both methods in an effort to identify genes involved in the adaptation of *D. melanogaster* to the European climate. First, we selected a genomic region containing ecologically important genes and searched for footprints of selection at the DNA level. We used re-sequencing and advanced population genetics tests to confirm that this region evolved under positive selection. In the ancestral African population, we were able to identify a 2.7-kb region as the putative target of selection that contains the last exon of *HDAC6* harboring a ubiquitin-binding domain. HDAC6 is an unusual histone deacetylase with two catalytic domains and is localized in the cytoplasm. Its activities (ubiquitin binding and tubulin deacetylase) mark a distinct departure of HDAC6 from the known function of other HDACs. Recent discoveries have shown that HDAC6 is a key regulator of cytotoxic stress resistance (reviewed in Matthias *et al.* 2008). It appears to be both a sensor of stressful environmental stimuli and an effector, which mediates and coordinates appropriate cell responses. Nevertheless the precise ecological factors driving selection on *HDAC6* remained unknown.

This motivated investigations at the phenotype level. As cold tolerance might be an ecologically relevant for adaptation to the European climate, we decided to search for the genetic factors that allowed the European *D. melanogaster* population to adapt

to the colder temperatures found at higher latitudes. First, we estimated the naturally occurring variation in cold tolerance by applying the CCR test to *D. melanogaster* lines from two geographical regions with different contrasted climates. It revealed that cold tolerance is an ecologically relevant trait as it seems to be under selective constraints in the European population. We then measured the phenotype-to-genotype association by Composite Interval Mapping in an effort to localize the responsible genetic factors. Finally, comparisons of our results with those of other analyses - including association and selection mapping studies from Ometto *et al.* (2005) and Li and Stephan (2006) - allowed us to compile a list of candidate genes that may affect cold tolerance in *D. melanogaster*. Since this list is very short, it can be of great value for future studies that attempt to determine the genomic target and the form of selection operating on the trait.

7. Materials and methods

7.1 Selective sweep mapping

***Drosophila* lines and DNA sequencing:** DNA sequence data were collected from 12 highly inbred lines sampled in Africa (Lake Kariba, Zimbabwe). Furthermore, sequence data were obtained from 12 inbred European lines from The Netherlands. Both samples are described in detail in Glinka *et al.* (2003). All *Drosophila* strains were kept at 23°C in glass bottles of 250 ml containing 80 ml standard cornmeal and yeast medium under a 6-18 dark-light cycle with 45% humidity.

DNA primers were designed based on the *D. melanogaster* genome sequence (http://flybase.org/) and obtained from Metabion (Martinsried, Germany). Genomic DNA from each line was extracted from pools of 20 females using the Puregene DNA isolation kit (Gentra System, Minneapolis). Short DNA fragments of about 300 to 700bp long were amplified by standard PCR using the Taq DNA polymerase recombinant kit (Invitrogen, Carlsbad, USA). PCR products were purified using the Exosap-It kit (USB, Cleveland) and sequence reactions were conducted with ABI PRISM Big Dye Terminator v1.1. Sequence data were then obtained by an ABI 3730 DNA analyzer (Applied Biosystems +Hitachi, Foster City, USA).

Sequence edition and alignments were performed with the DNAstar software package, including Editseq, Seqman and Megalign (DNASTAR, Madison, USA). Alignments were performed using the ClustalV option of Megalign. However, in cases of ambiguous alignments, we manually chose the most parsimonious scenario. Insertion and deletion polymorphisms were excluded from further analysis. Absolute positions of the DNA sequence follow the Flybase release 5.10.

Mapping strategy: To identify and map the target of selection, we proceeded as follows. First, we selected a subgenomic region of about 70 kb on the X chromosome that contained several ecologically interesting genes, including a gene encoding a putative antifreeze protein (*CG6227*). This region partially overlaps with the window 47 in Li and Stephan (2006). Re-sequencing an additional (limited) number of short fragments of 500-600 bp in the 70-kb subgenomic region, we found very low levels of variation across most of the region in the European sample (data not shown), while the valley of reduced variation in the African sample appeared much narrower; i.e., the situation was similar as in the case of the *roughest* and *wapl* regions (Pool *et al.* 2006; Beisswanger *et al.* 2006). To be able to localize the target of selection as precisely as possible, we therefore decided to follow the same strategy as in the *wapl* analysis (Beisswanger and Stephan 2008) and concentrated on the African sample (see "Standard analyses of a candidate region of selection" in the Results section). In a second step, we narrowed this 70-kb region down to 22 kb, re-sequenced this segment completely, and applied the specific tests for selective sweeps to this region (see Results).

Outlier analysis: We used DnaSP 4.50.3 (Rozas *et al.* 2003) to calculate the basic summary statistics π, θ_W, Tajima's *D* (Tajima 1989), divergence, Fu and Li's *D* (Fu and Li 1993) and Fay and Wu's *H* (Fay and Wu 2000). Divergence was calculated between the sample from the African population of *D. melanogaster* and the available online release of the *D. simulans* sequence (Flybase consortium; http://www.flybase.org). The ancestral states were defined using either *D. simulans* or (when not available) its close relative *D. sechellia*.

We compared the mean value of each summary statistic of the 70-kb candidate region to its average value obtained for the whole X chromosome (Ometto *et al.* 2005). For each summary statistic, we used the Mann-Whitney test to infer whether the region represents an outlier compared to the rest of the X chromosome.

Ascertainment bias correction: Thornton and Jensen (2007) describe an approach that generates a uniform distribution of p-values when some of the assumptions of the neutrality tests are violated. They study cases when past demographic events have shaped the polymorphism patterns of a subgenomic region, which is a biased sample based on a priori information (for example, from a genome scan). The *HDAC6* subgenomic region was selected based on the genes in this region that may contribute to the ecological adaptation of *D. melanogaster*. Even if such a sampling is not random, it is unclear whether it generates any bias on selective sweep scanning and how to sample conditional on this biological information.

Performing a genome scan analysis, Li and Stephan (2006) discovered a 100-kb fragment that overlaps with the *HDAC6* region and showed evidence of recent positive selection in the European population of *D. melanogaster*. Among the fragments Li and Stephan (2006) analyzed was a 560-bp fragment located within the *HDAC6* subgenomic region that contained no polymorphic sites. This information was not considered important for the initial choice of the 70-kb region. However, we decided to include it into the analysis as a priori information making this analysis more conservative. Thus, we simulate a sample of 24 lines (12 European and 12 African ones) according to the demographic scenario inferred by Li and Stephan (2006).

Conditioning on the existence of a monomorphic 560-bp fragment within the European sample, we create the null distribution of the neutrality test statistics used in this paper.

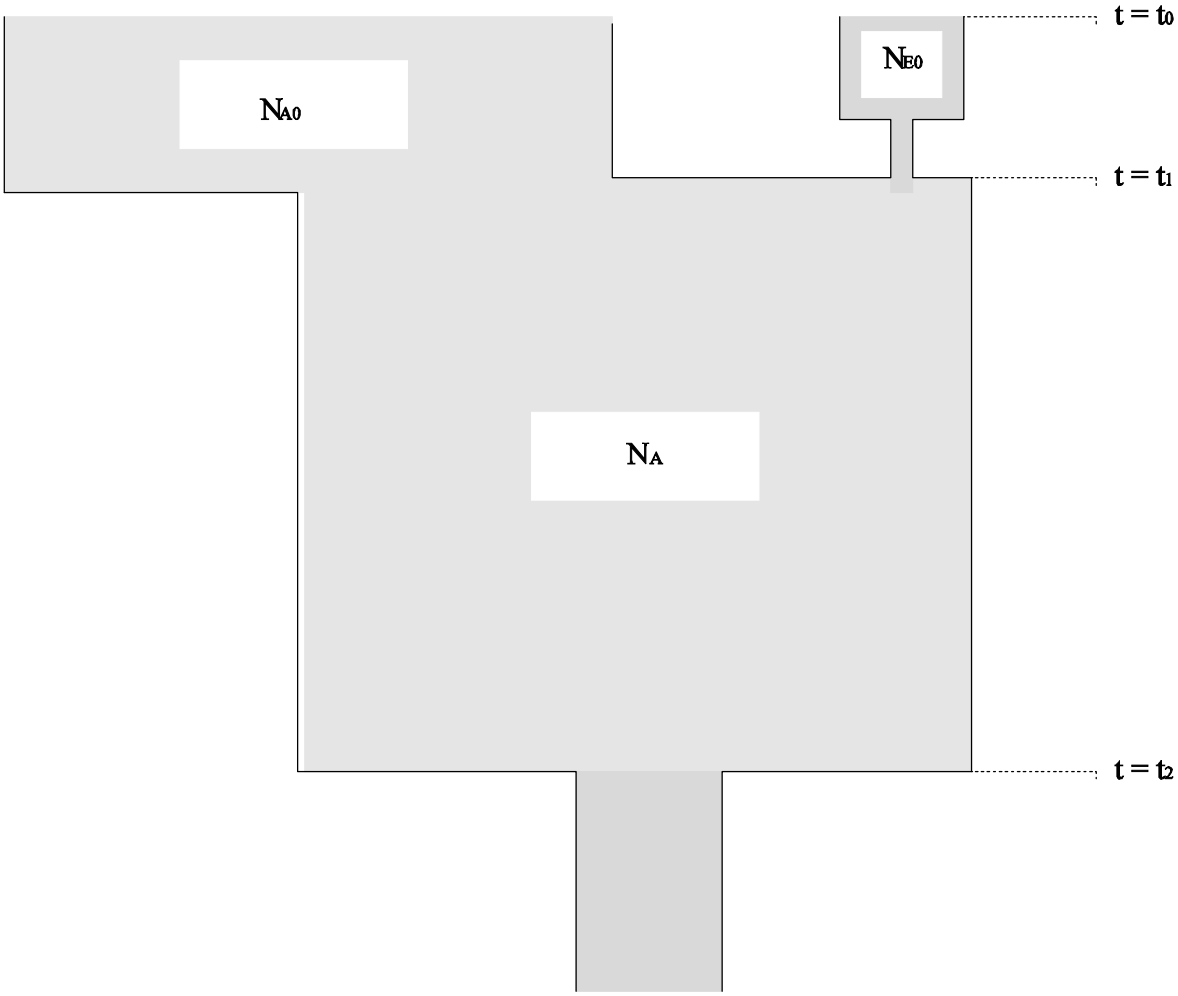


Figure 2: The demographic model of the European and African population of *D. melanogaster* as it was inferred by Li and Stephan (2006) and used in this study. The present European effective population size is approximately $N_{E0} = 10^6$ whereas the African population (N_{A0}) is 8 times larger. Backward in time the model can be described by a severe bottleneck in the European population that took place $t_1 = 15460$ years ago and lasted for ~340 years. During the bottleneck the effective population size of the European population was decreased to 2200. Approximately at $t_2 = 15800$ years ago the European population merges with the African population forming the ancestral population ($N_A = N_{A0}$). Finally, the ancestral population decreases to a fifth of the present day African population at $t_3 = 60000$ years ago.

***CLR* test:** The *CLR* test (Kim and Stephan 2002) was used to infer selection. It computes the composite-likelihood ratio (Λ_{CLR}) between a standard neutral model and a selective sweep model. The null distribution of the statistic is derived using the approach described in the "Ascertainment bias correction" section (see also figure 2). This modification follows a suggestion of Thornton and Jensen (2007) who showed that the false positive rate can be controlled if the correct demographic null model is used. For the generation of the simulated datasets we used the estimated value of the parameter θ_W (0.0499) under the demographic scenario of figure 2. Furthermore, the B test of the Kim and Stephan (2002) method was performed because it is more conservative. The *CLR* test was also used to estimate the target site of selection. However, its confidence interval could not be determined (in contrast to Beisswanger and Stephan 2008), as population recombination rate was too high to run simulations of the sweep model in reasonable times.

***SweepFinder* test:** To infer selection, we also used the *SweepFinder* test. It takes into account the SFS of the whole chromosome (background SFS) in order to calculate the likelihood of the neutral model. Non-polymorphic sites were excluded from the analysis, as Nielsen *et al.* (2005) suggest. *SweepFinder* uses the same principles as the *CLR* test: by comparing two hypotheses, a model of neutral evolution and a model of a selective sweep that just completed, it calculates the maximum likelihood estimates of the position of the beneficial allele as well as the strength of selection. Additionally, it reports the likelihood ratio Λ_{SF} between the null and the alternative model. Similarly to the *CLR* test, a null distribution is required to decide about the statistical significance of the selective sweep hypothesis. The main advantage

of the *SweepFinder* is that a specific population genetic model is not considered in the null hypothesis, but the SFS is derived from the whole-chromosomal pattern of variation; i.e., from the data itself.

We have extended the original approach for calculating the significance threshold for the *SweepFinder*. According to Nielsen *et al.* (2005) the 95th percentile of the statistic Λ_{SF} denotes the threshold value. Our approach, however, splits the region of interest into k fragments and for each one the $100 - \frac{5}{k}$ percentile is used as the cut-off value, resulting in a variable region-specific threshold. This approach helps to remedy the tendency of the *SweepFinder* to produce higher Λ_{SF} values at the borders of the region under study (P. P., unpublished results). Here we chose k=10. The demographic model of figure 2 (Li and Stephan 2006) with the ascertainment bias described in the "Ascertainment bias correction" section is used to create the null distribution of the test statistics for all performed neutrality tests.

Estimation of the time since fixation of the beneficial allele: The time since the fixation of the beneficial allele was estimated by the methods described in Przeworski (2003) and Slatkin and Hudson (1991). For the Przeworski test, mutation rate μ = 1.45 .10^{-9}/ bp / gen (Li and Stephan 2006) and recombination rate r = 4.718 .10^{-8} / bp / gen (Comeron *et al.* 1999) were used. The local parameters were estimated from a 925-bp long region located between the 7th and 9th exon of *HDAC6* [as exon 8 is very short (88bp), it has presumably no special effect on the parameter estimates, and was thus kept in the analysis]. This region contains 10 segregating sites forming 8 haplotypes, and Tajima's D = -1.74221. Two positions of the beneficial mutation were tested: one in the last exon of *HDAC6* and one in the last exon of *CG9123*.

We also used the Slatkin-Hudson method (Slatkin and Hudson 1991) assuming a star-like genealogy since the fixation of the beneficial allele. We based this estimation on the DNA region between positions 9.865 and 12.443 kb. In this region 19 segregating sites were detected and divergence to *D. simulans* is 0.056. To convert the obtained estimates into years, we assumed 10 generations per year for both methods.

7.2 QTL mapping experiments

***Drosophila* Lines:** All *Drosophila* lines were kept at 23°C in glass bottles of 250 mL containing 80 mL standard cornmeal and yeast medium under a 6-18 dark-light cycle with 45% humidity. We used four isofemale lines (A95, A157, A186, A337) that were sampled in Africa (Lake Kariba, Zimbabwe) and four isofemale lines (E11, E13, E14, E20) collected in Europe (Leiden, The Netherlands). All these wild lines are highly inbred. They are described by Glinka *et al.* (2003). Line 5905 is a lab strain of American origin carrying an X-linked w^{1118} mutation. The 6418 strain is derived from 5905 such that the X chromosome is replaced by an FM7j balancer chromosome. Both lines, 5905 and 6418 were obtained from the Bloomington Stock Center. 5905 was used to control for the effect of the genetic background on cold tolerance, and 6418 was required for substitution lines construction.

Substitution lines were constructed to assess the presence of cold tolerance factors on the X chromosome of *D. melanogaster*. To this end, a wild X chromosome was introgressed into a 6418 genetic background as follows. We first isolated a single female offspring from a A157×6418 cross and then backcrossed it for 7 generations to 6418 males (see crossing scheme in Appendix Figures B1 and B2). The resulting A* line carries the X chromosome of A157 and the mitochondria and the 2^{nd} and 3^{rd} chromosomes from 6418. Following the same crossing scheme, the European substitution line E* was constructed using the E14 line instead of A157. We subsequently checked whether A* and E* carry the same autosomal background by sequencing several short DNA fragments on the autosomes (data not shown). The result was that E* and A* were completely identical except for their X chromosomes.

For QTL mapping, a set of 186 RILs (Recombinant Inbred Lines) was constructed using the two substitution lines (E* and A*) as parental lines as follows. Offspring of the two reciprocal crosses between E* and A* were placed into a population box containing standard cornmeal and yeast medium. Flies were allowed to mate freely for 5 generations such that a sufficient number of recombination events could occur. During this phase, generations were strictly separated by removing adults from the population cage just before their offspring emerged. F5 recombinant X chromosomes were then isolated in specific lines by crossing and backcrossing independently single F5 offspring males from the population box to 6418.

Phenotyping: As a measurement of cold tolerance we used a modified protocol from the Chill Coma Recovery experiment of David *et al.* (1998). The chill coma recovery test was used because it presents some advantages (see also Macdonald *et al.* 2004). First, the CCR test was shown to be tightly linked to cold tolerance ability as this trait also follows latitudinal clines (David *et al.* 1998; Gibert *et al.* 2001; Bubliy *et al.* 2002). Second, the protocol is simple and can be performed on a large number of individuals at the same time. Finally, the test is non-lethal, and the time metric corresponds to a quantitative variable.

To remove any influence of breeding conditions (David *et al.* 1998), all tested lines were maintained in low density. This limited competition between individuals and allows them to reach their maximal body size. In order to avoid any possible effect of the photoperiod and of the circadian rhythm (Sørensen and Loeschcke 2002), the flies that were used in the CCR experiments were maintained under the same conditions as during breeding (see above). Under slight CO_2 anesthesia, young (less than one day old) males were selected and placed into 50-mL vials containing 10 mL of standard

cornmeal food. At 4 to 5 days old, they were separated without anesthesia and placed into individual empty 8-mL vials. As there is a controversy about the effect of CO_2 anesthesia on temperature experiments (Milton and Partridge 2008), such treatment has the advantage not to expose the tested individuals to CO_2 immediately before the cold shock. The vials were then placed into an ice-water bath (temperature was controlled to be between 0 and 1°C). After 7 hours of cold exposure, vials were removed from the ice-water bath and placed back at room temperature (23 ±1°C). The CCR time of each individual was recorded as the time required standing up again (David *et al.* 1998; Macdonald *et al.* 2004; Morgan and Mackay 2006). The experiment was stopped after 75 min of recording. Males that were still alive but did not recover were assigned a CCR time of 75 min. Dead individuals were not included into the analysis. They represented less than 1% of the total flies tested, and were randomly distributed among the lines tested. The results were analyzed with an ANOVA and its Tahmane post-hoc test.

Genotyping: More than 250 DNA fragments distributed along the X chromosome were sequenced in lines E14 and A157 by Ometto *et al.* (2005) revealing hundreds of fixed nucleotide differences between these two lines. These DNA polymorphisms are perfect genotypic markers for a QTL experiment. A first set of 19 of them, uniformly spread along the X chromosome, was selected as markers. These makers were named according to the fragment in which they were detected in Ometto *et al.* (2005). After a preliminary QTL analysis, 4 additional markers were designed in the regions of interest in order to increase the density of the marker map and the mapping precision (see Appendix Table B3 for maker list and positions)

We genotyped each RIL by extracting DNA from pools of 20 females with a Puregene DNA isolation kit (Gentra System, Minneapolis). Then 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 allele at each marker location was then determined with the 7500 Fast Software provided by ABI.

QTL Analysis: CCR time was recorded for 7 to 13 individuals of each RIL. Broad-sense heritability was computed as $H^2 = \sigma_L^2 / (\sigma_L^2 + \sigma_E^2)$, where σ_L^2 is the among-RILs and σ_E^2 the within-RIL variance component. For the marker map construction, we converted the physical distances (d in bp) between markers found on the Flybase web site (*D. melanogaster* genome release 5.4) to relative distances (C in centimorgans, cM) using the following formula $C = d\ r$. This calculation is an approximation but as the distances are small the use of the Haldane map function only slightly modified the C values. The recombination rates (r) were obtained from the Recombrate software (Comeron *et al.* 1999).

Mean CCR time and genotype data were compiled and the position of the QTL was determined using the Composite Interval Mapping (CIM) method (Zeng 1994) implemented in the Windows QTL Cartographer version 2.5. CIM tests the hypothesis that the interval between two adjacent markers contains a QTL which influences the trait, while simultaneously controlling for the effects of other QTL located outside the test interval. A conditioning window size was defined such that only markers 10 cM away from the borders of the test interval were included in the analysis. We used Haldane's map function to correct for the occurrence of possible double recombinations. The likelihood-ratio (LR) test statistic is $-2ln(L_0/L_1)$ where L_0/L_1 is the

ratio of the likelihood under the null hypothesis L_0 (there is no QTL in the interval) to the alternative hypothesis L_1 (there is a QTL in the test interval). The LR statistic was evaluated every 0.5 cM. The significance level for each analysis to infer the presence of a QTL was determined by permutations. For this, empirical distributions of the LR test statistics under the null hypothesis of no association between test intervals and mean trait values were obtained by randomly permuting the trait data 1000 times and calculating the maximum LR statistic across all intervals for each permutation. The significance threshold was calculated with a 5% risk.

8. Results

8.1 Selective sweep mapping

Standard analyses of a candidate region of selection: The region analyzed here is about 70 kb long. It is located in a highly recombining portion of the X chromosome (r = 4.718 $.10^{-8}$ / bp / gen.) and is relatively gene dense. This region contains 12 genes, five of which have unknown molecular functions (*CG15032*, *CG9114*, *CG9123, CG12608*, and *CG9164*). The other genes have been functionally characterized (*gce*, *Top1, dah*, *HDAC6*, *CG6227*, *acj6*, and *Pp1*). In order to perform a fine-scale analysis of the African sample, we sequenced 15 non-coding (intronic or intergenic) DNA fragments of 511 bp on average, in addition to the four already sequenced by Ometto *et al.* (2005) (figure 3). For each of these 19 fragments, basic summary statistics were calculated, averaged over the whole candidate region, and then compared to the chromosomal average. Only 15 of the 19 fragments could be aligned with *D. simulans*.

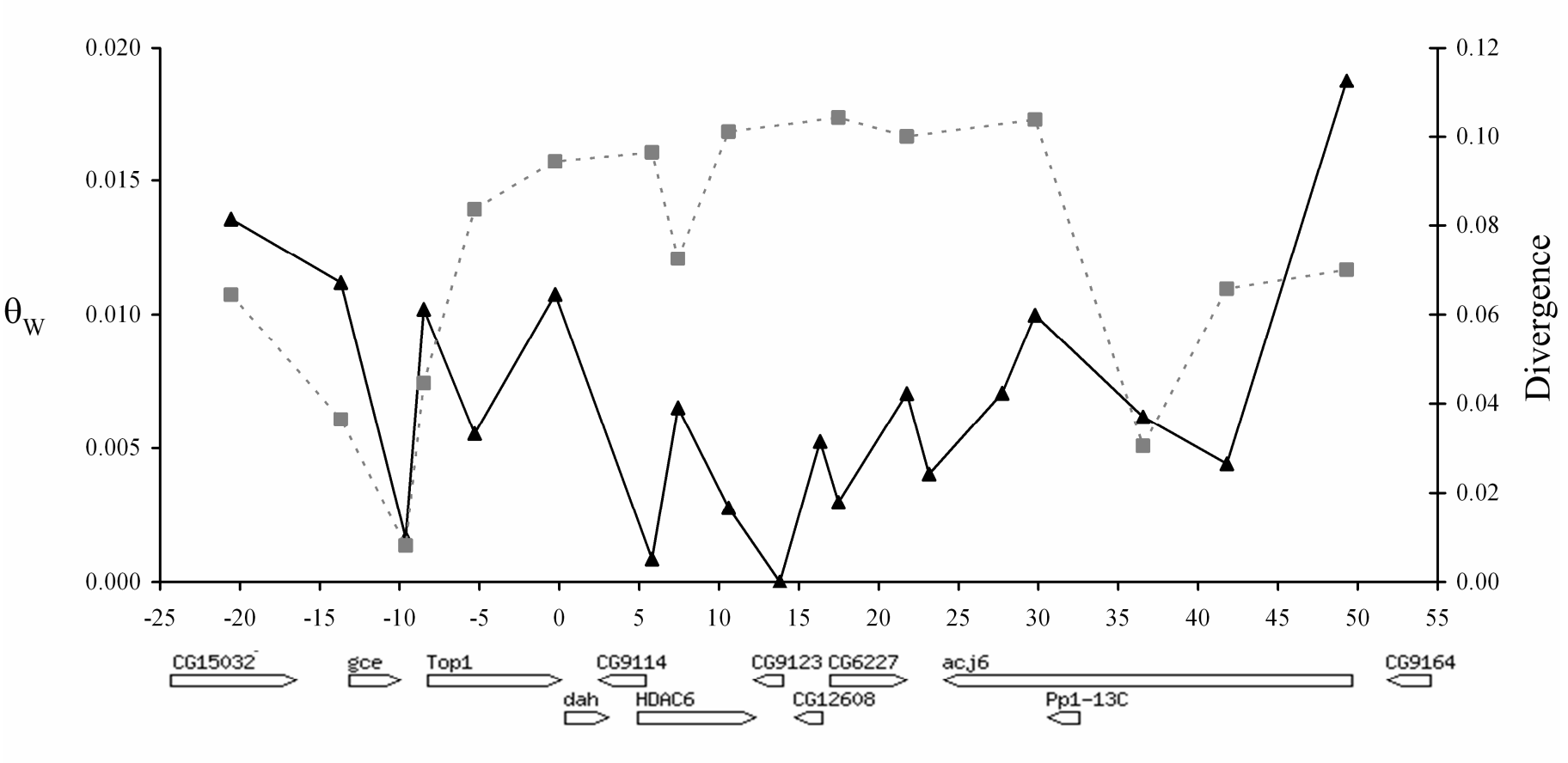


Figure 3: Nucleotide diversity θ_W (solid line) and divergence to *D. simulans* (dashed line) across the candidate region for selection. The relative positions in kb are on the X-axis. Gene spans (according to Flybase) are at the bottom of the graph.

