Universidade da Madeira

A multilocus comparative study of nucleotide variation in *Drosophila madeirensis* and the close relative *D. subobscura*: Insights into the speciation process of *D. madeirensis.*

Clévio David Rodrigues Nóbrega

2006

A multilocus comparative study of nucleotide variation in *Drosophila madeirensis* and the close relative *D. subobscura*: Insights into the speciation process of *D. madeirensis.*

Clévio David Rodrigues Nóbrega

Tese submetida de acordo com o regulamento em vigor na Universidade da Madeira para obtenção do grau de Doutor em Ciências Biológicas, especialidade Citogenética e Biologia Molecular.

Universidade da Madeira
2006

A multilocus comparative study of nucleotide variation in *Drosophila madeirensis* and the close relative *D. subobscura*: Insights into the speciation process of *D. madeirensis.*

Clévio David Rodrigues Nóbrega

Tese supervisada pela Dra Carmen Segarra da Universitat de Barcelona e pela Dra Mahnaz Khadem da Universidade da Madeira.

Universidade da Madeira
2006

But more, much more than this,
I did it my way...

To my parents, my sister and my brother...

Thanks to...

Carmen and Nazy, my two thesis directors

for the support, direction... and mainly for the friendship and comprehension. I will be always indebt to both...

Barcelona's crew, all the people in EVO group and others

for the oportunity and pleasure of working with you. Thanks to all, because in a given moment you were important to me...

Alex Perez, my good friend

for all... in the bad days and in the good moments...

Gema, my coffee and laboratory partner

for the daily smile and laugh that helped me everyday in these last 4 years...

Inês, "portuga" like me

for the cumplicity, the daily foods and the Portugal matches...

Anna, another laboratory partner

for working beside me, for the daily cofee and talk...

My phd partners in the lab (Eva, Eva P., Dani, Ursula)

for the talks, the foods, the cofees and more...

Madeira's crew, all people that helped in these 4 years

for being there, even when I was far away...

Ana, my carrer friend

for all the help, tips, friendship, advices and for being there...

Roberto Camacho, my lab partner in Madeira

for all the inestimable experimental help...

Aletta Camacho, my english revisor

for the english correction, the tips and the patience...

To all my friends around the world

for being there, for helping me... and mainly for being my friends...

To my close family: parents, sister, brother, nephew, brother and sister-in-law

for being my life support and my daily pillars, for filling my life with joy and happiness... This is also their merit...

A multilocus comparative study of nucleotide variation in *Drosophila madeirensis* and the close relative *D. subobscura*: Insights into the speciation process of *D. madeirensis.*

Experimental work done in the Biology Department, Universidade da Madeira and in MolEvol group, Genetics Department, Biology Faculty, Universitat de Barcelona.

Thesis submitted in Universidade da Madeira.

Work supported by grant SFRH/BD/6583/2001 to Clévio Nóbrega from Fundação para a Ciência e Tecnologia, Ministério da Ciência e do Ensino Superior, República Portuguesa.

The grant SFRH/BD/6583/2001 was co-financed by the POCI2010 and the FSE from the European Union.

CONTENTS

A multilocus comparative study of nucleotide variation in *Drosophila madeirensis* and the close relative *D. subobscura*: Insights into the speciation process of *D. madeirensis.*

1. PREFACE

1.1. Nota introdutória

Esta tese de doutoramento está redactada em inglês. Assim, e de acordo com o ponto terceiro do 15º artigo do regulamento de doutoramentos da Universidade da Madeira, apresento em seguida um resumo da dissertação em português.

1.2. Resumo

O estudo da divergência entre espécies relacionadas, utilizando dados de vários loci é uma das formas de estudar a especiação e permite distinguir as forças que actuam em todo o genoma daquelas, como por exemplo a selecção natural, que afectam apenas os loci individualmente. Assim sendo, um dos grandes objectivos deste trabalho foi estudar a história da divergência entre duas espécies relacionadas, *Drosophila madeirensis* e *D. subobscura*, utilizando dados de sequências de DNA. O polimorfismo das inversões cromossómicas é uma característica comum do genoma do género *Drosophila*, no qual cerca de 60% das espécies são polimórficas para as inversões pericêntricas em populações naturais. O outro grande objectivo deste trabalho foi estudar o nível de variação nucleotídica em duas diferentes ordenações do cromossoma X de populações de *D. subobscura* (A_2 e A_{st}). Assim, a análise da variação nucleotídica ao longo da inversão pode indicar o papel desempenhado pela selecção natural na formação e manutenção dos polimorfismos cromossómicos. Para atingir estes dois principais objectivos foram estudadas cinco regiões (que se distribuíam ao longo da inversão A_2) do cromossoma X. De uma forma geral, o nível de polimorfismo nucleotídico encontrado foi similar nas duas espécies, apesar de *D. madeirensis* ser uma espécie endémica da ilha da Madeira e as populações de *D. subobscura* estudadas serem também insulares, e por tal se espera que possuam um tamanho populacional efectivo pequeno. Estimou-se o tempo de divergência entre *D. subobscura* e *D. madeirensis* em 640.000 e 1.400.000 anos e o modelo de isolamento sem fluxo génico após a divergência como causa de especiação não foi rejeitado. Em relação ao estudo das ordenações cromossómicas de *D. subobscura*, verificou-se que existe uma grande diferenciação genética em todas as regiões estudadas entre as duas ordenações. Os dados apontam para a existência de fenómenos de conversão génica e pouca evidência de troca de material genético na parte central da inversão.

1.3. Abstract

An approach to analyze the genetic changes that occur in populations during speciation is to examine the level and pattern of DNA sequence variation in species that recently shared a common ancestor. This approach becomes even more informative when data from multiple loci are available and thus they can be used to distinguish forces that act on all genes from those, like natural selection or gene flow, that affect individual loci. Therefore, one of the main goals of this study was shed some light into speciation divergence process between *D. subobscura* and *D. madeirensis*, using data from five gene regions of X chromosome. Chromosomal inversion polymorphism is a common feature of the genome in the *Drosophila* genus. About 60% of the *Drosophila* species are polymorphic for paracentric inversions in natural populations. *D. subobscura* is a species with a very rich inversion polymorphism, thus, another objective of this study was to analyze the levels of nucleotide variation in two different chromosomal arrangements of this species: A_2 and A_{st}. The levels of nucleotide polymorphism found were very similar between the two species, although *D. madeirensis* is expected to have a much lower effective population size. A divergence process in allopatry without gene flow could be a likely explanation for the speciation between *D. madeirensis* and *D. subobscura*, as the isolation model without migration was not rejected. A strong genetic differentiation in all studied regions was detected between the two chromosomal arrangements of *D. subobscura*, and evidence of gene conversion was also detected. These data, with those previously reported for the O_3 inversion, if extended to other inversions of *D. subobscura* might indicate that the genome of this species is highly structured.

2. INTRODUCTION

2.1. Species and speciation

There are two main approaches to study the genetic basis of species divergence. The classical approach is to map genetically the loci controlling traits that are important in speciation. The other approach is to explore the genetic changes that occur in populations during the speciation process. This approach focus on the study of the level and pattern of DNA sequence variation in species that recently shared a common ancestor (divergence population genetics). Levels of divergence, and mainly of nucleotide variation shared between species, can be interpreted in the light of different speciation models, particulary when the data from more than one region of the genome are available.

Most genetic traits studied according to the classical approach, can be included in three different classes. First, traits for which species exhibit differences that probably represent species-specific adaptations. Major lifestyle or life history adaptations can, *a priori* , play a direct role in speciation, mainly if these changes are first polymorphic within the ancestral species (Rice and Hostert 1993; Jones 1998). A second class of speciation traits are those related to mating. In the 90's decade, a host of interesting *Drosophila* mating phenotypes have came under focus, including species-specific mate detection pheromones (Coyne *et al.* 1994; Coyne and Charlesworth 1997), sperm competition (Snook *et al.* 1994; Price *et al.* 1999), and female mediation of sperm competition (Price 1997). Finally, the third class of speciation traits are those that appear in interspecific hybrids, such as hybrid inviability and hybrid sterility. In general, these traits can be genetically studied only when postzygotic isolation between species is incomplete and thus, it is possible to recover progeny at least in backcrosses.

2.1.1. Mapping of speciation traits

Speciation is known to involve genes that prevent individuals from mating, or, if they do mate, from producing viable and fertile offspring. These genes contribute for instance, to mate discrimination, sucess in fertilization, adaptation to particular environments, and physiological defects of the hybrid when two species interbreed. However, the identity of such genes, their normal functions and the forces that have shaped their evolution are largely unknown. There are two major difficulties in tracking them down. First, species that fail to produce viable and fertile hybrids cannot be studied through genetic crosses; and second, mate discrimination and hybrid sterility often result from the action of many genes, sometimes interacting through complex networks. Therefore, the identification of the genes that cause speciation has proved difficult, even in species with complete sequenced genome. Nevertheless, unraveling the genetic basis of reproductive isolation is crucial for

understanding speciation. Species are recognized on the basis of differences that have evolved between them. Differences that limit gene flow are particulary pertinent to the speciation process because they promote the subdivision of the species in different groups, and subsequently allow them to continue to evolve independently. A variety of characters can contribute to reproductive isolation, such as hybrid sterility and hybrid inviability.

Most studies that foused on the analysis of interspecific crosses tried to map genes involved in interspecific hybrid male sterility (Dobzhansky 1936; Coyne 1984, 1985; Vigneaut and Zouros 1986; Orr 1987; Coyne and Charlesworth 1989; Khadem and Krimbas 1991a; Perez *et al.* 1993; Cabot *et al.* 1994; Noor *et al.* 2001a); but studies on female hybrid sterility (Orr 1987, 1989) and hybrid male inviability (Orr *et al.* 1997) have been also made. These studies aimed primarily to know the number and the identity of the loci that, in some way, contributed to speciation (Wu and Palopoli 1994; Coyne and Orr 1998; Ting *et al.* 1998; Presgraves *et al.* 2003). In the *subobscura* cluster, these kind of studies have been made between *Drosophila subobscura* and *D. madeirensis* (Khadem and Krimbas 1991a, 1991b, 1993, 1997; Papaceit *et al.* 1991), and analyzed hybrid male sterility, viability, and abnormal characters of hybrids.

The gene mapping studies brings some light on the genetic architecture of the phenotypes that may have been important in speciation. However, these genes bear no direct connection to the demographic factors that have originated species, and they may have not been involved in speciation. Some of the speciation phenotypes may arise during or following speciation, which is primarily caused by selection on other phenotypes. Thus, for example, hybrid sterility and inviability may arise as epistatic by-products of independent adaptations of the separate incipient species (Dobzhansky 1936; Muller 1940). In contrast to the gene mapping approach,

divergence studies can focus directly on the evolutionary forces, particulary the demographic factors, that affect all genes in the genome.

2.1.2. Divergence population genetics

A different genetic approach to understand speciation is to study the history of species divergence as it is revealed in the nucleotide polymorphism pattern at randomly selected genes. Studies on gene flow via interspecific hybridization can be extremely valuable for understanding the role of natural selection during the formation of new species and for identifying genomic regions involved in reproductive isolation. When reproductive isolation is not complete, genes can be transfered between species. Therefore, incipient or hybridizing species can exchange genes and thus, they can share genetic variation. Gene flow between incipient species is a component of the divergence-with-gene-flow models of speciation (Maynard Smith 1966; Endler 1977; Felsenstein 1981; Rice and Hostert 1993). According to these models, incipient or hybridizing species can become divergent over some regions of the genome at the same time that they share variation at others (Wang *et al.* 1997). This is because some regions of the genome may introgress more readily than others (Clarke *et al.* 1996; della Torre *et al.* 1997; Wang *et al.* 1997; Rieseberg *et al.* 1999; Jiang *et al.* 2000; Noor *et al.* 2001b). Natural selection is expected to preclude gene flow at regions of the genome associated with species-specific adaptations. Thus, natural selection can maintain interspecific divergence at some genes, in spite of persistent gene flow at other genes. Under divergence-with-gene-flow models, natural selection has a direct role in generating and strengthening barriers to prevent gene flow and therefore, a direct role in generating species. The role of natural selection in these models differs from that in the classical genetic model of speciation (Dobzhansky 1937; Muller 1940), in which natural selection plays an indirect role in speciation. In this model, reproductive isolation is simply the result of

incompatibilities between gene variants that have arisen independently in each species and that are deleterious in a different genetic background.

Recently, speciation studies have taken advantage of several population genetic and phylogenetic methods developed to analyze multilocus DNA sequence data (Bernardi *et al.* 1993; Hey and Kliman 1993; Burton and Lee 1994; Hey 1994; Hilton and Hey 1997; Wang *et al.* 1997; Hare and Avise 1998; Kliman *et al.* 2000; Machado *et al.* 2002; Hey and Nielsen 2004; Ramos-Onsinis *et al.* 2004; Stadler *et al.* 2005). The methods are a direct extension to the species level of basic population genetics questions (*i.e.*, questions about population subdivision, gene flow, and natural selection). However, the use of DNA sequence data also permits to apply genealogical coalescent models, which incorporate classical population genetics parameters (such as population size and migration rate) within a gene tree framework (Hudson 1990).

These methods are even more informative when data from multiple loci are available and thus, they can distinguish forces that act on all genes from those that affect individual loci (Hudson *et al.* 1987; Hey 1994). Therefore, it is possible to detect whether different regions of the genome of incipient species have undergone to a differential gene flow. The divergence population genetics approach is thus a powerful tool to assess the importance of gene flow and natural selection during species divergence.

2.2. The inversion polymorphism

The presence of inverted chromosome segments as an usual occuring genetic change in *Drosophila* was first postulated by Sturtevant (1917). Inversion

polymorphism occurs when a large inverted region of a chromosome segregates together with its non-inverted (called standard) counterpart within populations. Inversions inhibit recombination between standard and inverted chromosomes because single crossover events within the inverted segment produce inviable and unbalanced gametes. Thus, inversions reduce recombination dramatically in some regions of the genome, although they can also increase recombination in other regions through an unknown mechanism (Krimbas and Powell 1992).

Inversions are a privileged system either to study diverse evolutionary topics such as phylogenies, geographical clines, temporal cycles and meiotic drive, or to look for evidence of natural selection (reviewed in Krimbas and Powell 1992). The first studies on the role of natural selection in the maintenance of genetic polymorphisms were performed on polymorphic inversions, because inversions could be detected by means of simple cytological techniques and their changes in frequency through generations could be easily followed (Dobzhansky 1970; Lewontin 1981). The development of the electrophoresis technique and the onset of the allozymes era was followed by an intense search for linkage disequilibrium between allozyme loci associated with inversions, and between allozyme loci and the inversions themselves. Indeed, it was expected that epistatic selection generated these desiquilibria (Prakash and Lewontin 1968; Zapata and Álvarez 1987, 1992, 1993; Krimbas and Powell 1992; Schaeffer *et al.* 2003). With the oncoming of the DNA sequencing technology, nucleotide variation at loci associated with inversions could be studied. Indeed, inversions are a good system to detect the action of selection (Kreitman and Wayne 1994; Depaulis *et al.* 1999). As supressors of the recombination, inversions can act as amplifiers of the effects on nucleotide polymorphism caused by selective events such as hitchhiking (Kaplan *et al.* 1989; Aquadro and Begun 1993; Aquadro *et al.* 1994) or deleterious background selection (Charlesworth *et al.* 1993; Charlesworth 1994; Hudson 1994; Hudson and Kaplan 1995a, 1995b). The relation between inversion polymorphism and fertility, viability,

and sexual isolation in *Drosophila* has also been a major field of investigation related to inversions (Stalker 1976; Zouros 1981; Anderson 1989; Coyne *et al.* 1991, 1993; Tyler *et al.* 1993; Kamping and Van Delden 1999; Singh and Singh 2001).

2.2.1. Distribution of inversions frequencies

Inversions can be pericentric or paracentric depending on whether the inverted segment includes the centromere or not, respectively. Inversion polymorphisms usually involve paracentric inversions (Krimbas and Powell 1992). Paracentric inversion polymorphisms are common in dipterans other than *Drosophila*. More than 120 polymorphic inversions occur in natural populations of the *Anopheles gambie* mosquitoes species complex (Coluzzi *et al.* 2002). Paracentric inversions have also been detected in mammals, including humans. In humans, they can occur on all chromosome arms (Pettenati *et al.* 1995), and usually are associated with diseases. Inversions together with translocations are six times more frequent in individuals with abnormal physical and mental development than in control groups (Bugge *et al.* 2000). Inversions are particulary common in individuals with haemophilia A (Deutz-Terlouw *et al.* 1995) and leukaemias (Rowley 1998).

In *Drosophila* most of species present naturally occuring inversion variation. Among a 182 studied species, only 46 were found to be monomorphic (Sperlich and Pfriem 1986). These estimates seems to point out that three quarters of all species in the genus *Drosophila* are polymorphic for inversions (Krimbas and Powell 1992; Powell 1997). Some polymorphic species, such as *D. willistoni*, *D. paulistorum*, and *D. subobscura*, have all chromosomal arms highly polymorphic, while other species, such as *D. pseudoobscura* and *D. persimilis*, have the inversion polymorphism restricted to only one chromosome of the complement (Krimbas and Powell 1992).

2.2.2. Inversions origin

It is usually accepted that inversions are generated by two consecutive events. First, two breaks in the chromosome generate a chromosome segment that is then reattached in an inverted orientation with respect to the flanking regions (Stadler 1932). According to this process, inversions are generated by only two breaks at a time. Another important evolutionary consequence of this process is that naturally occuring inversions are unique in origin: all copies of a particular gene arrangement can be traced back to a single event and thus to a single chromosome. The proposal that inversions have an unique origin was first made by Sturtevant and Dobzhansky (1936) and later elaborated by Dozhansky (1937). This assertion is based on two assumptions: the generation of an inversion under natural conditions is a rare event and the probability that two such rare events involving exactly the same piece of chromosome occur is negligible. The uniqueness of the inversion origin may be questioned by the presence in some *Drosophila* species of “hot spots”, where the breakpoints of different inversions have been mapped. This occurs for example in *D. subobscura* (reviewed in Krimbas and Loukas 1980) and in *D. melanogaster* (Grossman 1967). The existence of these “hot spots” argues against the random occurrence of breaks, that is one of the assumptions of the inversion uniqueness proposal. The discovery of transposable elements and the evidence of their ubiquity in *Drosophila* (Green 1980) could explain the existence of such “hot spots”. Although some studies revealed the implication of transposable elements in the origin of polymorphic inversions (Cáceres *et al.* 1999; Casals *et al.* 2003), others studies failed to detect transposable elements around breakpoints (Wesley and Eanes 1994; Cirera *et al.* 1995). Even in cases where transposable elements seem to played a role in the origin of the inversion, it has been confirmed that the inversion was monophyletic (Cáceres *et al.* 2001).

2.2.3. Changes in inversions frequencies

Clinal variation for traits and genetic variants along latitudinal or altitudinal gradients provides strong evidence for the action of natural selection in relation to climatic factors. Stable latitudinal clines for inversion frequencies in *Drosophila* are repeatable across continents (reviewed in Krimbas and Powell 1992). The existence of these clines may suggest that there are probably beneficial effects associated with inversions. In Europe, *D. subobscura* displayed similar inversion clinal patterns for more than 20 years (Sole *et al.* 2002). The colonization of the America by *Drosophila subobscura* provided a great opportunity to examine the speed at which latitudinal clines in inversion frequencies develop. After colonization, latitudinal inversion clines developed in North and South America in less than 5 years (Prevosti *et al.* 1988). Balanya *et al.* (2003) confirmed that the frequencies of most inversions still change with latitude and tend to parallel the latitudinal clines present in Europe. Altitudinal clines for some *Drosophila* species have also been reported (Dobzhansky 1948). Seasonal changes in inversion frequencies were also documented in *D. pseudoobscura* (reviewed in Krimbas and Powell 1992) and in other *Drosophila* species (Rodriguez-Trelles *et al.* 1996; Rodriguez-Trelles and Rodriguez 1998).

2.2.4. Inversions, traits and fitness

In *Drosophila*, traits that have been linked to inversion polymorphism include viability (Fernandez Iriarte and Hasson 2000; Zivanovic and Marinkovic 2003), development time (Betran *et al.* 1998; Fernandez Iriarte and Hasson 2000), longevity (Rodriguez *et al.* 1999), mating success and female fecundity (Brockett *et al.* 1996), resistence to thermal extremes (McColl and McKechnie 1999; Hoffman *et al.* 2002; Weeks *et al.* 2002; Anderson *et al.* 2003; Frydenberg *et al.* 2003) and body size (Bitner-Mathe *et al.* 1995; Norry *et al.* 1995; Fernandez Iriarte and Hasson 2000;

Orengo and Prevosti 2002; Weeks *et al.* 2002; Fernandez Iriarte *et al.* 2003; Santos *et al.* 2004). In mosquitoes, inversions have also been linked to some of these traits, as well as resistence to DDT and dieldrin pesticides (Brooke *et al.* 2002).

Associations between body size and inversion arrangements are particulary common in *Drosophila* species, such as *D. ananassae* (Yadav and Singh 2003), *D. subobscura* (Orengo and Prevosti 2002) and *D. buzzatii* (Fanara *et al.* 1997). Many *Drosophila* species exhibit adaptive clines in body size, with a large size found at relative cold latitudes (Huey *et al.* 2000; Gockel *et al.* 2001; Hallas *et al.* 2002). Clines in body size might be influenced by inversion frequencies, and several studies point to this hypothesis (Huey *et al.* 2000; Gockel *et al.* 2002; Weeks *et al.* 2002; Calboli *et al.* 2003a, 2003b).

Currently there is little information about the genes located within inversions that influence fitness traits. However, Dobzhansky's experiments (1970) provided indirect support that inversions might affect fitness by maintaining combinations of favourable genes by means of reducing recombination between them (coadaptation). A study by Schaeffer *et al.* (2003) provided direct support at molecular level for tight associations among alleles in different gene arrangements, which is consistent with the Dobzhanksy's coadaptation hypothesis. Inversions may also have direct effects on fitness by affecting the expression of genes that are closely linked to inversion breakpoints (Wesley and Eanes 1994). The disruption of genes due to chromosome inversion has also been reported (Matzkin *et al.* 2005).

2.2.5. Patterns of variation within inversions

Inversions have been shown to have several effects on recombination rates (reviewed in Roberts 1976). First, inversions relocate genes along a chromosome

potentially modifying their recombination context. Second, recombination is reduced by the inhibition of chiasmata in heterokaryotypes (Roberts 1976; Coyne *et al.* 1993; Navarro and Ruiz 1997). This effect is likely to be most marked near the breakpoints of paracentric inversions (Novitski and Braver 1954; Grell 1962). Finally, recombination is reduced in heterokaryotypes because crossing over events within the inverted region give rise to non-viable unbalanced meiotic products. The main consequence of reduced recombination along inversion is to subdivide the population into two classes for the inverted segment: standard and inverted. These two classes, however, are not completely isolated. Genetic exchange between chromosomal arrangements are possible, because viable recombinant gametes arise by multiple crossing over (reviewed in Ashburner 1989) and by gene conversion (Chovnick 1973). Given that inversions inhibit recombination in heterokaryotypes, loci within newly arisen monophyletic inversions will be in strong linkage disequilibrium (nonrandom association between alleles from two or more different loci). Double crossover events will break down through time the associations between loci centrally located in the inversion. Strong linkage disequilibrium is still expected around the breakpoints and in regions where selection mantains the associations among alleles (Navarro *et al.* 1997).

The effect of inversions on nucleotide variability depends on the age of inversions and the rate of genetic exchange between the inverted and standard arrangements (Navarro *et al.* 2000). As inversions spread, they will initially reduce genetic variability within a population because of the strong disequilibrium between the inversion and genes within it. After a long period of time, inversions can increase genetic variability near breakpoints because they subdivide the population and allow the independent accumulation of genetic variability in the different arrangements. However, inversions will tend to reduce variability if present for several generations fewer than the effective population size (Navarro *et al.* 2000). Empirical data suggest that nucleotide variability increases rapidly away from the breakpoints of inversions

(Andolfatto *et al.* 2001; Laayouni *et al.* 2003). This pattern suggests that the studied inversions are not ancient balanced polymorphisms.

2.3. Nucleotide variation

An accurate description of the level and the pattern of genetic variation in natural populations is a pre-requisite to understand the forces driving evolution. Kreitman (1983) published the first description of DNA sequence variation in a sample of alleles from natural populations of *D. melanogaster*. Since then, a great number of studies have helped to present a reasonably detailed picture of DNA sequence variation from multiple loci in natural populations of *Drosophila*. One important observation to emerge from these studies is that there is a considerable heterogeneity among genes in the level of naturally occuring DNA polymorphism. In addition, it was evident that both natural selection and recombination are important in determining levels of variation at nuclear genes, and that different genes have different evolutionary histories and experience different effective population sizes depending on their recombinational environment.

2.3.1. The neutral theory of molecular evolution

The neutral theory of molecular evolution asserts that the great majority of evolutionary changes at the molecular level (as revealed by comparative studies of protein and DNA sequences), are not caused by Darwinian selection but by random drift of selectively neutral or nearly neutral mutants (Kimura 1983). According to the neutral theory of molecular evolution, the nucleotide substitution rate (k) is equal to neutral mutation rate (μ_N),

$$k = \mu_N$$

while the intraspecific variability or heterozygosity per nucleotide (θ) is directly proportional to effective population size (N_e) and to the neutral mutation rate,

$$\theta = 4 N_e \mu_N$$

The hypothesis that molecular polymorphism and divergence were mainly due to neutral mutations and genetic drift, was first proposed in 1968 by Kimura. Although there were previous suggestions that molecular evolution was neutral (Freese 1962; Sueoka 1962), Kimura was the first author, who combine population genetics theory with molecular evolution data. He developed a theory that proposes that genetic drift is the main force changing allele frequencies. In 1969, Kimura published a paper on the rate of molecular evolution, in which he argued that the rate of amino acid substitutions of homologous proteins is almost constant. The same year, King and Jukes, published a paper entitled "Non-Darwinian Evolution", and that independently proposed that most amino acid substitutions are neutral. The authors suggested that "proteins, and sites within proteins, differ with regard to the stringency of their requirements". Another achievement that permited the formulation of the neutral theory of molecular evolution was the infinite-allele model (Kimura and Crow 1964).

As more protein sequence data were available, it became clear that the rate of protein evolution differed greatly between proteins (Dickerson 1971). This pattern of variation became one of the principles of Molecular Evolution (Kimura and Ohta 1974): "Functionally less important molecules or parts of a molecule evolve faster than more important ones". Important parts of proteins were said to be selectively constrained because they could not be changed without a severe and detrimental impact on fitness. Those mutations in constrained regions were rejected by natural selection because of their deleterious effects, while mutations that become fixed in populations were considered neutral. It was thought, that advantageous mutations were so rare that they make only a negligible contribution to the totality of

substitutions. Kimura (1971) and Ohta (1972a, 1973, 1974) proposed that slightly deleterious mutations might be quite common among amino acid substitutions. At that time, some results of the studies on molecular evolution seemed to contradict the predictions of the neutral theory. Hence, the upper limit of the heterozygosity estimates determined by protein electrophoresis (Lewontin 1974). Under the neutral theory, the heterozygosity is expected to increase with the species effective population size. Ayala *et al.* (1972), argued that the effective size of *Drosophila* species must be much larger than predicted by the neutral theory. They considered that the effective size of for *D. willistoni* could be 10^9 and thus, that the observed average heterozygosity was much lower than the expected by neutral predictions. An explanation for low heterozygosities may be hitchhiking of neutral variation in chromosome regions due to directional selection (Maynard Smith and Haigh 1974; Aquadro 1992). The interest for the hitchhiking effect has recently increased because of the reduction in silent variation detected in regions of low recombination in *Drosophila* (Langley 1990). Similarly, population size fluctuations may affect heterozygosity due either to a lowering of the effective population size by relatively rapid fluctuations in size or to severe bottlenecks whose effects may last for many years (Nei *et al.* 1975). Both factors may keep populations out of mutation-drift equilibrium and thus, cause that populations do not fullfil neutral predictions.

With the improvement of the DNA sequencing techniques, a great number of studies that focused on the comparison of DNA sequences were published. Pseudogenes were shown to evolve rapidly and thus, they provided further support to the neutral theory (Li *et al.* 1981; Miyata and Yasunaga 1981). At that time, it was considered that substitutions in non-coding DNA and synonymous substitutions in coding regions were neutral, that amino acid substitutions were deleterious or nearly neutral, and that advantageous substitutions make up a minor fraction of all substitutions. Ikemura (1981) proposed that codon bias was correlated with transfer RNA abundance in the cell. Kimura (1981) proposed that synonymous mutations

were not strictly neutral and that codon usage could be explained according to the neutral theory in terms of selective constraint. As a consequence the rate of synonymous substitutions would be slightly reduced compared to that of strictly neutral mutation, and codon usage may be highly biased at equilibrium (Kimura 1981; Li 1987; Bulmer 1991). Takahata (1987) proposed a variant of the simple neutral model, the fluctuating neutral space model, which assumes that the neutral mutation rate changes with each neutral substitution.

As the rapid accumulation of DNA sequences continued, more progress was made in contrasting the dynamics of synonymous and nonsynonymous substitutions. In general, nonsynonymous substitutions are thought to be more strongly influenced by selection than synonymous substitutions. Therefore, the generation-time effect should be more easily detected in synonymous than in nonsynonymous substitutions, as the former more faithfully reflect the mutation rate. DNA sequence analysis confirmed this prediction (Li *et al.* 1981; Ohta 1993, 1995). The absence of the generation-time effect in nonsynonymous substitutions can be explained according to the nearly neutral model of molecular evolution (Ohta 1992). Indeed, the generation-time effect would be cancelled by the population-size effect if most nonsynonymous mutations are slightly deleterious, and the function of the protein has been conserved for a long time. On the other hand, an acceleration of nonsynonymous substitutions is often observed in genes that acquire a new function (Ohta 1994).

Several theories of molecular evolution, where natural selection rather than genetic drift is the main force driving evolutionary changes, have also been proposed. The mutational landscape model (Gillespie 1984, 1991), was proposed to explain the episodic nature of amino acid substitutions. In this model, molecular evolution was pointed as generally stagnated at a local optimum due to the very low mutation rates, an to the fact that sequences are more than a mutational step away

from the locally optimal sequence. Other model, the TIM model (Takahata *et al.* 1975) is similar to the nearly neutral model, except that the fitness of genotypes changes slowly through time. The SAS-CFF model has been developed to explain patterns of proteins evolution mainly when selection is strong and mutation is weak (Gillespie 1991).

2.3.2. Linkage and selection in molecular evolution

2.3.2.1. Balancing selection

Balancing selection maintaining two alleles has been investigated according to the infinite sites model using a coalescent approach, either without recombination (Kaplan *et al.* 1988; Kelly and Wade 2000) or with arbitrary recombination (Hudson and Kaplan 1988). Given a random sample of n alleles, the expected number of segregating sites, S, at a locus in the sample is

$$E(S) = \mu E(T)$$

where μ is the mutation rate per generation to neutral alleles at the locus and $E(T)$ is the expected sum of the branch lengths in the gene tree.
According to the standard neutral model, without balancing selection

$$E(T) = \sum_{i=1}^{n-1} (2 / i)$$

where, T is measured in units of $2N$ generations (N is the population size). For a sample of size two, $E(T) = 2$ and $E(S) = 4N\mu$. The selective model assumes that there is recurrent mutation at the site giving rise to the balanced polymorphism. With a low recurrent mutation rate between the two selected alleles, for example $0.01/2N$ and no recombination, the expected total branch length of the tree becomes very

large, in this case $E(T) = 52.0$ (Kaplan *et al.* 1988). Therefore, a sample of size two will have, on average, 26 times more segregating neutral variants under balancing selection than in the strictly neutral case. Adding recombination ($R = 2Nr$, where r is the recombination rate between the site under selection and a neutral site), $E(T)$ will be much larger than in the strictly neutral case if the recurrent mutation rate is low and $R < 1$. Therefore, an excess of neutral variation is expected to accumulate between the selected alleles for a tightly linked region around a site under balancing selection.

Balancing selection, or heterozygous advantage, can thus elevate the level of linked neutral variation. A balanced polymorphism if ancient will appear as a region containing an unusually high level of silent variation. Several loci seem to show the pattern of variation expected under balancing selection: the *Adh* locus in *D. melanogaster* (Kreitman and Hudson 1991), the *Gpdh* locus in *D. melanogaster* (Lindsey and Zimm 1992), revealed a significant higher level of variation in the coding region than in flanking regions by HKA test (Hudson *et al.* 1987); in *D. simulans* the *Est-6* locus also seems to include a balanced polymorphism (Karotam *et al.* 1993).

2.3.2.2. Directional positive selection

The fixation of a favored mutation by selection can pull linked neutral mutations along with it. This effect is known as the hitchhiking effect or selective sweep. The hitchhiking effect depends on the strength of selection and on the rate of recombination between the selected site and the neutral linked sites (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989; Gillespie 1997, 2000a). The expected reduction in the level of linked neutral polymorphism will also depend on the elapsed time since the most recent selective event. This implies that the mutation rate to neutral alleles,

μ, will also affect the magnitude of hitchhiking effect. The hitchhiking model predicts that loci in regions of low recombination will harbor lower levels of variation than loci in regions of normal recombination and an excess of low frequency variants at polymorhic sites. In fact, the physical length of the hitchhiking region depends on the strength of selection relative to the recombination rate (Kaplan *et al.* 1989). If selection coefficients are similar across the genome, loci in regions of low recombination will be affected by more selective sweeps per unit of time and hence, are more likely to be sampled shortly after a sweep. In the recovery phase, most polymorphisms will be young and the new variants present at low frequencies.

2.3.2.3. Background selection

An alternative to the hitchhiking model is the background selection model that is based on the action of purifying selection against strongly deleterious mutations (Charlesworth *et al.* 1993, 1995; Hudson and Kaplan 1994, 1995a, 1995b). In this model, a neutral allele will persist in the population only if it finds itself on a chromosome (or segment of chromosome) free of deleterious mutations, either when it first arises in the population or by recombination. If selection coefficients and deleterious mutation rates are the same in different regions, the rate of recombination will determine the extent of the reduction in neutral diversity, that is, the extent of which neutral alleles can escape from background selection.

Despite the fact that purifying selection undoubtedly occurs, the uncertainty about key parameters (such as the distribution of selection coefficients and the deleterious mutation rate), questions the importance of background selection in reducing levels of variability. Similar uncertainty exists for the importance of positive selection. Consequently a intense debate has revolved around the relative importance of background selection and hitchhiking in shaping patterns of variability.

While positive selection models predict an excess of low frequencies variants at linked neutral sites relative to a model of no selection (Braverman *et al.* 1995; Gillespie 2000a), the background selection model does not, if the population is large and the deleterious mutation rate is not extremely high (Hudson and Kaplan 1994; Charlesworth *et al.* 1995). In *D. melanogaster*, these two conditions are likely to be met (Li *et al.* 1999; McVean and Vieira 2001). Thus, polymorphism data from this species provide an opportunity to distinguish between both models. Some surveys of loci in *D. melanogaster* did not detect a skew toward rare variants in regions with low recombination (Begun and Aquadro 1993; Charlesworth *et al.* 1995). These observations suggested that background selection might be a sufficient explanation for the correlation between diversity levels and the recombination rate, that is, that there is no unequivocal evidence for positive selection. However, other stuides supported the directional selection model either at particular loci (Langley *et al.* 2000) or in multilocus approaches (Andolfatto and Przeworski 2001). This last survey in 29 loci of *D. melanogaster* showed that in African populations, a summary of the frequency spectrum of polymorphic mutations is positively correlated with the meiotic rate of crossing over. It was thought that this pattern was unlikely under a model of background selection and thus, that hitchhiking due to the recurrent fixation of advantageous variants was the most plausible explanation for the data.

2.3.3. Variation and recombination

The rates of recombination vary dramatically across the genome for *D. melanogaster*, but not the levels of divergence with its sibling species. Several studies (Aguadé *et al.* 1989; Berry *et al.* 1991; Begun and Aquadro 1992; Aguadé and Langley 1994; Aquadro *et al.* 1994; Pritchard and Schaeffer 1997; Andolfatto and Przeworski 2001) have shown that, in regions of extremely low recombination, levels of DNA sequence variation are dramatically reduced within *D. melanogaster*. This trend has also been gathered in a wide variety of organisms (Nachman 1997;

Dvorak *et al.* 1998; Stephan and Langley 1998; Przeworski *et al.* 2000). The detection of a genome-wide positive correlation between DNA sequence variation and regional rate of recombination further demonstrated that this pattern extends to the whole genome (Begun and Aquadro 1992; Aquadro *et al.* 1994; Moriyama and Powell 1996). Available data also suggests that a relation between recombination and variation is a feature of *D. simulans* and *D. ananassae* (Martin-Campos *et al.* 1992; Stephan 1994). The correlation between variation and recombination rate seems to be a general feature in different species, in which rates of recombination vary across the genomes too. Several studies have demonstrated a positive correlation between recombination rates and variation for some genes in the X chromosome of *Mus domesticus* (Nachman and Churchill 1996; Nachman 1997) and for several X-linked loci in humans (Nachman *et al.* 1998; Przeworski *et al.* 2000; Nachman 2001). There are four non-exclusive explanations for this correlation (Aquadro 1997). First, the mutation rate might be high in regions of high recombination, likely due to the recombination process itself. Second, functional constraints may be negatively correlated with recombination, due, for example, to the evolutionary sequestering of genes coding for highly constrained proteins into regions of the genome with low recombination. Third, the correlation could reflect the cumulative footprints of selective sweeps associated with new advantageous mutations (directional selection). Fourth, the correlation could result from a reduction in the gene-specific effective population size due to the elimination of linked deleterious mutations (background selection). The first and second explanations predict a positive correlation between recombination and sequence divergence between species. No such correlation has been detected in *Drosophila*, mice or humans (Begun and Aquadro 1992; Aquadro *et al.* 1994; Moriyama and Powell 1996; Nachman and Churchill 1996; Nachman 1997; Nachman *et al.* 1998; Przeworski *et al.* 2000; Nachman 2001).

Both selective sweeps (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989; Gillespie 1997, 2000a) and background selection (Charlesworth *et al.* 1993, 1995;

Hudson and Kaplan 1994, 1995a, 1995b) could cause the detected variation/recombination correlation by reducing the effective population size at chromosomal regions experiencing low rates of recombination. Both models predict that selection reduces polymorphism at linked neutral sites, and the lower the recombination rate, the greater the magnitude of this reduction. In addition, according to both models, it is not expected that variation in recombination rates affects divergence between species at neutral sites (Birky and Walsh 1988). The similar predictions of the two models have made it difficult to determine their relative importance. The background selection model can provide a reasonably good fit to the polymorphism data, given certain parameters of deleterious mutation rate and recombination rate in *D. melanogaster* (Hudson and Kaplan 1995a; Charlesworth 1996). Similary, a simple hitchhiking model also fits the *D. melanogaster* polymorphism data quite well (Wiehe and Stephan 1993; Stephan 1995; Andolfatto and Przeworski 2001).

This correlation between variation and recombination indicates that effective population size (N_e) is not the same for all genes across a genome. In regions of low recombination, N_e can be dramatically reduced below the species-effective population size. In regions of the genome with low N_e, genetic drift becomes more important, and selection less efficient. Therefore, rates of fixation caused by positive selection are reduced and rates of fixation of slightly deleterious mutations are increased in regions of low recombination (Charlesworth 1996).

2.3.4. X-chromosome versus autosomes

If nucleotide polymorphism is entirely or predominantly neutral, the level of diversity is expected to be directly proportional to effective population size (Kimura 1983). X-linked genes have an effective population size that is ¾ the effective size of

autosomal genes (assuming a 1:1 sex ratio). Thus, sex-linked genes should exhibit a level of neutral variation equal to ¾ the level present in autosomal genes (Moriyama and Powell 1996). However, if sexual selection in males is prevalent in natural populations of *Drosophila* (Andersson 1994), then the ratio of effective sizes of the X chromosome and autosomes may be closer to the unity (Caballero 1995). Laboratory measurements have suggested that the effective population size is greater in females than in males (Crow and Morton 1954). Thus, the relation in the level of nucleotide variation between X-chromosomes and autosomes may be smaller than ¾ (Andolfatto 2001).

First studies on the X chromosome of *D. melanogaster* (Langley and Aquadro 1987; Miyashita and Langley 1988; Schaeffer *et al.* 1988; Aguadé *et al.* 1989) have not shown the expected reduction in nucleotide variation. However, posterior studies (Aquadro *et al.* 1994; Moriyama and Powell 1996) showed that the differences in the level of nucleotide variation between X-linked and autosomal gene was close to predictions, especially in *D. melanogaster*. In *D. simulans* X-linked genes have shown in average about the half of the variation of autosomal genes (Begun and Whitley 2000). On the contrary, another study has shown that nucleotide variation on the X chromosome is twice the level on autosomes in African populations of *D. melanogaster* (Andolfatto 2001). However, more data obtained consistently from the same sampled populations are needed to establish the statistical significance of this trend.

Many factors, both selective and demographic, can contribute to the difference in the level of nucleotide diversity between the X chromosome and the autosomes (Aquadro *et al.* 1994; Caballero 1995; Charlesworth 1996; Fay and Wu 1999). For instance, if most advantageous alleles are recessive, the hitchhiking model (Maynard-Smith and Haigh 1974) predicts reduced diversity on the X chromosome relative to

autosomes, whereas the background selection model (Charlesworth *et al.* 1993) predicts the opposite pattern (Aquadro *et al.* 1994). Hitchhiking is more effective in the X chromosome than in the autosomes, because the time of a beneficial recessive mutation on its way to fixation is shorter in the X chromosome (Avery 1984; Aquadro *et al.* 1994). Thus, the adaptive substitution rate may be higher on the X chromosome than on autosomes. On the other hand, background selection is more effective on the autosomes, as the effect of background selection is proportional to the frequency of deleterious alleles under purifying selection (Charlesworth *et al.* 1993; Charlesworth 1994). Recessive deleterious alleles can reach higher frequencies on the autosomes than on the X, tus, background selection eliminates more variation from the autosomes than from the X chromosome.

2.4. *Drosophila* genus

The Drosophilidae is one of the most diverse and widely distributed families of Diptera. This family includes more than 3000 species (Wheeler 1981, 1986; Ashburner 1989) grouped into more than 60 genera (Wheeler 1986; Remsen and O'Grady 2002). The origin of the family occured some 80-120 million years ago (Throckmorton 1975). Despite extensive research, the phylogeny and taxonomy of the family remains controversial. Wheeler's (1981, 1986) standard classification seems inconsistent with the phylogenetic relationships among the species (Kwiatowski and Ayala 1999), on either morphologic traits (Throckmorton 1975; Grimaldi 1990) or molecular data (DeSalle 1992; Pélandakis and Solignac 1993; Thomas and Hunt 1993; Kwiatowski *et al.* 1994, 1997; Russo *et al.* 1995; Powell 1997; Remsen and DeSalle 1998; Tatarenkov *et al.* 1999).

Drosophila, with 1600 species, is the Drosophilidae genus with more species (Wheeler 1986; Remsen and O'Grady 2002). Throckmorton (1975) proposed that the genus originated some 60-80 million years ago, probably in southeast Asia. The principal split of the genus occurred aproximately 50 million years ago is Old World tropics giving rise to the *Sophophora* and *Drosophila* subgenera. However, there is not yet a clear consensus on the number of *Drosophila* subgenera (Throckmorton 1975; Wheeler 1981; Grimaldi 1990; Kwiatowski and Ayala 1999; Tatarenkov *et al.* 1999).

2.4.1. *Sophophora* subgenus

This subgenus was proposed by Sturtevant (1939, 1942) when he attempted to subdivid the genus *Drosophila* into subgenera and species groups. He proposed four species groups in the *Sophophora* subgenus: *melanogaster*, *obscura*, *saltans* and *willistoni*. As the Australasian and Afrotropical fauna was studied in more detail, three additional groups, *dispar*, *fima* and *dentissima*, were proposed (Burla 1954; Mather 1954; Tsacas 1979, 1980). Currently, the seven species groups of the *Sophophora* subgenus contain aproximately 300 species (Wheeler 1981, 1986; Lemeunier *et al.* 1986). Throckmorton (1975) suggested that the species of the *Sophophora* subgenus are the result of a large radiation of flies, that predated the radiation that gave rise to the other species of the genus *Drosophila*. Posterior molecular studies supported this idea (Remsen and DeSalle 1998; Kwiastowski and Ayala 1999; Remsen and O'Grady 2002).

Morphology, DNA-DNA hybridization and phylogenetic reconstructions from DNA sequences support the monophyly of the *Sophophora* subgenus (Throckmorton 1975; DeSalle 1992; Thomas and Hunt 1993; Russo *et al.* 1995; O'Grady and Kidwell 2002). The four largest species groups of this subgenus (*melanogaster*, *obscura*,

saltans, and *willistoni*) also seem to be monophyletic (reviewed in Powell 1997). The *melanogaster* species group includes mainly Old World tropical species, although some species are cosmopolitan (Leumeunier *et al.* 1986; Lachaise *et al.* 1988). The *obscura* species group is distributed primarily in the Holarctic region, but some species inhabit tropical regions (Lakovaara and Saura 1982; Heed and O'Grady 2000). The *saltans* and *willistoni* groups include a Neotropical clade of species.

2.4.2. *Obscura* group

According to biogeographical data, the *melanogaster* and *obscura* species groups originated from a common ancestral lineage in southeast Asia during the mid-Oligocene (Throckmorton 1975). The species of the *Drosophila obscura* group has been used in several studies of evolutionary topics such as reproductive isolation, mating preference, inversion polymorphism and genetic variation in natural populations (reviewed in Dobzhanksy and Powell 1975; Lakovaara and Saura 1982; Powell 1997). The phylogeny of the species group has also been extensively analyzed on the basis of morphologic traits (Sturtevant 1942; Buzzati-Traverso and Scossiroli 1955), allozymes (reviewed in Lakovaara and Saura 1982), RFLP analysis of the mitochondrial DNA (Latorre *et al.* 1988; Gonzalez *et al.* 1990; Barrio *et al.* 1992), polytene chromosomes (Brehm *et al.* 1991; Bondinas *et al.* 2001, 2002), DNA-DNA hybridization (Goddard *et al*. 1990) and nucleotide sequences (Beckenbach *et al.* 1993; Barrio *et al.* 1994; Russo *et al.* 1995; Barrio and Ayala 1997; O'Grady 1999). However, the phylogeny and taxonomy of the group still remains controversial.

Based on morphological traits, Sturtevant (1942) originally divided the *obscura* species group into two subgroups: the *affinis* subgroup including mostly Nearctic species and the *obscura* subgroup containing Palearctic and Nearctic species. Later, Buzzati-Traverso and Scossiroli (1955) concluded that, the *obscura* subgroup,

included two different lineages of Nearctic species and several lineages of Palearctic species. In the review of the *obscura* group phylogeny, published by Lakovaara and Saura (1982), the *obscura* group was divided into two subgroups: the *obscura* and the *pseudoobscura* subgroups. The last subgroup contained only Nearctic species and is closely related to the *affinis* subgroup (Lakovaara and Saura 1982; Barrio *et al.* 1992; Barrio and Ayala 1997). The common ancestor of the *pseudoobscura* and *affinis* subgroups likely colonized the western Nearctic region aproximately 20 million years ago (Throckmorton 1975). The *obscura* group also includes a set of African species, most of them discovered after 1985 and classified in the *microlabis* subgroup (Tsacas *et al.* 1985; Cariou *et al.* 1988; Brehm and Krimbas 1990a, 1992, 1993; Brehm *et al.* 1991; Bachmann *et al.* 1992; Ruttkay *et al.* 1992). Finally, Barrio *et al.* (1994) proposed the division of *obscura* subgroup proposed by Lakovaara and Saura (1982) into two subgroups: the *subobscura* and the *obscura* subgroups. The *subobscura* subgroup includes the widespread Palearctic species *D. subobscura* and two endemic species: *D. madeirensis* (Madeira Island) and *D. guanche* (Canary Islands). The revised *obscura* subgroup includes all the other Palearctic species. Currently, the *Drosophila obscura* group includes of more than 35 species (Heed and O'Grady 2000) and the classification of these species into five subgroups (*affinis, pseudoobscura, obscura, subobscura* and *microlabis*) is widely accepted.

2.5. *Drosophila subobscura*

The first description of *D. subobscura* appeared in a short note published by Collin (1936) as an addendum in Gordon (1936). Collin recognized *D. subobscura* as a different species from *D. obscura*. Later, Pomini (1940), Smart (1945) and Burla (1951) provided a more detailed description of the species.

2.5.1. Phylogenetic relationships

D. subobscura does not cross with the other species of the *obscura* group (Wallace and Dobzhansky 1946; Buzzati-Traverso and Scossiroli 1955) except with *D. madeirensis*. In laboratory, both reciprocal crosses between *D. madeirensis* and *D. subobscura* render progeny. Hybrid females are fertile, but hybrid males are sterile (Krimbas and Loukas 1984; Khadem and Krimbas 1991a, 1991b; Papaceit *et al.* 1991). *D. madeirensis* also crosses, with some difficulty with *D. guanche*. Both female and male hybrids are sterile (Krimbas and Loukas 1984). These three species are closely related, and constitute the *subobscura* cluster of species. The complete homologies of the segments of their chromosomes was determined (Krimbas and Loukas 1984; Molto and Martinez-Sebastián 1986; Molto *et al.* 1987; Papaceit and Prevosti 1989; Brehm and Krimbas 1990a, 1990b, 1992, 1993; Segarra and Aguadé 1992).

2.5.2. Geographic distribution and population size

D. subobscura is a Palearctic species that recently colonized North and South America. It is one of the most common *Drosophila* species in many countries of its distribution area. *D. subobscura* is present all over Europe. The north limit of its distribution is not clear, since it seems that the species is extending towards north, maybe due to a recent climatic change. The Eastern limit of the distribution seems to be Kazakhstan and Iran (Krimbas 1993). To the south, the species is found in Morrocos, Egypt, Algeria and Tunisia. *D. subobscura* is included in the fauna of many Mediterranean islands such as Balearic Islands, Corsica, Sardinia, Sicily, Crete and Cyprus. To the west, the species has been collected in Madeira, Azores and the Canary Islands; but until now it has not been found in the Cape Verde Islands (Krimbas 1993). In the late 70's, *D. subobscura* colonized South America, covering an

area from South Chile to approximately 30ºS latitude to the north and to the east reaching the Atlantic coast of Argentina. It has also colonized the Pacific coast of North America from 50ºN to approximately 34ºN latitude (Brncic *et al.* 1981; Beckenbach and Prevosti 1986).

The estimates of the effective population size of *D. subobscura* vary from 93.000 individuals in a Greek population (Begon *et al.* 1980) to 35.000-70.000 individuals in a population of the Mediterranean spanish coast (Mestres and Serra 1991), although the estimate in a previous study of the same population using a different method was 120.000-240.000 individuals (Serra *et al.* 1987). The individual density of *D. subobscura* varies between 1.3 and 750 flies for 100 m^2 in England and between 10-190 flies for 100 m^2 in Greece (Begon 1978; Begon *et al.* 1980).

2.5.3. Genetic variation in natural populations

The first studies of genetic variation in natural populations of *D. subobscura* were performed by Monclús (1953) and Prevosti (1954), who examined the geographic pattern of the number of teeth in male sex combs and in claspers. It was found that the number of sex comb teeth decreased with continental climate, and the number of teeth claspers exhibited an opposite tendency. These clines were interpreted as the result of natural selection. Prevosti (1954, 1955) demonstrated that wing size follows the July isotherms in England and Spain and related this result with a possible selective advantage, as mating speed is positively correlated with wing size (Monclús and Prevosti 1971). Other quantitative characters were also studied (reviewed in Krimbas 1993).

After the publication of the paper by Loukas *et al.* (1979), several studies on genetic variation at allozyme loci in natural populations of *D. subobscura* from many

geographic locations were performed (reviewed in Krimbas 1993). These studies differed substantially in methodology, the allozyme loci analyzed and the *D. subobscura* populations sampled. However, some general conclusions can be drawn from the obtained results. The species is quite polymorphic, with a mean heterozygosity around 15-20% and a proportion of polymorphic loci of 61%. This high variability could be expected for a common and successful species. In addition, some electrophoretic markers showed linkage disequilibrium with polymorphic inversions. In general, populations with a low level of chromosomal inversion polymorphism seem to be as much variable for electrophoretic markers as populations having a high level of inversion polymorphism (Loukas and Krimbas 1980).