Table 1: Mann-Whitney comparisons of different summary statistics between the candidate region and the chromosomal mean

Summary statistics	Chromosomal mean	Candidate region mean	p value
Recombination rate	3.515	4.714	< 0.0001
Sample size	11	11	0.81
Fragment length	501	511	0.63
Segregating sites	19	11	< 0.0001
θ_W	0.013	0.007	< 0.0001
π	0.012	0.005	< 0.0001
Tajima's *D*	- 0.667	-1.143	0.0001
Number of haplotypes	9.655	7.348	0.0001
Haplotype diversity	0.93	0.785	0.0001
ZnS	0.14	0.12	0.44
Divergence	0.064	0.068	0.36
Fay and Wu's *H*	- 0.26	-0.32	0.56

The region exhibits a strong reduction in nucleotide polymorphism. On average the 259 fragments sequenced by Ometto *et al.* (2005) for the African population contained twice as many segregating sites as the 70-kb candidate region ($p < 0.0001$; see table 1). θ_W and π were significantly lower than the chromosomal average ($p < 0.0001$ for both). As can be seen in figure 3, the θ_W curve is roughly U shaped (with a minimum between 10 and 15 kb), except for two positions at -10 and around 40 kb where divergence is very low. In general, divergence is rather high in the region of reduced variation between positions 0-22 kb (~0.09).

Furthermore, the region shows deviations from the chromosomal expectation with regard to the SFS. Indeed, Tajima's *D* values are more negative than the X chromosome average (- 1.143 vs. - 0.667), which is highly significant ($p = 0.001$). Four fragments show significantly negative Tajima's *D* values (data not shown). In contrast, Fay and Wu's *H* statistic does not depart from the chromosomal average. This illustrates that the SFS is lacking intermediate frequency variants and shows an excess of low frequency SNPs.

The number of haplotypes ranges from 1 to 12 in the candidate region, but its mean is significantly lower than the chromosomal average ($p < 0.001$). Similarly, haplotype diversity is significantly lower ($p < 0.001$). LD as measured by the *ZnS* statistic is relatively constant over the whole region (< 0.3) and does not deviate from the chromosomal average.

The genes *CG9123* and *CG12608* are paralogs. Among the 12 *Drosophila* genomes examined (*Drosophila* 12 Genomes Consortium 2007) this duplication is present only in *D. melanogaster*. Both copies are highly diverged from *D. simulans*. Investigating the pattern of polymorphism at both genes, we did not find evidence for extensive gene conversion; for instance, there is only one SNP shared between both copies (out of 48 SNPs in total). *CG9123* contains many non-synonymous SNPs in relatively high frequency, most of which produce drastic amino acid changes (see Appendix table A1 part 8-11). In addition, we observed some deletions in the coding region, one of which causes a frame shift change. This may suggest that *CG9123* is under weak functional constraints or even a pseudogene.

Application of the *CLR* and *SweepFinder* tests: In order to perform more advanced neutrality tests, we defined a region of about 22 kb (corresponding to the segment between absolute positions 15,222,319 and 15,244,496 in Flybase release 5.10, and to positions 0 to 22 kb in figure 3). This region was then completely sequenced and subjected to the *CLR* and *SweepFinder* tests. The *CLR* test was marginally significant ($p = 0.048$) when the null distribution of the statistic Λ_{CLR} was constructed from the demographic scenario of the African population inferred by Li and Stephan (2006) (figure 2). Figure 4A shows Λ_{CLR} along the region. The beneficial mutation is estimated to have occurred at position 11.378 kb relative to the beginning of the 22-kb region, and $\alpha = 2Ns$ is approximately 13076 (where N is the effective population size and s the selection coefficient). This value is much higher than most other reported estimates, which is consistent with the observed width of the valley of reduced variation and the fact that population recombination rate $4Nr$ is very high in this part of the genome.

The *SweepFinder* test was also significant ($p = 0.034$) for the 22-kb completely sequenced region. In figure 4B we show the Λ_{SF} values along the region. Consistent with the result of the *CLR* test, three positions (11.315, 12.474 and 13.110 kb) show the highest Λ_{SF} values. The high value around position 1.0 kb is probably not a target of selection as it is not confirmed by Tajima's *D* and the *CLR* test.

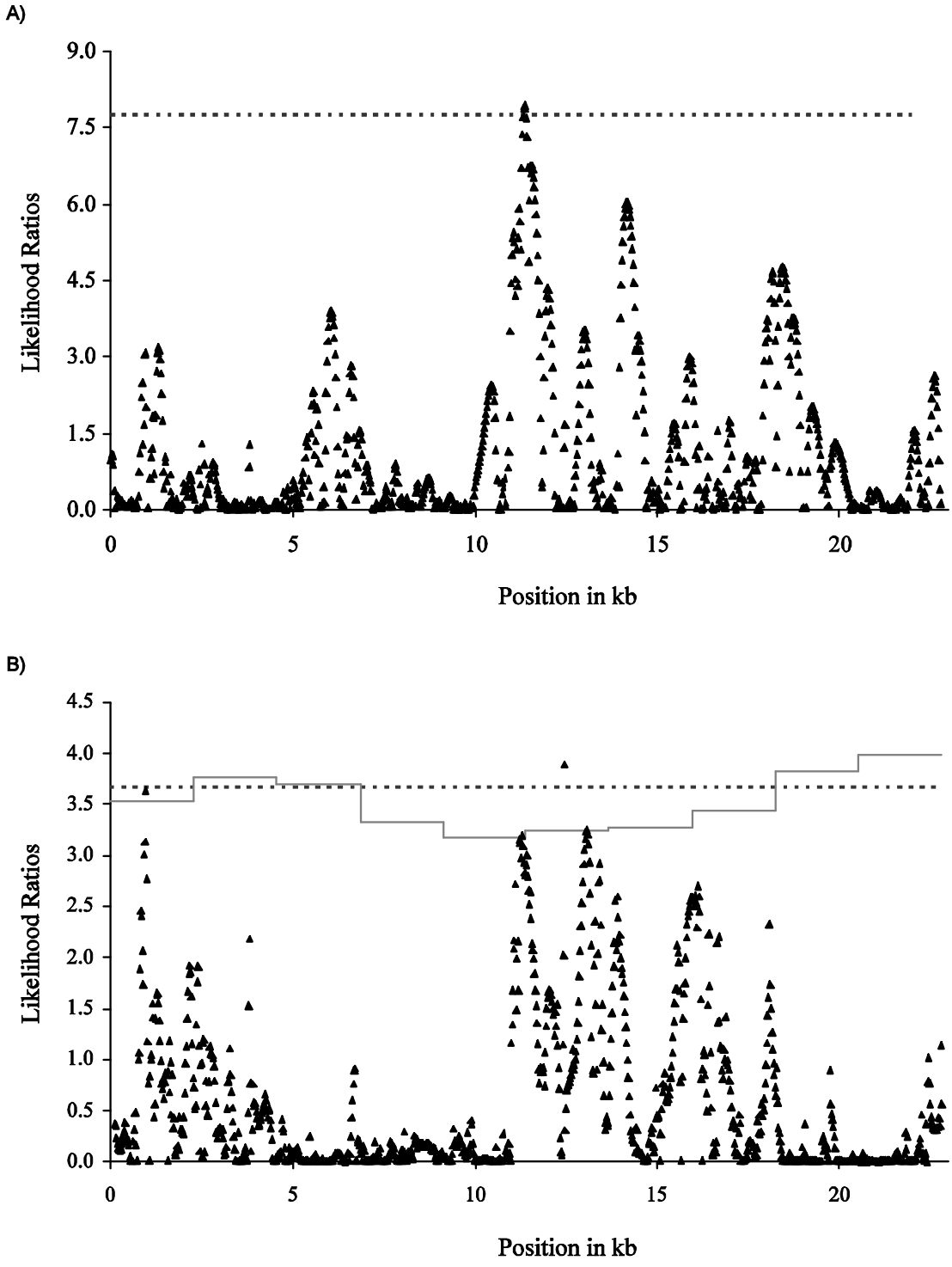


Figure 4: The likelihood-ratio values calculated by A) the *CLR* and B) the *SweepFinder* tests for a 22-kb subregion of the 70-kb region (for 1000 bins). Each triangle denotes the value of the test statistics for a selective sweep model for which the beneficial mutation occurred at that specific position. In B) the dashed line depicts the constant threshold calculated according to Nielsen *et al.* (2005), whereas the solid line shows the variable threshold (see Materials and Methods).

Sliding window analysis: To corroborate our mapping results, we also performed a sliding window analysis on the SFS of the sequenced 22-kb region (figure 5). Indeed, significantly negative Tajima's *D* and Fu and Li's *D* values were found near the estimated targets of selection, consistent with the *CLR* and *SweepFinder* results. The analysis revealed a small segment showing a local reduction of nucleotide diversity and a SFS shifted toward low-frequency variants despite normal levels of divergence. This region coincided with the 2.7-kb fragment mentioned above. Four exons lie in this region: the three last exons of *HDAC6* and a portion of the last exon of *CG9123*. The relatively low value of Tajima's *D* around position 19 kb is probably due to purifying selection (causing the observed low divergence in the helicase functional domain of *CG6227*; data not shown). In order to identify candidate substitutions under selection, we aligned the 2.7-kb region of *D. melanogaster* to that of *D. sechellia*, *D. simulans*, *D. erecta*, and *D. yakuba*. As the 2.7-kb region centers on *HDAC6*, we focused our investigations on this gene. The *HDAC6* introns were poorly conserved between species but we obtained a good alignment of the 3' UTR and of the three last exons of the gene. In the 3' UTR, we found 6 nucleotide substitutions specific to the *D. melanogaster* lineage. In exon 7, we identified three non-synonymous substitutions specific to *D. melanogaster*. All of them cause non-polar to non-polar amino-acid replacements. We also found a deletion of 9 nucleotides that is specific to *D. melanogaster* at the end of exon 9 (see Appendix table A2). This exon also carries two non-synonymous substitutions. One of them generates a drastic amino-acid change: a valine to glutamic acid substitution (see Appendix table A2). In addition, this substitution is in a region predicted by the program MyHits (http://myhits.isb-sib.ch) to be the ubiquitin-binding site of HDAC6.

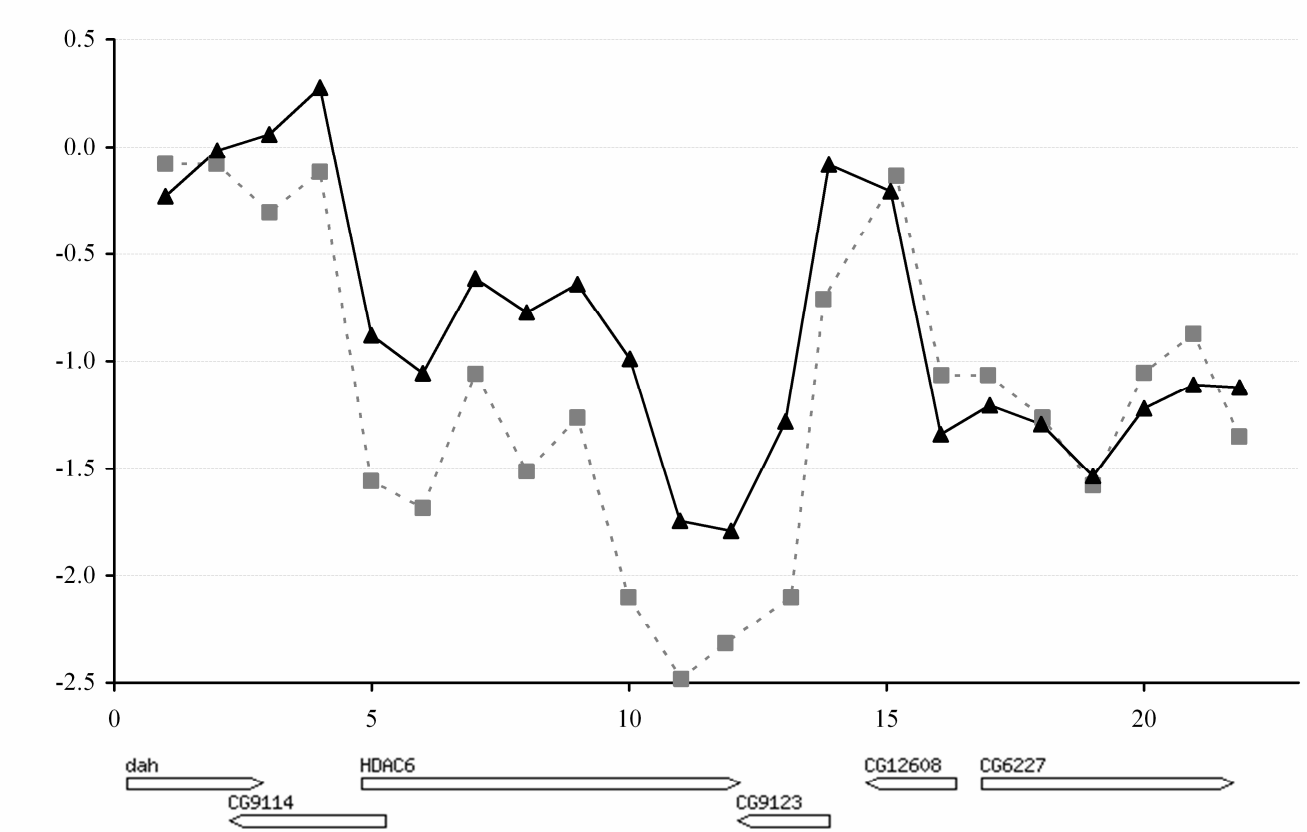


Figure 5: Sliding window analysis of the fully sequenced 22-kb region. Tajima's *D* and Fu and Li's *D* are represented by solid lines (black triangles) and dashed lines (grey squares), respectively. Each data point represents the midpoint of a 2000-bp long window and the step size is 1000 bp. In order to describe the neutral frequency spectrum we excluded the non-synonymous sites from this analysis

Age of the selective sweep: The age of the sweep in the 22-kb region was estimated by the Przeworski and Slatkin-Hudson methods (cf. Material and Methods). We used Przeworski's approach with two positions as input parameter values that are near the estimated selected sites: position 11.787 kb gave a time since fixation of the beneficial allele of 63,334 years (95% C.I.: 23,382-628,432 years), while position 12.787 kb gave 56,770 years (95% C.I.: 21,121-577,307 years). Using the Slatkin-Hudson method the age of the sweep was estimated as 50,047 years.

These estimates suggest that the sweep occurred before the European lineage split off from the African one (about 16,000 years ago; Li and Stephan 2006). In order to confirm this hypothesis, we re-sequenced the region between position 8.0 and 15.0 kb in 12 lines of a European sample from The Netherlands (Materials and Methods). We found that the European lines were identical with those of the African sample in a limited segment of approximately 2.7 kb from position 9.8 to 12.5 kb (except for three derived singletons and one doubleton; see Appendix table A1). This suggests, in conjunction with the estimated age of the sweep, that the selected allele has been exported to Europe during the colonization process.

8.2 QTL mapping

Wild Lines: As a first step in our analysis of the adaptation of *D. melanogaster* to the European climate we compared the cold tolerance of the four European to the four African wild lines. Cold tolerance is evaluated by measuring the CCR times of males kept at 0°C during 7 hours. Figure 6 shows the cumulative percentage of recovered individuals over time. Generally the behavior of the lines within a population is rather homogenous. The European lines clearly show significantly higher cold tolerance than the African ones. The European lines start much earlier to recover from chill coma than the African ones and the mean onset time of the European population is 25 min *vs* 32 min for the African one (Mann-Whitney test: $p = 0.019$). The mean CCR time of the European lines is also significantly shorter than that of the African ones (37 min *vs* 58 min; Mann-Whitney test: $p = 0.021$). In addition, the two populations differ by their phenotypic variability. Indeed the African lines have a significantly higher variance in CCR times than the European ones. This is not due to the effect of the mean as the Coefficients of Variation (CV = Standard Error / Mean) follow the same tendancy (Mann-Whitney U test, $p = 0.043$).

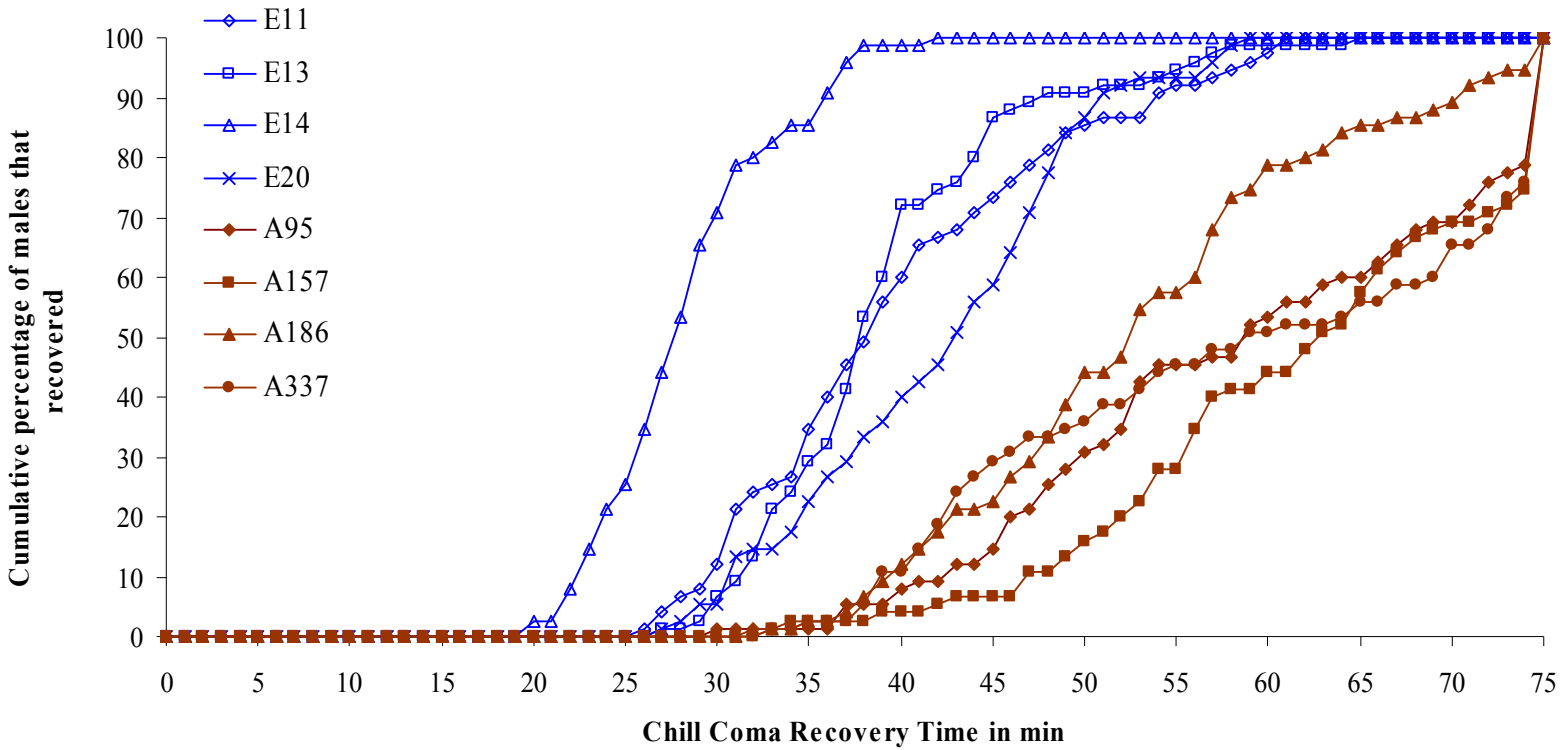


Figure 6: Cumulative percentage of males that recovered over time in a sample of wild inbred lines. The European lines are represented by blue open symbols (E11, E13, E14, E20), and the African ones by brown solid symbols (A95, A157, A186, A337). Sample size N=75 for each line.

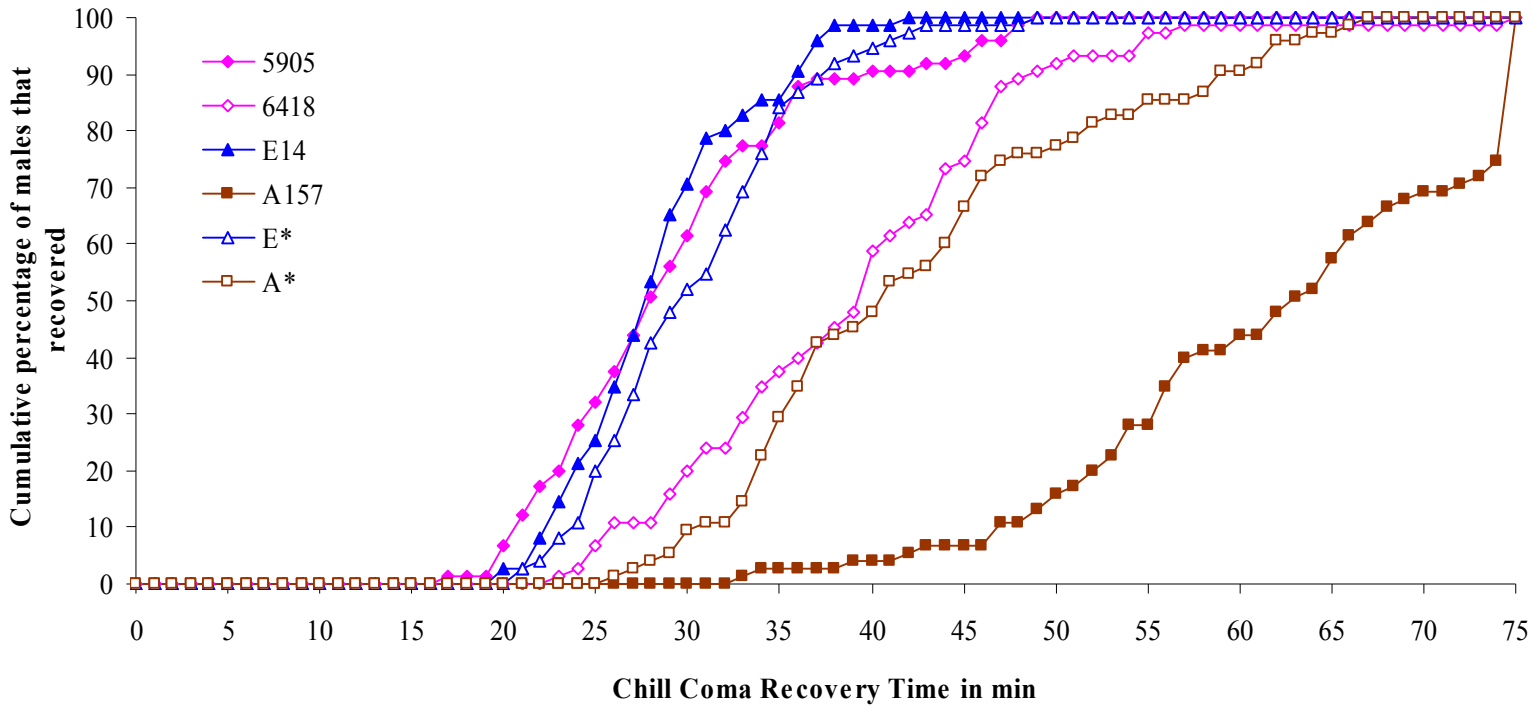


Figure 7: Cumulative percentage of males that recovered over time. The lines with open symbols share the same genetic background as 5905. The lines shown in blue share the same X chromosome as E14. The lines shown in brown share the same X chromosome as A157. Sample size N=75 for each line.

Substitution lines: As the substitution lines A* and E* differ only by their X chromosome, the comparison of their CCR times should reveal the effect of the X chromosome on cold tolerance. E* and A* show significantly different mean CCR times (respectively 30 vs 42 min) suggesting the presence of cold tolerance factors on the X chromosome (Figure 7). Tahmane ANOVA post hoc tests at a risk of 0.0005 clustered the lines into three groups: the first one contains E14, 5905 and E*, the second 6418 and A*, and the third one A157 (Figure 7). The within-line variances in CCR time of E* and A* are not significantly different.

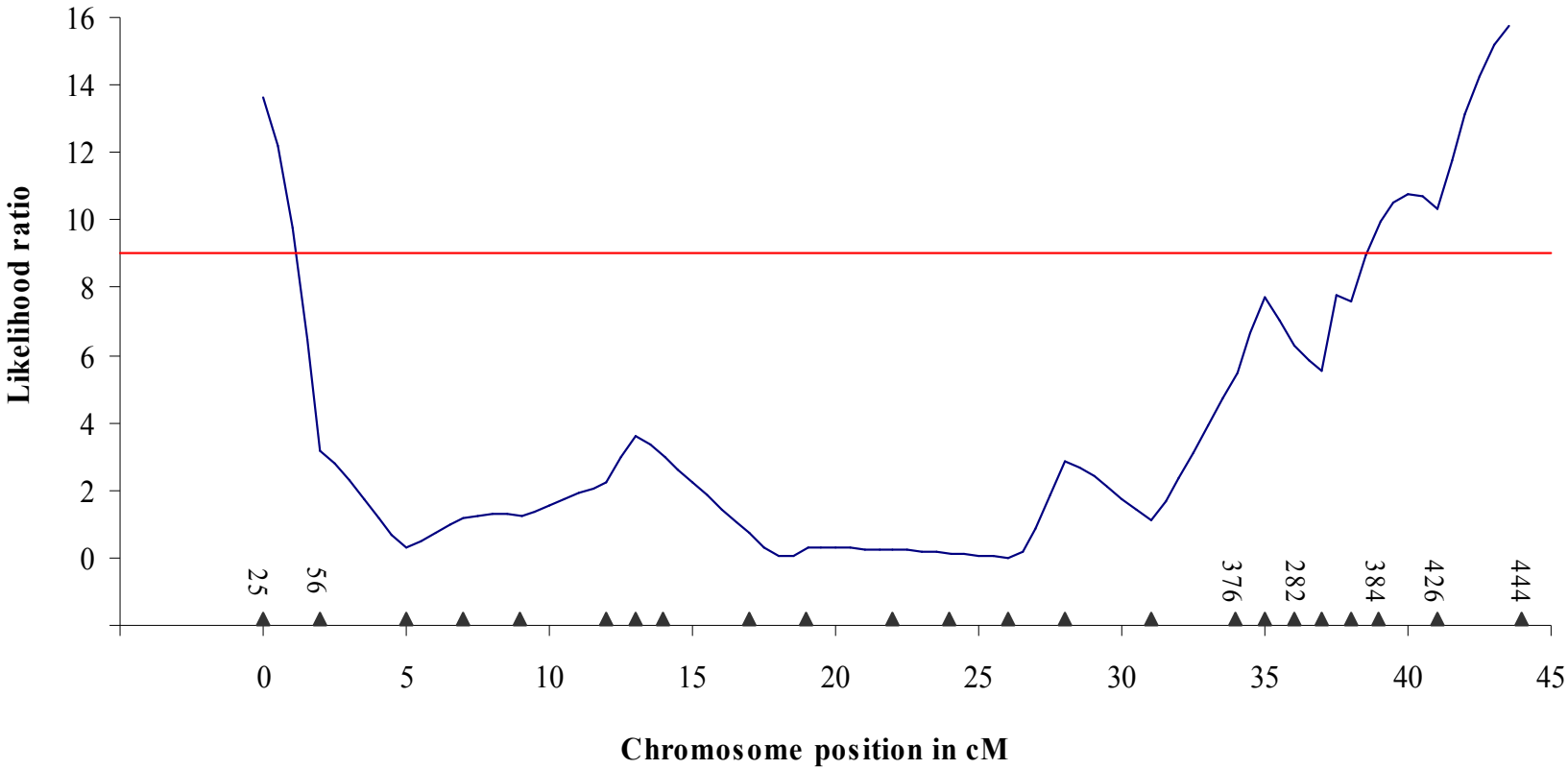


Figure 8: X chromosome scan for cold tolerance factors. The likelihood ratio scores from the composite interval mapping analysis are plotted against map position in cM. The regions where the LR curve is above the threshold (red line) are significantly associated with the trait. The triangles on the X-axis are referring to marker positions.