Studies on allozyme polymorphism were gradually replaced by studies of polymorphism at the DNA level. Kittel and Sperlich (1989) described the restriction fragment length polymorphism at the *rosy* locus in a population of *D. subobscura* from Tübingen, Germany. Rozas and Aguadé (1990) identified 14 restriction fragment length polymorphisms and 8 length polymorphisms in a 1.6 kb region including the *rp49* gene. This gene was used as a marker of the O_3 inversion of *D. subobscura*, since it is located very close to the proximal breakpoint of this inversion. Posterior studies of the same region in samples from the Canary Islands, Madeira and Europe revealed that some polymorphisms were shared between the O_{st}, O_{3+4} and O_{3+4+8} gene arrangements (Rozas and Aguadé 1991a, 1991b; Rozas *et al.* 1995; Khadem *et al.* 1998). This result was interpreted as an evidence of genetic exchange, likely by gene conversion between these gene arrangements.

The great advances in the DNA techniques, mainly the DNA amplification by Polymerase Chain Reaction (PCR) and the automatic sequencing facilities permited extensive studies of genetic variation in several natural populations of *D. subobscura*

and close relatives (*D. madeirensis* and *D. guanche*) in many gene regions (Rozas and Aguadé 1993, 1994; Cirera and Aguadé 1998; Rozas *et al.* 1999; Navarro-Sabaté *et al.* 1999a, 2003; Llopart and Aguadé 2000; Munté *et al.* 2000, 2005; Perez *et al.* 2003).

2.5.4. Chromosomes

The configuration of the mitotic chromosomes in *D. subobscura* females has five pairs of acrocentric chromosomes and one pair of dot-like chromosomes. The Y chromosome is also acrocentric (Krimbas 1993). Mainx *et al.* (1953) identified the different chromosomes by the vowels of the alphabet, *A* being attributed to the sex chromosome (X). The chromosomes homologies between different species of the *obscura* group are shown in Table 2.1. The *D. subobscura* genome is divided in one hundred sections distributed along the different chromosomes: A (sections 1 to 16), J (17-35), U (36-53), E (54-74), O (75-99) and dot (100). According to Kunze-Muhle and Muller (1958), each section contains several subdivisions. There are a total of 405 subsections in the citological map.

Table 2.1.
Chromosomal homologies in different *Drosophila* species.

General symbol	*D. subobscura*		*D. melanogaster*	*D. pseudoobscura*
(Muller)	(Mainx)	(Frizzi)		*D. affinis (miranda)*
A	A	1	X	XL
B	U	2	2L	4
C	E	3	2R	3, (X_2)
D	J	4	3L	XR
E	O	5	3R	2
F	Dot	6	4	5

Modified from Krimbas, 1993.

2.5.5. The inversion polymorphism of *D. subobscura*

D. subobscura is characterized for having a very rich inversion polymorphism. In an early publication (Sokolov and Dubinin 1941), it was stated that several collected stocks contained heterozygous flies at a 100% frequency, and it proved to be impossible to obtain homozygous stocks even after 8 generations of inbreeding. Thus, it was considered that the species had systems of balanced lethals. Philip *et al.* (1944) and Buzzatti-Traverso and Scossiroli (1955) arrived at a similar conclusion. The genetic studies of Maynard Smith and co-workers (reviewed in Krimbas 1993) were important for showing that the effects of inversions on fitness were not due to a position effect. Later, it was reported that homozygous flies for gene arrangements are present in natural populations in the expected frequencies under panmixia, but that these frequencies are so small that is very unlikely to collect a fly homozygous for gene arrangements in all chromosomes of the complement. At present, 67 polymorphic inversions have been described in *D. subobscura*. These inversions form 93 different chromosomal arrangements (Table 2.2). In addition, two duplications that likely originated by a crossing over effect in a inversion heterozygous, have been described.

In contrast to the association observed between inversions of the same chromosome, there is no association between inversions located in different chromosomes, both in samples of natural populations and in samples from laboratory population cages (Frutos and Aguilar 1978). Given that *D. subobscura* harbors a very rich inversion polymorphism, several studies related with chromosome variation have been made in the species. These studies have analyzed, for instance, the relationship between inversion polymorphism and body traits such as wing size and shape or heat resistence (Prevosti 1967; Quintana and Prevosti 1991; Orengo and Prevosti 2002; Santos *et al.* 2004), viability and lethal-gene allelism (Mestres *et al.* 1995; Sole *et al.* 2000; Zivanovic *et al.* 2000). The seasonal variation of the inversion frequencies has

also been surveyed (Krimbas 1967; Gosteli 1991; Rodriguez-Trelles *et al.* 1996; Zapata *et al.* 2000). Other characteristics of inversion polymorphism such as the distribution of the inversion frequencies and their long-term change have been also the focus of several studies (Krimbas 1964; Jungen 1968a, 1968b; Frutos *et al.* 1987a, 1987b; Gosteli 1990; Peixoto and Klackzo 1991; Rozas and Aguadé 1994; Mestres *et al.* 1998; Sole *et al.* 2002; Balanya *et al.* 2003; Andjelkovic *et al.* 2003). Inversions have also been used in phylogenetic studies (Brehm *et al.* 1991). More recently, comparative studies of nucleotide variation of markers associated with inversions in different chromosomal arrangements have also been published (Rozas and Aguadé 1990; Rozas *et al.* 1999; Munté *et al.* 2000, 2005).

Table 2.2.
Inversions, duplications and gene arrangements per chromosome in *D. subobscura*.

Chromosome	Inversions	Duplications	Gene arrangements
A	8	0	11
J	4	0	4
U	12	1	15
E	16	0	17
O	27	1	46
Total	**67**	**2**	**93**

Modified from Krimbas, 1993.

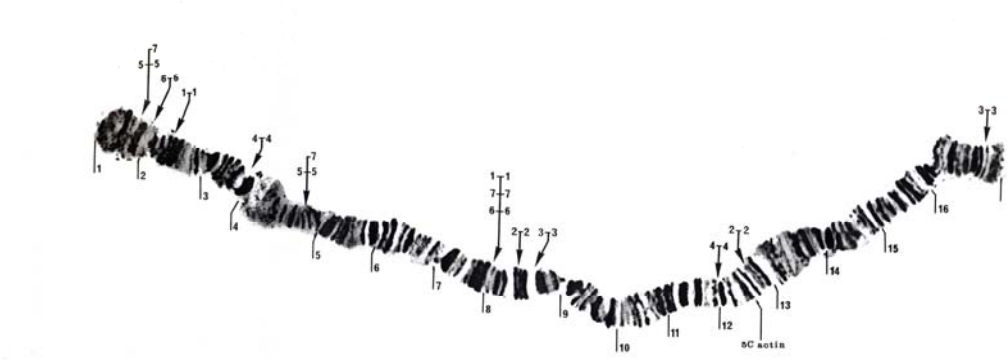
2.5.6. The A (=X) chromosome

The sex chromosome (A) is divided into two parts (Figure 2.1): the first segment (divisions 8 to 16) may carry inversions A_2, A_3 and A_8, while the second segment (divisions 1 to 7) the inversions A_1, A_5, A_6 and A_7. Inversion A_4 is an exception, it extends to both segments (Krimbas 1993). These inversions form complex gene arrangements: A_{2+6}, A_{2+3+6}, A_{2+3}, $A_{2+3+5+7}$, A_{2+4}, A_{2+3+5}, A_{2+5+7}, A_{2+8}. Recombination between segment I and segment II is possible. However, it seems that some

combinations of inversions are completely prevented. Sperlich and Feuerbach-Mravlag (1974), crossed A_1 and A_2 flies and although they recovered A_{st} recombinants, no A_1+A_2 recombinants were obtained. The A_1+A_2 arrangement has never been found in nature.

Some of the A chromosome inversions show latitudinal clines. The A_2 gene arrangement, at least in the western part of Europe, shows a north-south cline, increasing its frequency towards south. In northwest Africa, the A_2 arrangement is partly replaced by more complex arrangements derived from it (A_{2+3}, A_{2+4}, A_{2+6}, A_{2+3+6}, A_{2+3+5}, $A_{2+3+5+7}$, A_{2+5+7}). These arrangements are restricted to this region, except for a rare appearance of A_{2+6} in Israel and in Atlantic islands. The, A_{st} arrangement in west Europe has a clear north-south cline, decreasing in frequency towards south. Finally, the A_1 arrangement shows an east to west cline. Krimbas and Loukas (1980) interpreted this cline by a re-colonization phenomena of Europe after the last glaciation (10.000 years ago) from populations of Iran and Turkey, where this inversion is very frequent.

Figure 2.1.
Photomap of chromosome A of *Drosophila subobscura*.

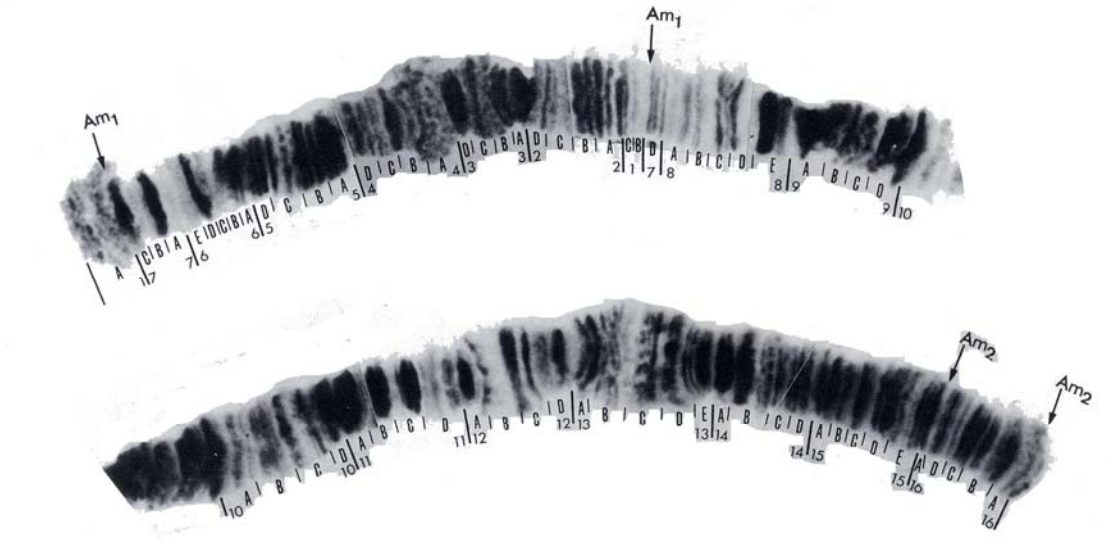


The different inversions breakpoints are indicated, as well as the several chromosomal sections (modified from Krimbas 1993).

2.6. *Drosophila madeirensis*

Drosophila madeirensis (Monclús 1984) is an endemic species of the Madeira Island and closely related to *D. subobscura*. There is a great similarity between the chromosomes of the two species (Krimbas and Loukas 1984; Papaceit and Prevosti 1989, 1991). However, they show two main differences: the chromosomes are thicker *D. madeirensis* than in *D. subobscura*, probably due to an additional step of polyteny, and the chromosomes are more fragile, which causes that they appear more frequently broken at different points in *D. madeirensis* than chromosomes in *D. subobscura* (Papaceit and Prevosti 1991). The gene arrangement of the A chromosome of *D. madeirensis* (Figure 2.2) is different from any known arrangement of *D. subobscura*, both for segment I and segment II (Papaceit and Prevosti 1989).

Figure 2.2.
Photomap of chromosome A of *Drosophila madeirensis*.



The breakpoints of the A_{m1} and A_{m2} inversions are indicated, as well as the several chromosomal sections (modified from Papaceit and Prevosti 1991).

Two nonoverlapping inversions differentiate the A chromosome of *D. madeirensis*: *Am*$_1$, with breakpoints in subsections 7C/7D and 1A/1B (Papaceit and Prevosti 1989), and *Am*$_2$. First, it was proposed that this latter inversion involved the terminal subsections 16BCD (Krimbas and Loukas 1984). However, a posterior study suggested that it also involved subsection 16A (Papaceit and Prevosti 1991). Brehm and Krimbas (1990b) mapped the distal breakpoint of the A_{m1} inversion at 6E/7A, but the results of Segarra and Aguadé (1992) confirmed the breakpoints at the subsection 7C/7D proposed by Papaceit and Prevosti (1989).

Chromosomal homologies between *D. madeirensis* and other species of the *obscura* group have been analyzed in detail (Krimbas and Loukas 1984; Papaceit and Prevosti 1989, 1991; Brehm and Krimbas 1990a, 1990b, 1991, 1992; Brehm *et al.* 1991; Segarra and Aguadé 1992). Reproductive isolation, mainly the genetic basis of male hybrids sterility, has also been studied (Papaceit *et al.* 1991; Khadem and Krimbas 1991a, 1991b, 1993, 1997). Nucleotide divergence between *D. madeirensis* and close relatives has also been analyzed at multiple loci (Ramos-Onsins *et al.* 1998; Cirera and Aguadé 1998; Llopart and Aguadé 1999; Navarro-Sabaté *et al.* 1999b; Munté *et al.* 2001). However, nucleotide variation in natural populations of *D. madeirensis* has been analyzed only at the *rp49* gene region (Khadem *et al.* 2001).

3. OBJECTIVES

The main objective of this study was to analyze nucleotide polymorphism and divergence in the two closely related species *D. madeirensis* and *D. subobscura*. These species are expected to have strong differences in the effective size due to marked differences of their distribution areas. *D. madeirensis* is endemic of the Madeira Island and *D. subobscura* is widely distributed and abundant in Europe and North Africa, where it shows a rich chromosomal polymorphism. Nucleotide variation in five X-linked gene regions was analyzed in 12 *D. madeirensis* lines and 18 *D.* subobscura lines. In addition, the *D. subobscura* lines differed in their arrangement for the X (=A) chromosome: 12 lines were A_2 and 6 A_{st}.

The obtained data allowed us to make inferences about the level of nucleotide variation in an endemic species in contrast to a continental close relative, the

speciation process that originated *D. madeirensis*, and the effect of chromosomal polymorphism on nucleotide variation in *D. subobscura*.

These main aims of the study can be summarized in the following more specific objectives.

1 – To contrast whether *D. madeirensis* can be considered an endangered species due to the progressive destruction of the Laurisilva in Madeira Island that is the natural habitat of the species.

2 – To contrast whether the level of polymorphism reflects the strong differences in effective size expected between *D. madeirensis* and *D. subobscura*.

3 – To compare the level of variation detected in the studied X-linked regions with those previously reported for an autosomal region in *D. madeirensis*.

4 – To compare the level of variation detected in the studied X-linked regions in a natural populations of *D. subobscura* from Madeira with those previously reported for other X-linked and autosomal regions in natural populations of *D. subobscura* from Europe.

5 – To contrast whether the pattern of nucleotide polymorphism in *D. madeirensis* and *D. subobscura* is consistent with the neutral model of molecular evolution or reflects the action of natural selection.

6 – To study the level of genetic differentiation along the chromosome segment affected by inversion 2 that differentiates the A_{st} and A_2 arrangements of *D. subobscura*.

7 – To contrast whether the genetic exchange between the A_{st} and A_2 arrangements of *D. subobscura* is higher in the central part of inversion 2 than near the breakpoints.

8 – To estimate how old is the A_2 arrangement of *D. subobscura*.

9 – To contrast whether *D. madeirensis* originated in allopatry from ancestral populations of *D. subobscura* that colonized the Madeira Island.

10 – To analyze whether there was genetic introgression between natural populations of *D. madeirensis* and *D. subobscura* in Madeira.

4. MATERIAL AND METHODS

4.1. Genome regions studied

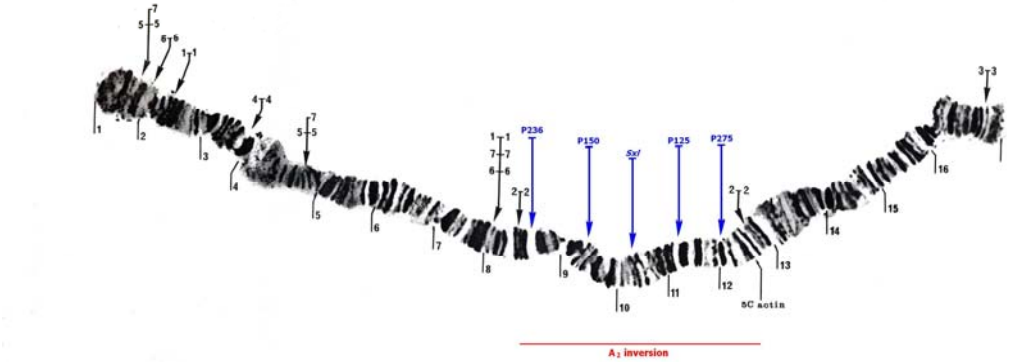
Five gene regions of the X chromosome were chosen for this study. They mapped inside the A_2 inversion of *D. subobscura*, and were chosen to span all the inversion length (Figure 4.1). Four regions were isolated from a phage library of *D. subobscura* (P236, P150, P125 and P275). The library was obtained using λ EMBL4 cloning vector (Frischauf *et al.* 1983). The fifth region, corresponded to a fragment of the *Sex-lethal* gene region, that was cloned and sequenced in *D. subobscura* by Penalva *et al.* (1998) and the sequence is available in EMBL database with the acession number X98370.

These genome regions were chosen because they were distributed along the A_2 inversion, permiting spanning the whole inversion. The distance of each region to the

nearest breakpoint of the inversion was estimated assuming a homogenous distribution of DNA along the A_2 inversion and the data published for the related species *D. pseudoobscura* (Richards *et al.* 2005). These distances expressed in centiMorgan and in Megabases are shown in Table 4.1.

Figure 4.1.

Photomap of X chromosome of *D. subobscura* showing the regions studied.



The five regions studied are indicated in the image, as well as several inversion breakpoints identified for this chromosome. The chromosome corresponds to the A_{st} arrangement (modified from Krimbas 1993).

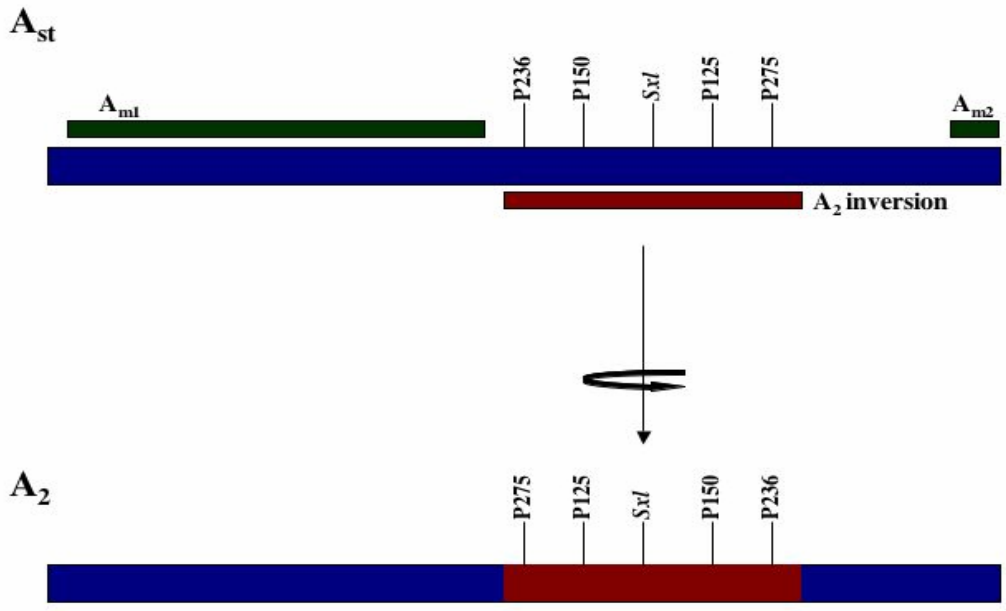
Table 4.1.

Estimates of distance between each studied region and the nearest inversion breakpoint.

		cM (centiMorgan)	Mb (Megabases)
X chromosome total length		150cM (Spurway 1945)	25.8 Mb (Richards *et al.* 2005)
Inversion A_2 length		41.3	7.1
Distance to nearest breakpoint	**P236**	0.5	0.08
	P150	8.8	1.5
	Sxl	16.4	2.8
	P125	16.6	2.9
	P275	7.0	1.2

The distribution of the five regions in the different chromosomal arrangements of the *D. subobscura* X chromosome is shown in Figure 4.2. None of the fixed inversions between *D. madeirensis* and *D. subobscura* affected the distribution of the genome regions studied.

Figure 4.2.
Location in the different chromosomal arrangements of *D. subobscura* of the five genome regions studied.



A_{m1} and A_{m2} are the two inversions fixed on the X chromosome between *D. madeirensis* and *D. subobscura*.

4.2. Experimental procedure

4.2.1. Bacterial strains

Two bacterial strains of *Escherichia coli* were used in this study. The MRA(P2) strain was used to prepare an exponential bacterial culture in order to perform the

infection and amplification of the phages. The XL1-Blue strain was used to carry out transformations with recombinant plasmids.

4.2.2. Amplification of phages and DNA extraction

Two different procedures were used to infect bacteria with phages: the phage inoculation procedure and the phage dilution plating procedure. When infections were sucessfully acchieved and debris were evident, phage DNA was extracted using the Qiagen® Lambda kit following the manufacturer's instructions.

Phage inoculation procedure

- Add 100-125 μl of a exponential bacterial culture with $MgSO_4$ to a plastic tube and 25-50 μl of the phage solution.
- Make a control with no phages
- Incubate at 37ºC for 20 minutes.
- Add 10 ml of NYCYM medium and incubate overnight.
- Add approximately 20 μl of chloroform.
- Incubate at 37ºC for 15 minutes and verify whether there are debris.
- Centrifuge at 4ºC for 10 minutes at 3.000 rpm.
- Transfer the supernatant to a 30 ml plastic tube.

Phage dilution plating procedure

- Add 2.5 ml of preboiled top agarose in a glass reaction tube and keep it at 42ºC.
- Add 100 μl of exponential bacterial culture (with $MgSO_4$) and the corresponding volume of different phage dilutions:
    - 1 μl of a 1/10 dilution in SM gel of the phage solution.
    - 1 μl phage solution.
    - 10 μl phage solution.
    - no phages (control).

- Incube at 37ºC for 20 minutes.
- Add the 2.5 ml of top agarose to the exponential+phage solution and immediately throw it on a Petri plat with LB agar medium.
- Keep 5 minutes at room temperature.
- Incubate plates at 37ºC overnight.

4.2.3. Checking the DNA extraction

An electrophoresis with 0.8% agarose (0.8 gr of agarose to 100 ml of TBE 100 mM) gels was performed to visualize the phage DNA extraction. A mix of phage DNA (1 μl phage DNA, 3 μl loading buffer and 6 μl of distilled water) was loaded into the gel wells. The 1 kB ladder was used as a fragment size marker. After running in a electrophoresis apparatus for approximately one hour, the gel was stained with ethidium bromide and visualized in a UV transiluminator.

This procedure was repetead to verify results obtained in DNA extractions, digestions, and PCR-amplifications and purifications. The amount of DNA loaded into the gel depended on the process. For instance, after PCR-amplification, the volume loaded in each well was at least 4 μl.

4.2.4. Phage DNA digestion and purification

Purified phage DNA was digested with six-cutter restriction enzymes (Sambrook and Russell 2001). First, a small digestion of 10 μl total volume was made with the *EcoRI*. Depending on the results of this digestion, additional digestions were made with other enzymes. The enzyme that cutted the phage DNA in several fragments with different size, was the enzyme chosen to perform a digestion of 25 μl. P236, P275 and P125 were digested with *EcoRI* and P150 with *SalI*. The vector used for the subcloning procedure (pBluescript® II SK^+ phagemid, Stratagene) was also digested

independently with these two enzymes. The phage and vector digestions were purified with Microcon-PCR columns (Millipore) following the manufacturer's instructions.

4.2.5. Ligation with the cloning vector

A 10 μl ligation reaction was performed using the purified digested phage DNA, the purified digested and dephosphorilated vector and DNA ligase. Ligation was incubated overnight at 14ºC.

4.2.6. Bacteria transformation

Competent bacteria (strain XL1-Blue) were prepared by the protocol of Hanahan (1983). The transformation of competent cells was performed by the cold-shock procedure as described by Sambrook and Russell (2001). The selective X-Gal/IPTG system was used to distinguish the recombinant colonies (white) from those that only contained the vector (blue colonies).

4.2.7. Screening of the recombinant colonies and amplification of the inserts for sequencing

The procedure followed was a modification of the Kilger and Schimd (1994) protocol. At least 12 recombinat colonies from each phage were screened to determine the length of the insert. Inserts were amplified by a PCR reaction using the universal primers (T7 and T3) of the vector. An overnight culture of each PCR-amplified colony was also prepared.

Choosing the colonies procedure

- In a sterile environment, a colony was picked with a small stick and spread on a 0.5 ml PCR tube.
- Put the stick in a glass reaction tube with 5 ml of LB liquid and 5 μl of ampicillin.
- Heat the PCR tubes on a microwave for 3 minutes, then add 25 μl of PCR mix.
- Incubate the reaction tubes (containing the stick) at 37ºC overnight.

A 25 μl PCR amplification reaction was performed in a Perkin Elmer GeneAmp PCR system 2400 or Applied Biosystems GeneAmp PCR system 2700), using the 12 bacterial colonies picked for each one of the phage and using the universal primers T3 (5'AATTAACCCTCACTAAAGGG3') and T7 (5'GCCCTATAGTGAGTCGTATTAC3'). The 25 μl amplification reactions contained: 0.5 μl of primer T3 (100 ng/μl), 0.5 μl of primer T7 (100 ng/μl), 2 μl of dNTP's (100 mM), 2.5 μl of buffer (with $MgCl_2$), 0.125 μl of *Taq polymerase* (Pharmacia) and 19.3 μl of distilled water. The amplification conditions were: an initial cicle 94ºC – 45'', 46ºC – 45'', 72ºC – 2'; then 28 cicles 94ºC – 45'', 46ºC – 45'', 72ºC – 4'; and a final cicle 94ºC – 45'', 46ºC – 45'', 72ºC – 8'.