Analysis of the RILs: The distribution of the mean CCR times in RILs population ranges from 22.9 to 42.8 min (see appendix figure B4). This distribution is very regular and do not exhibit a stepwise pattern. Interestingly, the A* line is among the highest CCR time recorded whereas E* is not among the lowest. The estimated heritability (in the broad sense) of CCR times in the RILs is 0.2. Based on the CIM method, two regions of the X chromosome show significant association with CCR time. The first QTL (which will be referred to as QTL 1) is located between markers 25 and 56 (Figure 8). It is close to the telomere (spanning from cytological position 1A to 3A, which is about 2.5 Mb). The second one (which will be referred to as QTL 2) is located close to the centromere, between markers 384 and 444 (Figure 8). It spans from cytological positions 14B to 20F, which corresponds to about 4 Mb. Each of the two QTL explains 7 to 9% of the total phenotypic variance. The shape of the LR curve suggests the presence of two loci influencing CCR time in the latter region, where the first one would be located between cytological bands 14B and 16B (between markers 384 and 426) and the second one between 16B and 20F (between markers 426 and 444). Furthermore, there is a third marginally significant peak between cytological position 12F and 13C (between markers 376 and 282).

It is interesting to note that the QTL have been detected at the telomere and the centromere; i.e., at the ends of the X chromosome. This may be the result of a bias due to some (unknown) boundary effects. However, in both regions the power to detect QTL is higher than in the middle of the X because the frequencies of the African - and European - derived marker variants in the RILs population are closer to 0.5 at the tips of the chromosome.

Candidate genes: To obtain a first insight into possible candidate genes affecting chill coma recovery we scanned the regions we identified in our QTL analysis

for genes with relevant Gene Ontology (GO) features using Flybase (*D. melanogaster* genome release 5.4). Only 14 X-linked genes fall into the "response to temperature stimulus" GO class. Only two of these genes are located under the identified QTL: *Adar*, an adenosine deaminase, and *per*, a transcription factor; both co-localize with QTL 1. In QTL 2 there is nothing known. This means that the present work might have identified new genetic factors involved in cold tolerance that have not been described before.

In order to find additional candidate genes, we compared our results with those of an association mapping study by Ayroles *et al.* (2009). She studied CCR in a population from Raleigh (North Carolina). North American populations mostly derive from the European one and are thus genetically close. We thus assumed that their results are comparable to our. Ayroles *et al.* (2009) detected 15 X-linked single feature polymorphisms (SFP; i.e. changes in the coding region) that are associated with CCR. Among these genes 8 are located in the QTL detected in the present study (see Table 2), such that four co-localize with QTL 1 and four with QTL 2.

We compared the map of candidate regions for selective sweeps in the European population obtained by Li and Stephan (2006) with the QTL detected in the present study. QTL 2 co-localizes with two putative selective sweeps: selective sweep windows 51 and 55. The *CG9911* gene, coding for a protein disulfide isomerase detected in Ayroles *et al.* (2009) study, is located in QTL 2 and window 51 which makes this gene a particularly interesting candidate. Window 55 does not overlap with any of SFPs identified. But Ayroles *et al.* (2009) also defined a list of genes exhibiting expression variations putatively associated with chill coma recovery. Two of these genes are located in window 55: *CG16700* and *CG4991*.

Table 2: X-linked genes containing single feature polymorphisms associated with chill coma recovery in a Raleigh population (Ayroles *et al.* 2009) and their co-localization with the QTL identified in the present study and the candidate selective sweep regions detected in the European population by Li and Stephan (2006).

Gene Name	Cytological position	QTL (this study)	Selective sweep window (Li and Stephan 2006)
skpA	1B14	1	
Ribosomal Protein L22	1C4	1	
CG4045	2C10	1	
VhaC39	3F4	1	
Ribosomal Protein L17	6C10		
CG14439	6C11		
CG32710	8B2		
Enhancer of rudimentary	8B7		
CG7267	8C4		
CG15209	9F5		19
CG4404	11D1		
CG9911	14B9	2	51
Furin2	14C1	2	
TH1	14C2	2	
Lethal(1)G0269	19F4	2	

9. General discussion

The general aim of the present work was to study adaptation in *Drosophila melanogaster*. The historical biogeography of this species provides evidence for the feasibility of such investigations. In addition, the broad range of genetic tools available in *Drosophila* makes studies of adaptation possible. Here, we combine selection mapping with QTL analysis to identify genes that could have played an adaptive role.

9.1 Detection of footprints of selection on the DNA

Evidence for a selective sweep in the *HDAC6* region of African *D. melanogaster*: By completely re-sequencing a 22-kb region around *HDAC6* in a sample of 12 African *D. melanogaster* X chromosomes and applying two likelihood tests (*CLR* and *SweepFinder*), we found evidence consistent with the presence of a selective sweep in this region. Furthermore, our mapping showed that the target of selection is most likely located in a 2.7-kb DNA region, centering on the last exon of *HDAC6*.

The expected age of the sweep was estimated as 50,000 to 63,000 years, depending on the method and input parameter values. This suggests that the sweep occurred before the European lineage split off from the African one (which occurred about 16,000 years ago; Li and Stephan 2006). Our age estimates are consistent with the observation that the sequences of the *HDAC6* alleles from our European sample are identical with that of the African haplotype in the swept region of approximately 2.7 kb (except for some derived low-frequency variants). Consistent with the relatively old age of the selective sweep, we did not identify any pattern of LD that is characteristic of a sweep (according to the predictions by Pfaffelhuber *et al.* 2008). Interestingly, a PAML

analysis (Yang 2007) of HDAC6 sequences from five species of the *D. melanogaster* subgroup found no evidence of selection (data not shown). This suggests that, prior to the inferred selective sweep, *HDAC6* has not undergone accelerated evolution in the past few million years.

It is clear that the evidence we provided is subject to some uncertainty. First, the results inferred by the *CLR* and *SweepFinder* tests may depend to some degree on demography. In particular, complex demographies could be a confounding factor (for instance, population size bottlenecks; Pavlidis *et al.* 2008). However, the demographic history of the African population we inferred previously is probably relatively simple and may be summarized by an expansion model (Li and Stephan 2006; Hutter *et al.* 2007). Furthermore, we have improved the original *CLR* test by Kim and Stephan (2002) and have now taken demography into account. Finally, the problem of demography is alleviated by applying *SweepFinder*, because the chromosome-wide background SFS is used rather than a specific model. Second, a more general concern may be that if selection is a frequent and major pervasive force our two-step approach for inferring selection may not work (Hahn 2008). Then a joint inference of selective and demographic parameters would be a more appropriate approach. However, we emphasize that we search for very strong selection. In such a case, our method of separating demography from selection is expected to be a reasonable first approximation. Third, the uncertainty in the estimates of the target site of selection needs to be mentioned. Unlike Beisswanger and Stephan (2008), we were not able to obtain confidence intervals of our estimates, as the rate of recombination in the *HDAC6* region is too large. However, based on the site frequency spectrum we were able to support our conclusion that the most likely target of selection is located in a 2.7-kb region (between positions 9.8 and 12.5 kb; see figure 5). This result is consistent with

the observation that the European alleles are identical in this region with the selected African allele. This latter argument, however, requires that the sweep occurred in Africa before the African and European lineages split, which is indeed supported by the estimated lower bound of the age of the sweep of >20,000 years.

Can the polymorphism patterns in the *HDAC6* region be explained by selective pressures other than positive directional selection? It is possible that at least part of the polymorphism pattern is associated with the action of purifying selection. The entire 70-kb region contains several functional elements that give rise to low divergence levels (figure 3). In the identified 2.7-kb region between positions 9.8 and 12.5 kb, however, divergence is everywhere in the range of 5-8% and thus comparable to the average of the whole 70-kb region of 6.8%. This suggests that purifying selection is not likely a major cause of the observed pattern of variation in the 2.7-kb region.

Significance of the selective sweep in relation to the function of HDAC6: The 2.7-kb region we mapped by the selection approach overlaps with the last exons of two genes, *HDAC6* and *CG9123*. The latter is a duplicate of *CG12608*. According to the alignment of the 12 fully sequenced *Drosophila* species (*Drosophila* 12 Genomes Consortium 2007), this duplication event occurred in the *D. melanogaster* lineage. From an evolutionary perspective, such a duplication is important, even if *CG9123* seems not to have any coding function (based on the polymorphism pattern mentioned above, *CG9123* is probably a pseudogene or on its way to becoming one), the duplication could have altered some regulatory regions or some UTRs. Another possibility is, for example, that the transcript of *CG9123* could behave like an RNAi and play a role in the regulation of the *CG12608* gene. The functional consequences of the duplication should be further investigated. But, we concentrate the following

discussion on *HDAC6* because *CG9123* is located at the boundary of the identified 2.7-kb region.

HDAC6 is a unique member of the histone deacetylase family harboring a ubiquitin-binding site and two catalytic deacetylase domains (Verdel *et al.* 2000; Khochbin *et al.* 2001). In addition, its localization in the cytoplasm is very unusual for an histone deacetylase (Verdel *et al.* 2000). It has been shown that its role is not limited to gene regulation. Rather, it is also important for the general cytotoxic stress response. It is involved in the two major cellular mechanisms degrading misfolded protein aggregates: autophagy and the ubiquitin-proteasome system (Pandey *et al.* 2007). HDAC6 detects and mediates the cytotoxic stress response at three different levels. First, its strong ubiquitin-binding ability coupled with its ability to move along microtubules allows HDAC6 to transport ubiquitinated protein aggregates, thus favoring the formation of aggresomes. Second, HDAC6 is able to stimulate autophagy when the ubiquitin-proteasome system is impaired (Pandey *et al.* 2007), and finally it mediates the activation of heat shock proteins (Boyault *et al.* 2007). More generally, HDAC6 is believed to be involved in several other cell stress response pathways such as antiviral responses (Boyault *et al.* 2006). In *D. melanogaster*, *HDAC6* is mainly expressed in an insect specific organ: the Malpighian tubule (Chintapalli *et al.* 2007). Its tissues might be exposed to a broad range of cellular stress as it carries out most of the osmoregulation and the excretion of organic solutes as well as xenobiotics (Dow and Davies 2006)

To identify possible targets of selection, we aligned the *HDAC6* sequence of five *Drosophila* species. It revealed that *HDAC6* carries a limited number of *D. melanogaster*-specific changes. But we could neither confirm nor exclude that any of them is a positively selected substitution. Indeed, any nucleotide change in the introns

or 3' UTR could affect *HDAC6*'s regulation or expression and any of the non-synonymous changes observed in the exons could modify the protein's properties. However, in the last exon of *HDAC6* one non-synonymous substitution may well have significant functional consequences: a valine-to-glutamic acid replacement that occurred in the *D. melanogaster* lineage and is located in the ubiquitin-binding site of *HDAC6*. Could this substitution affect the ubiquitin-binding affinity of HDAC6 and thus the response of cells to stress? Ubiquitin-binding assays (Boyault *et al.* 2006) comparing the *D. melanogaster* and *D. simulans* alleles may provide an answer to this question.

9.2 Cold tolerance as an adaptive phenotype

Very little is known about the genetic factors contributing to adaptation to new climatic conditions. The aim of the present experiment was to identify genes involved in the adaptation of *D. melanogaster* to the European climate. When *D. melanogaster* expanded its range to northern latitudes, it had to adapt to colder climates. As the minimum winter temperature is an important factor controlling the distribution of insect species, the colonization of Europe is expected to drive strong selective pressure toward higher cold tolerance (Izquierdo 1991, Hoffman *et al.* 2003).

Naturally occurring variation in Chill Coma Recovery time: In the present study, the wild African genotypes are clearly more sensitive to cold than the European ones. This can be observed with regard to the onset of CCR as well as to mean CCR time. These findings are in agreement with previous results (Arthur *et al.* 2008; Ayrinhac *et al.* 2004): temperate populations of *Drosophila* are more cold-tolerant than tropical ones, and this pattern follows a latitudinal cline. Similar observations have been made in other studies using fitness traits or cold shock survival (Trotta *et al.* 2006; Bubliy *et al.* 2001).

Interestingly, the variance in CCR time was significantly higher in the African population than in the European one. A possible explanation may be provided by the canalization theory developed by Waddington in the 1940s (reviewed in Flatt 2005). The theory is based on the observation that some genotypes are more stable in producing a given phenotype than others. Here, the observation that the variance in CCR time is lower in the European population than in the African one suggests that the cold tolerant phenotype in Europe is canalized. The canalization phenomenon can be

produced by selective forces driving a phenotype to an optimum value that maximizes the fitness and thus decreases the variation around that optimum (see figure 11.10 in Lynch and Walsh 1997, p. 314). According to this theory, natural selection has shaped the European genotypes to produce an optimal cold tolerance phenotype, whereas little or no selection of this type occurred on this trait in Africa. Moreover, morphological measurements made on the same lines (S. Hutter, pers. comm.) did not showed evidence of environmental canalization. Theses observations rule out any possible effect of the breeding conditions on developmental instability and confirm our selection hypothesis.

Some cold tolerance factors are located on the X chromosome: The two substitution lines A* and E* differ from each other only by their X chromosome. This means that all observed phenotypic differences may be attributed to genetic differences carried by their respective X chromosomes. Significant differences in CCR between these two lines were found, confirming the presence of cold tolerance factors on the X chromosome. This has not been shown in previous studies (Morgan and Mackay 2006; Norry *et al.* 2007a; Norry *et al.* 2008). The X-chromosome effect on cold tolerance can also be seen by comparing the CCR profiles of 6418 and 5905 in Figure 7. Both lines differ only in their X chromosome (5905 carries a wild type chromosome whereas 6418 carries a balancer X chromosome). However, our observations suggest that genetic factors controlling cold tolerance are also present on the autosomes, because lines A* and A157 differ significantly in their CCR profile.

RILs analysis: As the studied European and the African X chromosomes carry some segregating factors affecting cold tolerance, we decided to map them more precisely using a QTL approach.

The general goal of quantitative genetics is to identify the genetic factors (Quantitative Trait Loci or QTL) that determine a complex phenotype. Many of such QTL have been identified in *Drosophila melanogaster* for a broad variety of phenotypes (Mackay 2001). Most of these studies started from parental lines artificially selected for divergent phenotypes (Long *et al.* 1995; Moehring and Mackay 2004; Norry *et al.* 2008) in order to maximize the initial phenotypic variance (and thus the chances to detect significant QTL). However, with the recent advance of statistical methods and molecular marker techniques, it is now possible to detect QTL based on much lower levels of phenotypic variation. This makes of the QTL approach an interesting tool for evolutionary biologists.

As shown by our heritability results, about 20% of the variation we observe in the RILs has a genetic origin, which is in the range of previous studies (Morgan and Mackay 2006). The distribution of the mean CCR times (see Appendix figure B3) in the RILs population does not show a biphasic response expected under a major QTL with two allele scenario. This suggests that in the present case the observed phenotypic variation is due to several loci with medium to low contribution to the trait. Interestingly, there are some RILs exhibiting higher cold tolerance than the wild European parental line. This means that it is possible to generate African-European mosaic X chromosomes that confer a higher cold tolerance than a full European X chromosome. Some African alleles can thus contribute positively to cold tolerance when placed in a specific genetic background.

QTL analysis: The CIM mapping analysis detected the presence of only two significant QTL at both ends of the chromosome. We believe that this result is rather conservative because the threshold level is calculated from data permutations. Furthermore, because of the bias of African to European allele frequencies the power to detect QTL in the middle of the chromosome is relatively lower than at the ends, making the overall threshold higher. Other studies (Morgan and Mackay 2006; Norry *et al.* 2008) did not detected cold tolerance factors on the X chromosome. There could be several reasons for that. In the present study we focused our efforts on the X chromosome, and removed the effect of the rest of the genome. In addition we used more markers and more RILs. And, most importantly, we used lines that were not artificially selected for any phenotype. It might well be that the genes underlying phenotypic adaptation to cold are not those that respond to artificial selection in lab conditions.

Candidate genes: According to the GO database, the X chromosome is rather poor in genes involved in temperature-related phenotypes. There is a well known QTL for knockdown resistance to high temperature in the middle of the chromosome between bands 10A-12D (Norry *et al.* 2007b; Norry *et al.* 2008). However this region is relatively distant to the identified QTL. Moreover, all of the cold tolerance genes previously identified (like *ppk*, *trap1*, *Fst*, *Sas*, *desat2*, *Catsup*, *Ddc*, *trap1*, *nompA*, *hsp 70*, *Dca*, and *hsr-omega*) are located on the autosomes (Goto 2000; Goto 2001; Morgan and Mackay 2006; Norry *et al.* 2007a; Sinclair *et al.* 2007; Norry *et al.* 2008). From Gene Ontologies, two candidate genes for QTL 1 are *Adar* and *per*. Some *Adar* alleles have been shown to increase sensitivity to heat shocks (Ma *et al.* 2001). The *per* gene has been shown to regulate the circadian rhythms and being involved in diapause. This

phenomenon can be described as an energy allocation shift from reproduction to survival triggered by decreasing temperature and day length (Saunders *et al.* 1989). The energy reallocation seems to trigger the expression of numerous protective proteins like Heat shock proteins (HSP). However a broad variety of proteins from different pathways (like the glycolytic pathway) can be involved in the recovery of cold shock (Clark and Worland 2008).

If the cold shock response in *Drosophila* has some similarities with bacteria, cold shock exposure produces a drastic reprogramming of gene expression to allow survival under the new unfavorable conditions (Giuliodori *et al.* 2007). Under these conditions, precise candidates might have been detected by Ayroles *et al.* (2009). Some genes of Table 2 are thus potential good candidates as *RpL22*, *CG4045* and *TH1* are coding for a ribosomal protein, a tRNA methyltransferase, and an mRNA binding protein, respectively.

Some other potential candidates are *VhaC39* and *Lethal(1)G0269*. They code for an hydrogen exporting ATPase and a phosphatase, respectively. Very little is known about these genes. All of them could be involved either in the cold sensing pathways or into the physiological response to cold. Unlike heat response very little is known about the precise molecular pathways involved in cold tolerance. The present study may have identified new factors related to cold tolerance (probably into QTL 2). We combined QTL locations with additional information from selection and association mapping studies to compile a candidate gene list. Of the new genes identified *CG9911* appears to be the most likely candidate. That gene is located into QTL 2 region, has been associated with CCR in previous studies and is located into one of Li and Stephan (2006) candidate sweep region. Interestingly, this gene codes for a disulfide isomerase. This could be of importance as some of the thermal hysteresis (antifreeze) proteins

(THP) that have been described in insects are more cystein rich (Duman and Horwath 1983) than alanine rich (like in fishes). Cysteins are well known for their ability to form disulfide bonds that are the substrate of CG9911. Additional selective sweep investigations on this gene have to be done. Similarly, the fine-scale analysis of the window-55 region surrounding *CG16700* and *CG4991* will confirm or infer the presence of footprints of selection. Complementary quantitative PCR experiments could precisely determine their expression pattern in our populations and their contribution to cold tolerance.

Perspectives: QTL mapping experiment requires that the measurements are conducted in a special context. In the present case, we measured cold tolerance in terms of CCR in a special experimental design (7 hours cold shock at 0°C). A broad range of phenotypic tests exists (Hoffmann *et al.* 2003b; Sinclair and Roberts 2005) using survival, fertility or the diapause phenomenon, and these tests can be performed with or without cold hardening and in different temperature conditions (freezing or non-freezing). All these tests may give different results and any one may not tell the whole story. The general cold tolerance ability is something broad and any or many of its small phenotypic manifestations might be selected in natural populations.

In the present work, we investigated the genetic factors affecting cold tolerance in males. The same experiment carried out in females could confirm or infer the QTL we found, and may detect new ones. Indeed, it is likely that in cosmopolitan populations the female is the overwintering sex (Izquerido 1990; Hoffmann *et al.* 2003). They have a larger body size than males making them more cold tolerant (David *et al.* 1998; Morgan and Mackay 2006). In addition, females seem to be able to store sperm during the warm season, survive winter using a reproductive diapause, and lay

fertile eggs when the climatic conditions improve again (Izquierdo 1991; Hoffman *et al.* 2003). Finally, the X chromosome is more likely to carry factors beneficial to females (reviewed in Ellegren and Parsch 2007)

More generally, we show that QTL mapping can be an interesting approach in an evolutionary context. Usually, QTL analyses are performed on artificially selected lines with the aim of identifying the genes that code for a particular trait. The present QTL mapping experiment successfully detected some segregating genetic factors responsible of cold tolerance variation between two natural populations of *D. melanogaster*. The phenotypic variation produced by natural selection is probably lower than the one that can be produced by artificial selection. Nevertheless, with the use of a dense marker map and a large RIL population, the QTL approach is able to detect significant marker-trait associations on such subtle phenotypic variability. This makes QTL mapping a valuable approach for studying adaptive phenotypes in natural populations.

9.3 The combination of selective sweep and QTL mapping

At the first glance, the selective sweep and the QTL approaches do not converge in their conclusions and might not appear complementary. Nevertheless, some interesting parallels can provide significant insights but also reveal the limitations of such an approach. First, the QTL analysis detected a marginally significant QTL around the position of the *HDAC6* selective sweep region (see Figure 8). If this is not an artifact, could the *HDAC6* region contain candidate genes for adaptation? HDAC6 is implied in a general cell stress response pathway and could thus play a cell protective role under cold stress conditions. However, the fact that the selective sweep is older than the out-of-Africa migration event argues against this hypothesis. Indeed, the selective sweep has probably no link with the non-significant QTL mapped. The two lines, A157 and E14, used for the QTL experiment do not carry noticeable nucleotide differences within the 2.7 kb candidate region identified in the selective sweep mapping study (see Appendix table A1 part 7-8). Thus, none of the sites segregating between A157 and E14 in the last exon of *HDAC6* can explain the observed phenotypic differences. However, would a second selective event private to the European population in that region be possible? Such an event could generate nucleotide differences influencing phenotypic changes. This hypothesis cannot be completely ruled out: the 5' upstream region of *HDAC6* contains some derived variants that are fixed in Europe (see Appendix table A1 part 3-4). Nevertheless, such a signal can also be generated by the strong population bottleneck experienced by the European flies (simulation data not shown). In addition, such nucleotide changes in the 5' upstream region should probably affect the expression of *HDAC6*. The micro-arrays analysis of Hutter *et al.* (2008), however, did not detect expression changes for *HDAC6*.

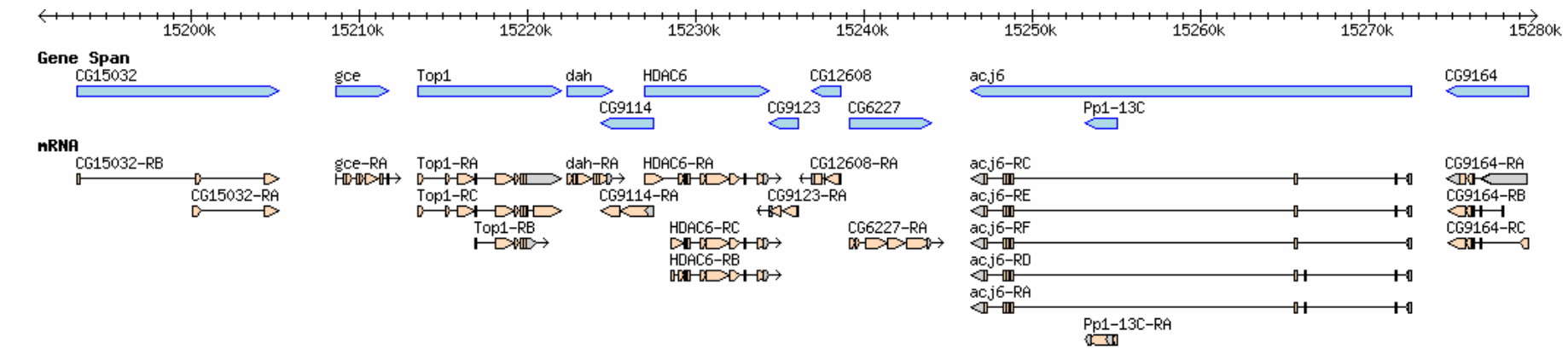


Figure 9: The genes located in the *HDAC6* region according to *D. melanogaster* genome release 5.1. The figure was obtained from the Flybase Genome Browser (http://flybase.org/cgi-bin/gbrowse/dmel/).

Then, could other genes be responsible for the signal detected in the QTL experiment? Several genes located in the *HDAC6* region could play a role in the cold-stress response (see figure 9). First, the *pp1* gene codes for a protein that is implied in the regulation of polyol synthesis. These molecules are alcohols derived from different sugars and they are able to lower the freezing point of a solution. In biological tissues, they are thus supposed to play a cryoprotectant role. And, it has been shown in several species that the polyol concentrations in the body (and noticely in the hemolymph) increase in winter (for example Hayakawa 1985; Storey and Storey 1986). Another gene in the region that may play a role in cold tolerance is *CG6227*. The predicted amino-acid product of this gene contains very long alanine chains. This special amino-acid feature makes of CG6227 a putative anti-freeze protein (AFP). Indeed, these alanine chains have the ability to bind to ice crystals in formation. By doing so, AFPs thus reduce the cell damages caused by freezing and strongly lower the freezing point of organic tissues. Such AFPs have been detected in insects (see Duman *et al.* 2004 for a list) but it is not known whether CG6227 codes for one. Interestingly, the A157 and E14 alleles of *CG6227* code for protein products that differ in their alanine chain length. Further investigations are needed to demonstrate the AFP function of CG6227. If the thermal hysteresis role of this protein is confirmed, the functional differences coded by the two alleles of the gene could be tested by using transgenic flies. Transgenic lines carrying the two different alleles could be constructed and compared for their cold resistance ability. Additionally, the two allozymes can be compared for thermal hysteresis performances (Duman and Horwath 1983).

The above discussion also sheds light on the limitations of the combined usage of selection and QTL mapping. Indeed, finding a co-localization of a selective sweep

and a QTL for a relevant phenotype does not necessarily mean that the two signals are linked. The selective sweep can be due to selection on a trait that has nothing to do with the QTL. Conversely, selection on the trait (even if ecologically relevant) used for the QTL mapping may not have necessarily created selective sweeps. But, some additional experiments could show whether cold tolerance has been positively selected. As several genes putatively implied in cold tolerance have already been mapped (like *Fst*, *Sas*, *desat2*, *Catsup*, *Ddc*, *trap1*, *nompA*, *hsp70*, *Dca*, and *hsr-omega;* Goto 2000; Goto 2001; Morgan and Mackay 2006; Norry *et al.* 2007a; Norry *et al.* 2008), it would be possible to check whether these genes show footprints of selection in our populations. The general approach would be similar to the one used in the selective sweep study. The detection of footprints of selection in the cosmopolitan *D. melanogaster* populations around cold tolerance factors could then confirm the selective importance of cold adaptation in the colonization of the northern latitudes.

In the effort of mapping adaptive genes, the combined approach has the big advantage of using both genetic (selection) and phenotypic information. However, the link between the two approaches has to be done with complementary data. Such information could come from genome annotations or micro-arrays experiments (like those of Ayroles *et al.* 2009). In the present case, micro-arrays experiments would be a very interesting complementary study. It would determine whether differently expressed genes between these populations could be responsible of the cold tolerance differences between Europe and Africa. Gene expression could be compared between different contexts: under different photoperiods, before and after cold shock, using cold hardening or not, or using different temperature conditions or cold shock durations. In bacteria, cold shock exposure produces a drastic reprogramming of gene expression (Giuliodori *et al.* 2007). If it is also true for *Drosophila*, the present set of RILs is

perfectly suited for the search of regulatory elements controlling the expression of one or several genes (called eQTL mapping). For eQTL mapping, the present RILs have the advantage of originating from X substitution lines (which have an homogeneous genetic background). The effects of *trans*-acting factors on the expression of the X-linked genes are thus very limited, which increases the detection power of *cis*-acting elements.

Whatever method is used for establishing a link between selection at the DNA level and a phenotype, the final answer, however, has to come from the direct comparison of the candidate alleles in a transgenic experiment.

9.4 Theoretical consideration about the combined approach

The common QTL analysis searching for the genes that control the expression of a special trait use complementation mapping as a subsequent fine mapping strategy (reviewed in Mackay 2001). For QTL analysis performed in an evolutionary context, selective sweep mapping might be a powerful fine mapping strategy. Indeed, it is a very precise DNA based mapping strategy that is largely documented and widely used. However, the assumption that the trait studied created selective sweeps is not trivial. The detection of a selective sweep and QTL at the same location might require very special biological conditions (Chevin and Hospital 2008). In many cases, one expects that both signals will not co-localize. In particular, complex traits may be under the control of "many" genes, each having a small contribution to the trait. The prediction of the relative response to selection of the genetic factors contributing to a quantitative trait is difficult. It somehow relies on how the selective pressure will be distributed among all the contributing genes. For example the genes that have small effects on the trait or those that are strongly constrained by pleiotropy in some other pathways might be nearly masked for selection. However taking the reasoning in the opposite way, if there are many reasons that QTL and selective sweep do not co-localize, finding such case might be of importance. Another problem common to selective sweep and QTL mapping is the confounding effects due to the surrounding genes. Indeed, because of their physical linkage on the DNA, some genetic factors might affect the detection of sweeps as well as QTL.

9.5 The perspectives of studying the dynamics of complex traits in natural populations

Generally, the study of adaptation contributes to the understanding of the evolutionary processes that influence genetic variation for phenotypic traits. Although the effects of demographic and selective pressures on DNA are well understood, much less is known about their influence on the phenotypic traits themselves. The study of the combined effects of drift, gene flow, and the mutation-selection balance on the complex traits will certainly be a key future research objective for the deeper understanding of evolution. Nowadays, the difficulty of such studies lies in both the complexity of the gene interactions (epistasis) that contribute to the expression of a given trait and the complexity of the combined effects of the evolutionary forces shaping DNA evolution (demography, selection…). Moreover, adaptation can be understood only in a particular environment (some genes may be adaptive in a given environment but can be neutral or even deleterious in other conditions). Local adaptation of populations and the subsequent occurrence of migration between them might even increase the complexity of such studies. Nevertheless, more and more studies were successful in finding the evolutionary forces shaping phenotypic evolution. For example certain nucleotide variants in the *Mc1r* gene are responsible for the coat color polymorphism among different populations of beach mices (Hoekstra *et al*. 2006), and the foraging gene in *Drosophila* controls the larval locomotion behavior and seems to evolve under frequency-dependent selection (Fitzpatrick *et al*. 2007). If in these examples, the traits have a rather simple genetic determinism, with the increasing technical and statistical progress, we might be able to understand more complex cases.

Thus, we will enhance our understanding of the processes that shaped life for billion years.

10. Bibliography

AGUADE, M., N. MIYASHITA and C. H. LANGLEY, 1989. Reduced Variation in the *Yellow-Achaete-Scute* Region in Natural Populations of *Drosophila melanogaster*. Genetics 122: 607-615.

AKEY, J. M., M. A. EBERLE, M. J. RIEDER, C. S. CARLSON, M. D. SHRIVER *et al*., 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol 2: e286.

ARNHEIM, N., C. STRANGE and H. ERLICH, 1985. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of the HLA class II loci. Proc Natl Acad Sci U S A 82: 6970-6974.

ARTHUR, A. L., A. R. WEEKS and C. M. SGRO, 2008. Investigating latitudinal clines for life history and stress resistance traits in *Drosophila simulans* from eastern Australia. J Evol Biol 21: 1470-1479.

AYRINHAC, A., P. DEBAT, P. GIBERT, A.-G. KISTER, H. LEGOUT *et al*., 2004. Cold adaptation in geographical populations of *Drosophila melanogaster*: phenotypic plasticity is more important than genetic variability. Functional Ecology 18: 700-706.

AZEVEDO, R. B., V. FRENCH and L. PARTRIDGE, 1997. Life-history consequences of egg size in *Drosophila melanogaster*. Am Nat 150: 250-282.

AYROLES, J. F., M. A. CARBONE, E. A. STONE, K. W. JORDAN, R. F. LYMAN *et al*., 2009. Systems genetics of complex traits in *Drosophila melanogaster*. Nat Rev Genet 41: 299-307.

BEGUN, D. J., and C. F. AQUADRO, 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature 356: 519-520.

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

BEISSWANGER, S., W. STEPHAN and D. DE LORENZO, 2006. Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. Genetics 172: 265-274.

BOTSTEIN, D., R. L. WHITE, M. SKOLNICK and R. W. DAVIS, 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314-331.

BOYAULT, C., B. GILQUIN, Y. ZHANG, V. RYBIN, E. GARMAN *et al*., 2006. *HDAC6-p97/VCP* controlled polyubiquitin chain turnover. Embo J 25: 3357-3366.

BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics 140: 783-796.

BUBLIY, O. A., A. RIIHIMAA, F. M. NORRY and V. LOESCHCKE, 2002. Variation in resistance and acclimation to low-temperature stress among three geographical strains of *Drosophila melanogaster*. Journ. of Therm. Biol. 27: 337-344.

BURR, B., F. A. BURR, K. H. THOMPSON, M. C. ALBERTSON and C. W. STUBER, 1988. Gene mapping with recombinant inbreds in maize. Genetics 118: 519-526.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289-1303.

CHARLESWORTH, D., B. CHARLESWORTH and M. T. MORGAN, 1995. The pattern of neutral molecular variation under the background selection model. Genetics 141: 1619-1632.

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

CHINTAPALLI, V. R., J. WANG and J. A. DOW, 2007. Using *FlyAtlas* to identify better *Drosophila melanogaster* models of human disease. Nat Genet 39: 715-720.

CLARK, M. S., and M. R. WORLAND, 2008. How insects survive the cold: molecular mechanisms-a review. J Comp Physiol [B] 178: 917-933.

COMERON, J. M., M. KREITMAN and M. AGUADE, 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics 151: 239-249.

DARWIN, C., 1859. *The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. Murray, London.

DAVID, J. R., and P. CAPY, 1988. Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet 4: 106-111.

DAVID, J. R., P. GIBERT, E. PLA, G. PETAVY, D. KARAN *et al*., 1998. Cold stress tolerance in *Drosophila*: analysis of chill coma recovery in *D. melanogaster*. Journ. of Therm. Biol. 23: 291-299.

DE LUCA, M., N. V. ROSHINA, G. L. GEIGER-THORNSBERRY, R. F. LYMAN, E. G. PASYUKOVA *et al*., 2003. *Dopa decarboxylase* (*Ddc*) affects variation in *Drosophila* longevity. Nat Genet 34: 429-433.

DERRY, J. F., 2009. Darwin in disguise. Trends Ecol Evol 24: 73-79.

DOBZHANSKY, T., 1937. *Genetics and the Origin of Species*. Columbia University Press, New York.

DOEBLEY, J., 1992. Mapping the genes that made maize. Trends Genet 8: 302-307.

DONIS-KELLER, H., P. GREEN, C. HELMS, S. CARTINHOUR, B. WEIFFENBACH *et al*., 1987. A genetic linkage map of the human genome. Cell 51: 319-337.

DOW J.A., and S.A. DAVIES, 2006. The Malpighian tubule: rapid insights from post-genomic biology. J. Insect. Physiol. 52: 365-378.

DROSOPHILA 12 GENOMES CONSORTIUM. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature. 450: 203-218.

DUMAN, J.G., T. BENNETT, T. SFORMO, R. HOCHSTRASSER and B.M. BARNES, 2004. Antifreeze proteins in Alsakan insects and spiders. J. Insect. Physiol. 50: 259-266.

DUMAN, J., and K. HORWATH, 1983. The role of hemolymph proteins in the cold tolerance of insects. Annu Rev Physiol 45: 261-270.

DUMONT, V. B., and C. F. AQUADRO, 2005. Multiple signatures of positive selection downstream of *notch* on the X chromosome in *Drosophila melanogaster*. Genetics 171: 639-653.

ELLEGREN, H., and J. PARSCH, 2007 The evolution of sex-biased gene and sex-biased expression. Nat. Rev. Genet. 8: 689-698.

FALCONER, D. S., and T. F. C. MACKAY, 1996. *Introduction to Quantitative Genetics*, Essex.

FAY, J. C., and C. I. WU, 2000. Hitchhiking under positive Darwinian selection. Genetics 155: 1405-1413.

FITZPATRICK, M. J., E. FEDER, L. ROWE and M. B. SOKOLOWSKI, 2007. Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. Nature 447: 210-212.

FLATT, T., 2005. The evolutionary genetics of canalization. Q Rev Biol 80: 287-316.

FRARY, A., T. C. NESBITT, S. GRANDILLO, E. KNAAP, B. CONG *et al*., 2000. *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85-88.

FU, Y. X., and W. H. LI, 1993. Statistical tests of neutrality of mutations. Genetics 133: 693-709.

GIBERT, P., B. MORETEAU, G. PETAVY, D. KARAN and J. R. DAVID, 2001. Chill-coma tolerance, a major climatic adaptation among *Drosophila* species. Evolution 55: 1063-1068.

GIULIODORI, A. M., C. O. GUALERZI, S. SOTO, J. VILA and M. M. TAVIO, 2007. Review on bacterial stress topics. Ann N Y Acad Sci 1113: 95-104.

GLINKA, S., L. OMETTO, S. MOUSSET, W. STEPHAN and D. DE LORENZO, 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics 165: 1269-1278.

GOTO, S. G., 2000. Expression of *Drosophila* homologue of senescence marker protein-30 during cold acclimation. J Insect Physiol 46: 1111-1120.

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

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

HAHN, M.W., 2008. Toward a selection theory of molecular evolution. Evolution. 62: 255-265.

HAYAKAWA, Y., 1985 Activation of insect fat-body phosphorylase by cold - phosphorylase-kinase, phosphatase and ATP level. Insect Bio Chem 15: 123-128.

HOEKSTRA, H.E., R.J. HIRSCHMANN, R.A. BUNDEY, P.A. INSEL and J.P. CROSSLAND, 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313: 101-104.

HOFFMANN, A. A., M. SCOTT, L. PARTRIDGE and R. HALLAS, 2003. Overwintering in *Drosophila melanogaster*: outdoor field cage experiments on clinal and laboratory selected populations help to elucidate traits under selection. J Evol Biol 16: 614-623.

HOFFMANN, A. A., J. G. SØRENSEN and V. LOESCHCKE, 2003. Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular appraoches. Journ. of Therm. Biol. 28: 175-216.

HUEY, R. B., G. W. GILCHRIST, M. L. CARLSON, D. BERRIGAN and L. SERRA, 2000. Rapid evolution of a geographic cline in size in an introduced fly. Science 287: 308-309.

HUTTER, S., H. LI, S. BEISSWANGER, D. DE LORENZO and W. STEPHAN, 2007. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosome wide single nucleotide polymorphism data. Genetics 177: 469-480.

HUTTER, S., S.S. SAMINADIN, W. STEPHAN and J; PARSCH, 2008 Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biol. 9: R1-12

IHLE, S., I. RAVAOARIMANANA , M. THOMAS, and D. TAUTZ, 2006. An analysis of signatures of selective sweeps in natural populations of the house mouse. MBE 23: 790-797.

INTERNATIONAL HAPMAP CONSORTIUM. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851-861.

IZQUIERDO, J. I., 1991. How does *Drosophila melanogester* overwinter? Entomol. exp. appl. 59: 51-58.

JAMES, A. C., R. B. AZEVEDO and L. PARTRIDGE, 1995. Cellular basis and developmental timing in a size cline of *Drosophila melanogaster*. Genetics 140: 659-666.

JAMES, A. C., R. B. AZEVEDO and L. PARTRIDGE, 1997. Genetic and environmental responses to temperature of *Drosophila melanogaster* from a latitudinal cline. Genetics 146: 881-890.

JENSEN, J. D., V. L. BAUER DUMONT, A. B. ASHMORE, A. GUTIERREZ and C. F. AQUADRO, 2007. 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, J. D., Y. KIM, V. B. DUMONT, C. F. AQUADRO and C. D. BUSTAMANTE, 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics 170: 1401-1410.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989. The "hitchhiking effect" revisited. Genetics 123: 887-899.

KELLER, A., 2007. *Drosophila melanogaster*'s history as a human commensal. Curr Biol 17: R77-81.

KHOCHBIN, S., A. VERDEL, C. LEMERCIER and D. SEIGNEURIN-BERNY, 2001. Functional significance of histone deacetylase diversity. Curr Opin Genet Dev 11: 162-166.

KIM, Y., and R. NIELSEN, 2004. Linkage disequilibrium as a signature of selective sweeps. Genetics 167: 1513-1524.

KIM, Y., and W. STEPHAN, 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160: 765-777.

KIMURA, M., 1968. Evolutionary rate at the molecular level. Nature 217: 624-626.

KINGMAN, J. F. C., 1982. The coalescent. Stochast. Proc. Appl. 13: 235-248.

LACHAISE, D., M.-L. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS *et al*., 1988. Historical Biogeography of *Drosophila melanogaster* Species Subgroup. Evolutionary Biology 22: 159-225.

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

LAMARCK, J.-B., 1809. *Philosophie Zoologique*. Dentu, Paris.

LANDER, E. S., and D. BOTSTEIN, 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. J. DALY *et al*., 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

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

LONG, A. D., R. F. LYMAN, C. H. LANGLEY and T. F. C. MACKAY, 1998. Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. Genetics 149: 999-1017.

LONG, A. D., S. L. MULLANEY, L. A. REID, J. D. FRY, C. H. LANGLEY *et al*., 1995. High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. Genetics 139: 1273-1291.

LYNCH, M., and B. WALSH, 1998. Genetics and Analysis of Quantitative Traits, Sunderland.

MA, E., X. Q. GU, X. WU, T. XU and G. G. HADDAD, 2001. Mutation in pre-mRNA adenosine deaminase markedly attenuates neuronal tolerance to O2 deprivation in *Drosophila melanogaster*. J Clin Invest 107: 685-693.

MACDONALD, S. S., L. RAKO, P. BATTERHAM and A. A. HOFFMANN, 2004. Dissecting chill coma recovery as a measure of cold resistance: evidence for a biphasic response in *Drosophila melanogaster*. J Insect Physiol 50: 695-700.

MACKAY, T. F.C., 2001. Quantitative trait loci in *Drosophila*. Nat Rev Genet 2: 11-20.

MACKAY, T. F.C., 2004. The genetic architecture of quantitative traits: lessons from *Drosophila*. Curr Opin Genet Dev 14: 253-257.

MATTHIAS, P., M. YOSHIDA and S. KHOCHBIN, 2008. HDAC6 a new cellular stress surveillance factor. Cell Cycle 7: 7-10.

MAYNARD-SMITH, J., and J. HAIGH, 1974. The hitch-hiking effect of a favourable gene. Genet Res 23: 23-35.

MICHELMORE, R. W., I. PARAN and R. V. KESSELI, 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci U S A 88: 9828-9832.

MILTON, C. C., and L. PARTRIDGE, 2008. Brief carbon dioxide exposure blocks heat hardening but not cold acclimation in *Drosophila melanogaster*. J Insect Physiol 54: 32-40.

MITROVSKI, P., and A. A. HOFFMANN, 2001. Postponed reproduction as an adaptation to winter conditions in *Drosophila melanogaster*: evidence for clinal variation under semi-natural conditions. Proc Biol Sci 268: 2163-2168.

MOEHRING, A. J., J. LI, M. D. SCHUG, S. G. SMITH, M. DEANGELIS *et al*., 2004. Quantitative trait loci for sexual isolation between *Drosophila simulans* and *D. mauritiana.* Genetics 167: 1265-1274.

MORGAN, T. J., and T. F. MACKAY, 2006. Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. Heredity 96: 232-242.

NIELSEN, R., S. WILLIAMSON, Y. KIM, M. J. HUBISZ, A. G. CLARK *et al*., 2005. Genomic scans for selective sweeps using SNP data. Genome Res 15: 1566-1575.

NORRY, F. M., F. H. GOMEZ and V. LOESCHCKE, 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*. Mol Ecol 16: 3274-3284.

NORRY, F. M., P. SAMBUCETTI, A. C. SCANNAPIECO, F. H. GOMEZ and V. LOESCHCKE, 2007. X-linked QTL for knockdown resistance to high temperature in *Drosophila melanogaster*. Insect Mol Biol 16: 509-513.

NORRY, F. M., A. C. SCANNAPIECO, P. SAMBUCETTI, C. I. BERTOLI and V. LOESCHCKE, 2008. QTL for the thermotolerance effect of heat hardening, knockdown resistance to heat and chill-coma recovery in an intercontinental set of recombinant inbred lines of *Drosophila melanogaster*. Mol Ecol 17: 4570-4581.

OMETTO, L., S. GLINKA, D. DE LORENZO and W. STEPHAN, 2005. Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. Mol Biol Evol 22: 2119-2130.

ORENGO, D. J., and M. AGUADE, 2004. Detecting the footprint of positive selection in a european population of *Drosophila melanogaster*: multilocus pattern of variation and distance to coding regions. Genetics 167: 1759-1766.

ORENGO, D. J., and M. AGUADE, 2007. Genome scans of variation and adaptive change: extended analysis of a candidate locus close to the phantom gene region in *Drosophila melanogaster.* Mol Biol Evol 24: 1122-1129.

PANDEY, U. B., Z. NIE, Y. BATLEVI, B. A. MCCRAY, G. P. RITSON *et al*., 2007. HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. Nature 447: 859-863.

PAVLIDIS, P., S. HUTTER and W. STEPHAN, 2008. A population genomic approach to map recent positive selection in model species. Mol Ecol. 16: 3585-98

PFAFFELHUBBER, P., A. LEHNERT, and W. STEPHAN, 2008. Linkage disequilibrium under genetic hitchhiking in finite populations. Genetics. 179: 527-537.

POOL, J. E., V. BAUER DUMONT, J. L. MUELLER and C. F. AQUADRO, 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., 2003. Estimating the time since the fixation of a beneficial allele. Genetics 164: 1667-1676.

ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.

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

SAX, K., 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8: 552-560.

SINCLAIR, B. J., and S. P. ROBERTS, 2005. Acclimation, shock and hardening in the cold. Journ. of Therm. Biol. 30: 557-562.

SINCLAIR, B. J., A.G. GIBBS and S. P. ROBERTS, 2007. Gene transcription during exposure to, and recovery from, cold and desiccation stress in *Drosophila melanogaster* Insect Mol. Biol. 16: 435-443.

SLATKIN, M., and R. R. HUDSON, 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129: 555-562.

SØRENSEN, J. G., and V. LOESCHCKE, 2002. Natural adaptation to environmental stress via physiological clock-regulation of stress resistance in *Drosophila*. Ecology Letters 5: 16-19.

STEPHAN, W., and C. H. LANGLEY, 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., Y. S. SONG and C. H. LANGLEY, 2006. The hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics 172: 2647-2663.

STINCHCOMBE, J. R., and H. E. HOEKSTRA, 2007. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100: 158-170.

STOREY, J. M., and K. B. STOREY, 1986. Winter survival of the gall fly larvae, *Eurosta solidaginis*: profiles of fuel reserves and cryoprotectants in a natural population. J Insect Physiol 63: 549-556.

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.

THODAY, J. M., 1961. Location of polygenes. Nature 191: 368-379.

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

TROTTA, V., F. C. CALBOLI, M. ZIOSI, D. GUERRA, M. C. PEZZOLI *et al*., 2006. Thermal plasticity in *Drosophila melanogaster*: a comparison of geographic populations. BMC Evol Biol 6: 67.

VERDEL, A., S. CURTET, M. P. BROCARD, S. ROUSSEAUX, C. LEMERCIER *et al.*, 2000. Active maintenance of mHDA2/mHDAC6 histone-deacetylase in the cytoplasm. Curr Biol 10: 747-749.

VISSCHER, P. M., W. G. HILL and N. R. WRAY, 2008. Heritability in the genomics era--concepts and misconceptions. Nat Rev Genet 9: 255-266.

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

WILLIAMSON, S. H., M. J. HUBISZ, A. G. CLARK, B. A. PAYSEUR, C. D. BUSTAMANTE *et al*., 2007. Localizing recent adaptive evolution in the human genome. PLoS Genet 3: e90.

YANG, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591.

ZENG, Z. B., 1994. Precision mapping of quantitative trait loci. Genetics 136: 1457-1468.