After the selection of the colonies according to the insert length, a plasmid DNA extraction was made from the overnight cultures. About 2 ml of bacterial culture were centrifuged, the supernatant was discarded and again 2 ml of the culture were added to the tube and centrifuged. The bacterial pellet was the starting matherial for the plasmid DNA extraction, that was performed with the Qiagen® Lambda kit according to manufacturer's instructions. A modification was made to the standard protocol: DNA was eluted in water pre-heated for 10 minutes at 75ºC.

The purified plasmid DNA (3μl), or alternatively 1 μl of the bacterial culture dried for 3 minutes in the microwaves were used as templates for the amplification reactions. These reactions were performed as previously described but in a volume of

100 μl. The PCR amplification conditions were the same previously described. The PCR amplification products were purified with Microcon-PCR columnes (Millipore) and eluted in 50 μl of distilled water. This procedure for PCR-amplification ad purification of the PCR-products was repetead along the study changing some conditions and concentrations and the purification protocol depending on the amplification.

4.2.8. Phage insert sequencing

The chosen inserts from each phage were sequenced using the universal primers SK (5'CGCTCTAGAACTAGTGGATC3') and T7. The amount of the purified DNA used in each sequencing reaction depended on the intensity of the band visualized in the agarose gel. The Bigdyes 3.0 or 3.1 kit (Applied Biosystems) was used for the sequencing reactions. The 10 μl sequencing reactions contained: 3.5–6.5 μl of DNA, 0–3 μl of distilled water, 1.5 μl of primer (10 ng/μl) and 2 μl of BigDye. The amplification conditions were: 94ºC – 4', then 25 cicles of 96ºC – 10'', 50ºC – 5'', 60ºC – 4'.

The precipitation of the amplification products was made using the Applied Biosystems protocol. Sequencing was performed in a ABI 377 or ABI 3700 (Applied Biosystems) automated DNA sequencer in Serveis Científico-Tècnics de la Universitat de Barcelona.

4.2.9. BLAST search in the *Drosophila melanogaster* genome

The obtained sequences were used to perform a BLAST search (http://flybase.net/blast) in the *D. melanogaster* genome (release 3.1). From the multiples homologies found for each one of the phages, an approximately 2 kb

region was chosen for the posterior study of nucleotide variation in the *Drosophila* lines. For the selected region, amplification primers were designed based in the obtained *D. subobscura* sequences. The region was PCR-amplified in the *chcu* strain and both strands of the amplification product were sequenced by primer walking. Thereafter, the amplification and sequencing of the selected regions in the *Drosophila* lines was started.

4.2.10. *Drosophila* lines studied

The *Drosophila subobscura* flies were collected in the localities of Ribeiro Frio and Cancela in the Madeira Island, and *D. madeirensis* flies were sampled in Ribeiro Frio. Lines were estabilished from gravid females and maintained in standard agar medium at 18ºC. Twelve generations of inbreeding by sib mating were perfomed in order to try to reduce heterozigosity. The chromosomal arrangement for the X chromosome of the *D. subobscura* lines was determined by crosses with the *chcu* strain wih the A_{st} arrangement or with an A_{2+6} strainkindly provided by M. Papaceit. Table 4.2 shows the lines studied for each species, and in *D. subobscura* for each one of the chromosomal arrangements. Twelve *D. madeirensis* lines and 18 *D. subobscura*, 12 with the A_2 arrangement and 6 A_{st}, were studied.

4.2.11. Chromosome slides

This method allowed a quick and quite simple observation of polytene chromosomes of third-instar *Drosophila* larvae, in order to determine the chromosome arrangement in a optic microscope (the classification of lines according to their chromosomal arrangement was done by Dra Carmen Segarra and Dra Mahnaz Khadem.

Chromosome slides protocol

- Put a larvae of 3th development stage in a slide and add a drop of saline solution.
- Determine the larvae sex. Males were discarded.
- Extract the salivary glands after dissecting the larvae.
- Clean the salivary glands from other tissues.
- Transfer the glands to a drop of acetic orcein in a clean slide.
- Keep it for 20 minutes.
- Cover the slides with a cover slip.
- Observe the slides in a optic microscope.

Table 4.2.
***Drosophila* lines studied.**

	Lines	Total
***D. sub* A_2**	1, 3, 8, 9, 10, 12, 15, 18, 02-a, 02-1, 02-4, 97-1	**12**
***D. sub* A_{st}**	can 19, can 20, can 39, can 50, can 51, can 72	**6**
D. mad	02-I, V, 01.27, 01.2, 01.37, 01.13, 01.17, 01.10, 01.56, 01.8, 01.30, 01.23	**12**

D. sub A_2, *D. subobscura* lines with the A_2 arrangement; *D. sub* A_{st}, *D. subobscura* lines with the A_{st} arrangement and *D. mad*, *D. madeirensis* lines.

4.2.12. *Drosophila* gDNA extraction

A single male from each one of the lines studied was used for DNA extraction and posterior PCR-amplification and sequencing. The extraction of DNA from a single male was decided in order to avoid possible problems of heterozigosity in the lines and to be sure of the interlocus linkage. The procedure used for the extraction was a modification of Ashburner (1989) protocol 48.

DNA extraction protocol

- Homogenize a single fly in a microcentrifuge tube with 100 μl of solution A (for 1 ml: 100 μl Tris HCl 1M pH 9.0; 200 μl EDTA 0.5M pH 8.0; 50 μl SDS 20% and 650 μl distilled water).
- Incubate for 20-30 minutes at 65ºC.
- Add 14 μl of potassium acetate 8M and keep 30 minutes in ice.
- Centrifuge 10 minutes at 10.000 rpm
- Transfer the supernatant to a fresh microcentrifuge tube, taking care of not to remove the pellet.
- Adjust the volume at 300 μl with TE (for 150 ml: 1.5 ml Tris HCl 1M pH 8.0; 300 μl EDTA 0.5M and adjust the volume to 150 ml with distilled water).
- Add one volume of phenol/chlorophorm (1:1), mix very well by inversion and centrifuge 10 minutes at 12.000 rpm.
- Transfer the supernatant to a new microcentrifuge tube, add a volume of chlorophorm, mix very well by inversion and centrifuge for a minute at 12.000 rpm.
- Transfer the supernatant to a microcentrifuge tube, add 0.5 volumes of amonium acetate and 2,5 volumes of absolute ethanol cold. Mix it very well by inversion and let it overnight at –20ºC or, alternatively, for an hour at –80ºC.
- Centrifuge at 4ºC for 15 minutes at 12.000 rpm.
- Throw out the supernatant, wash the pellet with 1.5 ml of 70% ethanol and centrifuge 5 minutes at 12.000 rpm.
- Throw out the supernatant and dry the pellet in a vacuum centrifuge for approximately 5 minutes.
- Ressuspend the pellet with 20 μl of distilled water.

4.2.13. Amplification of the genome regions in *Drosophila*

PCR-amplifications were performed in reactions of 25 or 100 μl. The 25 μl reactions contained: 0.5 μl of primer 1 (100 ng/μl), 0.5 μl of primer 2 (100 ng/μl), 2 μl of dNTP's (100 mM), 2.5 μl of buffer (with $MgCl_2$), 0.125 μl of Taq polymerase (Pharmacia), 18.875 μl of distilled water and 0.5 μl of gDNA. For the 100 μl reactions the concentrations were as follow: 2 μl of primer 1 (100 ng/μl), 2 μl of primer 2 (100

ng/μl), 8 μl of dNTP's (100 mM), 10 μl of buffer (with $MgCl_2$), 0.5 μl of Taq polymerase, 75.5 μl of distilled water and 2 μl of gDNa. The sequences of primers used for the PCR-amplifications are shown in appendix A.

The amplification conditions used were similar for all regions with 28 cycles of amplification, an extension temperature of 68ºC for two minutes in each cycle, and an anneling temperature that ranged from 50ºC (in P125) to 58ºC (in *Sex-lethal*). A detailed description of the amplification conditions for each one of the genome regions studied is shown in the appendix B.

4.2.14. Purification of PCR products

Three different procedures were used to purify the PCR-amplified fragments along the study: the Microcon-PCR kit (Millipore), the MiniElute Purification kit (Qiagen) and a modification of a purification protocol from Dean *et al*. (2003). The two kits were used following the manufacturer's instructions.

Purification protocol (modified from Dean *et al.* 2003)

- For a 25 μl of PCR reaction, add 10.5 μl of NH_4Ac 7.5M and 31.5 μl of cold 100% ethanol.
- Centrifuge at 2200 xg for 15 minutes.
- Aspirate carefully the supernadants with a Pasteur pipette.
- Add 200 μl of 70% ethanol and centrifuge for 1 minute at 2200 xg.
- Aspirate the supernadant like in step 3.
- Dry the sample in a vacuum centrifuge for 3-5 minutes.
- Ressuspend the pellet in 25 μl of deionized water and keep at –20ºC.

4.2.15. Sequencing of PCR-amplified regions

The genome regions were sequenced in *Drosophila* using the BigDyes 3.0 or 3.1 (Applied Biosystems) kit. Sequencing reactions had a total volume of 10 μl: 3.5 – 6.5 μl of DNA, 0 – 3 μl of distilled water, 1.5 μl of primer (10 ng/μl) and 2 μl of BigDye 3.0 or 3.1. The sequencing conditions were: 94ºC – 4' and then 25 cicles of 96ºC – 10'', 50ºC – 5'', 60ºC – 4'. A set of different primers were used in order to sequence completely both strands of each region. The sequences of the primers designed for each region are shown in appendix C.

4.3. Data analysis

4.3.1. Computer programs

4.3.1.1. Sequence alignment

- *Megalign (Lasergene v6.0)* – Multiple and pairwise sequence alignment.
- *ClustalX v1.81* (Thompson *et al.* 1997) – Multiple and pairwise sequence alignment.
- *MacClade v4.05* (Maddison and Maddison 2002) – Multiple sequence alignment edition.

4.3.1.2. Primers design

- *Oligo v4.05* (Rychlik 1992) – Primer design based on termodynamic parameters and nucleotide sequence complementarity.

4.3.1.3. Nucleotide variation and interspecific divergence analyses

- *DnaSP v4.10* (Rozas *et al.* 2003) – Estimates of nucleotide variation, DNA population differentiation, gene conversion, gene flux, linkage disequilibrium, recombination and performance of neutrality tests.
- *Sites* (Hey and Wakeley 1997) – Estimates of the γ recombination parameter.
- *HKA* – Computer program to carry out the multilocus HKA (Hudson *et al.* 1987), Tajima's *D* (Tajima 1989) and the Fu and Li's *D* (Fu and Li 1993) tests.
- *WH* (Wakeley and Hey 1997; Wang *et al.* 1997) – Computer program to carry out the isolation model of speciation test and the linkage disequilibrium test of gene flow.

4.3.1.4. Genealogy reconstruction

- *Mega v3.0* (Kumar *et al.* 2004) – phylogenetic reconstruction by the *neigbour-joining* method.
- *TreeView v1.6.6* (Page 2000) – phylogenetic trees edition

4.3.2. Estimates

4.3.2.1. DNA polymorphism within populations

Average number of nucleotide differences (*k*)

The average number of nucleotide differences is defined by

$$\hat{k} = \sum_{i<j} k_{ij} / \left(\frac{n}{2}\right) \qquad \text{(Tajima 1983, equation A3)}$$

where k_{ij} is the number of nucleotide differences between the *i*th and *j*th sequences in a sample of n DNA sequences.

Nucleotide diversity (π)

The nucleotide diversity is the average number of nucleotide differences per site between two sequences in a randomly mating population. It is estimated by

$$\hat{\pi} = \frac{n}{n-1}\left[\sum_{ij} x_i x_j \pi_{ij}\right] \qquad \text{(Nei 1987, equation 10.5)}$$

where n, is the number of sequences, x_1 and x_j, the frequency of *i*th and *j*th type of DNA sequence, respectively, and π_{ij} is the proportion of nucleotide differences between the *i*th and *j*th type of DNA sequences.

Nucleotide heterozygosity (θ)

This measure takes into account the number of polymorphic sites in n DNA sequences, and assumes that the population is at equilibrium and that all variants are neutral. When population is at mutation-drift equilibrium, θ is equal to the nucleotide diversity. For autosomal loci, the θ parameter is equal to $4N_e\mu$, where N_e is the effective population size and μ the neutral mutation rate. This parameter is estimated by:

$$\hat{\theta} = \frac{S_i}{m_i \sum_{i=1}^{n} \frac{1}{i}} \qquad \text{(Watterson 1975, equation 1.4a)}$$

where S_i is the number of polymorphic sites and m_i is the total number of sites.

4.3.2.2. DNA divergence between populations

Average number of nucleotide differences between populations (D_{xy})

The average number of nucleotide substitutions between populations (D_{xy}) is estimated by:

$$D_{XY} = \sum_{ij} x_i y_j d_{ij}$$ (Nei 1987, equation 10.20)

where x_i and y_i are the frequency of haplotype ith and jth for populations X and Y respectively, and d_{ij} is the proportion of nucleotide differences between the ith haplotype from X and the jth haplotype from Y.

Net number of substitutions between two populations (D_a)

The net number of nucleotide substitutions between two populations (D_a) is estimated by:

$$D_a = D_{xy} - \frac{(\pi_x + \pi_y)}{2}$$ (Nei 1987, equation 10.21)

where π_x and π_y are the nucleotide diversity for the X and Y populations, respectively.

4.3.2.3. Recombination

The minimum number of recombination events in the history of a sample (R_m) and was estimated by the four-gamete test as proposed by Hudson and Kaplan (1985). The recombination parameter R (Hudson 1987) was also estimated. This estimator is based on the variance of the average number of nucleotide differences between pairs of sequences. Usually, the estimate of R (=4Nr) is given per gene (r,

is the recombination rate per generation between the most distant sites). A third recombination parameter was estimated: γ is also an estimate of the population recombination rate *4Nc*, where *c* is the recombination rate per generation per base pair (Hey and Wakeley 1997).

4.3.2.4. Linkage disequilibrium

Linkage disequilibrium analysis

An important characteristic of the pattern of nucleotide polymorphism is the linkage disequilibrium or putative association between variants at different polymorphic sites. Linkage disequilibrium (LD), was estimated by the following parameters: *D* (Lewontin and Kojima 1960), *D'* (Lewontin 1964), *R* and R^2 (Hill and Robertson 1968).

Both the two-tailed Fisher's exact test and the chi-square test were performed to determine whether the associations between polymorphic site were significant. The Bonferroni correction for multiple tests was also performed. This procedure tries to avoid spurious rejections of the null hypothesis when performing multiple tests.

Linkage disequilibrium test of gene flow

Machado *et al.* (2002) introduced a test of gene flow based on patterns of linkage disequilibrium among specific classes of segregating sites, that is, using a subset of the total intragenic linkage disequilibrium. Under a scenario of gene flow, linkage disequilibrium among pairs of shared polymorphisms (average = DSS) in the recipient species should tend to be positive, and linkage disequilibrium among pairs of sites where one member is a shared and the other an exclusive polymorphism

(average = DSX) should tend to be negative. Then for species *I*, which shares some polymorphisms with species *j* and for locus *k*, the following parameter was proposed:

$$x_{(I,J)k} = DSS_{(I,J)k} - DSX_{(I,J)k}$$

In principle, *x* can be estimated for any measure of linkage disequilibrium. D' was used in this study. D' is equal to the conventional measure of linkage disequilibrium divided by the maximum value given by the allele frequencies (Lewontin 1964). To test whether the observed values of *x* were consistent with the isolation model, computer simulations were performed (these simulations were implemented in the isolation model program that is described later).

4.3.2.5. Genetic differentiation test between populations

Permutation test

This statistical test was proposed by Hudson *et al.* (1992b) and is based on the *Kst** statistic that measures the genetic differentiation between two or more subpopulations from the nucleotide variation of DNA sequences. The statistical significance of *Kst** is determined by Monte Carlo simulations.

Hypergeometric distribution

The hypergeometric distribution allows estimating the probability that the observed number of shared polymorphic sites between gene arrangements has arisen by chance (by independent accumulation of mutations in each gene arrangement). Assuming that all sites have the same probability to be polymorphic, this probability is:

$$P(k) = \frac{\left(\frac{m}{k}\right)\left(\frac{3N-m}{n-k}\right)}{\left(\frac{3N}{n}\right)}$$

where *m* and *n* are the observed number of polymorphic sites in each gene arrangement, *k* is the number of shared polymorphic sites between the two gene arrangements, and *N* is the total number of silent sites.

4.3.2.6. Gene flow

Average level of gene flow (N_m)

The average level of gene flow was computed from the F_{ST} estimate assuming the island model of population structure (Wright 1951, Hudson *et al.* 1992a). The average level of gene flow was estimated by

$$\langle N_m \rangle_F = \left(\frac{1}{F_{ST}} - 1 \right)$$ (Hudson *et al.* 1992a, equation 2)

Proportion of nucleotide diversity due to variation between populations (F_{ST})

F_{ST} is estimated by

$$\langle F_{ST} \rangle = 1 - \frac{H_w}{H_b}$$ (Hudson *et al.* 1992a, equation 3)

where H_w is mean number of differences between different sequences sampled from the same population, and H_b is the mean number of differences between sequences sampled from the two different subpopulations.

4.3.2.7. Neutrality tests

Tajima's test (1989)

Tajima (1989) proposed a statistical method to test the neutral model of molecular evolution by using intraspecific DNA polymorphism data. This test consider several assumptions, random mating in a population of *N* diploid individuals, no selection and no recombination between DNA sequences. It also assumes that the number of sites on a DNA sequence is so large that a newly arisen mutation takes place at a site different from the site where the previous mutations have occurred (infinite sites model – Kimura 1969). Tajima's test (1989) is based on a comparison between two estimators of the heterozygosity per sequence: the number of segregating (or polymorphic) sites (*S*) in a sample

$$E(S)=a_1M \quad \text{(Tajima 1989, equation 1)}$$

and the average number (*k*) of pairwise nucleotide differences between the DNA sequences

$$E(k)=M \quad \text{(Tajima 1983, equation 23)}$$

Thus, Tajima's *D* statistic allows determining if the frequency distribution of the polymorphic variants is significantly different from the distribution expected under the neutral model

$$D=\frac{d}{\sqrt{\hat{V}(d)}}=\frac{k-\hat{\frac{S}{a_1}}}{\sqrt{e_1S+e_2S(S-1)}} \quad \text{(Tajima 1989, equation 38)}$$

where a_1, e_1 and e_2 are functions of the sample size (*n*).

Fu and Li's tests (1993)

Fu and Li's tests (1993) are based on comparison of different estimates of the heterozygosity per sequence inferred from the number of mutations in the external and internal branches in the genealogies of neutral alleles. In a genealogy, "old" mutations will tend to acumúlate in the internal branches, while "new" mutations will tend to accumulate in the external branches It is expected to be an excess of mutations in the external branches of a genealogy if an advantageous allele has recently become fixed in the population, because then the majority of the mutations in the population are expected to be relatively young. On the other hand, if balancing selection is operating at a locus, then some alleles may be old and thus there may be a deficiency of mutations in the external branches. These tests are based on the assumption of no recombination, no migration and a constant population size.

Fu and Li's proponed two tests statistics (*D* and *F*) that can be estimated when the sequence o fan outgroup species is available or without outgroup. The tests without outgroup are based on the number of singletons in the sample, while the tests with an outgroup are based on the number of mutations in the external branches of the genealogy. The *D* statistic is based on the total number of mutations and the number of mutations in the external branches of the genealogy (ot the number of singletons), while the *F* statistic is based on the average number of nucleotide differences between sequences and the number of mutations on the external branches of the genealogy (or the number of singletons).

D-statistic test with outgroup,

$$D = \frac{\eta - a_n \eta_e}{\sqrt{u_D \eta - v_D \eta^2}}$$

where η and η_e are the total number of mutations and the number of mutations in the external branches, respectively, and a_n, u_D and v_D are functions of the simple size.

*D**-statistic test without outgroup,

$$D^* = \frac{\left(\frac{n}{n-1}\right)\eta - a_n\eta_s}{\sqrt{u_D^*\eta - v_D^*\eta^2}}$$

where η and η_s are the total number of mutations and the number of singletons mutations in the sample, respectively, and a_n, u_D^* and v_D^* are functions of the sample size.

F-statistic test with outgroup,

$$F = \frac{k - \eta_e}{\sqrt{u_F\eta - v_F\eta^2}}$$

where *k* and η_e are the average number of nucleotide differences between sequences and the number of singletons in the sample, respectively, and u_F and v_F are functions of the simple size.

*F**-statistic test without outgroup,

$$F^* = \frac{k - \frac{n-1}{n}\eta_s}{\sqrt{u_F^*\eta - v_F^*\eta^2}}$$

where k and η_s are the average number of nucleotide differences between sequences and the number of singletons in the sample, respectively and u_F^* and v_F^* are functions of the sample size.

Hudson, Kreitman and Aguadé (HKA)test

The neutral theory of molecular evolution predicts that the amount of within-species diversity should be correlated with the level of between-species divergence, since both depended on the neutral mutation rate (Kimura 1983). The Hudson, Kreitman and Aguadé test (1987) evaluates the fit of polymorphism and divergence data to this prediction. The test requires data on the level of within-species polymorphism data and between-species divergence at least at two loci. Therefore, the relatives amounts of polymorphism and divergence can be compared across loci. This test assumes that all loci are unliked, that species are stationary at the time of sampling and that the two species sampled derived from a single ancestral population T generations ago.

4.3.2.8. Speciation models

Wakeley and Hey isolation test

This test was developed to fit a general model of speciation via isolation to polymorphism data from two closely related populations or species (Wakeley and Hey 1997, Wang *et al.* 1997). The isolation model assumes that two descendent populations formed from an ancestral population at a single point in the past and that there was no gene flow between the populations alter that point. Each one of the three populations have constant sizes, although the size may differ between them. The input data are the counts of four types of polymorphic sites: polymorphisms that are exclusive to species 1, polymorhisms exclusive to species 2,

polymorphisms that are shared by the two species and polymorphisms that appear as fixed differences between the two species. The recombination parameter γ (Hey and Wakeley 1997) estimates for each species are also used in the input data. The method yields estimates of the population mutation parameter θ, which is equal to *4Nu*, where *N* is the effective population size and *u* is the neutral mutation rate. Since there are three species (species 1, species 2, and the ancestral species) and they may have different effective population size, there are three population mutation parameters estimates: θ_1, θ_2 and θ_A. The method also yields an estimate of the time since isolation *T*, in units of $2N_1$ generations.

The fitting of the data with the isolation model without gene flor was tested alter 10.000 coalescent simulations in base of the tests statistics χ^2 and *wwh* (Wang *et al.* 1997).

The χ^2 statistic is based on the discrepancies between observations and expectations for each locus and each type of polymorphic sites. If we denote the counts of the four types of polymorphic sites for a locus *u* as $S_{i,j}$ with $j = 1...4$, and if there are *L* loci, then

$$\chi^2 = \sum_{i=1}^{L} \sum_{j=1}^{4} \frac{(S_{i,j} - E(S_{i,j}))^2}{E(S_{i,j})}$$

When there is some genetic Exchange between incipient species due to hybridization, gene flor is expected to affect differentialy to different loci. Indeed, gene flor is expected to be low in loci directly implied in the speciation process. In contrast, it is not expected gene flor to be prevented in loci that are not directly envolved in speciation. Therefore, according to a model of speciation without isolation, great differences in the number of shared polymorphic sites and in the number of fixed differences are expected among loci. This reasoning suggests a test

statistic that would have a high value when there is lots of variation among loci for fixed polymorphisms, and when there is lots of variation among loci for shared polymorphisms. The *wwh* test (Wang *et al.* 1997) is a measure of variation in fixed and shared differences. It can be defined as the difference between the highest and lowest values of fixed differences among the loci plus the difference between the highest and lowest values of shared polymorphism.

5. RESULTS

5.1. Genomic regions analysis

5.1.1. P236 region

5.1.1.1. Studied Region

Phage P236 was digested with *EcoRI* and four fragments were obtained: two of about 5 kb, and two of 3 kb. These fragments were subcloned and sequenced using T7 and SK universal primers. The BLAST program was used to search for conserved sequences in the *D. melanogaster* genome (release 3.1). Multiple sequences with a high similarity were found in a 15 kb fragment, that included the putative gene *CG12625*. A region of about 2 kb was chosen for further analysis. This region included the unique intron, the second exon of *CG12625*, as well as a part of its 3'

flanking region. The selected region was PCR amplified in the *chcu* strain of *D. subobscura* and progressively sequenced using internal primers. Alignment of the obtained sequence with 2kb sequence of *D. melanogaster* showed that the second exon of *CG12625* was not conserved in *D. subobscura*. The same result was found in *D. pseudoobscura*, after a BLAST search with *CG12625* sequence of *D. melanogaster*.

5.1.1.2. Nucleotide polymorphism

The multiple alignment of the P236 region in 12 *D. madeirensis* lines and 18 *D. subobscura* lines (12 with the A_2 arrangement and 6 A_{st}) included a total of 1818 sites. This number dropped to 1531 when sites with alignment gaps were not considered. All sites were silent as P236 corresponded to a non-coding region. A total of 114 polymorphic sites (52 singletons) were detected in the data set. The minimum number of mutations was 118, as there were sites that segregated for more than two variants. Nucleotide variation of P236 region at polymorphic sites is shown in Figure 5.1. Sites with alignment gaps were completely excluded from all analysis.