11. Appendix A: selective sweep mapping additional figures

Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6* part 1

	pos.	115	149	176	417	419	420	424	425	426	494	679	733	737	750	770	825	902	917	992	1001	1031	1071	1121	1216	1454	1562	1597
	Line																											
African	82	A	A	G	G	T	T	C	C	G	T	T	G	C	T	C	T	C	A	C	C	G	T	C	T	G	T	T
	84	A	G	G	G	T	T	C	C	C	T	T	G	C	C	T	C	G	A	T	C	A	C	T	T	G	T	T
	95	G	G	G	G	T	T	C	T	C	T	T	G	C	C	T	C	G	A	T	C	A	C	T	T	G	T	T
	131	A	G	G	A	G	G	C	C	C	T	T	C	C	T	C	C	G	A	T	C	A	C	T	T	G	T	T
	145	A	G	G	G	T	T	C	C	C	T	T	G	C	C	T	C	G	A	T	C	G	T	C	T	G	T	T
	157	A	G	G	A	G	G	T	C	C	T	T	G	C	C	T	C	G	A	T	C	A	C	T	T	A	T	T
	186	A	G	G	G	T	T	C	C	C	T	T	G	C	C	T	C	G	A	T	C	A	C	T	T	G	T	T
	191	A	G	G	A	G	G	C	C	C	T	T	C	C	C	T	C	G	A	T	C	A	C	T	A	G	C	T
	229	A	G	G	A	G	G	C	C	C	T	T	G	C	C	T	C	G	A	T	G	A	C	T	T	G	T	C
	337	A	G	G	A	G	G	C	C	C	T	T	G	C	C	T	C	G	A	T	C	A	C	T	A	G	T	T
	384	A	G	G	A	G	G	C	C	G	T	C	G	C	C	T	C	G	A	T	C	A	C	T	T	G	T	T
	398	A	G	G	G	T	T	C	C	C	T	T	G	C	C	T	C	G	A	T	C	A	C	T	T	G	T	C
European	1	A	G	G	A	G	G	C	C	C	C	T	G	T	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	2	A	G	G	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	11	A	G	G	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	12	A	G	G	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	13	A	G	G	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	14	A	G	G	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	15	A	G	G	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	16	A	G	G	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	17	A	G	G	A	G	G	C	C	C	C	T	G	T	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	18	A	G	A	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	19	A	G	G	A	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	20	A	G	G	A	G	G	C	C	C	C	T	G	T	T	C	C	C	G	T	C	A	C	T	T	G	T	T
	Anc. St.	A	G	G	G	G	G	C	C	C	C	T	G	C	T	C	C	C	G	T	C	A	C	T	T	G	T	G
		utr	utr	1	I	I	I	I	I	I	2	I	I	3	3	3	3	3	3	3	3	3	3	3	3	3	3	I
				S							S			S	S	S	S	S	S	S	S	S	S	S	S	NS Val=>Glu	S	S	

dah

Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*

part 2

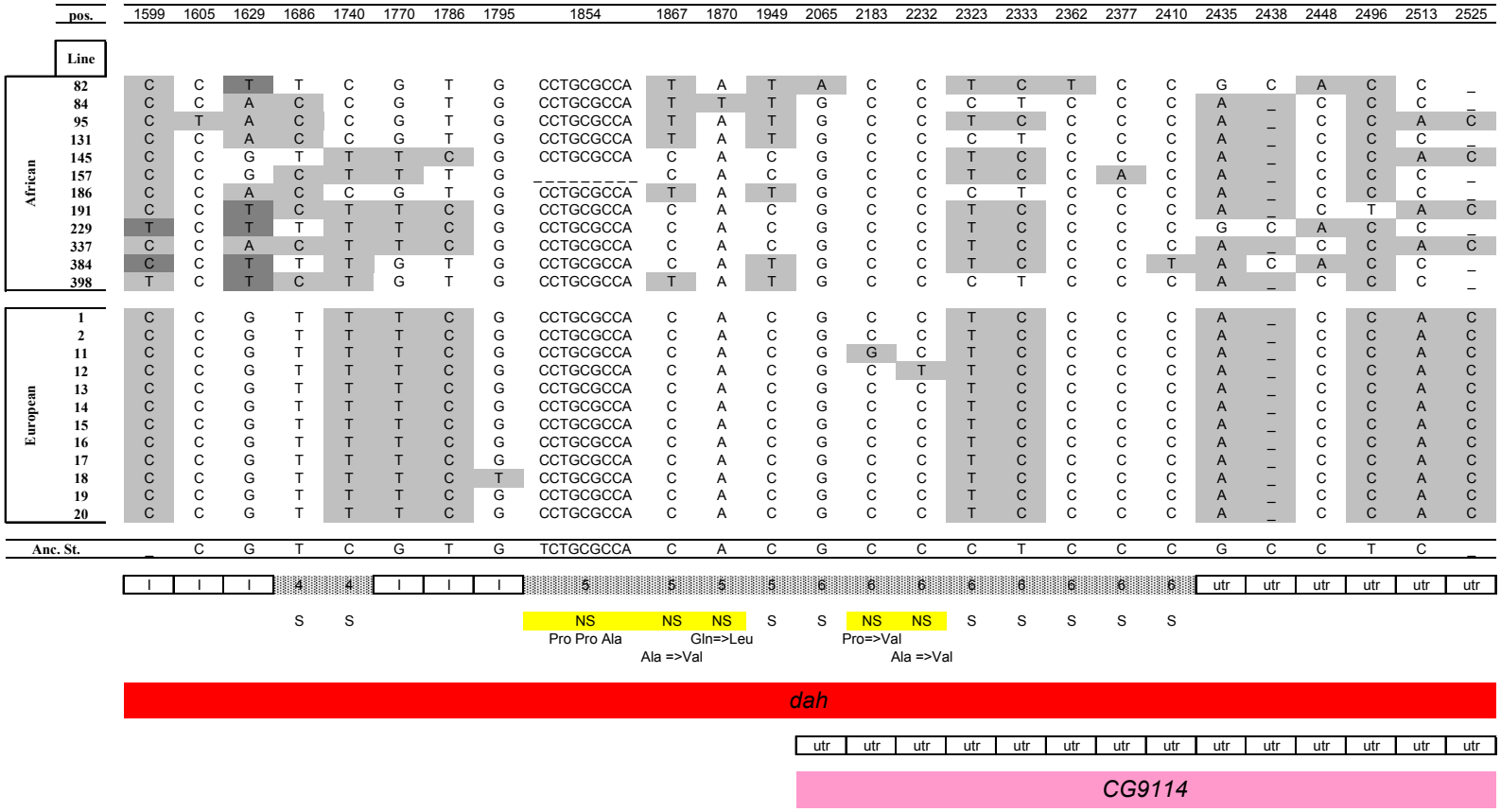


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6* part 3

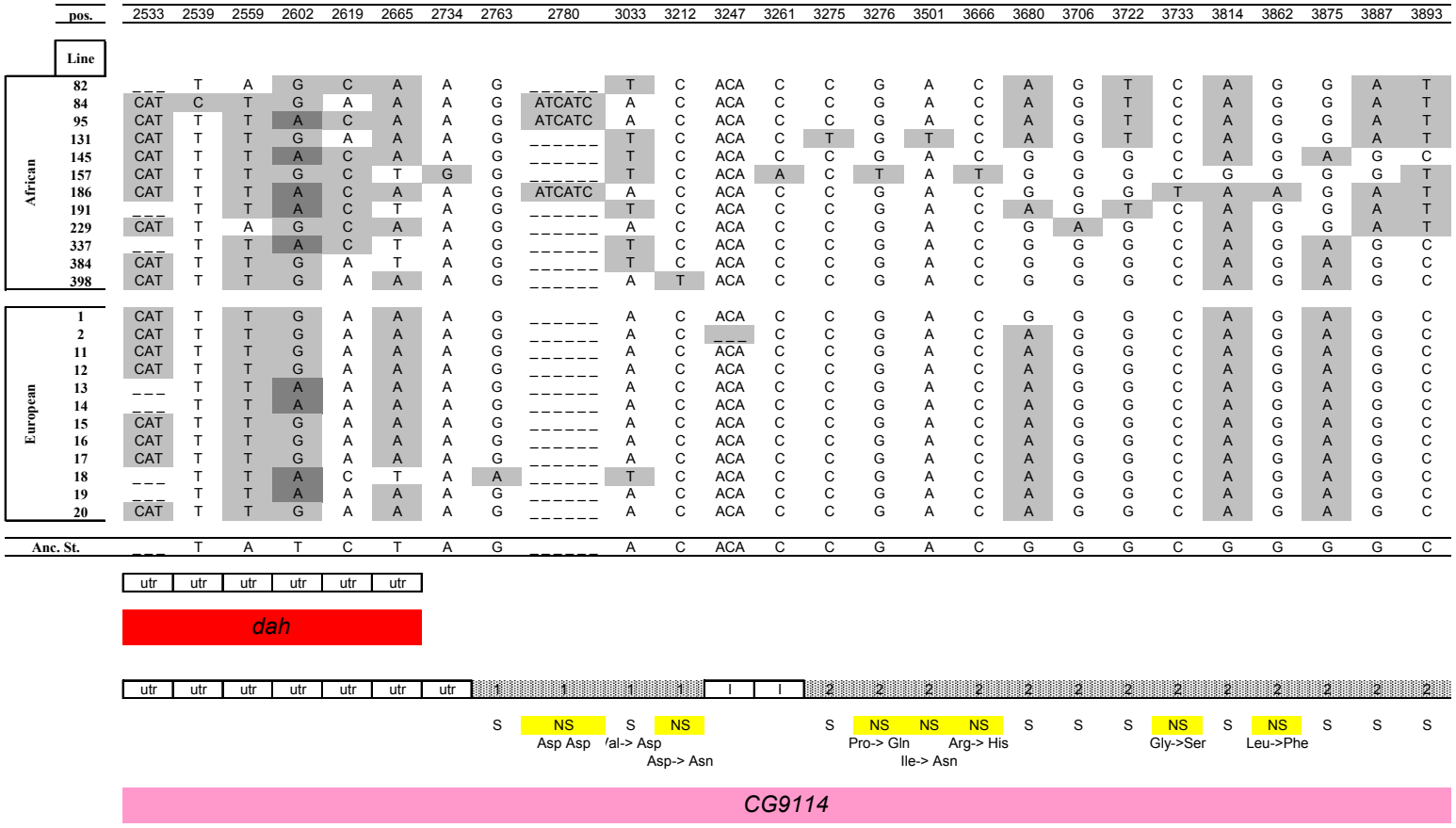


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6* part 4

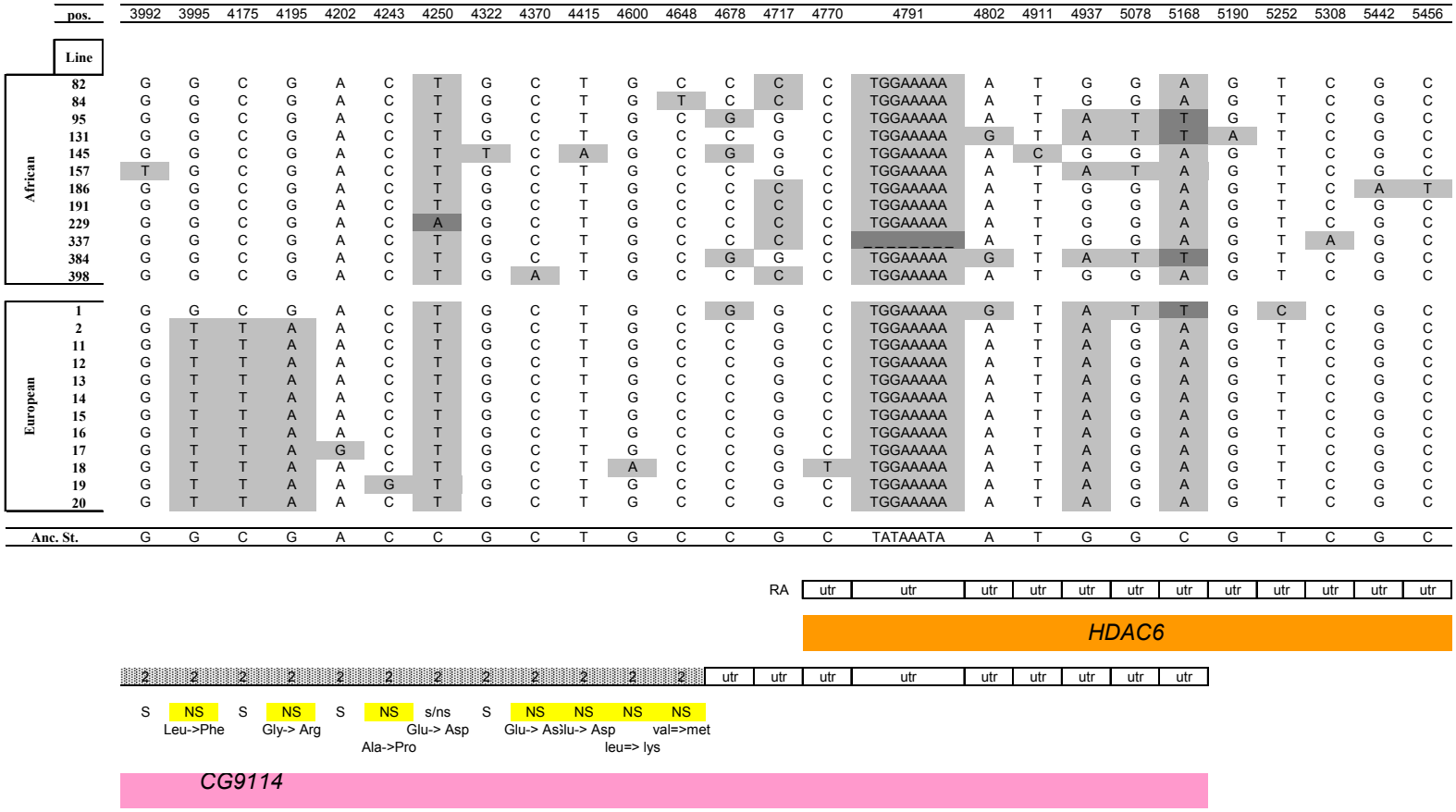


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6* part 5

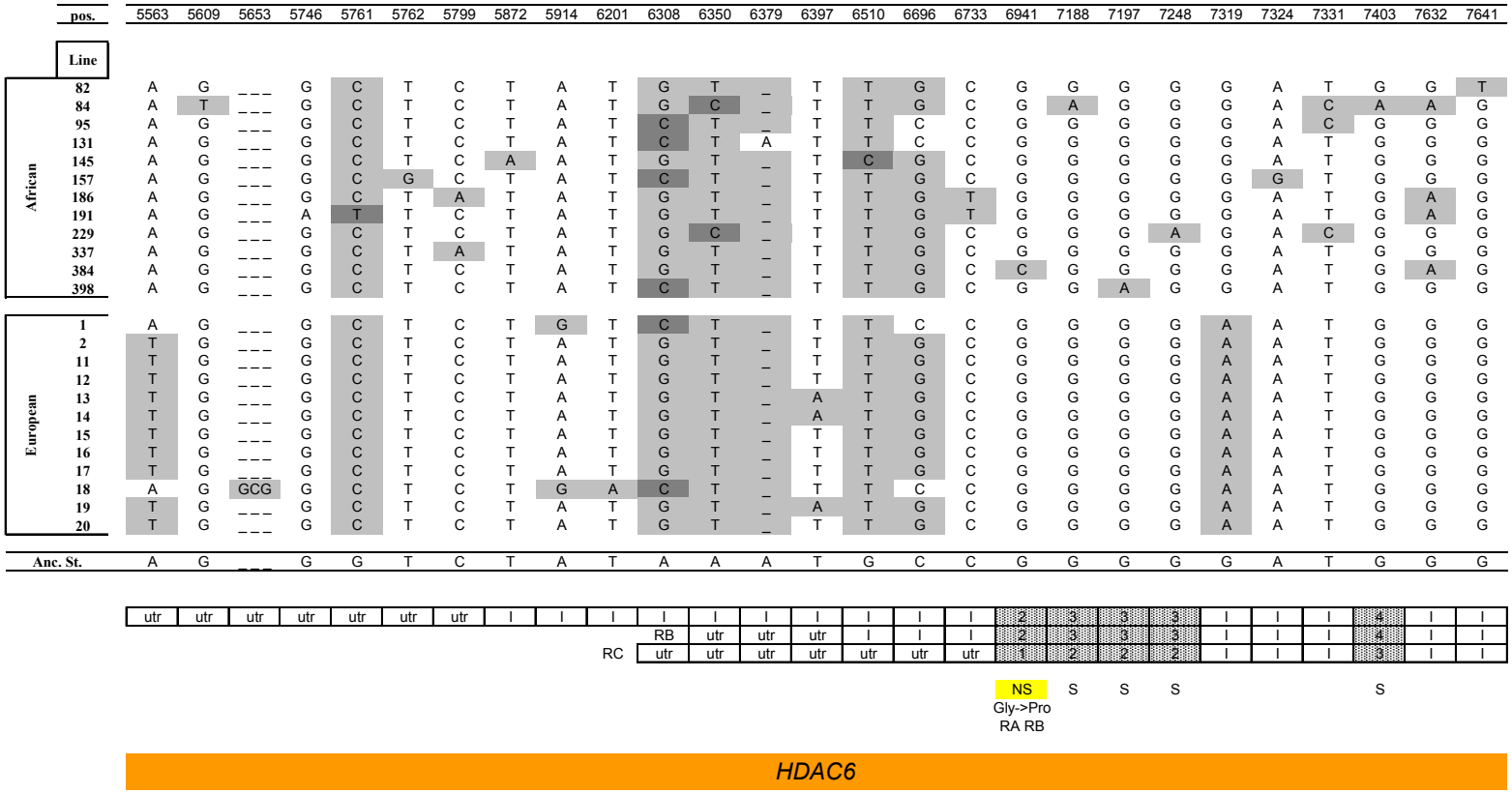


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*

part 6

	pos.	7657	7705	7857	7910	7997	7998	8001	8036	8047	8068	8150	8194	8328	8477	8485	8875	9049	9099	9117	9416	9529	9552	9630	9657	9789
	Line																									
African	82	A	G	GCTATATGGTATC	T	A	C	T	A	A	G	C	A	A	G	C	G	C	G	T	C	C	C	C	T	G
	84	A	G	GCTATATGGTATC	T	A	C	T	A	A	G	C	A	A	G	A	G	C	A	T	C	C	C	C	T	G
	95	A	A	_____	C	A	G	T	A	A	G	C	A	A	G	C	G	C	G	T	C	C	C	C	T	G
	131	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	C	G	C	G	C	T	C	T	T	C	G
	145	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	C	G	C	G	T	C	C	C	T	C	G
	157	T	G	_____	T	A	C	T	A	A	A	C	A	A	G	C	A	C	G	T	C	C	C	T	C	G
	186	A	G	_____	T	T	C	C	A	G	G	C	G	A	G	A	G	C	G	T	C	C	C	T	C	G
	191	A	G	_____	T	T	C	C	A	A	G	C	G	A	G	A	G	C	G	T	C	C	C	C	C	G
	229	A	A	GCTATATGGTATC	T	T	C	C	G	A	G	C	G	A	G	A	G	C	G	T	C	C	C	T	C	G
	337	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	A	G	C	A	T	C	C	C	T	C	G
	384	A	G	GCTATATGGTATC	T	T	C	T	A	A	G	C	A	G	G	C	G	C	G	T	C	C	C	C	T	A
	398	A	A	_____	C	A	C	T	A	A	G	C	A	A	G	A	G	C	A	T	C	C	C	C	T	G
European	1	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	A	G	T	G	C	C	T	C	C	T	G
	2	A	A	_____	T	A	C	T	A	A	G	G	A	A	G	A	G	T	G	C	C	C	C	C	T	G
	11	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	A	G	T	G	C	C	C	C	C	T	G
	12	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	C	G	T	G	C	C	C	C	C	T	G
	13	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	A	G	T	G	C	C	C	C	C	T	G
	14	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	A	G	T	G	C	C	C	C	C	T	G
	15	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	C	G	T	G	C	C	C	C	C	T	G
	16	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	C	G	T	G	C	C	C	C	C	T	G
	17	A	A	_____	T	A	C	T	A	A	G	C	A	A	C	A	G	T	G	C	C	C	C	C	T	G
	18	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	A	G	T	G	C	C	C	C	C	T	G
	19	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	A	G	T	G	C	C	C	C	C	T	G
	20	A	A	_____	T	A	C	T	A	A	G	C	A	A	G	A	G	T	G	C	C	C	C	C	T	G
	Anc. St.	A	G	_CAGTATTGTATC	T	A	C	T	C	A	G	C	A	A	G	C	G	C	G	T	C	C	C	C	C	C
		I	I	I	I	I	I	I	I	I	I	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6
		I	I	I	I	I	I	I	I	I	I	6	5	5	6	6	6	6	6	6	6	6	6	6	6	6
		I	I	I	I	I	I	I	I	I	I	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5
												S	NS	NS	NS	S	S	S	S	S	NS	S	S	S	S	S
													Lys-> Arg		Glu->Gln						Pro->Leu					
														Iso->Val												

HDAC6

Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*

part 7

	pos.	9891	9903	10573	10644	10688	10769	10832	10878	10958	11131	11193	11270	11327	11500	11863	12235
	Line																
African	82	C	C	A	G	T	G	G	A	T	T	G	A	T	T	T	A
	84	C	C	A	G	T	G	G	T	A	T	G	A	T	C	G	A
	95	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	131	C	C	A	G	T	A	G	A	A	A	A	A	T	C	T	A
	145	C	C	A	G	T	G	G	T	A	T	G	T	T	C	G	A
	157	C	C	A	G	A	G	G	A	A	T	G	A	T	C	G	A
	186	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	191	C	C	G	G	T	G	G	A	A	T	G	A	T	C	G	A
	229	C	C	A	G	T	G	G	A	A	T	G	A	T	C	T	A
	337	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	C
	384	C	A	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	398	T	C	A	A	T	G	G	T	A	T	G	A	T	C	G	A
European	1	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	2	C	C	A	G	T	G	G	A	A	T	G	A	A	C	G	A
	11	C	C	A	G	T	G	T	A	A	T	G	A	T	C	G	A
	12	C	C	A	A	T	G	G	A	A	T	G	A	T	C	G	A
	13	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	14	C	C	A	G	T	G	G	A	A	T	G	A	T	C	T	A
	15	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	16	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	17	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	18	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	19	C	C	A	G	T	G	G	A	A	T	G	A	T	C	T	A
	20	C	C	A	G	T	G	G	A	A	T	G	A	T	C	G	A
	Anc. St.	C	C	A	G	T	G	G	?	A	T	A	G	T	C	T	A

I	I	I	I	I	I	E	I	I	I	I	I	I	I	I	I	E	utr
I	I	I	I	I	I	E	I	I	I	I	I	I	I	I	I	E	utr
I	I	I	I	I	I	E	I	I	I	I	I	I	I	I	I	E	utr
						S										NS Val-> Gly	

HDAC6

Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*

part 8

	pos.	12277	12280	12363	12396	12409	12429	12448	12478	12481	12493	12501	12502	12503	12507	12508	12509	12653	12657	12701	12734	12735	12736	12746	12749
	Line																								
African	82	G	T	G	T	C	C	A	A	A	---------	-	T	T	G	T	G	T	TCT	A	T	C	AGACTC	TTC	G
	84	G	T	G	T	C	C	C	A	A	TTGCTGCT	T	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	95	G	T	G	C	C	C	C	A	A	---------	-	T	T	T	T	G	T	TCT	A	-	-	_ _ _ _ _ _	_ _ _	-
	131	G	T	G	T	C	T	C	A	A	---------	-	T	T	T	T	G	T	TCT	A	T	C	AGACTC	TTC	G
	145	G	T	G	T	C	C	C	A	A	TTGCTGCT	T	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	157	A	A	G	C	C	C	C	A	A	---------	-	T	T	T	T	G	T	TCT	A	-	-	_ _ _ _ _ _	_ _ _	-
	186	G	T	G	T	C	C	C	G	G	---------	-	-	C	T	T	A	T	TCT	A	T	C	AGACTC	TTC	G
	191	G	T	G	T	C	C	C	A	A	---------	-	T	T	T	T	G	T	TCT	A	T	C	AGACTC	TTC	G
	229	G	T	C	T	T	C	C	A	A	TTGCTGCT	T	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	337	G	T	G	T	C	C	C	A	A	TTGCTGCT	T	G	C	T	C	A	T	TCT	G	T	C	AGACTC	TTC	G
	384	G	T	G	C	C	C	C	A	A	---------	-	T	T	T	T	G	T	TCT	A	-	-	_ _ _ _ _ _	_ _ _	-
	398	G	T	G	T	C	C	C	A	A	TTGCTGCT	T	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
European	1	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	2	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	11	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	T	AGACTC	TTC	G
	12	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	13	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	14	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	15	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	16	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	17	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	---	A	T	C	AGACTC	TTC	A
	18	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	G	TCT	A	T	C	AGACTC	TTC	G
	19	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	20	G	T	G	T	C	C	C	A	A	TTGCTGCT	C	G	C	T	C	A	T	TCT	A	T	C	AGACTC	TTC	G
	Anc. St.	G	T	?	?	?	T	C	G	G	TTGCTGCT	T	G	C	T	C	G	?	?	A	C	C	AGACTC	TTC	G
		utr	utr	utr	utr	utr	utr	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
								NS	NS	NS	NS	NS	NS		NS	NS		NS	NS	NS	NS	NS	NS	NS	NS
								Lys=> Asn					Arg=> Glu		Met=>Thr=>Ile					Phe=> Ser		Asp=>Asn			Ala=> Va
									!!!	!!!			!!!		Lys=> Gln			Lys=> Thr							