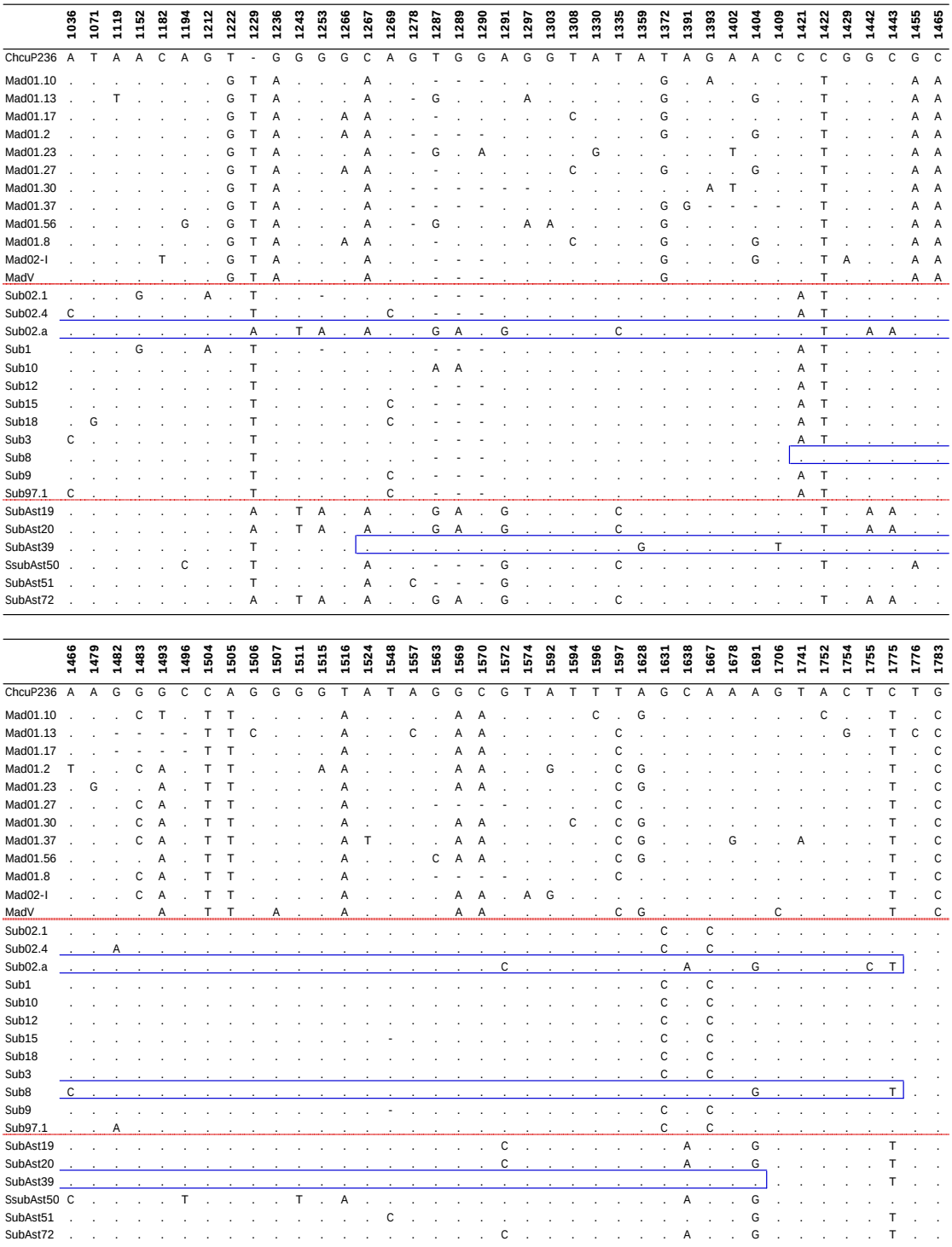
A general description of nucleotide polymorphism at P236 is shown in Table 5.1. The number of polymorphic sites was similar in *D. subobscura* and *D. madeirensis* despite their different sample size. When samples with the same number of lines were compared, the number of polymorphic sites was higher in *D. madeirensis* (12 lines) than in the A_2 arrangement of *D. subobscura* (12 lines). The percentage of singletons was higher in *D. madeirensis* than in *D. subobscura* (76% and 40% respectively). In the latter species, the A_{st} arrangement presented a higher percentage of singletons (64%) than the A_2 (49%). At P236 region, each line of *D. madeirensis* represented a single haplotype. In contrast, 4 different haplotypes were detected in the 6 A_{st} lines and 9 different haplotypes in the 12 A_2 lines. Therefore, haplotype diversity in the *D. subobscura* samples was lower than one.

Figure 5.1.
Polymorphic sites at P236 region (*continues next page*)**.**

	22	28	39	59	70	77	79	89	100	121	124	125	129	133	136	138	139	145	148	151	165	166	167	171	174	179	182	192	198	203	221	240	241	244	255	268	278	294	331
ChcuP236	A	C	G	C	T	G	A	C	T	A	A	A	A	T	A	C	A	T	T	C	T	T	C	C	A	T	G	A	T	T	G	A	A	G	C	A	C	A	C
Mad01.10	.	A	.	A	.	A	G	A	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	A	C	.	.	A	.	.	.	G	A
Mad01.13	.	A	.	A	.	A	G	A	.	.	.	G	.	.	.	.	.	.	A	A	.	C	A	A	.	.	A	.	.	A	C	.	.	.	.	.	.	G	A
Mad01.17	.	A	.	A	.	A	G	A	.	.	.	G	.	.	.	.	.	.	A	A	.	C	A	A	.	.	A	.	.	A	C	.	.	.	.	.	.	G	A
Mad01.2	.	A	.	A	.	A	G	A	.	.	.	G	.	.	.	G	C	.	.	A	.	.	.	A	.	.	A	.	.	A	C	.	.	A	.	.	.	G	A
Mad01.23	.	A	.	A	.	A	.	A	.	.	.	-	-	-	-	-	-	-	-	-	.	.	.	A	.	.	A	T	.	A	C	.	.	.	.	.	.	G	A
Mad01.27	.	A	.	A	.	A	.	A	.	.	.	G	.	.	.	.	.	G	.	A	.	.	.	A	.	C	A	.	.	A	C	.	.	A	.	.	.	G	A
Mad01.30	.	A	.	A	.	A	.	A	A	.	.	G	.	.	.	G	C	.	.	A	.	.	.	A	.	.	A	.	.	A	C	.	.	A	.	.	.	G	-
Mad01.37	.	A	.	A	.	A	.	A	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	A	T	.	A	.	.	A	C	.	.	-	.	.	.	G	A
Mad01.56	.	A	.	A	.	A	.	A	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	A	.	C	A	.	C	A	C	.	.	.	.	.	.	G	A
Mad01.8	.	A	.	A	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	A	.	.	.	.	.	T	.	G	A
Mad02-I	.	A	.	A	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	A	.	.	.	A	.	.	.	G	A
MadV	.	A	.	A	.	A	.	A	.	.	.	G	.	.	.	.	.	.	.	A	.	.	.	A	.	.	A	.	.	A	C	.	.	.	.	.	.	G	A
Sub02.1	G	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
Sub02.4	.	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
Sub02.a	.	A	.	.	.	A	.	A	G	.	C	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.
Sub1	G	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
Sub10	.	A	T	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
Sub12	G	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	-
Sub15	.	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
Sub18	G	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
Sub3	.	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
Sub8	.	A	.	.	.	A	.	A	G	.	C	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	.	.	A	.	A
Sub9	.	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
Sub97.1	.	A	.	.	.	A	.	A	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	G	.	A	.	.	G	.	.	.	.	.	A
SubAst19	.	A	.	.	.	A	.	A	G	.	C	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.
SubAst20	.	A	.	.	.	A	.	A	G	.	C	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.
SubAst39	.	A	.	.	.	A	.	A	G	.	C	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	.	C	.	.	.	.	.	.	.
SsubAst50	.	A	.	.	.	A	.	A	G	G	.	.	C	.	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A
SubAst51	.	.	.	.	C	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
SubAst72	.	A	.	.	.	A	.	A	G	.	C	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.

	362	425	435	446	457	460	466	469	474	519	528	534	586	590	627	628	631	636	641	642	645	659	660	661	672	673	740	760	797	804	829	834	844	890	964	972	973	1015	1028
ChcuP236	T	T	A	G	T	-	-	-	-	C	T	A	T	C	G	C	G	A	T	T	G	C	A	G	A	G	A	C	C	T	C	A	T	G	G	C	T	C	C
Mad01.10	C	.	.	.	.	T	-	-	-	T	A	C	.	.	.	.	A	.	.	C	C	.	.	.	-	-	.	.	A	.	.	.	.	.	T	A	.	.	T
Mad01.13	C	.	G	.	.	-	-	-	-	T	A	C	.	T	.	.	A	.	.	C	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	T	T
Mad01.17	C	.	.	.	.	-	-	-	-	T	A	C	.	.	.	.	A	.	.	C	C	.	.	.	-	-	.	.	A	.	.	.	.	.	.	A	.	.	T
Mad01.2	C	.	.	.	.	T	T	T	-	T	A	C	.	.	.	.	A	.	.	C	C	.	.	.	-	-	.	.	.	.	.	.	.	.	.	A	.	.	T
Mad01.23	C	.	.	.	.	T	-	-	-	T	A	C	.	.	.	.	A	G	.	C	C	.	.	.	-	-	.	.	A	.	.	.	.	.	.	A	.	.	T
Mad01.27	C	G	.	.	-	-	-	-	-	T	A	C	.	.	.	.	A	.	.	C	C	.	.	.	-	-	.	.	A	.	.	.	.	.	.	A	.	.	T
Mad01.30	C	.	.	.	.	T	T	T	C	T	A	C	.	.	.	.	A	.	.	C	C	.	.	.	-	-	.	.	.	.	.	.	.	A	.	A	.	.	T
Mad01.37	C	.	.	.	.	T	T	T	C	T	A	C	.	.	.	.	.	.	.	C	C	.	.	.	.	T	.	.	A	.	.	.	.	.	.	A	.	.	T
Mad01.56	C	.	.	.	.	T	T	-	-	T	A	C	.	.	.	.	A	.	.	C	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	T
Mad01.8	C	.	.	.	.	T	T	-	-	T	A	C	.	.	.	.	A	.	.	C	C	.	.	C	-	-	.	.	A	C	.	.	.	.	.	A	.	.	T
Mad02-I	C	.	.	.	.	-	-	-	-	T	A	C	.	.	.	.	A	.	.	C	C	.	.	.	-	-	.	.	A	.	.	.	.	.	.	A	.	.	T
MadV	C	.	.	.	-	-	-	-	-	T	A	C	A	.	.	.	A	.	.	C	C	.	.	.	-	-	T	A	.	.	C	.	.	.	.	A	.	.	T
Sub02.1	.	.	.	.	.	T	G	T	-	.	.	.	.	.	A	.	.	.	.	.	.	T	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub02.4	.	.	.	.	.	T	T	G	G	.	A	.	.	.	A	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub02.a	.	.	.	.	G	T	T	-	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub1	.	.	.	.	.	T	G	T	-	.	.	.	.	.	A	.	.	.	.	.	.	T	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub10	.	.	.	.	.	G	-	-	-	.	.	.	.	.	A	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub12	.	.	.	.	.	G	-	-	-	.	.	.	.	.	A	T	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub15	.	.	.	.	.	T	G	T	-	.	.	.	.	.	A	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub18	.	.	.	.	.	T	G	T	-	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub3	.	.	.	.	.	G	-	-	-	.	.	.	.	.	A	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub8	.	.	.	.	-	-	-	-	-	.	.	.	.	.	.	.	.	.	C	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T
Sub9	.	.	.	.	.	T	G	T	-	.	.	.	.	.	A	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Sub97.1	.	.	.	.	.	T	T	G	G	.	A	.	.	.	A	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	T
SubAst19	.	.	.	.	G	T	T	-	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
SubAst20	.	.	.	.	G	T	T	-	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
SubAst39	.	.	.	.	.	T	T	G	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	T
SsubAst50	.	.	.	A	.	T	T	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	.	.	.	.	T
SubAst51	.	.	.	.	.	T	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
SubAst72	.	.	.	.	G	T	T	-	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T

Figure 5.1. *continues from previous page.*



The multiple alignment is given relative to the reference sequence (*chcu* strain) of *D. subobscura*. The number above each site indicates its position in the multiple alignment. The dot (.) indicates the same nucleotide as in the reference sequence, and the dash (-) a deleted nucleotide. The blue boxes indicate the conversion tracts identified between the A_2 and A_{st} arrangements of *D. subobscura*.

Table 5.1.
Nucleotide polymorphism in P236 region.

	n	*len*	*len*^{w}	*S*	*mut*	*sin*	*%sin*	*inf*	*hap*	*H*
D. subobscura **A_{st}**	6	1764	1531	28	28	18	64	10	4	0.800
D. subobscura **A_2**	12	1773	1531	37	37	18	49	19	9	0.955
D. subobscura **Total**	18	1777	1531	50	50	20	40	30	13	0.961
D. madeirensis	12	1794	1531	55	55	42	76	13	12	1

n, sample size or number of sequences; *len*, total number of sites; *len*^{w}, number of sites excluding gaps; *S*, number of polymorphic sites; *mut*, number of mutations; *sin*, number of singletons; *%sin*, percentage of singletons; *inf*, number of parsimony informative sites; *hap*, number of haplotypes; *H*, haplotype diversity.

5.1.1.3. Nucleotide variation

The different estimates of nucleotide variation at P236 are shown in Table 5.2. Nucleotide variation was higher in *D. subobscura* than in *D. madeirensis* according to π (nucleotide diversity), but not according to θ (nucleotide heterozigosity). In *D. subobscura*, the level of variation was similar in both chromosomal arrangements according to θ, but was higher in A_{st} than in A_2 when π was considered.

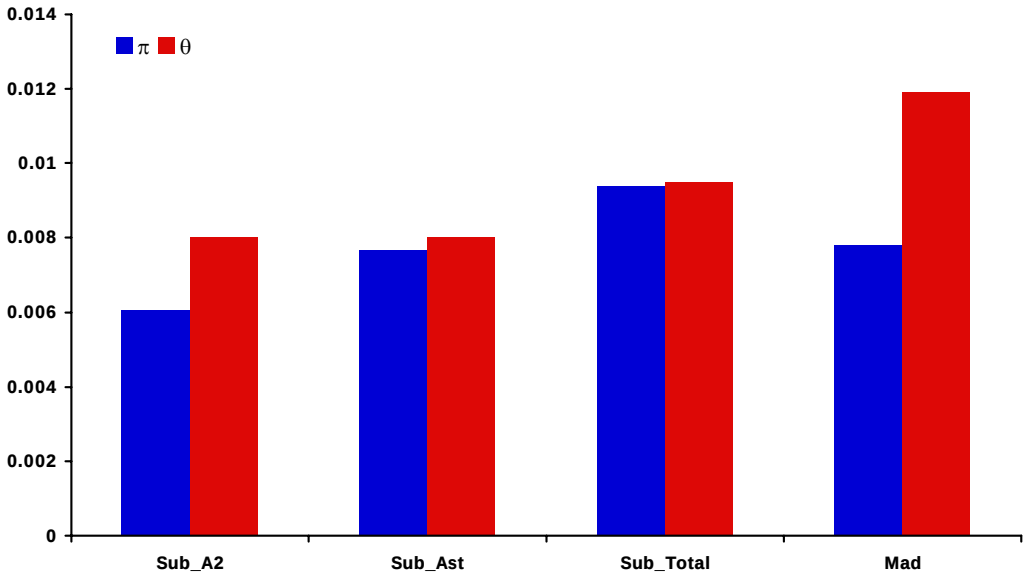
Table 5.2.
Estimates of nucleotide variation in P236 region.

	n	*S*	*k*	π	θ^{seq}	θ^{sit}
D. subobscura **A_{st}**	6	28	11.733	0.0076	12.263	0.0080
D. subobscura **A_2**	12	37	9.273	0.0060	12.252	0.0080
D. subobscura **Total**	18	50	14.366	0.0093	14.537	0.0094
D. madeirensis	12	55	11.955	0.0078	18.213	0.0119

n, sample size or number of sequences; *S*, number of polymorphic sites; *k*, average number of pairwise nucleotide differences; π, nucleotide diversity; θ^{seq}, nucleotide heterozigosity per sequence inferred from *S*; θ^{sit}, nucleotide heterozigosity per site inferred from *S*.

The levels of nucleotide diversity (π) and nucleotide heterozigosity per site (θ) are compared in Figure 5.2. The estimates of θ were higher than estimates of π in all samples, particularly in *D. madeirensis*. This result indicated a general excess of singletons in this region, mainly in the latter species.

Figure 5.2.
Nucleotide diversity (π) and nucleotide heterozigosity per site (θ) estimates.



Sub_A_2, *D. subobscura* lines with A_2 arrangement; Sub_A_{st}, *D. subobscura* lines with A_{st} arrangement; Sub_Total, all *D. subobscura* lines studied (A_2 + A_{st}); Mad, *D. madeirensis* lines; π, nucleotide diversity and θ, nucleotide heterozigosity per site.

5.1.1.4. Recombination

A minimum of 5 recombination events were detected in *D. madeirensis* and 6 in *D. subobscura*, according to the method proposed by Hudson and Kaplan (1985). In the latter species, two recombination events were detected within the A_2 and one within the A_{st} arrangement. The estimate of the recombination parameter *R* per site (Hudson 1987) was 0.8185 in *D. madeirensis*, and 0.0057 in *D. subobscura* (0.0001 and 0.0087 for A_2 and A_{st}, respectively).

5.1.1.5. Linkage disequilibrium

The results of the linkage disequilibrium analysis are shown in Table 5.3. After applying Bonferroni correction for multiple tests, 15 (1.2%) Fisher's tests and 64 (5.2%) chi-square tests remained significant for *D. subobscura*. The global level of linkage disequilibrium (*ZnS* and *Za* estimates) was higher in *D. subobscura* than in *D. madeirensis*. However, the highest level of linkage disequilibrium was detected in the A_{st} arrangement of *D. subobscura*.

Table 5.3.
Linkage disequilibrium analysis for P236 region.

	S_L	*comp*	*Fisher*	*%Fisher*	*chi-square*	*%chi-square*	*ZnS*	*Za*
***D. subobscura* A_{st}**	28	378	0	0	76	20.1	0.3489	0.4393
***D. subobscura* A_2**	37	666	59	8.8	293	43.9	0.3140	0.3342
***D. subobscura* Total**	50	1225	222	18.1	294	24	0.1824	0.2005
D. madeirensis	55	1485	3	0.2	128	8.6	0.0969	0.1051

S_L, number of segregating sites with two variants; *comp*, number of comparisons; *Fisher*, number of significant ($P < 0.05$) comparisons by the Fisher's test; *%Fisher*, percentage of significant ($P < 0.05$) comparisons by the Fisher's test; *chi-square*, number of significant ($P < 0.05$) comparisons by the chi-square test; *%chi-square*, percentage of significant ($P < 0.05$) comparisons by the chi-square test; *ZnS* (Kelly 1997) and *Za* (Rozas *et al.* 2001) are global measures of linkage disequilibrium.

5.1.1.6. Genetic differentiation between *D. subobscura* arrangements

The average number of nucleotide substitutions per site between the two arrangements was 0.0127 (Table 5.4). Putative significant genetic differentiation between the chromosomal arrangements of *D. subobscura*, was contrasted according to the Hudson *et al.* (1992b) permutation test. The statistical significance of the *Kst** statistic was obtained after 1000 replicates. The observed *Kst** value (0.16) was

statistically significant (P = 0.001), therefore, lines of *D. subobscura* with different gene arrangements were not pooled together. Genetic differentiation between arrangements prevailed despite the presence of 15 shared polymorphisms and the absence of fixed differences between them. According to the hypergeometric distribution, the high number of shared polymorphisms cannot be explained by recurrent mutation. So, the high number of shared polymorphisms between the two arrangements was due to genetic exchange between them. Indeed, 5 genes conversion tracts (4 in A_2 and 1 in A_{st}) were identified (Figure. 5.1) by the algorithm proposed by Betrán *et al.* (1997). The largest tract with 1608 nucleotides was detected in *sub02.a* line. Three tracts including 72, 343 and 356 nucleotides, respectively, were detected in line *sub8*. The unique tract identified in the A_{st} arrangement (*subAst39* line) had 407 nucleotides in length.

Table 5.4.
Genetic differentiation between species and chromosomal arrangements of *D. subobscura*.

	Fixed	*Shared*	S_{x1}	S_{x2}	*K*	*Dxy*
A_2 / A_{st}	0	15	22	13	19.583	0.0127
A_2 / *D. mad*	17	1	36	34	40.138	0.0262
A_{st} / *D. mad*	16	2	26	53	37.208	0.0243

A_2, *D. subobscura* lines with A_2 arrangement; A_{st}, *D. subobscura* lines with A_{st} arrangement; *D. mad*, *D. madeirensis* lines; Fixed, fixed differences between samples; Shared, polymorphic sites segregating for the same two variants in two samples; S_{x1}, exclusive polymorphisms sites in population 1; S_{x2}, exclusive polymorphisms sites in population 2; *K*, average number of nucleotide differences between arrangements or species; D_{xy}, average number of nucleotide substitutions per site between samples.

5.1.1.7. Genetic differentiation between species

The average number of nucleotide substitutions per site between the two species was similar, independently of the arrangement used to compare with *D.*

madeirensis (Table 5.4). There were only two shared polymorphisms between *D. madeirensis* and *D. subobscura*, but there were 15 fixed differences between them.

5.1.1.8. Patterns of polymorphism

Tajima's (1989) and Fu and Li's (1993) statistics were all negative, which indicated an excess of singletons variants (Table 5.5). Only the Fu and Li tests with an outgroup were statistically significant for *D. madeirensis*. Also, Tajima's *D* statistic proved to be significant (*P(D)* = 0.046) in this species, after computer simulations under the conservative assumption of no recombination.

Table 5.5.
Tajima's and Fu and Li's statistics for P236 region.

	Tajima's D	*Fu and Li's D*	*Fu and Li's F*	*Fu and Li's D**	*Fu and Li's F**
***D. subobscura* A_{st}**	-0.2729	-0.2924	-0.3488	-0.3831	-0.3929
***D. subobscura* A_2**	-1.1042	-0.5992	-0.9200	-0.5602	-0.8011
***D. subobscura* Total**	-0.0483	-0.4858	-0.4180	-0.5088	-0.4345
D. madeirensis	-1.5813	-2.3849*	-2.6225*	-1.8301	-2.0121

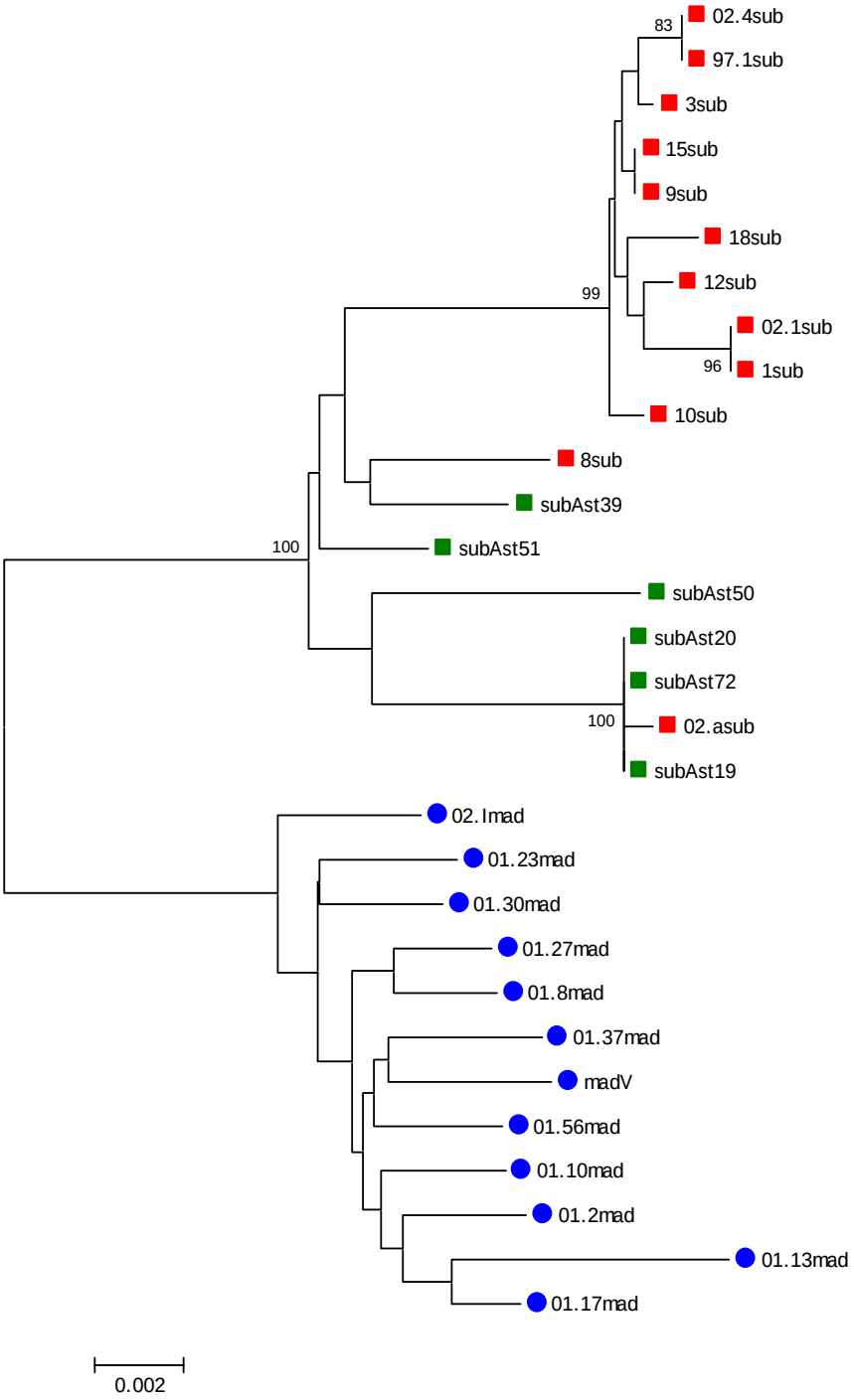
* significant for $P < 0.05$.

5.1.1.9. Gene genealogy

The gene genealogy reconstructed from variation at P236 is shown in Figure 5.3. All *D. madeirensis* lines grouped together in a single cluster. In *D. subobscura* a partial clustering of lines according to their gene arrangement was also detected. The A_{st} lines grouped together in a single cluster, although this cluster also included two A_2 lines (*sub8* and *sub02.a*). However, these two lines presented gene conversion tracts from A_{st}, which may explain their clustering with A_{st} lines.

Figure 5.3.

Gene genealogy reconstructed from nucleotide variation in P236 region.



The gene genealogy was obtained by the neighbour-joining method (Saitou and Nei 1987) using, as genetic distance, the number of substitutions per site, according to the Kimura 2-parameter method (Kimura 1980). Bootstrap confidence level (from 1000 replications) above 70% are shown below the branches. The horizontal bar at the bottom represents the distance scale of the branches.