CG9123

Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*

part 9

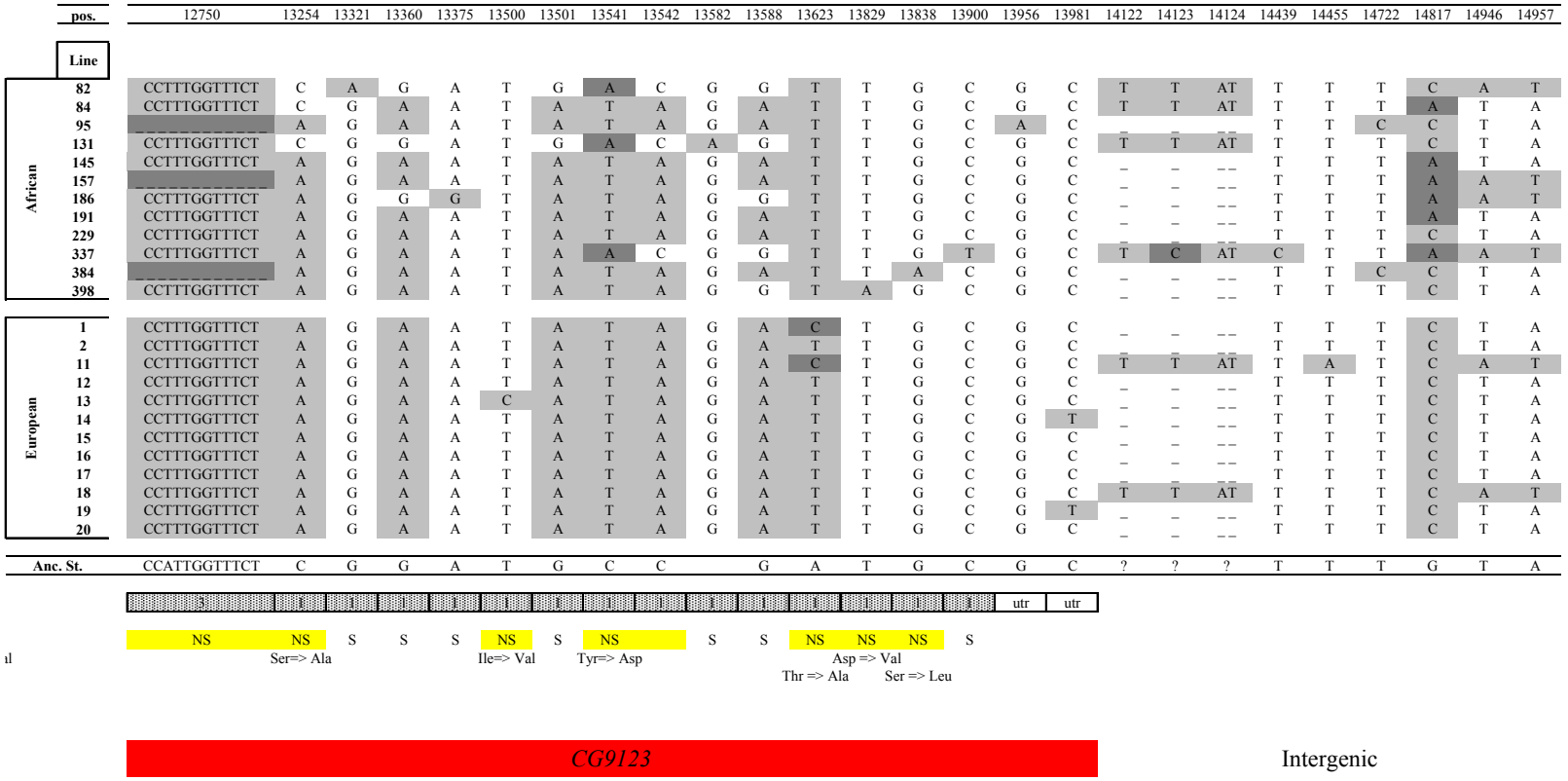


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6* part 10

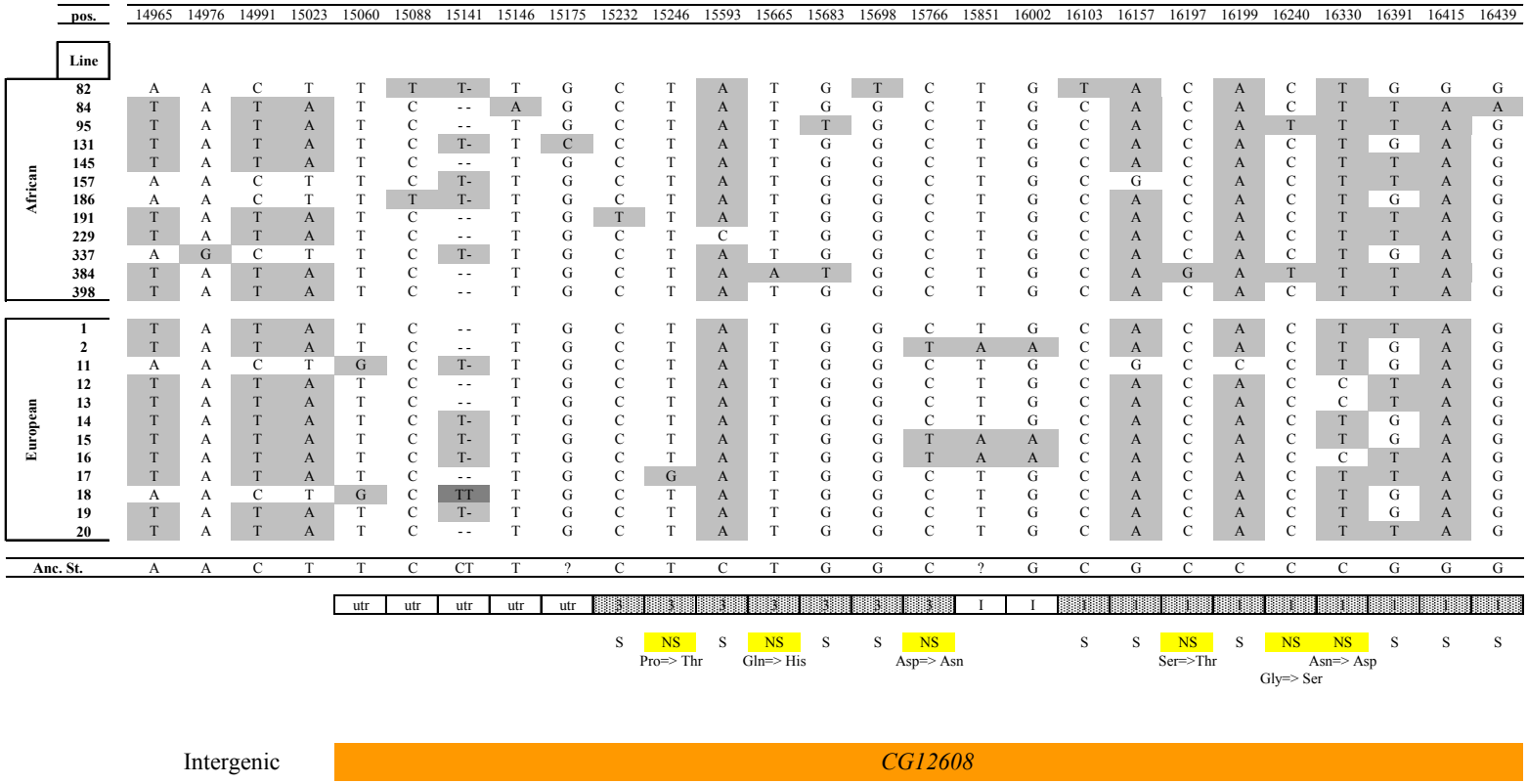


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6* part 11

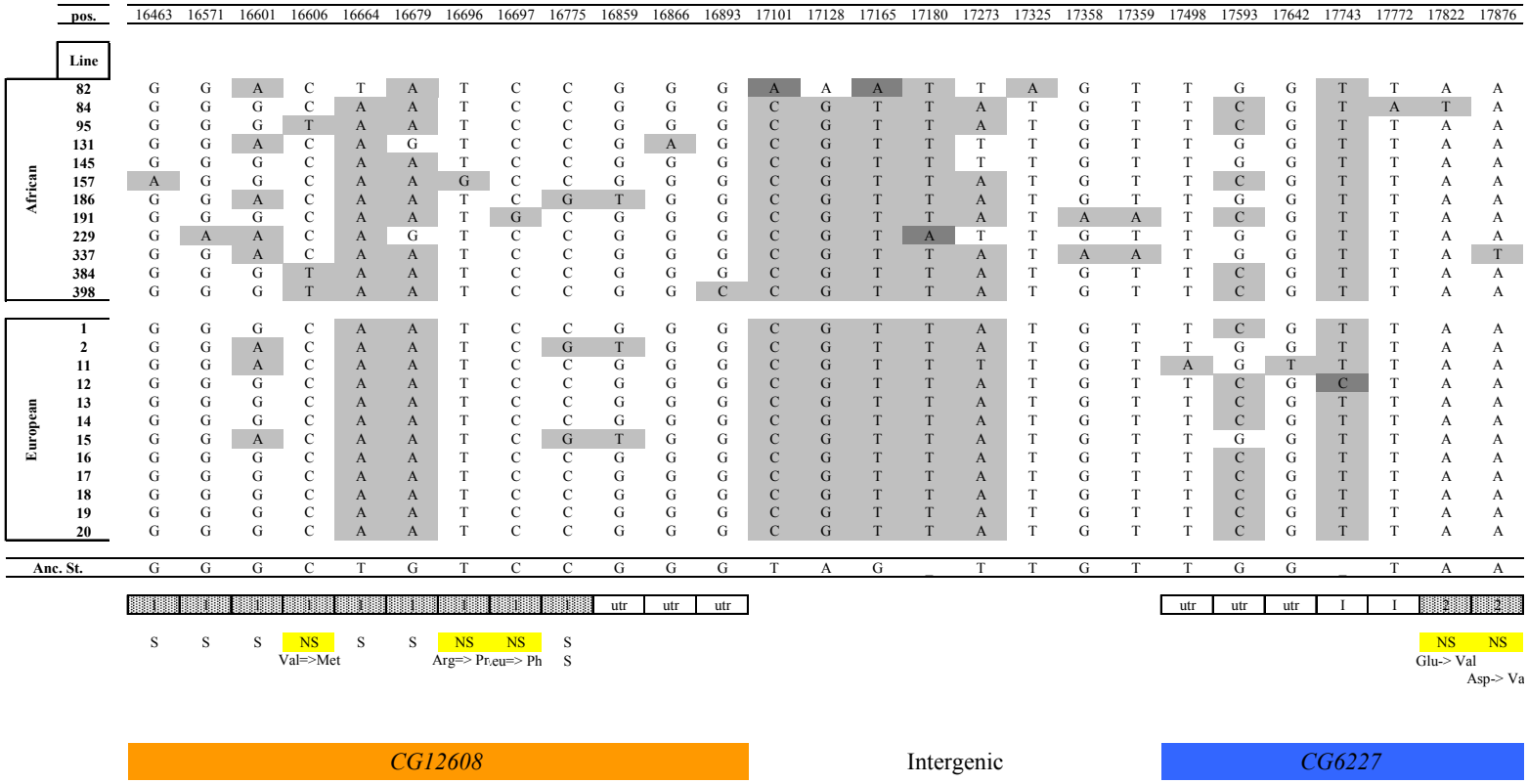


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*

part 12

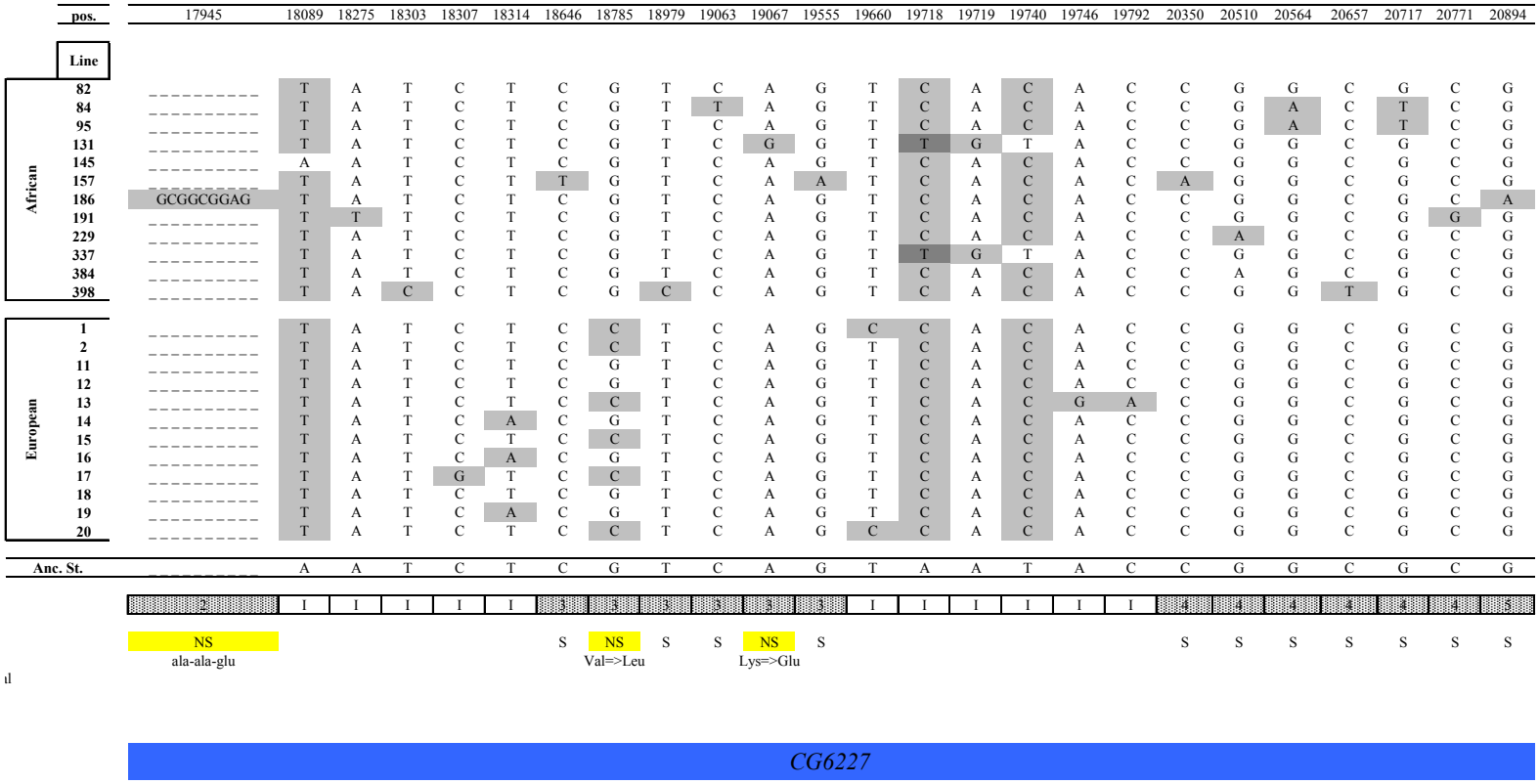


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*

part 13

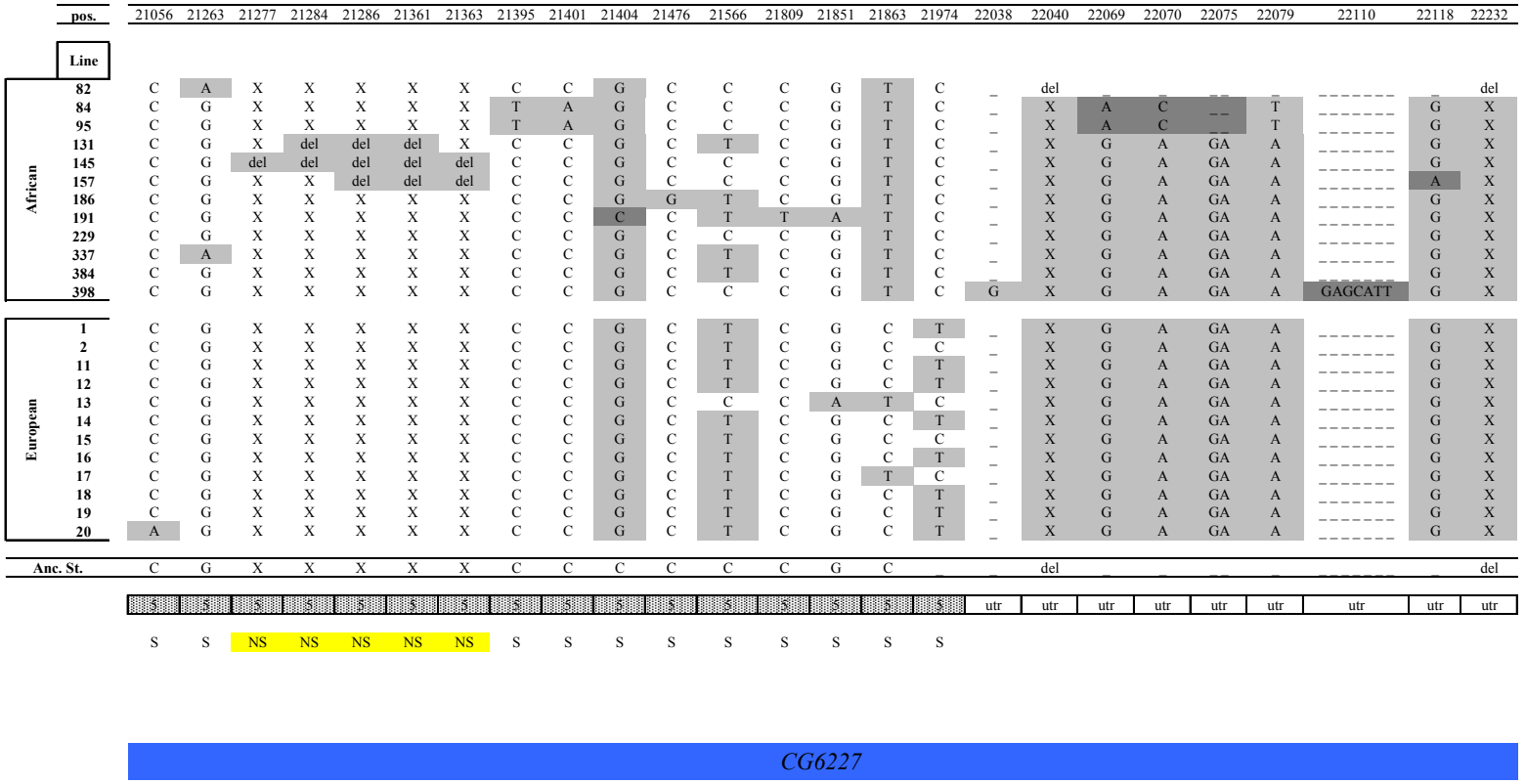


Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*

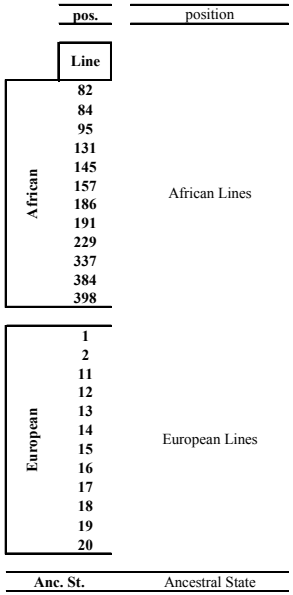
part 14

	pos.	22248	22253	22261	22330	22375	22399	22401	22405	22435	22441	22445	22809	22810	22811	22812
	Line															
African	82	T	T	A	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	84	T	A	G	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	95	T	A	G	TTAACTCAAT	A	A	C	T	T	T	C	T	T	A	A
	131	T	T	A	TTAACTCAAT	A	A	C	T	T	A	C	A	C	G	T
	145	T	A	G	TTAACTCAAT	A	A	G	T	C	T	C	A	C	G	T
	157	T	A	G	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	186	T	T	A	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	191	T	A	G	TTAACTCAAT	A	A	C	C	T	T	C	A	C	G	T
	229	A	A	G	TTAACTCAAT	A	A	C	T	T	T	T	A	C	G	T
	337	T	A	G	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	384	T	A	G	TTAACTCAAT	C	A	C	T	T	T	C	A	C	G	T
	398	T	T	A	TTAACTCAAT	A	A	G	T	C	T	C	A	C	G	T
European	1	T	A	G	TTAACTCAAT	A	G	C	T	T	T	C	A	C	G	T
	2	T	A	G	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	11	T	A	G	_ _ _ _ _ _ _ _ _ _	A	G	C	T	T	T	C	A	C	G	T
	12	T	A	G	TTAACTCAAT	A	G	C	T	T	T	C	A	C	G	T
	13	T	A	G	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	14	T	A	G	TTAACTCAAT	A	G	C	T	T	T	C	A	C	G	T
	15	T	A	G	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	16	T	A	G	TTAACTCAAT	A	G	C	T	T	T	C	A	C	G	T
	17	T	A	G	TTAACTCAAT	A	A	C	T	T	T	C	A	C	G	T
	18	T	A	G	TTAACTCAAT	A	G	C	T	T	T	C	A	C	G	T
	19	T	A	G	TTAACTCAAT	A	G	C	T	T	T	C	A	C	G	T
	20	T	A	G	TTAACTCAAT	A	G	C	T	T	T	C	A	C	G	T
	Anc. St.	T	T	G	TTAACTCAAT	A	_	C	T	C	G	C	A	C	G	T
		utr	utr	utr	utr											

CG6227

Intergenic

Table A1: Polymorphisms in the 22-kb region surrounding *HDAC6*



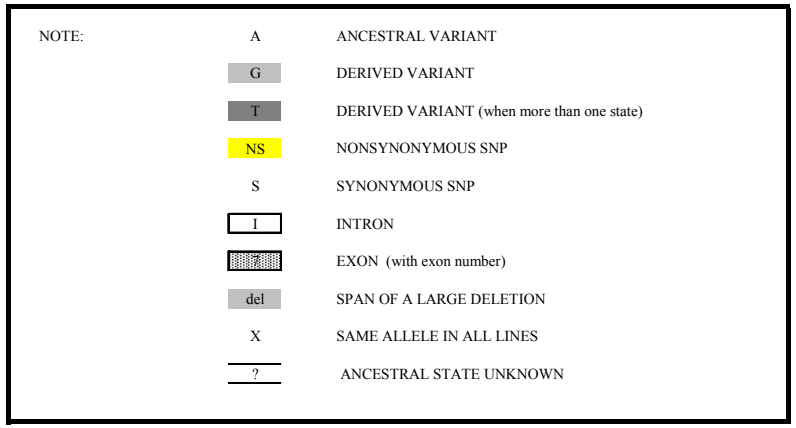


Table A2: Five-species alignment and annotation of the last exon of *HDAC6* part 1

		Span of the ubiquitin-binding site																																				
D. melanogaster	DNA sequence	cta	gac	agc	gga	gcg	gag	tgc	agc	gtt	tgc	gga	tcg	acg	ggg	gag	aac	tgg	gtg	tgc	ctg	agc	tgc	agg	cat	gtg	gcc	tgc	ggt	cgc	tat	gtg	aac	gcc	cac	atg
	Amino acid sequence	L	D	S	G	A	E	C	S	V	C	G	S	T	G	E	N	W	V	C	L	S	C	R	H	V	A	C	G	R	Y	V	N	A	H	M
D. sechellia	DNA sequence	cta	gac	agc	aga	gcg	gag	tgc	agc	gtt	tgc	gga	tcg	acg	ggg	gag	aac	tgg	gtg	tgc	ctg	agc	tgc	agg	cat	gta	gcc	tgc	ggt	cgc	tat	gta	aac	gcc	cac	atg
	Amino acid sequence	L	D	S	R	A	E	C	S	V	C	G	S	T	G	E	N	W	V	C	L	S	C	R	H	V	A	C	G	R	Y	V	N	A	H	M
D. simulans	DNA sequence	cta	gac	agc	aga	gcg	gag	tgc	agc	gtt	tgc	gga	tcg	acg	ggg	gag	aac	tgg	gtg	tgc	ctg	agc	tgc	agg	cat	gta	gcc	tgc	ggt	cgn	tat	gta	aac	gcc	cac	atg
	Amino acid sequence	L	D	S	R	A	E	C	S	V	C	G	S	T	G	E	N	W	V	C	L	S	C	R	H	V	A	C	G	X	Y	V	N	A	H	M
D. erecta	DNA sequence	cta	gac	agc	gga	gcg	gcg	tgc	agc	gtt	tgc	gga	tcg	acg	ggg	gag	aac	tgg	gtg	tgc	ctg	ggc	tgc	cgt	ctg	gtg	gcc	tgt	ggt	cgc	tat	atg	aac	gcc	cac	atg
	Amino acid sequence	L	D	S	G	A	A	C	S	V	C	G	S	T	G	E	N	W	V	C	L	G	C	R	L	V	A	C	G	R	Y	M	N	A	H	M
D. yakuba	DNA sequence	cta	gac	agc	gga	gcg	gcg	tgc	agc	gtt	tgc	gga	tcg	acg	ggg	gag	aat	tgg	gtg	tgc	ctg	agc	tgc	cgg	ctg	gtg	gcc	tgc	ggt	cgc	tac	atg	aac	gcc	cac	atg
	Amino acid sequence	L	D	S	G	A	A	C	S	V	C	G	S	T	G	E	N	W	V	C	L	S	C	R	L	V	A	C	G	R	Y	M	N	A	H	M
Nucleotide substitution					X		X										X					X		X	X	X		X		X	X	X					
AA change					X		X															X			X					X		X					
Nucleotide substitution private to *D. melanogaster*																																						
AA change private to *D. melanogatser*																																						

Note: X indicates the position of a nucleotide substitution, an amino acid change, a nucleotide substitution private to *D.melanogaster* or an amino acid change private to *D. melanogaster*

X indicates the only amino acid change private to *D. melanogaster* located in the ubiquitin binding site

Table A2: Five-species alignment and annotation of the last exon of *HDAC6* part 2

		Span of the ubiquitin-binding site																	
D. melanogaster	DNA sequence	gag	cag	cat	tca	gtg	gag	gag	cag	cat	ccg	ctg	gcg	atg	agc	acg	gct	gac	ctg
	Amino acid sequence	E	Q	H	S	V	E	E	Q	H	P	L	A	M	S	T	A	D	L
D. sechellia	DNA sequence	gag	cag	cat	tcg	gtg	gag	gag	cag	cat	ccg	ctg	gcg	atg	agc	acg	gcg	gac	ctg
	Amino acid sequence	E	Q	H	S	V	E	E	Q	H	P	L	A	M	S	T	A	D	L
D. simulans	DNA sequence	gag	cag	cat	tcg	gtg	gag	gag	cag	cat	ccg	ctg	gcg	atg	agc	acg	gcg	gac	ctg
	Amino acid sequence	E	Q	H	S	V	E	E	Q	H	P	L	A	M	S	T	A	D	L
D. erecta	DNA sequence	gag	cag	cac	tcg	gtg	gag	gcg	cag	cat	ccg	ctg	gca	atg	agc	acg	gcc	gac	ctg
	Amino acid sequence	E	Q	H	S	V	E	A	Q	H	P	L	A	M	S	T	A	D	L
D. yakuba	DNA sequence	gag	cag	cac	gcg	gtg	gag	tcg	cag	cat	ccg	ctg	gcg	atg	agc	acg	gcc	gac	ctg
	Amino acid sequence	E	Q	H	A	V	E	S	Q	H	P	L	A	M	S	T	A	D	L
Nucleotide substitution				X	X			X					X				X		
AA change					X			X											
Nucleotide substitution private to *D. melanogaster*																			
AA change private to *D. melanogatser*																			