5.1.2. P150 region

5.1.2.1. Studied region

P150 phage was digested with *SalI*, and a single 5.5 kb fragment was obtained. This fragment was subcloned and its ends were sequenced with T7 and SK universal primers. Two homologies were found in a 5 kb fragment using BLAST algorithm in *D. melanogaster* genome (release 3.1). This 5 kb fragment included part of putative gene *CG15364*. A 1.8 kb region including part of the intron of this gene, the complete second exon and some 3' flanking region was selected for further analysis.

5.1.2.2. Nucleotide polymorphism

The nucleotide variation of P150 region at polymorphic sites is shown in Figure 5.4. The multiple alignment of P150 region in 12 *D. madeirensis* lines, and in 18 *D. subobscura* lines (12 with A_2 arrangement and 6 A_{st}) included a total of 1313 sites. This number decreased to 1262 when sites with alignment gaps were excluded. The number of silent sites was 971, as this region included 291 sites from the second exon of *CG15364*. A total of 61 polymorphic sites (30 singletons) were detected in the data set. The minimum number of mutations was 62, as some sites segregated for more than two variants. Sites with alignment gaps were completely excluded from all analysis.

Figure 5.4.

Polymorphic sites in P150 region.

	49	59	65	73	132	136	137	144	157	159	185	198	218	266	303	324	417	436	441	463	545	547	587	597	600	614	651	666	674	681	699	713	715
ChcuP150	G	G	T	G	A	A	C	A	C	C	C	A	C	C	G	G	G	C	G	C	T	T	C	T	A	C	C	A	T	A	C	T	T
Mad01.10	.	.	.	.	.	.	.	-	.	.	T	C	A	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	.	.
Mad01.13	.	.	.	.	.	.	.	-	.	.	T	C	.	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	-	-
Mad01.17	T	.	.	.	.	.	.	.	.	.	T	C	.	.	.	A	.	.	.	A	C	.	.	.	T	T	.	.	.	.	.	C	A
Mad01.2	.	.	.	.	.	C	.	.	.	.	T	C	.	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	.	.
Mad01.23	.	.	.	.	.	.	.	-	.	.	T	C	.	.	.	A	.	.	.	A	.	.	.	.	C	T	.	.	.	.	.	.	.
Mad01.27	.	.	.	.	.	.	.	-	.	.	T	C	A	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	.	.
Mad01.30	.	.	.	.	.	C	.	.	.	.	T	C	.	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	.	.
Mad01.37	.	.	.	.	.	.	.	-	.	.	T	C	A	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	.	.
Mad01.56	.	.	.	.	.	.	.	-	.	.	T	C	.	T	.	A	.	.	.	A	.	.	.	.	T	T	.	C	.	.	A	.	.
Mad01.8	.	.	.	.	.	C	A	.	.	.	T	C	.	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	.	.
Mad02.I	.	.	.	.	.	.	.	-	.	.	T	C	.	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	-	-
MadV	.	.	.	.	.	C	.	.	.	.	T	C	.	.	.	A	.	.	.	A	.	.	.	.	T	T	.	.	.	.	.	.	.
Sub02.1	.	.	C	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
Sub02.4	T	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	C	A
Sub02.a	.	.	C	.	.	C	.	.	.	.	T	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
Sub1	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.
Sub10	.	.	.	.	.	.	.	.	.	.	T	C	.	.	C	.	.	A	.	A	C	.	.	.	.	.	.	.	.	.	.	C	A
Sub12	.	A	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	C	A
Sub15	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	.
Sub18	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.
Sub3	.	.	.	A	T	.	.	.	.	.	T	C	.	.	.	.	.	.	C	A	C	.	.	.	.	.	.	.	.	.	.	.	.
Sub8	.	.	C	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
Sub9	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	.	.	.	C	.	.	T	.	.	.	.	.	.
Sub97.1	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.
SubAst19	.	.	C	.	.	.	.	.	A	.	T	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	T	.	.	.
SubAst20	.	.	C	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	A	.	A	T	.	.	.	.	.	A	.	.	.	A
SubAst39	.	.	C	.	.	.	.	G	.	T	T	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst50	.	.	C	.	.	.	.	.	A	.	T	C	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	T	.	.	.
SubAst51	.	.	C	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
SubAst72	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

	726	735	755	776	835	853	888	899	901	915	953	997	1046	1047	1060	1073	1082	1088	1090	1091	1101	1105	1111	1114	1121	1140	1145	1168	1175	1200	1224	1232	1235	1240
ChcuP150	G	T	T	C	G	C	G	T	C	G	C	C	A	C	G	G	C	A	A	C	C	A	C	C	A	C	G	T	G	G	T	T	T	T
Mad01.10	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A
Mad01.13	-	-	C	.	A	.	.	.	T	.	.	.	G	.	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	A
Mad01.17	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Mad01.2	.	.	.	T	A	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	A
Mad01.23	.	.	.	.	A	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	A
Mad01.27	.	.	.	T	A	.	.	.	.	A	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	A
Mad01.30	.	.	.	T	A	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	A
Mad01.37	.	.	.	T	A	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	A
Mad01.56	.	.	.	.	A	.	.	C	.	.	.	.	.	.	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A
Mad01.8	.	.	.	T	A	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	A	C	.	.	.	.	A
Mad02.I	-	-	C	.	A	.	.	.	T	.	.	.	G	.	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	A
MadV	.	.	.	T	A	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	A
Sub02.1	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
Sub02.4	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Sub02.a	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sub1	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.
Sub10	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
Sub12	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Sub15	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.
Sub18	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.	.	.	C	.	.	.
Sub3	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	C	.	A	.	.	.	.	C	.	.	.	.	.	.	A	.	.
Sub8	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
Sub9	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.
Sub97.1	.	C	.	.	A	.	A	.	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst19	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst20	.	C	.	.	A	.	.	.	.	.	T	.	.	.	T	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
SubAst39	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst50	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst51	.	C	.	.	A	.	.	.	.	.	.	.	.	A	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst72	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

The multiple alignment is given relative to the reference sequence (*chcu* strain) of *D. subobscura*. The number above each site indicates its position in the multiple alignment. The dot (.) indicates the same nucleotide as in the reference sequence, and the dash (-) a deleted nucleotide. The red highlighted region indicates the polymorphic positions in the coding region.

A general description of nucleotide polymorphism in the P150 region is shown in Table 5.6. The number of polymorphic sites was higher in *D. subobscura* than in *D. madeirensis* when all *D. subobscura* lines or the A_2 lines were considered. The percentage of singletons was smaller in *D. madeirensis* than in *D. subobscura*, and in the latter species the percentage was higher in A_{st} (78%) than in A_2 (65%). The number of haplotypes in *D. madeirensis* (12 lines) was 9 and in *D. subobscura* (18 lines) this number was 16: 11 different haplotypes in 12 A_2 lines and 5 haplotypes in 6 A_{st} lines. Thus, the haplotype diversity was lower than one in both species.

Table 5.6.
Nucleotide polymorphism in P150 region.

	n	*len*	*len*^{w}	*S*	*mut*	*sin*	*%sin*	*inf*	*hap*	*H*
***D. subobscura* A_{st}**	6	1313	1262	18	18	14	78	4	5	0.985
***D. subobscura* A_2**	12	1313	1262	26	26	17	65	9	11	0.933
***D. subobscura* Total**	18	1313	1262	39	39	25	64	14	16	0.987
D. madeirensis	12	1313	1262	23	23	11	48	12	9	0.939

n, sample size or number of sequences; *len*, total number of sites; *len*^{w}, number of sites excluding gaps; *S*, number of polymorphic sites; *mut*, number of mutations; *sin*, number of singletons; *%sin*, percentage of singletons; *inf*, number of parsimony informative sites; *hap*, number of haplotypes; *H*, haplotype diversity.

A general description of the nucleotide polymorphism in the coding region at P150 is shown in Table 5.7. A single nonsynonymous polymorphism was detected in the coding region at P150 in *D. madeirensis*. In contrast, 5 polymorphic sites (2 synonymous and 3 nonsynonymous) were detected in the A_2 sample of *D. subobscura*, and 3 (2 synonymous and 1 nonsynonymous) in the A_{st} sample.

Table 5.7.

Nucleotide polymorphism in the coding region of P150 region.

	n	*len*	len^{w}	*S*	*syn*	*nonsyn*
D. subobscura **A_{st}**	6	294	291	3	2	1
D. subobscura **A_2**	12	294	291	5	2	3
D. subobscura **Total**	18	294	291	7	3	4
D. madeirensis	12	294	291	1	0	1

n, sample size or number of sequences; *len*, total number of sites in the coding region; len^{w}, number of sites excluding gaps in the coding region; *S*, number of polymorphic sites; *syn*, number of synonymous polymorphisms; *nonsyn*, number of nonsynonymous polymorphisms.

5.1.2.3. Nucleotide variation

The π estimates were similar in *D. madeirensis* and in *D. subobscura*, when either the total lines or each chromosomal arrangement class were considered (Table 5.8). In contrast, θ estimates were higher in the latter species than in the former, when the total lines of *D. subobscura* were considered.

Table 5.8.

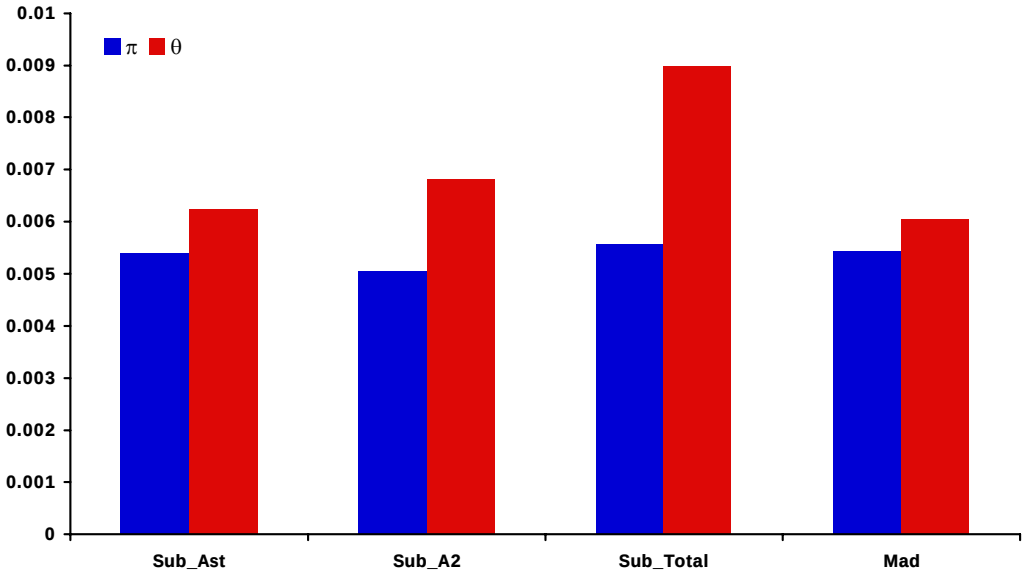
Estimates of nucleotide variation in P150 region.

	n	*S*	*k*	π	θ^{seq}	θ^{sit}
D. subobscura **A_{st}**	6	18	6.800	0.0053	7.883	0.0062
D. subobscura **A_2**	12	26	6.364	0.0050	8.610	0.0068
D. subobscura **Total**	18	39	7.033	0.0055	11.339	0.0089
D. madeirensis	12	23	6.848	0.0054	7.616	0.0060

n, sample size or number of sequences; *S*, number of polymorphic sites; *k*, average number of pairwise nucleotide differences; π, nucleotide diversity; θ^{seq}, nucleotide heterozigosity per sequence inferred from *S*; θ^{sit}, nucleotide heterozigosity per site inferred from *S*.

The levels of nucleotide diversity (π) and nucleotide heterozigosity (θ) are compared in Figure 5.5. For all samples, estimates of θ are greater than estimates of π, mainly in *D. subobscura* (when total lines were considered). This result indicated a general excess of singletons in all samples, mainly in *D. subobscura*.

Figure 5.5.
Nucleotide diversity (π) and nucleotide heterozigosity per site (θ) estimates.



Sub_A_2, *D. subobscura* lines with A_2 arrangement; Sub_A_{st}, *D. subobscura* lines with A_{st} arrangement; Sub_Total, all *D. subobscura* lines studied (A_2 + A_{st}); Mad, *D. madeirensis* lines; π, nucleotide diversity and θ, nucleotide heterozigosity per site.

The different estimates of nucleotide variation in functional regions (intron, exon and 3' flanking region) are shown in Table 5.9. The levels of nucleotide variation estimated either by π_s or θ_s were similar between the intron and the 3' flanking region in all samples. Estimates of π_a and θ_a were higher in *D. subobscura* than in *D. madeirensis*.

Table 5.9.

Estimates of nucleotide variation in functional regions of P150 region.

		Intron	Exon	3' region	Total
		1-288	**289-582**	**583-1313**	**1-1313**
	len	269	291	702	1262
	syn	269	59.5	702[a)]	1031.5
	nonsyn	-	228.5	-	228.5
D. madeirensis	π_s	0.0052	0.0000	0.0075	0.0065
	π_a	-	0.0007	-	0.0007
	θ_s	0.0062	0.0000	0.0080	0.0071
	θ_a	-	0.0015	-	0.0015
***D. subobscura* Total**	π_s	0.0060	0.0087	0.0058	0.0060
	π_a	-	0.0038	-	0.0038
	θ_s	0.0108	0.0147	0.0091	0.0099
	θ_a	-	0.0051	-	0.0051
***D. subobscura* A_2**	π_s	0.0046	0.0056	0.0057	0.0054
	π_a	-	0.0036	-	0.0036
	θ_s	0.0074	0.0111	0.0071	0.0074
	θ_a	-	0.0044	-	0.0044
***D. subobscura* A_{st}**	π_s	0.0069	0.0146	0.0053	0.0063
	π_a	-	0.0015	-	0.0015
	θ_s	0.0081	0.0147	0.0062	0.0072
	θ_a	-	0.0019	-	0.0019

a) for the A_{st} arrangement the number of synonymous sites was 703.

len, total number of sites in each region; *syn*, number of synonymous polymorphisms; *nonsyn*, number of nonsynonymous polymorphisms; π_s, nucleotide diversity in synonymous sites; π_a, nucleotide diversity in nonsynonymous sites; θ_s, nucleotide heterozigosity in synonymous sites and θ_a, nucleotide heterozigosity in nonsynonymous sites (both per site and inferred from *S*).

5.1.2.4. Recombination

A minimum of one recombination event was detected in *D. madeirensis*, and 6 in *D. subobscura*, according to the method proposed by Hudson and Kaplan (1985). In the latter species, 4 recombination events were detected within the A_2 and 1 within the A_{st} arrangement. The estimate of the recombination parameter *R* per site (Hudson 1987), was 0.0134 in *D. madeirensis*, and 0.4543 in *D. subobscura* (1.0633 and 0.1714 for A_2 and A_{st}, respectively).

5.1.2.5. Linkage disequilibrium

The analysis of linkage disequilibrium is summarized in Table 5.10. After applying Bonferroni correction for multiple tests, 28 (3.7%) chi-square tests remained significant for *D. subobscura*. The global level of linkage disequilibrium (*ZnS* and *Za* estimates) was higher in *D. madeirensis* than in *D. subobscura*. However, the highest value of linkage disequilibrium was detected in the A_{st} arrangement of *D. subobscura.*

Table 5.10.
Linkage disequilibrium analysis for P150 region.

	S_L	*comp*	*Fisher*	*%Fisher*	*chi-square*	*%chi-square*	*ZnS*	*Za*
***D. subobscura* A_{st}**	18	153	0	0	24	16	0.2378	0.3294
***D. subobscura* A_2**	26	325	3	1	27	8	0.1051	0.2024
***D. subobscura* Total**	39	741	7	1	59	8	0.0706	0.1184
D. madeirensis	23	253	26	10	46	18	0.1881	0.1914

S_L, number of segregating sites with two variants; *comp*, number of comparisons; *Fisher*, number of significant ($P < 0.05$) comparisons by the Fisher's test; *%Fisher*, percentage of significant ($P < 0.05$) comparisons by the Fisher's test; *chi-square*, number of significant ($P < 0.05$) comparisons by the chi-square test; *%chi-square*, percentage of significant ($P < 0.05$) comparisons by the chi-square test; *ZnS* (Kelly 1997) and *Za* (Rozas *et al.* 2001) are global measures of linkage disequilibrium.

5.1.2.6. Genetic differentiation between *D. subobscura* arrangements

The average number of nucleotide substitutions per site between the two arrangements was 0.007 (Table 5.11). Putative significant genetic differentiation between the chromosomal arrangements of *D. subobscura* was contrasted according to the Hudson *et al.* (1992b) permutation test. The statistical significance of the *Kst** statistic was obtained after 1000 replicates. The observed *Kst** value was 0.0402 and the result was statistically significant (P = 0.002). Therefore, the significant genetic differentiation between arrangements prevented pooling all *D. subobscura* lines in a single sample. Genetic differentiation between arrangements prevailed despite the presence of 7 shared polymorphisms and the absence of fixed differences between them. According to the hypergeometric distribution the high number of shared polymorphisms cannot be explained by recurrent mutation. So, this high number of observed shared polymorphisms has to be explained by genetic exchange between arrangements, although no conversion tract was identified in the P150 region.

5.1.2.7. Genetic differentiation between species

The average number of nucleotide substitutions per site between the two species was similar, independently of the arrangement used to compare with *D. madeirensis* (Table 5.11). There were 3 fixed differences and 3 shared polymorphisms between *D. madeirensis* and *D. subobscura*.

Table 5.11.

Genetic differentiation between species and chromosomal arrangements of *D. subobscura*.

	Fixed	*Shared*	S_{x1}	S_{x2}	*K*	*Dxy*
A_2 / A_{st}	0	7	22	15	9.194	0.0070
A_2 / *D. mad*	3	3	23	20	12.569	0.0099
A_{st} / *D. mad*	3	0	18	23	13.417	0.0106

A_2, *D. subobscura* lines with A_2 arrangement; A_{st}, *D. subobscura* lines with A_{st} arrangement; *D. mad*, *D. madeirensis* lines; Fixed, fixed differences between samples; Shared, polymorphic sites segregating for the same two variants in two samples; S_{x1}, exclusive polymorphisms sites in population 1; S_{x2}, exclusive polymorphisms sites in population 2; *K*, average number of nucleotide differences between arrangements or species; D_{xy}, average number of nucleotide substitutions per site between samples.

5.1.2.8. Patterns of polymorphism

Tajima's (1989) and Fu and Li's (1993) statistics were negative in all samples, indicating an excess of singletons variants (Table 5.12). Only the Fu and Li tests with an outgroup were statistically significant for *D. subobscura*. However, Tajima's *D* statistic was also significant (*P(D)* = 0.046) in this species after computer simulations under the conservative assumption of no recombination.

Table 5.12.

Tajima's and Fu and Li's statistics for P150 region.

	Tajima's D	*Fu and Li's D*	*Fu and Li's F*	*Fu and Li's D**	*Fu and Li's F**
***D. subobscura* A_{st}**	-0.8545	-1.2823	-1.4236	-0.8101	-0.8921
***D. subobscura* A_2**	-1.1648	-1.8375	-1.9981	-1.2818	-1.4250
***D. subobscura* Total**	-1.5473	-2.3949*	-2.5896*	-1.8110	-2.0125
D. madeirensis	-0.4468	-0.4342	-0.5042	-0.5081	-0.5603

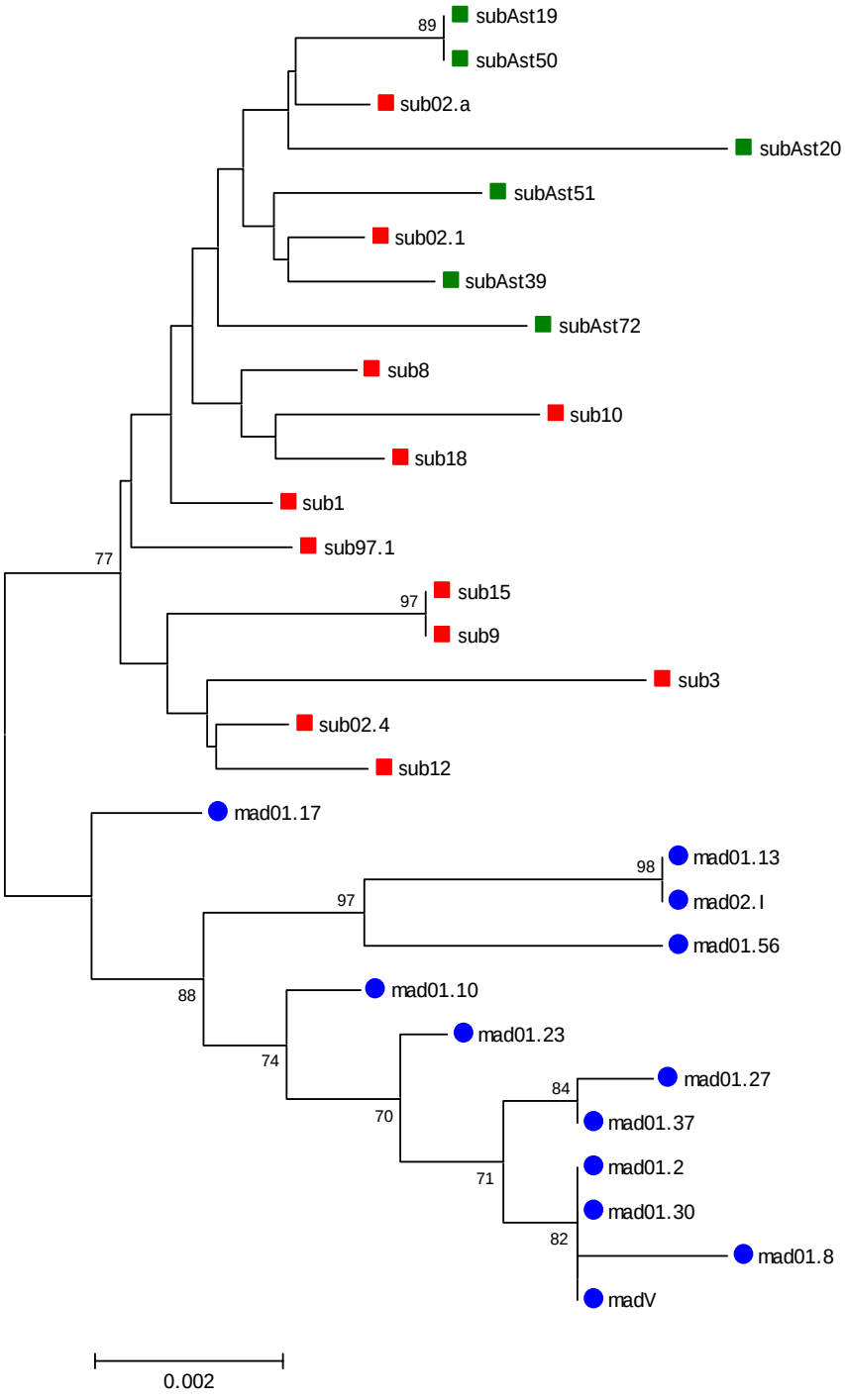
* significant for $P < 0.05$.

5.1.2.9. Gene genealogy

The gene genealogy reconstructed from variation in P150 region is shown in Figure 5.6. All *D. madeirensis* lines grouped together in a single cluster. In *D. subobscura* a partial clustering of lines, according to their gene arrangement was also detected. The A_{st} lines grouped together in a single cluster, although this cluster also included two A_2 lines (*sub02.a* and *sub02.1*).

Figure 5.6.

Gene genealogy reconstructed from nucleotide variation in P150 region.



The gene genealogy was obtained by the neighbour-joining method (Saitou and Nei 1987) using, as genetic distance, the number of substitutions per site, according to Kimura 2-parameter method (Kimura 1980). Bootstrap confidence level (from 1000 replications) above 70% are shown below the branches. The horizontal bar at the bottom represents the distance scale of the branches.

5.1.3. *Sex-lethal* gene region

5.1.3.1. Studied region

The S*ex-lethal* (*Sxl*) gene was cloned and sequenced in *D. subobscura* by Penalva *et al.* (1996). In this species, the gene maps at section 10B of the X chromosome. The sequence deposited in the EMBL database with acession number X98370 was used to select a region of about 2 kb for further analysis. The selected region included part of exon 4, part of exon 5 and the intron between them. This region was amplified with two primers drew in exons 4 and 5, respectively.

5.1.3.2. Nucleotide polymorphism

The multiple alignment of *Sex-lethal* gene region in 12 *D. madeirensis* lines, and in 18 *D. subobscura* lines (12 with A_2 arrangement and 6 A_{st}) included a total of 1975 sites. This number decreased to 1630 when sites with alignment gaps were not considered. From these, 209 corresponded to coding DNA (47 and 162 from exon 4 and 5, respectively), and the remain 1421 sites corresponded to non coding region (intron between the exon 4 and 5). A total of 98 polymorphic sites (58 singletons) were detected in the data set. The minimum number of mutations was 101, as there were sites that segregated for more than two variants. The nucleotide variation of *sex-lethal* gene region at polymorphic sites is shown in Figure 5.7. Sites with alignment gaps were completely excluded from all analysis. A single synonymous nucleotide polymorphism was found in the *sex-lethal* coding region (position 1920). This polymorphic site presented a unique variant (singleton) in *D. madeirensis* (*01.2* line).

Figure 5.7.

Polymorphic sites at *Sex-lethal* gene region (*continues next page*)**.**

	73	123	133	135	138	149	170	172	176	222	245	293	392	395	418	478	490	521	524	535	536	548	689	695	698	700	703	704	706	714	718	730	741	755	756	760	764	770
ChcuSXL	T	C	G	T	A	G	A	A	T	C	T	C	T	A	C	G	A	G	A	A	G	A	C	C	T	T	C	C	C	-	-	C	G	T	T	C	C	T
Mad01.10	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.	T	C	.	.	.	.	T	A	C
Mad01.13	.	.	C	.	G	.	.	.	.	.	.	.	.	.	T	.	.	A	.	.	A	.	.	.	C	.	T	A	.	T	C	.	.	.	.	T	A	.
Mad01.17	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.	T	C	.	.	.	.	T	A	C
Mad01.2	.	.	.	G	.	A	.	T	.	.	.	.	.	.	T	.	.	.	.	.	A	T	.	.	C	.	.	.	.	T	C	.	T	A	.	T	A	.
Mad01.23	.	.	.	G	.	A	.	T	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.	T	C	.	.	.	.	T	A	.
Mad01.27	.	.	.	G	.	A	.	T	A	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.	G	T	.	.	.	A	T	A	C
Mad01.30	A	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	G	G	.	.	.	T	C	A	.	.	.	T	A	.
Mad01.37	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.	G	T	.	.	.	.	T	A	.
Mad01.56	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	.	-	.	.	.	C	.	.	.	G	G	T	.	.	.	.	T	A	.
Mad01.8	.	.	.	G	.	A	.	T	A	.	.	.	.	.	T	.	.	.	.	.	A	T	.	.	C	.	.	.	.	T	C	.	.	.	.	T	A	.
Mad02.I	.	.	.	.	.	.	.	T	.	.	.	.	.	.	T	.	.	.	.	G	A	.	.	.	C	.	.	.	.	G	T	.	.	.	.	T	A	.
MadV	.	.	.	G	.	A	.	T	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	C	.	.	.	.	T	C	.	.	.	.	T	A	.
Sub02.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub02.4	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub02.a	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	-	.	.	.	A	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	A	A	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub8	.	.	.	.	.	.	T	.	.	.	A	.	.	.	.	C	.	.	.	.	-	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
Sub97.1	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
SubAst19	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
SubAst20	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	-	-	.	.	.	.	.	.	.
SubAst39	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	C	.	.	.	.	-	-	.	.	.	.	.	.	.
SubAst50	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	-	-	.	.	.	.	.	.	.
SubAst51	.	.	.	.	.	.	.	.	.	.	A	.	A	.	C	.	.	.	.	.	.	G	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.
SubAst72	.	.	.	.	.	.	.	.	.	.	.	.	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.

	801	807	810	814	836	837	843	846	848	851	854	856	865	867	869	898	903	908	909	925	937	938	948	977	979	980	984	985	988	991	994	1000	1001	1003	1008	1010	1037	1044	1062
ChcuSXL	T	T	A	C	A	C	A	G	C	A	T	A	G	G	T	T	T	T	C	T	A	G	G	A	A	C	C	C	G	G	A	A	A	G	A	A	A	A	A
Mad01.10	A	G	.	.	.	A	G	C	.	.	.	.	A	A	-	.	.	.	.	.	.	T	A	C	T	.	A	T	.	.	.	G	G	A	.	.	.	.	.
Mad01.13	.	.	.	.	.	A	G	C	.	.	.	.	A	A	-	.	.	G	.	.	.	T	A	C	.	.	.	.	.	.	.	G	G	A	.	.	.	.	.
Mad01.17	A	.	.	.	.	A	G	C	G	.	.	.	A	A	-	.	.	G	.	A	.	T	A	C	.	.	.	.	.	.	.	.	G	A	.	.	C	.	.
Mad01.2	.	.	.	.	.	A	G	C	.	.	.	.	A	A	-	.	.	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	G	A	.	-	.	.	.
Mad01.23	.	.	.	.	T	A	G	C	.	.	.	.	A	A	-	.	.	G	.	.	.	T	A	C	T	.	.	.	T	.	.	.	G	A	.	.	.	.	T
Mad01.27	A	.	.	T	.	A	G	C	.	.	.	G	A	A	-	.	.	G	.	.	.	T	A	C	T	.	.	.	.	.	.	G	G	A	.	.	.	.	.
Mad01.30	.	.	.	.	.	A	G	C	.	.	C	.	A	A	-	.	.	.	.	.	.	T	A	C	T	.	.	.	.	.	.	G	G	A	.	.	.	G	.
Mad01.37	.	.	C	.	.	A	G	C	.	.	.	.	A	A	-	.	.	.	.	.	.	T	A	C	T	.	.	.	.	.	.	G	G	A	.	G	.	.	.
Mad01.56	A	.	.	.	.	A	G	C	.	.	.	.	A	A	-	.	.	.	.	.	.	T	A	C	T	.	.	.	.	.	.	G	G	A	.	.	.	.	.
Mad01.8	.	.	.	.	.	A	G	C	.	.	C	.	A	.	.	.	.	.	.	.	G	T	A	C	T	.	.	.	.	.	.	.	G	A	.	.	.	.	.
Mad02.I	.	.	C	.	.	A	G	C	.	.	.	.	A	A	-	.	G	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	G	A	.	.	.	.	.
MadV	.	.	C	.	.	A	G	C	.	.	.	.	A	A	-	.	.	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	G	A	.	.	.	.	.
Sub02.1	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	T	.	.	.	.
Sub02.4	.	.	.	.	.	.	.	C	.	.	.	.	A	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sub02.a	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sub1	.	.	.	.	.	A	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	A	G	.	G	A	T	.	.	.	.
Sub10	.	.	.	.	.	.	.	C	.	C	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	T	.	.	.	.
Sub12	.	.	.	.	.	.	.	C	.	.	.	.	A	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sub15	.	.	.	.	.	.	.	C	.	.	.	.	A	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	.	.	.	.	.
Sub18	.	.	.	.	.	.	.	C	.	.	.	.	A	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	.	.	.	.	.
Sub3	.	.	.	.	.	.	.	C	.	C	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	T	.	.	.	.
Sub8	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	A	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	T	.	.	.	.
Sub9	.	.	.	.	.	.	.	C	.	.	.	.	A	.	G	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	.	.	.	.	.
Sub97.1	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	G	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	T	.	.	.	.
SubAst19	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	T	.	.	.	.
SubAst20	.	.	.	.	.	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst39	.	.	.	.	.	A	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	G	A	T	.	.	.	.
SubAst50	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst51	.	.	.	.	.	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SubAst72	.	.	.	.	.	A	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Figure 5.7. *continues from previous page.*

	1064	1094	1114	1127	1132	1149	1166	1167	1177	1188	1194	1195	1203	1228	1233	1239	1258	1274	1312	1356	1357	1361	1431	1447	1463	1488	1491	1518	1533	1573	1626	1633	1661	1666	1668	1682	1711	1920
ChcuSXL	A	A	T	A	T	A	A	G	A	A	G	T	G	G	C	C	-	-	-	-	-	-	-	C	C	C	T	-	C	G	C	C	C	C	A	G	A	T
Mad01.10	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	C	A	C	.	.	.	.	A	.	.	.	.	.	.	.	A	T	.
Mad01.13	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	.	A	.	.	.	.	.	.	.	A	T	.
Mad01.17	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	A	.	C	.	.	.	.	.	.	.	A	T	.
Mad01.2	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	A	A	C	A	A	.	.	.	.	A	G	.	.	.	.	.	.	A	T	C
Mad01.23	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	.	A	T	A	A	A	A	A	.	.	.	.	A	.	.	A	.	.	.	.	A	T	.
Mad01.27	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	A	.	.	.	.	.	A	.	A	T	.
Mad01.30	.	T	.	.	.	.	.	T	.	.	.	.	.	.	A	.	A	T	A	A	C	A	A	.	.	.	.	A	.	.	.	.	.	.	.	A	T	.
Mad01.37	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	A	G	A	.	T	.	.	.	A	T	.
Mad01.56	T	T	.	.	.	.	G	.	.	C	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	A	.	.	.	.	.	.	.	A	T	.
Mad01.8	T	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	.	A	.	.	.	T	.	.	.	A	T	.
Mad02.I	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	A	G	.	.	T	.	.	.	A	T	.
MadV	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	A	G	.	.	.	.	.	.	A	T	.
Sub02.1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub02.4	.	.	C	.	.	.	.	.	.	.	A	G	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub02.a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	C	-	.	.	.	.	.	.	.	A	.	.
Sub1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	T	A	.	.	.	.	-	.	.	.	.	.	.	G	A	.	.
Sub10	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	A	C	A	A	C	A	A	.	A	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub12	.	T	.	.	.	G	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub3	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	A	C	A	A	C	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub8	.	T	.	T	.	.	.	.	.	.	.	.	A	.	.	.	A	T	A	A	G	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	C	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
Sub97.1	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	T	A	T	A	T	C	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
SubAst19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	A	.	.	.	A	.	.	.	.	.	.	.	A	.	.
SubAst20	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.
SubAst39	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.
SubAst50	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	C	-	.	.	.	.	.	.	.	A	.	.
SubAst51	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	.	-	.	.	.	.	T	.	.	A	.	.
SubAst72	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	A	A	G	A	A	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.

The multiple alignment is given relative to the reference sequence (*chcu* strain) of *D. subobscura*. The number above each site indicates its position in the multiple alignment. The dot (.) indicates the same nucleotide as in the reference sequence, and the dash (-) a deleted nucleotide. The red highlighted position in white indicate the only polymorphic position found in the coding region.

A general description of nucleotide polymorphism at *Sex-lethal* gene is shown in Table 5.13. The number of polymorphic sites was higher in *D. madeirensis* (12 lines) than in *D. subobscura* (18 lines) despite their different sample size. When same size samples were compared, *D. madeirensis* still presented a higher number of polymorphic sites than the A_2 arrangement of *D. subobscura* (12 lines). The percentage of singletons was higher in *D. madeirensis* (72%) than in *D. subobscura* (57%). In the latter species, the A_2 arrangement presented a higher percentage of singletons than A_{st} (64% and 52%, respectively). In *sex-lethal* gene region, each line of *D. madeirensis* and of A_{st} arrangement of *D. subobscura* represented a single haplotype. In contrast, 10 different haplotypes were detected in the 12 A_2 lines, thus in this arrangement of *D. subobscura* haplotype diversity was lower than one.

Table 5.13.
Nucleotide polymorphism at *Sex-lethal* gene region.

	n	*len*	len^{w}	*S*	*mut*	*sin*	*%sin*	*inf*	*hap*	*H*
***D. subobscura* A_{st}**	6	1930	1630	21	22	11	52	10	6	1
***D. subobscura* A_2**	12	1926	1630	28	28	18	64	10	10	0.955
***D. subobscura* Total**	18	1950	1630	38	39	22	57	16	16	0.980
D. madeirensis	12	1921	1630	55	55	40	72	15	12	1

n, sample size or number of sequences; *len*, total number of sites; len^{w}, number of sites excluding gaps; *S*, number of polymorphic sites; *mut*, number of mutations; *sin*, number of singletons; *%sin*, percentage of singletons; *inf*, number of parsimony informative sites; *hap*, number of haplotypes; *H*, haplotype diversity.

5.1.3.3. Nucleotide variation

The different estimates of nucleotide variation at *Sex-lethal* gene region are shown in Table 5.14. Nucleotide variation estimated either by π or θ was higher in *D. madeirensis* than in *D. subobscura*. In the latter species, the level of variation was similar in both arrangements according to θ. However, when π estimates were considered, the A_{st} arrangement presented a higher level of variation than the A_2.

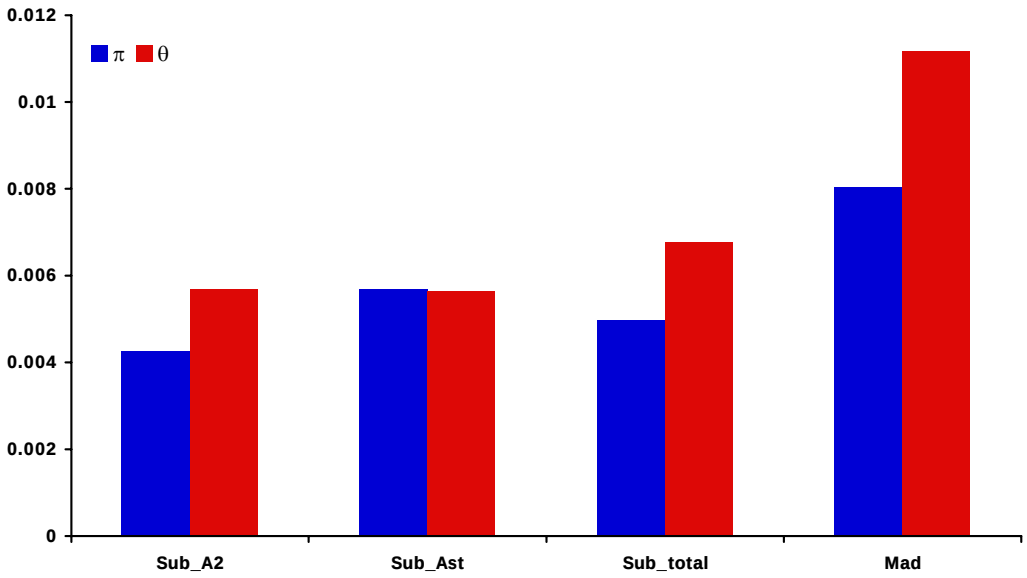
Table 5.14.
Estimates of nucleotide variation at *Sex-lethal* gene region.

	n	*S*	*k*	π	θ^{seq}	θ^{sit}
***D. subobscura* A_{st}**	6	21	9.267	0.0056	9.197	0.0056
***D. subobscura* A_2**	12	28	6.909	0.0042	9.272	0.0056
***D. subobscura* Total**	18	38	8.098	0.0049	11.048	0.0067
D. madeirensis	12	55	13.106	0.0080	18.213	0.0111

n, sample size or number of sequences; *S*, number of polymorphic sites; *k*, average number of pairwise nucleotide differences; π, nucleotide diversity; θ^{seq}, nucleotide heterozigosity per sequence inferred from *S*; θ^{sit}, nucleotide heterozigosity per site inferred from *S*.

The levels of nucleotide diversity (π) and nucleotide heterozigosity (θ) for the *Sex-lethal* gene region are compared in Figure 5.8. The estimates of θ were higher than estimates of π in all samples (indicating a general excess of singletons), except in A_{st}.

Figure 5.8.
Nucleotide diversity (π) and nucleotide heterozigosity per site (θ) estimates.



Sub_A_2, *D. subobscura* lines with A_2 arrangement; Sub_A_{st}, *D. subobscura* lines with A_{st} arrangement; Sub_Total, all *D. subobscura* lines studied (A_2 + A_{st}); Mad, *D. madeirensis* lines; π, nucleotide diversity and θ, nucleotide heterozigosity per site.

5.1.3.4. Recombination

A minimum of 8 recombination events were detected in the *D. madeirensis* and 5 in *D. subobscura*, according to the method proposed by Hudson and Kaplan (1985). Within *D. subobscura*, 3 recombination events were detected in the A_2 and 2 in the A_{st} arrangement.

5.1.3.5. Linkage disequilibrium

The results of the linkage disequilibrium analysis are shown in Table 5.15. After applying Bonferroni correction for multiple tests, 3 (0.45%) Fisher's tests remained significant and 17 (2.55%) chi-square tests remained significant for *D. subobscura*. The global level of linkage disequilibrium (*ZnS* and *Za* estimates) was higher in *D. madeirensis* than in *D. subobscura*. However, the highest level of linkage disequilibrium was detected in the A_{st} arrangement of *D. subobscura*.

Table 5.15.
Linkage disequilibrium analysis for *Sex-lethal* gene region.

	S_L	*comp*	*Fisher*	*%Fisher*	*chi-square*	*%chi-square*	*ZnS*	*Za*
***D. subobscura* A_{st}**	20	190	0	0	12	6.3	0.2181	0.3126
***D. subobscura* A_2**	28	378	6	1.6	42	11.1	0.1257	0.1702
***D. subobscura* Total**	37	666	16	2.4	60	9	0.0743	0.1068
D. madeirensis	55	1485	6	0.4	100	6.7	0.0923	0.1236

S_L, number of segregating sites with two variants; *comp*, number of comparisons; *Fisher*, number of significant ($P < 0.05$) comparisons by the Fisher's test; *%Fisher*, percentage of significant ($P < 0.05$) comparisons by the Fisher's test; *chi-square*, number of significant ($P < 0.05$) comparisons by the chi-square test; *%chi-square*, percentage of significant ($P < 0.05$) comparisons by the chi-square test; *ZnS* (Kelly 1997) and *Za* (Rozas *et al.* 2001) are global measures of linkage disequilibrium.

5.1.3.6. Genetic differentiation between *D. subobscura* arrangements

The average number of nucleotide substitutions per site between the two arrangements was 0.0054 (Table 5.16). Putative significant genetic differentiation between the chromosomal arrangements of *D. subobscura*, was contrasted according to the Hudson *et al.* (1992b) permutation test. The statistical significance of the *Kst** statistic was obtained after 1000 replicates. The observed *Kst** value was 0.033 and

the result was statistically significant (P = 0.02). Therefore, the significant genetic differentiation between arrangements, prevented pooling all *D. subobscura* lines in a single sample. Genetic differentiation between arrangements prevailed, despite the presence of 11 shared poly7morphisms and the absence of fixed differences between them. According to the hypergeometric distribution the high number of shared polymorphisms cannot be explained by recurrent mutation. So, the high number of shared polymorphisms between the two arrangements was due to genetic exchange between them, although no conversion tract was identified in *Sex-lethal* gene region.

Table 5.16.
Genetic differentiation between species and chromosomal arrangements of *D. subobscura*.

	Fixed	*Shared*	S_{x1}	S_{x2}	*K*	*Dxy*
A_2 / A_{st}	0	11	17	11	8.944	0.0054
A_2 / *D. mad*	8	0	28	55	23.250	0.0142
A_{st} / *D. mad*	7	0	22	55	23.486	0.0144

A_2, *D. subobscura* lines with A_2 arrangement; A_{st}, *D. subobscura* lines with A_{st} arrangement; *D. mad*, *D. madeirensis* lines; Fixed, fixed differences between samples; Shared, polymorphic sites segregating for the same two variants in two samples; S_{x1}, exclusive polymorphisms sites in population 1; S_{x2}, exclusive polymorphisms sites in population 2; *K*, average number of nucleotide differences between arrangements or species; D_{xy}, average number of nucleotide substitutions per site between samples.

5.1.3.7. Genetic differentiation between species

The average number of nucleotide substitutions per site between the two species was similar independently of the arrangement used to compare with *D. madeirensis* (Table 5.16). There were 7 fixed differences between *D. madeirensis* and *D. subobscura* and no shared polymorphism.

5.1.3.8. Patterns of polymorphism

Tajima's (1989) and Fu and Li's (1993) statistics were negative in all samples, which indicated an excess of singletons variants (Table 5.17). Only the Fu and Li's *D* resulted statistically significant for *D. madeirensis.* However, Tajima's *D* and Fu and Li's *D* statistics proved to be significant in *D. subobscura*, after computer simulations under the conservative assumption of no recombination.

Table 5.17.
Tajima's and Fu and Li's statistics for *Sex-lethal* gene region.

	Tajima's D	*Fu and Li's D*	*Fu and Li's F*	*Fu and Li's D**	*Fu and Li's F**
***D. subobscura* A_{st}**	-0.2397	-0.2860	-0.3324	-0.0644	-0.1111
***D. subobscura* A_2**	-1.1424	-1.7074	-1.9075	-1.2397	-1.3831
***D. subobscura* Total**	-1.1645	-1.8689	-2.0037	-1.3929	-1.5392
D. madeirensis	-1.2904	-2.1591*	-2.3187	-1.6648	-1.7859

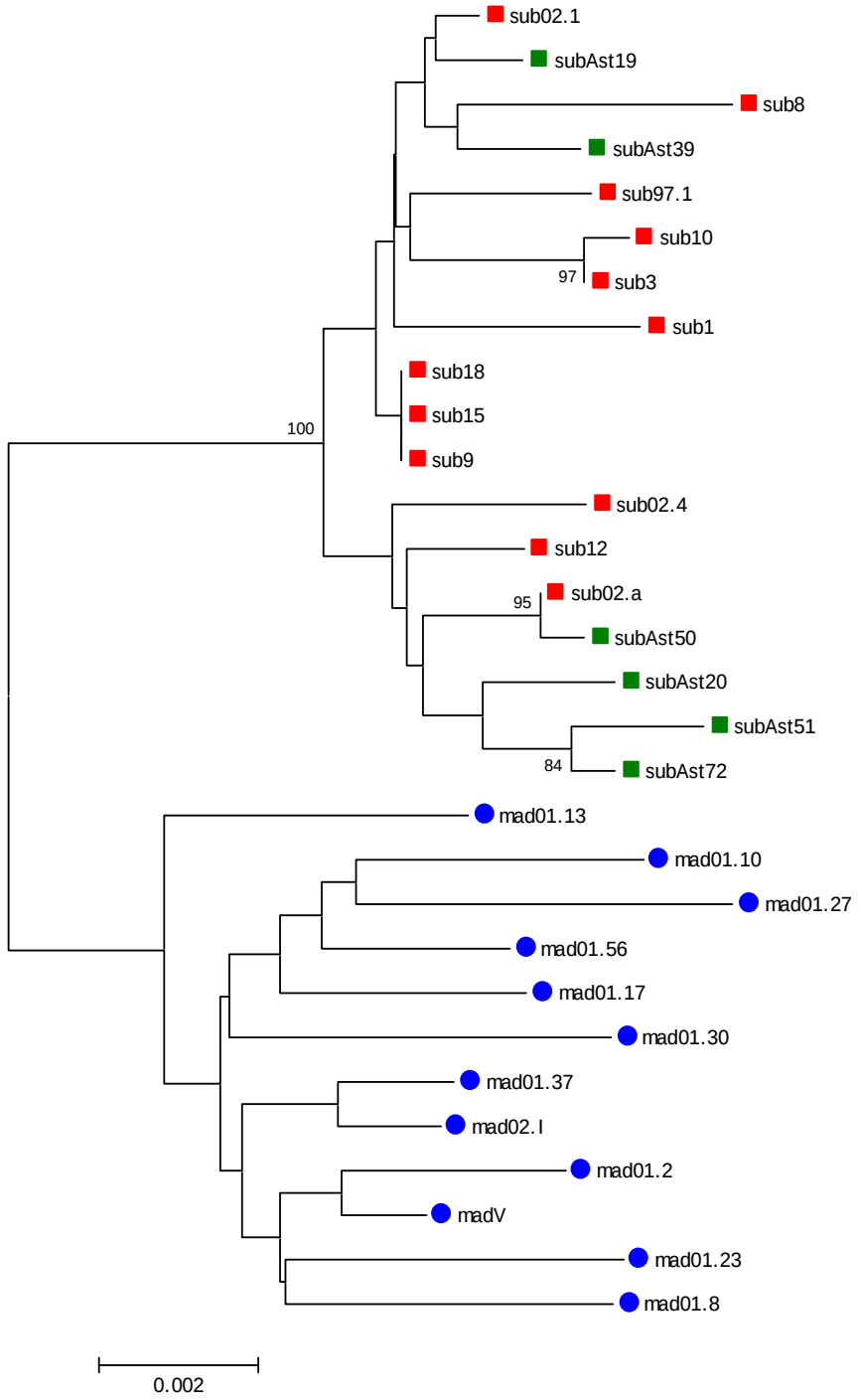
* significant for $P < 0.05$.

5.1.3.9. Gene genealogy

The gene genealogy reconstructed from variation in S*ex-lethal* gene region is shown in Figure 5.9. All *D. madeirensis* lines grouped together in a single cluster. In *D. subobscura* there was not a clear separation of the lines according to their chromosomal arrangement, although four A_{st} grouped together in a cluster with three A_2 lines.

Figure 5.9.

Gene genealogy reconstructed from nucleotide variation in *Sex-lethal* gene region.



The gene genealogy was obtained by the neighbour-joining method (Saitou and Nei 1987) using, as genetic distance, the number of substitutions per site, according to Kimura 2-parameter method (Kimura 1980). Bootstrap confidence level (from 1000 replications) above 70% are shown below the branches. The horizontal bar at the bottom represents the distance scale of the branches.

5.1.4. P125 region

5.1.4.1. Studied region

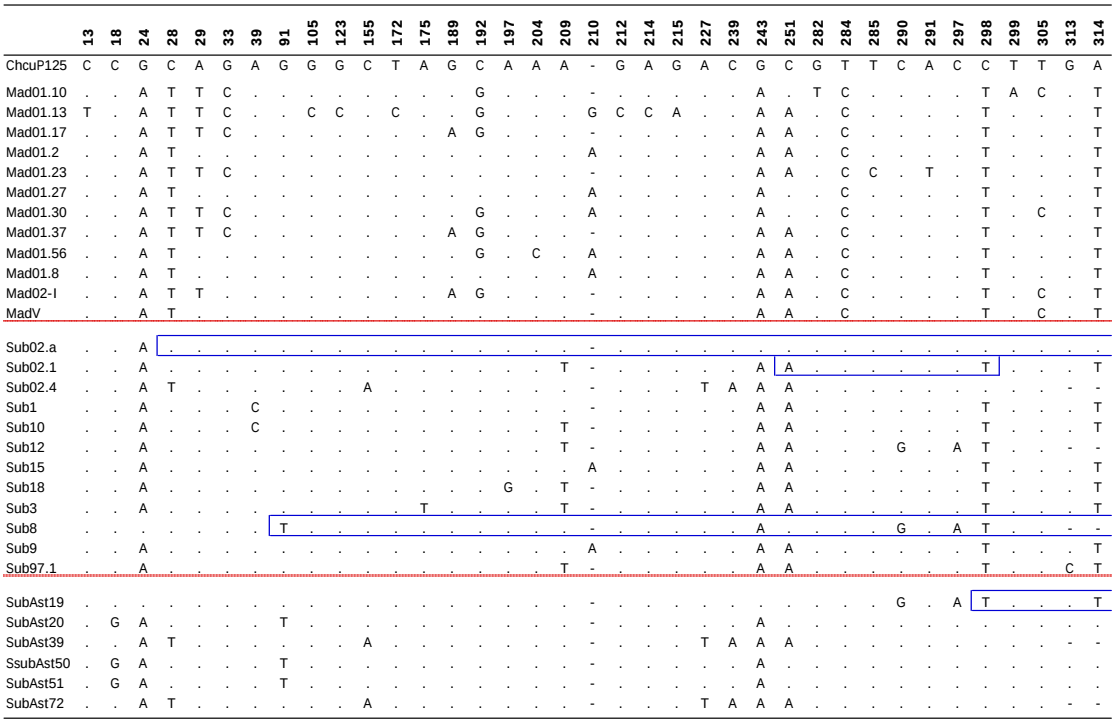
The digestion of P125 phage with *EcoRI* resulted in a single fragment with approximately 6 kb. This fragment was subcloned and its ends were sequenced. Conserved sequences, spanning of 6 kb, were detected after performing a BLAST search of these sequences with the *D. melangaster* genome (release 3.1). No gene was detected in this fragment, and a 2 kb region of non coding DNA was used for further analysis.