D. melanogaster	DNA sequence	tct	gtg	tgg	tgc	tac	gcg	tgc	tcc	gcg	tac	gtg	gac	cat	ccc	cgc	ctg	tat	gcc
	Amino acid sequence	S	V	W	C	Y	A	C	S	A	Y	V	D	H	P	R	L	Y	A
D. sechellia	DNA sequence	tct	gtg	tgg	tgc	tac	gcg	tgc	tcc	gcg	tac	gtg	gac	cat	ccc	cgc	ctg	tat	gcc
	Amino acid sequence	S	V	W	C	Y	A	C	S	A	Y	V	D	H	P	R	L	Y	A
D. simulans	DNA sequence	tct	gtg	tgg	tgc	tac	gcg	tgc	tcc	gcg	tac	gtg	gac	cat	ccc	cgc	ctg	tat	gcc
	Amino acid sequence	S	V	W	C	Y	A	C	S	A	Y	V	D	H	P	R	L	Y	A
D. erecta	DNA sequence	tcc	gtg	tgg	tgt	tac	gcg	tgc	tcg	gcg	tac	gtg	gac	cat	ccc	cgc	ctg	tat	gcc
	Amino acid sequence	S	V	W	C	Y	A	C	S	A	Y	V	D	H	P	R	L	Y	A
D. yakuba	DNA sequence	tcc	gtg	tgg	tgc	tac	gca	tgc	tct	gcg	tac	gtg	gac	cat	ccc	cgc	ctg	tat	gcc
	Amino acid sequence	S	V	W	C	Y	A	C	S	A	Y	V	D	H	P	R	L	Y	A
Nucleotide substitution		X			X		X		X										
AA change																			
Nucleotide substitution private to *D. melanogaster*																			
AA change private to *D. melanogatser*																			

Note:	X	indicates the position of a nucleotide substitution, an amino acid change, a nucleotide substitution private to *D.melanogaster* or an amino acid change private to *D. melanogaster*
	X	indicates the only amino acid change private to *D. melanogaster* located in the ubiquitin binding site

Table A2: Five-species alignment and annotation of the last exon of *HDAC6* part 3

Span of the ubiquitin-binding site

D. melanogaster	DNA sequence	tac	ctc	aat	ccg	ctg	cac	gag	gac	aag	ttc	cag	gaa	ccg	atg	gcc	tgg	aca	cat	ggg	tgt	gcg	tgg	cgc	gag	gat	ggc	tgc	tat	gcg	acg	ggg	ccc	gat	ggg	cga	gat
	Amino acid sequence	Y	L	N	P	L	H	E	D	K	F	Q	E	P	M	A	W	T	H	G	C	A	W	R	E	D	G	C	Y	A	T	G	P	D	G	R	D
D. sechellia	DNA sequence	tac	ctc	aat	ccg	ctg	cac	gtg	gac	aag	ttc	cag	gaa	ccg	atg	gcc	tgg	aca	cat	ggc	tgt	gcg	tgg	cgc	gag	gat	gga	tgc	tat	gcg	acg	ggg	ccc	gat	ggt	cga	gat
	Amino acid sequence	Y	L	N	P	L	H	V	D	K	F	Q	E	P	M	A	W	T	H	G	C	A	W	R	E	D	G	C	Y	A	T	G	P	D	G	R	D
D. simulans	DNA sequence	tac	ctc	aat	ccg	ctg	cac	gtg	gac	aag	ttc	cag	gaa	ccg	atg	gcc	tgg	aca	cat	ggc	tgt	gcg	tgt	cgc	gag	gat	gga	tgc	tat	gct	acg	ggg	ccc	gat	ggt	cga	gat
	Amino acid sequence	Y	L	N	P	L	H	V	D	K	F	Q	E	P	M	A	W	T	H	G	C	A	C	R	E	D	G	C	Y	A	T	G	P	D	G	R	D
D. erecta	DNA sequence	tac	ctg	aat	gcg	ctg	cac	gtg	cac	aag	ttc	cag	gag	ccg	atg	gcc	tgg	aca	cat	ggc	tgt	gcg	tgg	cgc	gag	gat	ggc	tgc	tat	gcg	gtg	gga	ccc	aat	ggg	cga	gac
	Amino acid sequence	Y	L	N	A	L	H	V	H	K	F	Q	E	P	M	A	W	T	H	G	C	A	W	R	E	D	G	C	Y	A	V	G	P	N	G	R	D
D. yakuba	DNA sequence	tac	ctc	aat	ccg	ctg	cac	gtg	gac	aag	ttc	cag	gag	ccg	atg	gcc	tgg	aca	cat	ggc	tgt	gcg	cgg	cgc	gag	gat	ggc	tgc	tat	gcg	ctg	ggc	ccc	gat	ggg	cga	gat
	Amino acid sequence	Y	L	N	P	L	H	V	D	K	F	Q	E	P	M	A	W	T	H	G	C	A	R	R	E	D	G	C	Y	A	L	G	P	D	G	R	D
Nucleotide substitution			X		X			X	X				X							X			X				X			X	X	X		X	X		X
AA change					X			X	X														X								X			X			
Nucleotide substitution private to *D. melanogaster*								X												X																	
AA change private to *D. melanogatser*								X																													

Note:	X	indicates the position of a nucleotide substitution, an amino acid change, a nucleotide substitution private to *D.melanogaster* or an amino acid change private to *D. melanogaster*
	X	indicates the only amino acid change private to *D. melanogaster* located in the ubiquitin binding site

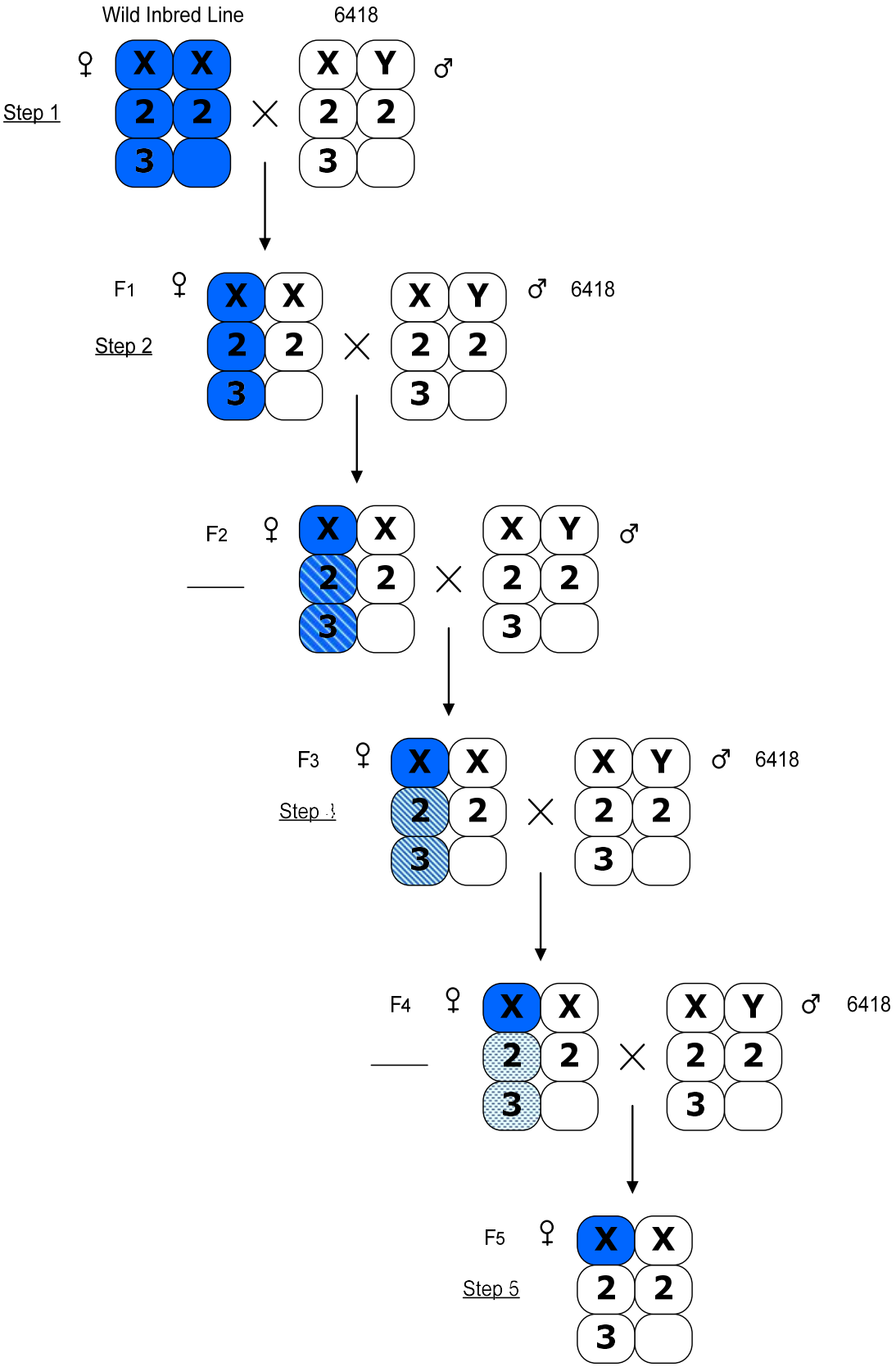
Table A2: Five-species alignment and annotation of the last exon of *HDAC6* part 4

D. melanogaster	DNA sequence	gag	gat	gac				gat	gat	gat	aat	ggc	gcc	ggt	agc	agc	att	tgc	ctg	cga	ctg	gag	cgc	aac	aac	tga
	Amino acid sequence	E	D	D				D	D	D	N	G	A	G	S	S	I	C	L	R	L	E	R	N	N	-
D. sechellia	DNA sequence	gag	gat	gat	gag	aat	aat	gat	gat	gac	aat	gtc	gcc	ggt	agc	agc	atc	tgc	ctg	cga	ctg	gag	cgc	aac	aac	tga
	Amino acid sequence	E	D	D	E	N	N	D	D	D	N	V	A	G	S	S	I	C	L	R	L	E	R	N	N	-
D. simulans	DNA sequence	gag	gat	gac	gag	aat	aat	gat	gat	gac	aat	gtc	gcc	ggt	agc	agc	atc	tgc	ctg	cga	ctg	gag	cgc	aac	aac	tga
	Amino acid sequence	E	D	D	E	N	N	D	D	D	N	V	A	G	S	S	I	C	L	R	L	E	R	N	N	-
D. erecta	DNA sequence	gag		gac	gcg	gat	cag	gat	gat	gcc		gtc	gcc	ggg	agc	agc	atc	tgc	ctg	cgc	ctg	gag	cgc	cac	aac	tga
	Amino acid sequence	E		D	A	D	Q	D	D	A		V	A	G	S	S	I	C	L	R	L	E	R	H	N	-
D. yakuba	DNA sequence	gag		gac	tcg		gag	gat		ggc		gtc	gcc	ggt	agc	agc	tac	tgc	ctg	cga	ctg	gag	cgg	aac	aac	tga
	Amino acid sequence	E		D	S		E	D		G		V	A	G	S	S	Y	C	L	R	L	E	R	N	N	-
Nucleotide substitution			X	X	X	X	X		X	X	X	X		X			X			X			X	X		
AA change			X		X	X	X		X	X	X	X					X							X		
Nucleotide substitution private to *D. melanogaster*												X														
AA change private to *D. melanogatser*												X														

Note: X indicates the position of a nucleotide substitution, an amino acid change, a nucleotide substitution private to *D.melanogaster* or an amino acid change private to *D. melanogaster*

X indicates the only amino acid change private to *D. melanogaster* located in the ubiquitin binding site

12. Appendix B: QTL mapping additional figures



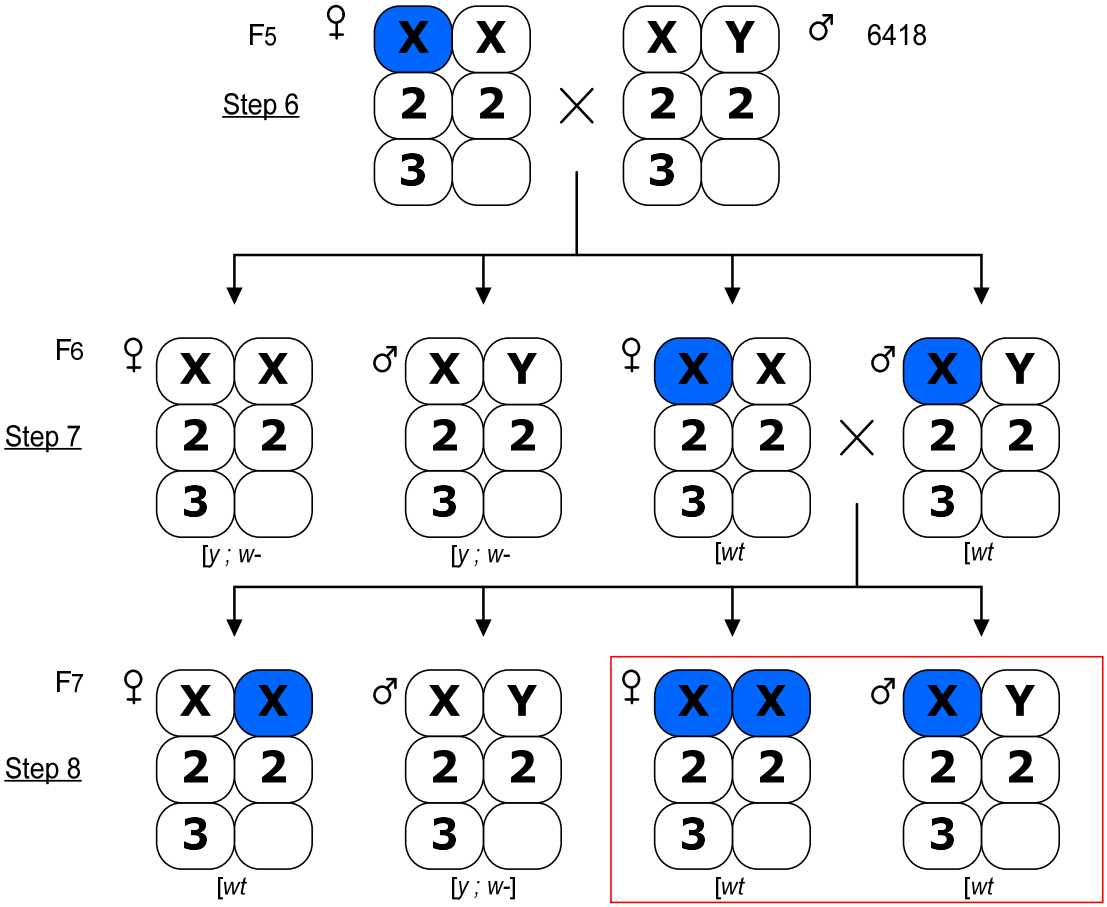


Figure B1: Construction of the substitution lines A* and E*. Both A* and E* were constructed according to the same crossing scheme: the X chromosome of E14 and A157, respectively, was introgressed into the 6418 genetic background. On the figure, wild origin chromosomes (WT) are shown in blue and 6418 chromosomes are in white. The line 6418 carries an FM7j balancer X chromosome which has two important properties: it is a very efficient suppressor of recombination and the homozygous females as well as hemizygous males are viable and fertile. That balancer chromosome can be traced thanks to two recessive mutations: w^{1118} and y^{93j} (respectively producing white eyes and yellowish body color; this phenotype is shown as [*y;w*-] on the figure). For steps 2 to 6, the wild type heterozygous X_{6418}/X_{WT} female offspring were 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 6418 balancer line. Each backcross divides by two the amount of wild-type genetic material. Statistically, at step 7 only 1.06% of the wild-type genetic material is left on the autosomes. Step 7 and 8 are brother-sister crosses of wild-type offspring. At step 8, a single brother-sister pair producing homogenous wild-type offspring was kept for establishing a line.

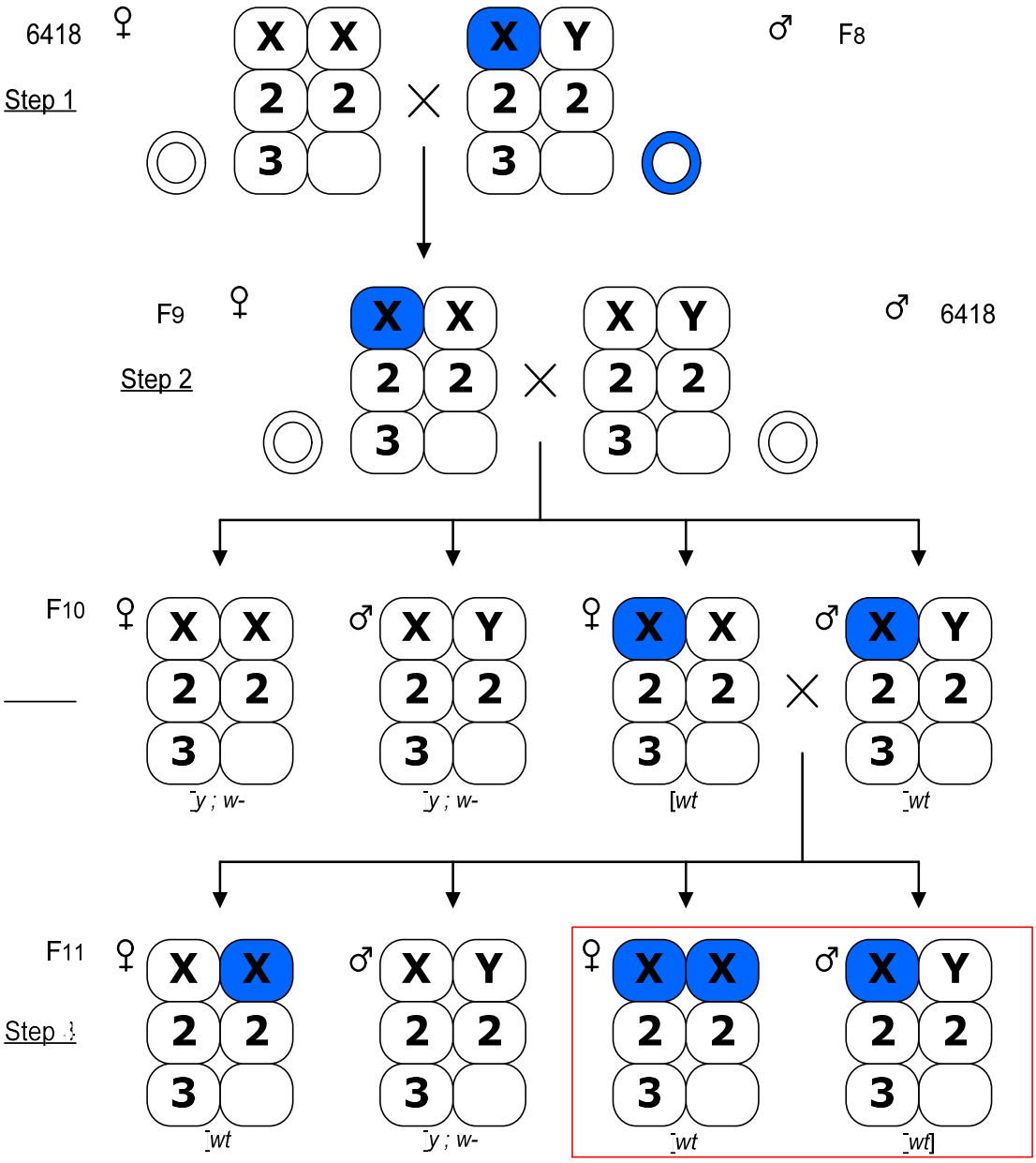


Figure B2: Homogenization of maternally inherited factors in A* and E*. The stabilized lines obtained previously derive from a wild-type female. The maternally inherited genetic material (like mitochondria) are thus different between A* and E*. The above crossing scheme homogenized the maternally inherited factors represented by a ring (blue for wild-type, white for 6418 origin). Step 2 produces individuals carrying maternally inherited factors of 6418 origin. Steps 3 and 4 are similar to steps 7 and 8 of Figure B1 and are carried out to establish a line. It is important to notice that this step added two more backcrosses to 6418 lowering the theoretical amount of wild type genetic material left on the autosomes to 0.26%.

Table B3: List of the molecular markers used for the QTL mapping. These marker polymorphisms were named according to the fragment in which they were detected in Ometto *et al.* (2005). The cytological position and absolute position in kb are given according to Flybase *D. melanogaster* genome release 5.4. The fixed polymorphisms that were used as makers are shown as A157/E14.

Marker name	Cytological position	Absolute position in kb	Polymorphism
025	3A2	2247.266	T/C
056	3F1	3629.942	G/A
067	4C14	4548.474	T/G
094	4F4	5126.621	C/A
109	5C7	5767.488	G/A
114	6C12	6605.700	A/T
126	7B2	7185.769	A/T
143	7E1	8118.991	C/T
167	8E1	9281.023	A/G
464	9B5	10104.098	T/C
210	10B8	11244.045	C/T
493	11A6	12015.826	TC/GG
250	11D1	12701.363	C/T
278	12A1	13318.463	ATGCCG/TTGCCC
326	12E2	13935.313	T/G
375	12F3	14561.277	TGG/GCT
386	13A12	14948.045	G/C
282	13C3	15282.518	C/T
296	13E15	15599.496	A/C
726	14B6	16096.635	A/T
364	15A1	16530.481	C/T
426	16B1	17353.968	T/C
444	17D3	18579.320	GCT/TCG

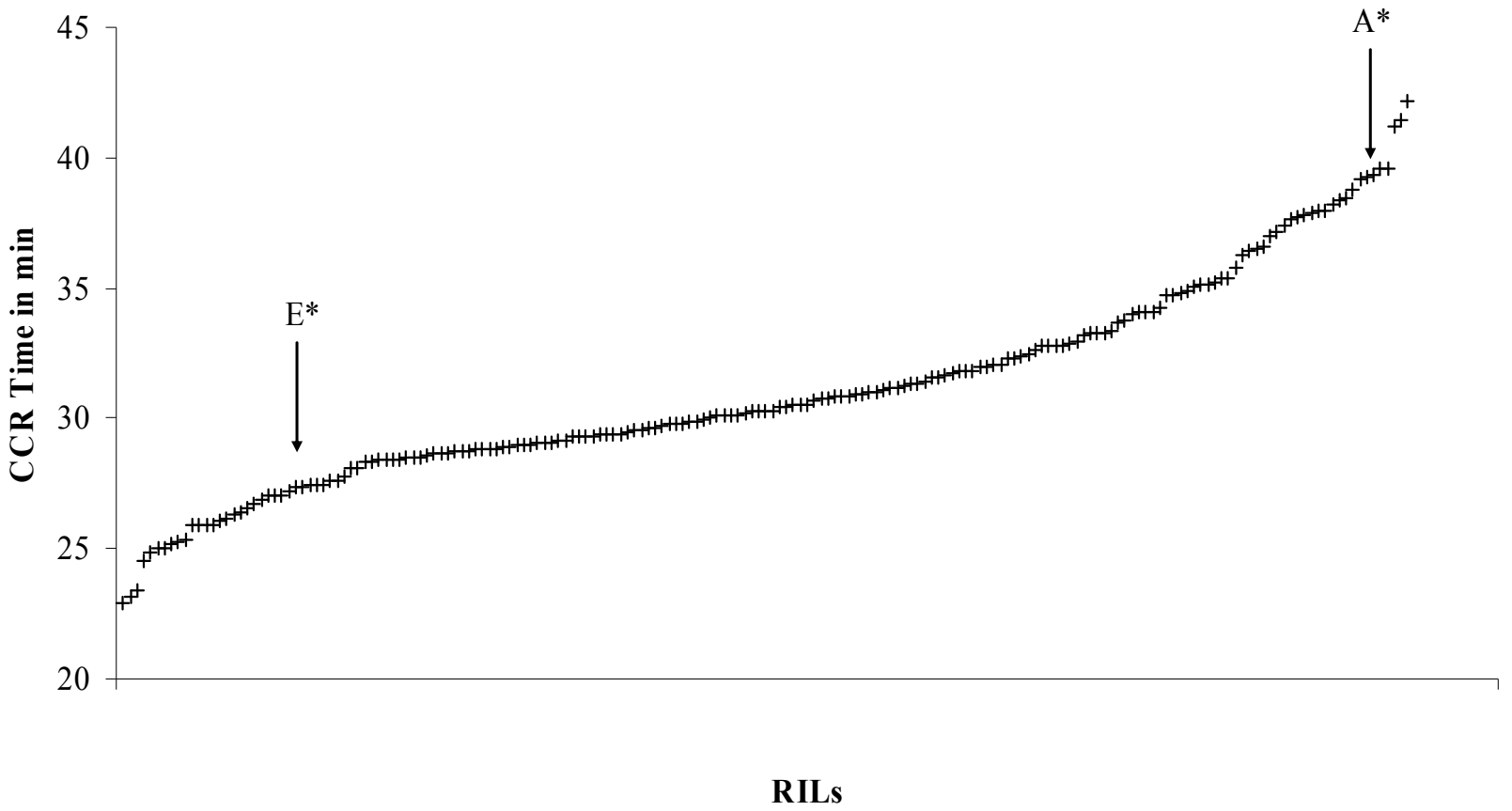


Figure B4: RILs sorted by increasing mean CCR times. Arrows indicate the mean CCR time of the parental lines.

13. Appendix C: Protocols

DNA Isolation From 10-15 *Drosophila melanogaster*

Expected Yield Range 5-20 µg DNA

Cell Lysis

1. Chill a 1.5 mL centrifuge tube containing 300 µL Cell Lysis Solution on ice. Please note that the solution will turn cloudy.
2. Add 10-15 flies (5-15 mg) to chilled Cell Lysis Solution, remove from ice, and homogenize thoroughly using a microfuge tube pestle. Place sample back on ice until next step.
3. Incubate lysate at 65°C for 15 minutes.

RNase Treatment

1. Add 1.5 µL RNase A Solution (4 mg/mL) to the cell lysate.
2. Mix the sample by inverting the tube 25 times and incubate at 37°C for 15 minutes.

Protein Precipitation

1. Cool sample to room temperature.
2. Add 100 µL Protein Precipitation Solution to the cell lysate.
3. Vortex vigorously at high speed for 20 seconds to mix the Protein Precipitation Solution uniformly with the cell lysate.
4. Centrifuge at 13,000-16,000x g for 3 minutes. The precipitated proteins and tissue particulates will form a tight pellet. If protein pellet is not tight, repeat step 3 followed by incubation on ice for 5 minutes, then repeat step 4.

DNA Precipitation

1. Pour the supernatant containing the DNA (leaving behind the precipitated protein pellet) into a clean 1.5 ml centrifuge tube containing 300 µL 100% Isopropanol (2-propanol).
2. Mix the sample by inverting gently 50 times.
3. Centrifuge at 13,000-16,000x g for 1 minute.
4. Pour off supernatant and drain tube on clean absorbent paper. Add 300 µL 70% Ethanol and invert tube several times to wash the DNA pellet.
5. Centrifuge at 13,000-16,000x g for 1 minute. Carefully pour off the ethanol. Pellet may be loose so pour slowly and watch pellet.
6. Invert and drain the tube on clean absorbent paper and allow to air dry 15 minutes.

DNA Hydration

1. Add 50 µL DNA Hydration Solution (50 µL will give a concentration of 200 µg/mL if the total yield is 10 µg DNA).

2. Allow DNA to rehydrate overnight at room temperature. Alternatively, heat at 65°C for 1 hour. Tap tube periodically to aid in dispersing the DNA.
3. If particulates are present in the rehydrated DNA sample, centrifuge at 13,000-16,000x g for 5-10 minutes and then transfer the supernatant containing the DNA to a clean tube.
4. Store DNA at 2-8°C.

PCR Reaction
For one reaction:
1. Mix 16.12 µL distilled water, 2.5µL Buffer10x, 1µL Mg 50mM, 0.25µL dNTP 100mM, 013µL Taq polymerase, 2µL forward primer solution and 2µL reverse primer solution.

Normal PCR Program

Cycle	Number of repetitions	Temperature in °C	Duration
1	x1	94	04:00
2	x33	94	00:30
		XX	00:35
		72	00:35
3	x1	72	04:00
4	x1	4	∞

2-Step PCR Program

Cycle	Number of repetitions	Temperature in °C	Duration
1	x1	94	04:00
2	x33	94	00:30
		60	00:30
3	x1	72	04:00
4	x1	4	∞

Grad PCR Program

Cycle	Number of repetitions	Temperature in °C	Time
1	x1	94	04:00
2	x33	94	00:30
		XX to XX	00:35
		72	00:30
3	x1	72	04:00
4	x1	4	∞

Exo sap
PCR reactions were cleaned using EXOSAP-IT (USB, Cleveland, USA)
1. Add 1 µL of ExoSAP-IT to 10 µL of PCR product.

Exosap Program

Cycle	Number of repetitions	Temperature in °C	Duration
1	x1	37	06:00
2	x1	80	15:00
3	x1	4	∞

Sequencing Reaction
Sequencing was performed separately for forward and reverse primers (both strands were sequenced), using the ABI BigDye Terminator v1.1 sequencing kit (Applied Biosystems, Buckinghamshire, UK):

1. Mix 1 µL 5x buffer with 2 µL of sequencing mix and 3 µL of distilled water.
2. Add 2 µL of forward OR reverse primer.
3. Add 2 µL of PCR product.

Seq Reaction Program

Cycle	Number of repetitions	Temperature in °C	Duration
1	x1	96	01:00
2	x40	96	00:10
		52	00:12
		60	04:00
3	x1	4	∞

4. After the PCR reaction, add 15 µL of distilled water prior to sequencing or store at –20°C.

Genotyping
Modified version of the Custom SNP Genotyping assay protocol:
(for one reaction):
1. Dilute SNP Genotyping assay solution (Primer) dilution by mixing 0.625µL TaqMan SNP genotyping assay 40x with 0.625µL 1xTE pH8.
2. Plate preparation:
Prepare 5µL master Mix TaqMan2x
Add 0.5µL 20x SNP Genotyping Mix (Primer)
Add 4.5µL DNA (ca. 5ng)
3. Inspect each well for volume uniformity, noting which wells do not contain the proper volume.
4. Cover the plate with MicroAmp Optical Adhesive Film
5. Centrifuge the plate briefly to spin down the contents and eliminate air bubbles from the solutions.

Genotyping Assay Program

Cycle	Number of repetitions	Temperature in °C	Duration
1	x1	95	10:00
2	x40	92	00:15
		60	01:00

14. Appendix D: PCR Primer list

FIN	Forward Primer sequence	Reverse Primer sequence	Ann. temp. in °C
X-01068	TCGGCTGACGTCTAATTG	AACCGGATGATTCACGAA	53.6
X-01069	TAGCAGCTGTTATCACA	GTGCCGTTATTTCGATAG	50.4
X-01070	ATCCGCATCATCGTCATC	CCAGCCAGGCATAACTCC	52.8
X-01071	CTTTGCCACGCTAATTCC	ATCATGGCCACTACATCC	50.3
X-01072	TTACCCGCCAAAGTATGC	TGTGGGTGGATGAGACGA	51.4
X-01073	GTCACGCCCGTGTAATCA	TCGGCTTCCAATTAGACC	51.7
X-01074	TCGCGGCGATATTGATAG	AAGCAGCGCCAATTACTG	53.7
X-01075	TGGCAGCGGAAAGGTAAC	TCCATCGCCTTGAGTAGC	52.3
X-01076	ACTGGCTGGTCATATTGC	ATCATGGCCACTACATCC	50.9
X-01077	TACGCCCAATTCGAGAAA	CCAAGGCCTTGTTAATGA	50.1
X-01078	TCGCCAGGGCATACAACA	TGGCCCGGACTGTACTTT	51
X-01079	AATGCCCAGAGAACCGAT	ACAATGCGCTCTAGTTCC	52.5
X-01080	GATCGGCTGGATAACGTA	CGCCAATGTTGTAGTTGT	51.1
X-01081	GTAATCCCGCAATCTCGT	GTGTGTGCGCATATCACC	50.9
X-01082	GGCGCGCCAATTAAGTGC	ACCCGAAGAAGCGTTGAAAA	53.8
X-01083	TGAAGCCCACCTTACACA	GATCTGCGGCTTACTTGC	51.8
X-01084	ACTGGCTGGTCATATTGC	ACCATGGCCACTACATCC	50.9
X-01085	GCTCCAGCGAACATCGTG	TCGCGCTTCTTTACCTATGC	56
X-01086	TAGCTGCGGTATTTACAATC	CTTTCGCGCACAGTATTCTA	51.7
X-01087	CTGCGAGAGAACGGATGA	GCGCCAGTCAATAACGTC	55
X-01088	GTACGCCACTGATCCAAC	TTGGCCATAAATCACGAG	52.3
X-01089	AGAGGGCGAATAGTTGGT	TTTGGGCGGTCAGTGTAA	52.3
X-01090	CGCTTGCCATATTATCCT	TATCCCGCCCACTAATTC	50.3
X-01091	AAGTGGCGGACATTGGTC	GGCGGGAAGAGAGTTGGT	54.2
X-01092	GACACCGCTTACTCGTAC	AAGGAGGCAACTAACACC	50.4
X-01093	CGTGCCACAAAGATACCT	CACGAACCCAAGTAGACA	50.9
X-01094	GAAGATCCTGTCGTCGGT	CAGTATGCTGCGTTTGGT	52.6
X-01095	TTGTTGGCTGGCGTCGTG	CGCGCTCCCTGTCTCGAT	55
X-01096	GGAACATGCGCGAAGACC	GGCCGCTCGTCTGCTATC	55.7
X-01097	CCCAAACGGATTTAACTC	GCCGACAAATTAGTGCAT	49.3
X-01098	CCAGCGGCAAGGACTATTC	AAGCAAATCACACGGCAGA	50.4
X-01099	CGGTGGCCAGTAATTCGT	CCTGACCTGTCGCACTTG	57.6
X-01100	CAATGTTGCCGTTGTCCA	CATCAAATCGCCGACTGA	58.1
X-01101	GGCAGCGAGAGAAGGAACGA	GACGGTGGGCTCTTCAACGA	60.5
X-01102	CAAGCGCTTCGACGATGA	CAGTTCCGGCACCTCGAC	58.5
X-01103	ACACGGCCGTTAATCTGG	ATGTCCTTGCCGATCTGC	56.6
X-01104	GCAGTGGCAAGACGTTGG	TTGGCATCGTCGTTGAGA	58.1
X-01105	GCAAATGGAGGCGTTAGC	CATGGCGCGGATAATGTC	57.4
X-01106	TGTGCCCAACCACTACGA	CCAGTGCCGGATTAGCTC	58.8
X-01107	CAAGCGCACGGTTAAGGA	TGCTGCGGGAAATCGTTA	60.2
X-01108	GGGCGGTGCTCTAAGCTCT	CCTTGATTAGCCGCGTGAT	59.2
X-01109	AAGGCCAAGCGCGAGATCA	CTCGAGTGCTGGCGGAGAT	55.1
X-01110	TTCTGTTCAAACGGCTGA	AACTCAAGCGCGAAATTC	49.7
X-01111	TTGTGCGATATCTGACGTTTA	TTTCACACATAACCCAATCAT	54.7
X-01112	CGTAAGCCGGCAGCGTAA	ATCCCGGATCCGTGATGT	55.3
X-01113	AAGGGCAATAGGTTGTGG	TTTCCGCTCGAGTGATAA	56.5
X-01114	GCATAGCGCGCAGTCAAC	ATATGCACGCCGCTGATT	56
X-01115	GGGCTTCAAGTTCGACGAG	TTGCTGCGGGAAATCGTTA	59.9
X-01116	AGGGCATTCGCATTGACC	TTCGTCTGCCTCGGCTCT	56
X-01117	GGAGGAGGTTTGGCGATG	CACCCAAGAGCGCAACAT	57.3

FIN	Forward Primer sequence	Reverse Primer sequence	Ann. temp. in °C
X-01118	GCCAGCAATTAGTTACCA	AAGTGAAGGCTCGATTTT	49.7
X-01119	TGCTTGCCAACTCGAAAT	ACGGCGATTGACATAGGA	52.5
X-01120	ATCCGCTCAAGACGTGTC	GTTCACATGCAGGCGATA	51
X-01121	TGGAGGCCCTAGTGAATTA	GGTTCTCAAATACGGCAAT	52.6
X-01122	CGGAGCTGGCGAAATCTG	TGCGGGCCTACGTGAATC	56.2
X-01123	GCTCGCAAATCGCCATGT	TGCGCGACTTGCCAATAA	56.9
X-01124	TTGCCAGATTCCATACGC	AGTTCCGCACACTTCGAT	50.6
X-01125	GTTCCGTTATCAATGGTATG	AAATGTTGCCAAACTAAACT	46.3
X-01126	GCTATGCCGGCGACATTAT	CCTTGATTAGCCGCGTGAT	60.2
X-01127	TTGTTGGCTGGCGTCGTG	GGCCGCTCGTCTGCTATC	55.8
X-01128	AACAAGCGAAATGATAAGAA	ACTTGTAATTCCCGTAATCC	51
X-01129	AAGCGCACGGTTAAGGATA	GGTAAACGGGTTGGTTGTG	59.8
X-01130	GGATTGGCTGGCGATACCC	ATGCCCACCATGCCCATAC	54.4
X-01131	CGGAGGCTCAGCGGACTG	GCGCAGGGTGGTTCCGTA	58.6
X-01132	TTCAAGGCGCACGCATTT	TGAAGCCGCGCTAATGAC	57.3
X-01133	TTTGCGGGCCATCGAGAG	ACCAGACGGCGTTTGTGC	54.7
X-01134	AAGGGCTTCAAGTTCGAC	CTGCGGGAAATCGTTAAT	58.2
X-01135	CCCTGCGCTCACACGAGT	TGGGCCATCAACCGAAAG	53.8
X-01136	TGAAAGCGGACCTAAGCA	ATAGGCCGCTGTACGATG	52.9
X-01137	TCTAGCCCGCTGTATCTG	TCTCCGTACCGAATATGC	53.9
X-01138	TCCCTTTCGTGCGAGTGT	GCGAGATGGCGGTAGAGC	51.1
X-01139	GCAGTCGGGCGGACATCT	TAACAACGCGCGACCAGA	52.8
X-01140	TTTGGCGGCTGATAGAAT	CCCTTCCAAATCTCGTCA	50.5
X-01141	TGTATCTTCGGGTTTTAACCA	GTTTTTCCGCCTCTAACTTTG	55.2
X-01142	CATCCCAGCCAAATCGAC	TTCATGGACTGCCGACCC	56.2
X-01143	CAGCAGGCCCAAGTCGGT	GCCGCGTCGCCTACAAGA	57.8
X-01144	ACCAATGGACCGCTAAGT	CATAGTGGGCACCTACGA	57.1
X-01145	CCCGCACCCACAACTGAT	TGCCGCTGGAACACGTTT	57.4
X-01146	ATTTCCCGGCTAGTGTGG	AGATGCTTGCCGTTCATA	55
X-01147	CAGTGGCATGGTCACGAG	GCTTGGGCTACGGTCAAA	57.2
X-01148	GTGTGCGGGAATTATCGT	CTGGGCGTTCCGAAAGTA	56.3
X-01149	GGGCTGGATAGGACAATC	AAAGCGCGAATGTAGAGT	52.2
X-01150	ACGGTTCGGCGAAGTGAT	TGCCAGAAACTCAACCCA	56.2
X-01151	TCAACCAGGATGCGAACT	AGGGCAGGCACAACGATA	56.6
X-01152	TTGTGGGCATACGTTTCC	GCTTAAATGGGCCGAATG	58.1
X-01153	AGTGGCCACCGCATCTAGTT	AGTTCGGCATTATCGCTGTA	57.4
X-01154	GGAGATCGTCGAGGACTC	ACAAATGGCCAGCGATGT	52.3
X-01155	TTCGGCAGCTTCGTTTGT	GGCCGAGGATCACGATGA	54.9
X-01156	TTATGGATTGCCTCATCGTGA	CGCCAGTTCTTCGGGTTAAAT	52.7
X-01157	ATTGCTTGCGTAATTGGT	ACTTTCCCGACACGTTTA	51
X-01158	TCTAGCCCGCTGTATCTG	TCCGTACCGAATATGCTC	53.9
X-01159	ATTCGCACGTCCGTCTGT	CGCTGGCGTCTTACACAA	56.5
X-01160	ATTTCCTGTTCGGCTTAGTTC	TTTAGCGTTTCGGTCTAGGAT	51.3
X-01161	CGCCAGCGTCAAGTTCGT	GCTCCCGCTCTAGGTGCC	55.7
X-01162	AGCGTCGCTCGAGTGTCC	AAATGGCCAGCGATGTGT	55.4
X-01163	GCCAAGCGCACGGTTAAG	TTGCTGCGGGAAATCGTT	60.4
X-01176	GCAAATCGCACAAGTCGG	CGATCGGCGCACTATGAG	54.7
X-01177	GAGGTCGCAATTGTAGCAC	CGTTTCCTCTCCCGATTTC	50.7
X-01178	GTCCGCTTGGTGTTCGTT	GGTCACCGCCATCTACCT	53.4
X-01179	TAGATGGCGGTGACCTTA	ATGGGCAAACACGTTAAT	52
X-01180	TGCGCCTTTAAATATCAGAC	TAAATAGCGCGGTTATCAAC	49.2
X-01181	CAGCAGCGGCGCATCAGC	AATCGCCAGCCGCAGTCA	57.5
X-01182	TCCGTTCTGGCGTACACA	CCAACCACCCGAGGACTA	57.4
X-01183	GAGACGCCTGCTATGCTG	AACGGGAGGGCATCTACC	57.8
X-01184	CTTGCCATTGCCGAACTC	TGGAGGGCACAGATTGGT	57

FIN	Forward Primer sequence	Reverse Primer sequence	Ann. temp. in °C
X-01185	GCGCCGGAATTCGTAGAA	CGGAGCGGAGTGCAATGT	56.4
X-01186	TTGGCCCTCGACTAACAC	ACCAAGCGAAATCCGTAT	54.3
X-01187	TTCTTTGGGTGCCTATTT	GTCGCATTTCCATACATT	51
X-01188	CACGGCCAAATAGCGACT	TGCGGGCAATTGATTGAA	55.5
X-01189	CCACCCAAGGTCGATAAT	TTTGCGACTTCCCTTATG	51.1
X-01190	AGTGCCACAGTGTGTTCGTGT	TCTAGCGATCTAGGCGAAAGA	51.4
X-01191	CATTCGGTGCGGCTGTAAGAG	TCAGCGGACAACTGTTTACCA	55.9
X-01192	TTCGGATCGGTTTAGCTTT	TTACGGCGATTGACATAGG	52.4
X-01193	ACGCGATATTTGGATACAGG	GTAGATCAGTGCCGTTGGTT	56.2
X-01194	TCTTGGAGGCGAAACGTC	GCAATGGGTCGCATCTAA	56.2
X-01195	CCGCAGTCCACATTCGAG	CAGGGAGCGCAGGGTTAG	59.7
X-01196	GACGCCGGCCTGCTATCC	ACATGGGCGCGAGTGTGG	62.2
X-01197	TGCACGACGACTACGGTC	ACCAGCTGCGGATTGAAC	59.1
X-01198	CACTTACGGGCCCAACTCAT	GTTCTCCGCCAAGTAGTCGC	60.4
X-01199	CACTGGCTGTCGGCTTTG	CGCCTCCTCCGGACGTAG	60.2
X-01200	CGTGTATCCGCTCAAGACGTG	ACTCCGCTCCGCTGTCTAGTG	53.7
X-01201	TCAGTATCCTTGCGTTTA	CAGAGCGCGCACAAGTAG	54.2
X-01202	GCCTGCGGTCGCTATGTG	TGTGGGTGCGGGCTAGAG	60.6
X-01203	AGCGCATTTACCTTAACC	AGAAACCAAAGGCGAAAG	52.8
X-01263	GCTGGAGCGCATTTACCTTAAC	TTCTCCGCCTCTAACTTCGAAT	57
X-01264	AGCCCGCACCCACAACTGAT	GGACTGCCGACCCACCGACT	57.6
X-01265	CATCCCAGCCAAATCGACAGTT	GGAGAAACCAAAGGCGAAAGTG	56.7
X-01266	GTCGCTGTCGGAGAAGTCGTC	TCGAGAACGCCCATGTTATGC	57.8
X-01267	CCAGGCCAGTTAAACACTCGTT	GCCGGAAATGGAGATCATAGAG	58.9
X-01268	GTTGGTGGCCTCTATGATCTC	AACCCGAAGATACATTGCTAA	49.7
X-01269	CCTCGTTGGTGGCCTCTATGATCTC	CGCATGGATTTCAATTCAGTAAGGA	52.4
X-01271	AAATCCATGCGAAAATTACGTT	AAAGGCGAAAGTGGAGTCTGAC	53.4
X-01272	TCATCCCAGCCAAATCGACAG	CACACGAGGCAAGGGTCAAGG	57.6
X-01273	CGCGTTTCCATGTCGATCTT	TCCAAGCAGGGCGATCACTT	56.4
X-01274	CCAGGCCAGTTAAACACTCGT	CGCCGGAAATGGAGATCATAG	58.9
X-01275	CTCGTAGGTGCCCACTATGA	TTTCGCTGGTCCAGTAATTG	51.5
X-01276	ATCCATGCGAAAATTACGTT	TTTCGCTGGTCCAGTAATTG	54.6
X-01277	GGCGACCAAAGGATGCTAATCT	CATTTGGCAATTTGGTAGCAGT	50.7
X-01278	CCGGATTTGGCCAGTAAGTTGA	AGATGGCGGTAGAGCGTAAGGG	51.8
X-01279	GCTTCCGCGAATATAGGTGC	TGAGCGGGAATCAATTAGCC	52.8
X-01280	CTCTGCCGCGTCACGATA	TATTCCCAACCCGCTCCC	53.6
X-01285	CTCTGCGCTTCTCAATCTAT	ATTTGCCAGATCTTAACGAA	51
X-01286	CTGATCCATGACGGCACATC	CTTGCTCGCTGTGAGTACGC	54
X-01287	AGAGCGCGACAGTAATCAGT	TCTCTCGTCGGCTCTAGAAG	50.7
X-01289	AACCCAAACCGATTAGAA	ACGGCAGGGTATAGCTTA	51.6
X-01290	ATAATTGCCAGCACACGACT	CACACTGGCAAGGGTACATC	51.6
X-01291	CCGCAGGATTGTGTACAG	GAATCGGAATCGGAAAGC	55.6

15. Curriculum Vitae

Nicolas SVETEC

Haslangstr. 30
D-80689 Munich
(Germany)
E-Mail: svetec@bio.lmu.de

Date of birth: 26th of December 1980
Place of birth: Colmar (France)
Nationality: French
Clean Driving Licence

Education

2005-present**: PhD in Evolutionary Biology** under the supervision of Prof. W. Stephan in Munich (Germany).

2003: Master's degree 2 in Ecophysiology, Energetics and Behaviour in Strasbourg (France).

2002: Master's degree 1 in Ecosystem's and Population's Biology in Dijon, (France) with the following optional lectures:

- Evolutionary Biology (Dr. P. Alibert)
- Animal Behaviour (Prof. F. Cézilly)
- Entomology (Prof. B. Frochot, Dr. M.J. Perrot-Minot).

2001: Bachelor's degree in Biology, in Grenoble (France).

1999-2000: University diploma in Biology in Grenoble (France).

Publications

SVETEC, N., and J.-F. FERVEUR. (2005)
Experience and pheromones can change male-male interactions in *Drosophila melanogaster*.
Journal of Experimental Biology 208: 891-898.

SVETEC, N., B. HOUOT, and J.-F. FERVEUR. (2005)
Effect of genes, social experience and their interaction on courtship behaviour of a transgenic *Drosophila* male.
Genetical Research 85: 1-11.

SVETEC, N., M. COBB, and J.-F. FERVEUR. (2005)
Chemical stimuli induce courtship dominance in *Drosophila*.
Current Biology 15: R790-2.

SVETEC, N., P. PAVLIDIS and W. STEPHAN. (2009)
Recent strong positive selection on *Drosophila melanogaster HDAC6*, a gene encoding a stress surveillance factor, as revealed by population genomic analysis. (in press)

HOUOT, B., N. SVETEC, and J.-F. FERVEUR. (2009)
Sexual characters vary between wild-type *Drosophila* populations before and after laboratory acclimation. **(**submitted)

International Conferences

NEUROFLY (2002) in Dijon (France)

Workshop: "Evolution across scales" (2008) in Potsdam (Germany)
Poster presentation.

XX International Congress of Genetics (2008) in Berlin (Germany)
Poster presentation.

Additional Information

- Good skills in written and oral English. German (adequate).

- Computing \ Office automation: good skills in Word, Excel, Photoshop, Lasergene package, QTLcartographer, Statview and SPSS. Notions in Arcview (G.I.S.)

- Hobbies: experienced tropical water aquariologist, hiking, skiing.

16. Acknowledgements

First I would like to thank Prof. Wolfgang Stephan for giving me the opportunity to conduct my PhD here in Munich and providing everything I needed to accomplish it. He contributed largely to improve my scientific skills and writing, and gave me efficient personal encouragements when needed. I also would like to deeply thank Pavlos Pavlidis for very fruitful collaborative work in a very nice atmosphere. Similarly, the QTL mapping could not have been performed without the theoretical input and advices of José Alvarez-Castro. Concerning the QTL analysis, I also thank T.F.C. Mackay for data sharing. I owe many thanks to David de Lorenzo for his large help to my adaptation to the German/Munich environment, for the initiation of my thesis projects and, of course, for the great atmosphere he brought in the department.

I would like to address a special thank to Anne Wilken for wonderful technical help and for teaching me everything about the DNA PCR and sequencing. Similarly I thank Simone Lange, Anica Vrljic, Susanne Voigt, Ricardo Wilches for their technical help and contribution to lab work and data collection.

From the Department of Evolutionary Biology I would like to thank Sasha Glinka, Stephen Beisswanger, Dirk Metzler, John Baines, John Parsch, and Joachim Hermisson, for very interesting scientific discussions and suggestions. Especially, I must thank Lino Ometto, Pleuni Pennings (special thanks for manuscript corrections), Stephan Hutter, Aurélien Tellier and mostly Stefan Laurent for the wonderful scientific input and brainstorming done in a very relaxed and friendly atmosphere. More generally, I would like thank everybody from the department for the nice working atmosphere namely Lisha Laduvilezhath (for all the entertaining discussions), Iris Fischer (the active member of the Witch-Club), Winfried Hense (for organization of lab

events and his sympathy), Sonja Grath (a noticeable great person), Anja Hörger (for the pleasant time spent together), Claus Kemkemer (for the funny flyroom discussions), Robert Piskol (the guy from the RNA world), Létizia Camus-Kulandaivelu (the authentic), Carlos Merino (the late guy), Katrin Kümpfbeck (for administrative help and support), Hilde Lainer (sorry for missing you on the motorway), Annegret Werzner (die West-Polisch Lara Croft), Sarah Peter (the Carribian girl), and Laura Rose (the defender of Plants' World). From working perspectives, I also have to thank all the hundred thousand flies that gave their life to science, Coca Cola for providing caffeine supply and ACDC for playing the greatest labwork music…

I address very special thanks to the Munich friends: Myriam, René, Benou, Marie-la-Vosgienne, l'Arno lyonnais, and Aurelaï, who encouraged me and were of great support/help in the hardest moments but also all the others for the Bierfests, Biergarten, barbecues and game evenings: Poupou, Mite, Christophe&Sarah, Phillippe, l'Edouard and Magali. The other French friends who, despite distance, are always there when needed: Dr. JB, Ze Grenoble Dream Team (namely le Ben -alias Tankeld-, Roolyo, Vince, Squale, Gaëlle, and Sam), Cathy, Flo, Ze Kiki (the best guy ever), Jerem, Fred&Céline (les inseparables), Benj and the Nice RP gamers (Méllu, Orca, Ake, Arma, Rhya, Liggur, Noso, Galaa and all the others…)

Finally, I address lots of love to my whole family who gave large moral and -a bit of- financial support (AM, JL, AC, Piti, Alex&Co, SalutAS, T. Tounet, T. Lolo and Mamie)