5.1.4.2. Nucleotide polymorphism

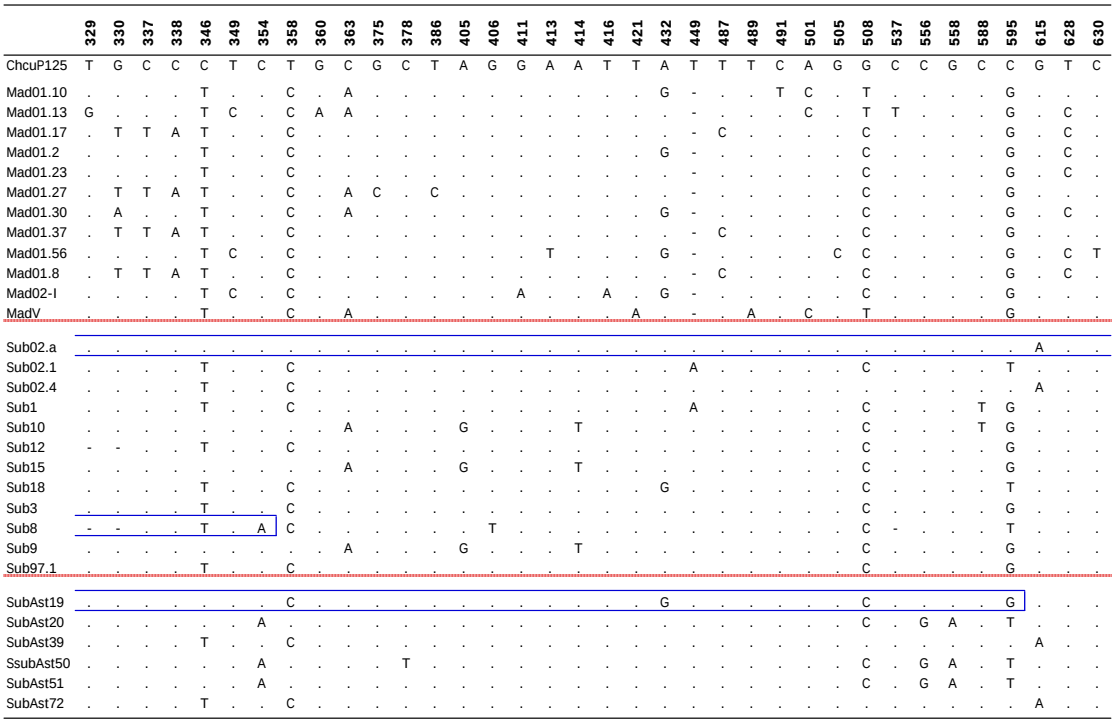
The multiple alignment of the P125 region in 12 *D. madeirensis* lines, and 18 *D. subobscura* lines (12 with the A_2 arrangement and 6 A_{st}) included a total of 1626 sites. This number decreased to 1460, when sites with alignment gaps were not considered. All sites were silent as P125 corresponds to a non coding region. A total of 129 polymorphic sites (52 singletons) were detected in the data set. The minimum number of mutations was 134, as some sites segregated for more than two variants. Sites with alignment gaps were completely excluded from all analysis. The nucleotide variation of P125 region at polymorphic sites is shown in Figure 5.10.

Figure 5.10.

Polymorphic sites in P125 region (*continues next page*)**.**



	13	18	24	28	29	33	39	91	105	123	155	172	175	189	192	197	204	209	210	212	214	215	227	239	243	251	282	284	285	290	291	297	298	299	305	313	314
ChcuP125	C	C	G	C	A	G	A	G	G	G	C	T	A	G	C	A	A	A	-	G	A	G	A	C	G	C	G	T	T	C	A	C	C	T	T	G	A
Mad01.10	.	.	A	T	T	C	.	.	.	.	.	.	.	.	G	.	.	.	-	.	.	.	.	.	A	.	T	C	.	.	.	.	T	A	C	.	T
Mad01.13	T	.	A	T	T	C	.	.	C	C	.	C	.	.	G	.	.	.	G	C	C	A	.	.	A	A	.	C	.	.	.	.	T	.	.	.	T
Mad01.17	.	.	A	T	T	C	.	.	.	.	.	.	.	A	G	.	.	.	-	.	.	.	.	.	A	A	.	C	.	.	.	.	T	.	.	.	T
Mad01.2	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	A	.	C	.	.	.	.	T	.	.	.	T
Mad01.23	.	.	A	T	T	C	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	A	.	C	C	.	T	.	T	.	.	.	T
Mad01.27	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	.	.	C	.	.	.	.	T	.	.	.	T
Mad01.30	.	.	A	T	T	C	.	.	.	.	.	.	.	.	G	.	.	.	A	.	.	.	.	.	A	.	.	C	.	.	.	.	T	.	C	.	T
Mad01.37	.	.	A	T	T	C	.	.	.	.	.	.	.	A	G	.	.	.	-	.	.	.	.	.	A	A	.	C	.	.	.	.	T	.	.	.	T
Mad01.56	.	.	A	T	.	.	.	.	.	.	.	.	.	.	G	.	C	.	A	.	.	.	.	.	A	A	.	C	.	.	.	.	T	.	.	.	T
Mad01.8	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	A	.	C	.	.	.	.	T	.	.	.	T
Mad02-I	.	.	A	T	T	.	.	.	.	.	.	.	.	A	G	.	.	.	-	.	.	.	.	.	A	A	.	C	.	.	.	.	T	.	C	.	T
MadV	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	A	.	C	.	.	.	.	T	.	C	.	T
Sub02.a	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sub02.1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	-	.	.	.	.	.	A	A	.	.	.	.	.	.	T	.	.	.	T
Sub02.4	.	.	A	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	-	.	.	.	T	A	A	A	.	.	.	.	.	.	.	.	.	-	-
Sub1	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	A	.	.	.	.	.	.	T	.	.	.	T
Sub10	.	.	A	.	.	.	C	.	.	.	.	.	.	.	.	.	.	T	-	.	.	.	.	.	A	A	.	.	.	.	.	.	T	.	.	.	T
Sub12	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	-	.	.	.	.	.	A	A	.	.	.	G	.	A	T	.	.	-	-
Sub15	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	A	.	.	.	.	.	.	T	.	.	.	T
Sub18	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	T	-	.	.	.	.	.	A	A	.	.	.	.	.	.	T	.	.	.	T
Sub3	.	.	A	.	.	.	.	.	.	.	.	.	T	.	.	.	.	T	-	.	.	.	.	.	A	A	.	.	.	.	.	.	T	.	.	.	T
Sub8	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	.	.	.	.	G	.	A	T	.	.	-	-
Sub9	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	A	.	.	.	.	.	.	T	.	.	.	T
Sub97.1	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	-	.	.	.	.	.	A	A	.	.	.	.	.	.	T	.	.	C	T
SubAst19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	G	.	A	T	.	.	.	T
SubAst20	.	G	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
SubAst39	.	.	A	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	-	.	.	.	T	A	A	A	.	.	.	.	.	.	.	.	.	-	-
SsubAst50	.	G	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
SubAst51	.	G	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
SubAst72	.	.	A	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	-	.	.	.	T	A	A	A	.	.	.	.	.	.	.	.	.	-	-



	329	330	337	338	346	349	354	358	360	363	375	378	386	405	406	411	413	414	416	421	432	449	487	489	491	501	505	508	537	556	558	588	595	615	628	630
ChcuP125	T	G	C	C	C	T	C	T	G	C	G	C	T	A	G	G	A	A	T	T	A	T	T	T	C	A	G	G	C	C	G	C	C	G	T	C
Mad01.10	.	.	.	.	T	.	.	C	.	A	.	.	.	.	.	.	.	.	.	.	G	-	.	.	T	C	.	T	.	.	.	.	G	.	.	.
Mad01.13	G	.	.	.	T	C	.	C	A	A	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	C	.	T	T	.	.	.	G	.	C	.
Mad01.17	.	T	T	A	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	C	.	.	.	.	G	.	C	.
Mad01.2	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	G	-	.	.	.	.	.	C	.	.	.	.	G	.	C	.
Mad01.23	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	C	.	.	.	.	G	.	C	.
Mad01.27	.	T	T	A	T	.	.	C	.	A	C	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	C	.	.	.	.	G	.	.	.
Mad01.30	.	A	.	.	T	.	.	C	.	A	.	.	.	.	.	.	.	.	.	.	G	-	.	.	.	.	.	C	.	.	.	.	G	.	C	.
Mad01.37	.	T	T	A	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	C	.	.	.	.	G	.	.	.
Mad01.56	.	.	.	.	T	C	.	C	.	.	.	.	.	.	.	.	T	.	.	.	G	-	.	.	.	.	C	C	.	.	.	.	G	.	C	T
Mad01.8	.	T	T	A	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	-	C	.	.	.	.	C	.	.	.	.	G	.	C	.
Mad02-I	.	.	.	.	T	C	.	C	.	.	.	.	.	.	.	A	.	.	A	.	G	-	.	.	.	.	.	C	.	.	.	.	G	.	.	.
MadV	.	.	.	.	T	.	.	C	.	A	.	.	.	.	.	.	.	.	.	A	.	-	.	A	.	C	.	T	.	.	.	.	G	.	.	.
Sub02.a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Sub02.1	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.	.	T	.	.	.
Sub02.4	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
Sub1	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.	T	G	.	.	.
Sub10	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	C	.	.	.	T	G	.	.	.
Sub12	-	-	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.
Sub15	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.
Sub18	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.
Sub3	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.
Sub8	-	-	.	.	T	.	A	C	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	C	-	.	.	.	T	.	.	.
Sub9	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.
Sub97.1	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.
SubAst19	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.
SubAst20	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	A	.	T	.	.	.
SubAst39	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.
SsubAst50	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	A	.	T	.	.	.
SubAst51	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	A	.	T	.	.	.
SubAst72	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.

Figure 5.10. *continues from previous page.*

	636	646	648	663	671	678	683	688	703	704	709	710	719	723	803	811	822	823	829	836	839	841	842	844	845	853	863	908	911	916	923	934	984	1098	1101	1102	1131
ChcuP125	C	G	G	A	C	A	C	C	C	T	-	-	G	G	T	G	C	-	T	C	C	C	C	G	C	G	C	A	A	A	C	G	C	C	A	G	G
Mad01.10	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	A	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	.
Mad01.13	.	C	.	.	.	.	.	.	.	.	A	-	.	.	.	.	.	-	.	A	.	.	.	A	.	T	.	.	.	G	G	.	.	.	.	.	.
Mad01.17	.	.	.	.	.	.	.	.	.	A	-	-	.	.	.	.	.	A	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	-
Mad01.2	.	.	.	.	.	.	T	.	.	A	-	-	.	.	.	.	.	-	C	A	.	.	.	A	.	T	.	.	.	G	G	.	.	.	.	.	.
Mad01.23	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	A	.	A	A	.	.	A	.	T	.	.	.	.	G	.	.	.	.	.	-
Mad01.27	.	.	A	.	.	.	.	.	.	A	-	-	.	.	.	.	.	-	.	A	.	T	.	A	.	T	.	.	.	G	G	.	.	.	.	A	.
Mad01.30	.	.	.	.	.	.	.	.	.	.	A	-	A	.	.	.	.	A	.	.	.	.	.	.	G	.	A	.	.	.	G	.	.	.	.	.	.
Mad01.37	.	.	.	.	.	.	.	.	.	A	-	-	.	.	.	.	.	A	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	-
Mad01.56	.	.	.	.	.	.	.	.	T	.	A	-	.	.	.	.	.	A	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	A
Mad01.8	.	.	.	.	.	.	.	.	.	A	-	-	.	.	.	.	.	A	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	-
Mad02-I	.	C	.	.	.	.	.	.	T	.	A	-	.	.	.	.	.	A	.	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	.
MadV	.	.	.	.	.	.	.	.	T	.	A	-	.	.	.	.	.	-	.	A	.	.	.	A	.	T	.	.	.	G	G	.	.	.	.	.	.
Sub02.a	.	.	.	.	.	.	.	.	-	-	-	-	.	.	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	-
Sub02.1	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	C	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	C	.	-
Sub02.4	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-
Sub1	.	.	.	.	.	T	.	.	.	.	-	-	.	.	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	.	C	.	.
Sub10	.	.	.	.	.	T	.	T	.	.	-	-	.	.	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Sub12	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	T	C	.	.	T	.	T	.	.	.	.	.	.	.	.	A	.	.	C	.	.
Sub15	.	.	.	.	T	.	.	.	.	.	-	-	.	.	.	.	T	A	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.
Sub18	.	.	.	.	.	.	.	.	-	-	-	-	.	.	.	.	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
Sub3	.	.	.	.	.	.	.	.	.	.	T	A	.	.	.	A	T	C	.	.	T	.	.	.	.	.	.	.	C	.	.	A	.	.	C	.	.
Sub8	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	T	C	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.	.
Sub9	.	.	.	.	T	.	.	.	.	.	-	-	.	.	.	.	T	A	.	.	T	.	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.
Sub97.1	T	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
SubAst19	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-
SubAst20	.	.	.	C	.	.	.	.	.	.	-	-	.	A	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	-
SubAst39	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-
SsubAst50	.	.	.	.	.	.	.	.	.	.	T	-	.	.	C	.	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	-
SubAst51	.	.	.	C	.	.	.	.	.	.	-	-	.	A	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	-
SubAst72	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	T	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-

	1132	1133	1152	1154	1156	1180	1235	1265	1266	1267	1274	1284	1314	1376	1388	1401	1409	1410	1421	1436	1443	1450	1453	1473	1475	1477	1478	1489	1492	1493	1494	1514	1532	1534	1543	1589
ChcuP125	T	G	G	G	A	A	G	C	T	G	T	G	A	C	G	T	T	A	A	T	T	T	A	G	T	T	T	T	C	C	G	C	T	C	A	A
Mad01.10	.	A	.	.	G	.	.	.	.	.	.	.	T	.	A	C	.	G	.	.	C	C	.	.	.	C	A	.	T	A	.	T	.	A	C	T
Mad01.13	.	.	.	.	G	.	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	G	.	.	.	A	.	.	.	.	.	.	.	C	.
Mad01.17	-	-	A	.	G	.	.	.	.	.	.	.	T	.	T	C	.	G	.	.	.	C	.	T	.	C	A	.	T	A	.	T	.	.	C	T
Mad01.2	.	.	.	.	G	.	.	.	.	.	.	.	T	.	A	C	.	G	.	.	.	.	.	.	.	C	A	.	T	A	.	T	.	A	C	T
Mad01.23	-	-	A	.	G	.	.	.	.	.	.	.	T	.	A	C	.	G	.	.	.	C	.	.	.	.	A	.	T	A	T	T	.	A	C	T
Mad01.27	.	.	.	.	G	.	.	.	.	.	.	.	T	.	A	C	C	G	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	C	.
Mad01.30	.	.	.	.	G	.	.	.	.	.	.	.	T	.	A	C	C	G	.	.	.	.	.	.	.	.	A	.	T	A	.	T	.	A	C	T
Mad01.37	-	-	A	.	G	.	.	.	.	.	.	.	T	.	T	C	.	G	.	.	.	C	.	T	.	C	A	.	T	A	.	T	.	.	C	T
Mad01.56	.	.	.	T	G	.	.	.	.	.	.	.	T	.	A	C	C	G	.	.	.	C	.	.	.	C	A	.	T	A	.	T	.	.	C	T
Mad01.8	-	-	A	.	G	.	.	.	.	.	.	.	T	.	A	C	.	G	.	.	.	C	.	.	.	C	A	.	T	A	.	T	.	A	C	T
Mad02-I	.	.	.	.	G	.	.	.	.	.	.	.	T	.	A	C	.	G	.	.	.	.	.	.	A	.	A	.	T	A	.	T	.	A	C	T
MadV	.	.	.	.	G	.	.	.	.	.	.	A	T	.	A	C	.	G	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	C	.
Sub02.a	-	-	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	G	.	C	.
Sub02.1	-	-	.	.	.	.	.	T	C	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
Sub02.4	-	-	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.
Sub1	-	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
Sub10	C	.	.	.	.	G	.	.	C	.	G	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	A	.	.	.	.	.	.	.	C	.
Sub12	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.
Sub15	.	.	.	.	.	.	.	T	C	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
Sub18	.	.	.	.	.	.	.	T	C	T	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
Sub3	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
Sub8	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	G	.	.	.	.	-	-	.	A	.	.	.	.	.	.	.	C	.
Sub9	.	.	.	.	.	.	.	T	C	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
Sub97.1	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
SubAst19	-	-	.	.	.	.	A	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	C	.
SubAst20	-	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
SubAst39	-	-	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.
SsubAst50	-	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
SubAst51	-	-	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.
SubAst72	-	-	.	.	.	.	.	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.

The multiple alignment is given relative to the reference sequence (*chcu* strain) of *D. subobscura*. The number above each site indicates its position in the multiple alignment. The dot (.) indicates the same nucleotide as in the reference sequence, and the dash (-) a deleted nucleotide. The blue boxes indicate the conversion tracts identified between the A_2 and A_{st} arrangements of *D. subobscura*.

A general description of nucleotide polymorphism at P125 is shown in Table 5.18. The number of polymorphic sites was similar in *D. madeirensis* and *D. subobscura* despite their different sample size. When samples with the same number of lines were compared, the number of polymorphic sites was higher in *D. madeirensis* (12 lines) than in the A_2 arrangement of *D. subobscura* (12 lines). The percentage of singletons was higher in *D. madeirensis* than in *D. subobscura* (55% and 32%, respectively). In the latter species, the A_2 arrangement presented a higher percentage of singletons (51%) than the A_{st} (32%). At P125, each line of *D. madeirensis* represented a single haplotype. In *D. subobscura*, 4 different haplotypes were detected in the 6 A_{st} lines and 11 different haplotypes in the 12 A_2 lines. Therefore, in this species haplotype diversity was lower than one.

Table 5.18.
Nucleotide polymorphism in P125 region.

	n	*len*	*len*^{w}	*S*	*mut*	*sin*	*%sin*	*inf*	*hap*	*H*
***D. subobscura* A_{st}**	6	1582	1460	31	32	10	32	21	4	0.867
***D. subobscura* A_2**	12	1618	1460	57	58	29	51	28	11	0.985
***D. subobscura* Total**	18	1619	1460	65	66	21	32	44	14	0.967
D. madeirensis	12	1597	1460	69	69	38	55	31	12	1.000

n, sample size or number of sequences; *len*, total number of sites; *len*^{w}, number of sites excluding gaps; *S*, number of polymorphic sites; *mut*, number of mutations; *sin*, number of singletons; *%sin*, percentage of singletons; *inf*, number of parsimony informative sites; *hap*, number of haplotypes; *H*, haplotype diversity.

5.1.4.3. Nucleotide variation

The different estimates of nucleotide variation at P125 region are shown in Table 5.19. Nucleotide variation estimated either by π or θ was higher in *D. madeirensis* than in *D. subobscura*. In the latter, the level of variation was similar in both

arrangements according to π. However, when θ estimates were considered, the levels of variation were higher in A_2 arrangement than in A_{st}.

Table 5.19.
Estimates of nucleotide variation in P125 region.

	n	*S*	*k*	π	θ^{seq}	θ^{sit}
***D. subobscura* A_{st}**	6	31	15.333	0.0105	13.577	0.0093
***D. subobscura* A_2**	12	57	16.045	0.0109	18.875	0.0129
***D. subobscura* Total**	18	65	17.170	0.0117	18.898	0.0129
D. madeirensis	12	69	20.684	0.0142	22.849	0.0156

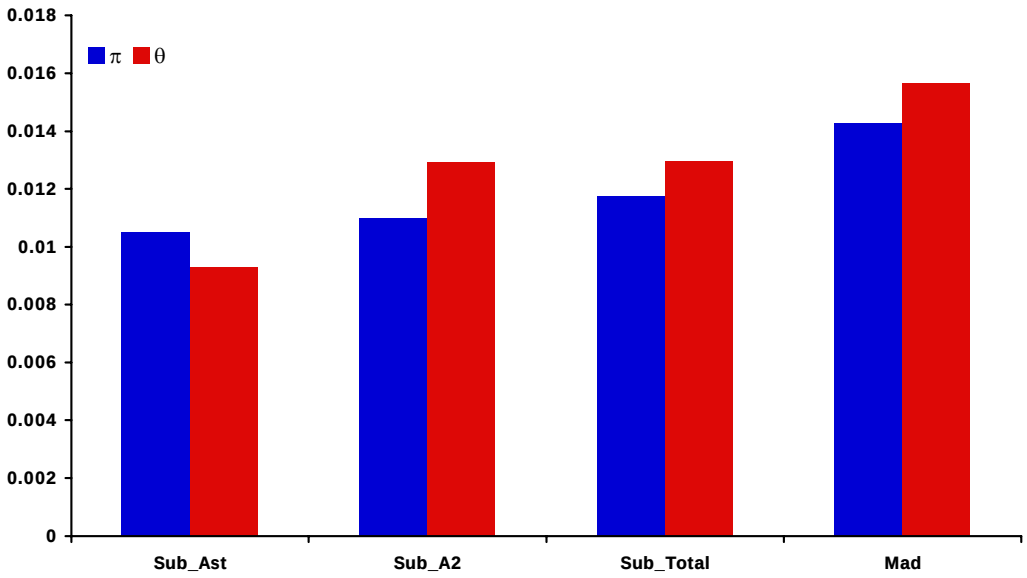
n, sample size or number of sequences; *S*, number of polymorphic sites; *k*, average number of pairwise nucleotide differences; π, nucleotide diversity; θ^{seq}, nucleotide heterozigosity per sequence inferred from *S*; θ^{sit}, nucleotide heterozigosity per site inferred from *S*.

The levels of nucleotide diversity (π) and nucleotide heterozigosity (θ) in the P125 region are compared in Figure 5.11. The estimates of θ were higher than estimates of π in all samples (indicating a general excess of singletons), except in A_{st} arrangement.

5.1.4.4. Recombination

A minimum of 15 recombination events were detected in the *D. madeirensis* sample and 12 in *D. subobscura*, according to the method proposed by Hudson and Kaplan (1985). In the latter species, 9 recombination events were detected in the A_2 and none in the A_{st} arrangement. The estimates of the recombination parameter *R* per site (Hudson 1987) were 0.0474 for *D. madeirensis*, and 0.0879 for *D. subobscura* (0.0795 and 0.0097 for A_2 and A_{st}, respectively).

Figure 5.11.
Nucleotide diversity (π) and nucleotide heterozigosity per site (θ) estimates.



Sub_A_2, *D. subobscura* lines with A_2 arrangement; Sub_A_{st}, *D. subobscura* lines with A_{st} arrangement; Sub_Total, all *D. subobscura* lines studied (A_2 + A_{st}); Mad, *D. madeirensis* lines; π, nucleotide diversity and θ, nucleotide heterozigosity per site.

5.1.4.5. Linkage disequilibrium

The results of the linkage disequilibrium analysis are shown in Table 5.20. After applying Bonferroni correction for multiple tests, 38 (1.9%) chi-square tests remained significant for *D. subobscura*. The global level of linkage disequilibrium (*ZnS* and *Za* estimates) was higher in *D. madeirensis* than in *D. subobscura*. However, the highest level of linkage disequilibrium was detected in the A_{st} arrangement of *D. subobscura*.

Table 5.20.
Linkage disequilibrium analysis for P125 region.

	S_L	*comp*	*Fisher*	*%Fisher*	*chi-square*	*%chi-square*	*ZnS*	*Za*
***D. subobscura* A_{st}**	30	435	0	0	97	22.3	0.4157	0.4824
***D. subobscura* A_2**	56	1540	41	2.7	174	11.3	0.1191	0.1737
***D. subobscura* Total**	64	2010	132	6.6	236	11.7	0.0951	0.1381
D. madeirensis	69	2346	50	2.1	179	7.6	0.1263	0.2369

S_L, number of segregating sites with two variants; *comp*, number of comparisons; *Fisher*, number of significant (P < 0.05) comparisons by the Fisher's test; *%Fisher*, percentage of significant (P < 0.05) comparisons by the Fisher's test; *chi-square*, number of significant (P < 0.05) comparisons by the chi-square test; *%chi-square*, percentage of significant (P < 0.05) comparisons by the chi-square test; *ZnS* (Kelly 1997) and *Za* (Rozas *et al.* 2001) are global measures of linkage disequilibrium.

5.1.4.6. Genetic differentiation between *D. subobscura* arrangements

The average number of nucleotide substitutions per site between the two arrangements was 0.0127 (Table 5.21). Putative significant genetic differentiation between the chromosomal arrangements of *D. subobscura*, was contrasted according to the Hudson *et al.* (1992b) permutation test. The statistical significance of the *Kst** statistic was obtained after 1000 replicates. The observed *Kst** value was 0.0437 and resulted statistically significant (P = 0.03). Therefore, the significant genetic differentiation between arrangements, prevented pooling all *D. subobscura* lines in a single sample. Genetic differentiation between arrangements prevailed, despite the presence of 24 shared polymorphisms and the absence of fixed differences between them. According to the hypergeometric distribution the high number of shared polymorphisms cannot be explained by recurrent mutation. So, the high number of shared polymorphisms between the two arrangements was due to genetic exchange between them. Indeed, 4 gene conversion tracts (3 in A_2 and 1 in A_{st}) were identified (Figure 5.11) by the algorithm proposed by Betrán *et al.* (1997). The largest tract with 1453 nucleotides was detected in the *sub02.a* line. Two tracts including 47 and

229 nucleotides were detected in *sub02.1* and *sub8*, respectively. The unique tract identified in the A_{st} arrangement (*subAst19* line) had 291 nucleotides in length.

Table 5.21.
Genetic differentiation between species and chromosomal arrangements of *D. subobscura*.

	Fixed	*Shared*	S_{x1}	S_{x2}	*K*	*Dxy*
A_2 / A_{st}	0	24	34	8	18.853	0.0127
A_2 / *D. mad*	3	4	54	65	33.840	0.0231
A_{st} / *D. mad*	6	2	30	67	36.361	0.0249

A_2, *D. subobscura* lines with A_2 arrangement; A_{st}, *D. subobscura* lines with A_{st} arrangement; *D. mad*, *D. madeirensis* lines; Fixed, fixed differences between samples; Shared, polymorphic sites segregating for the same two variants in two samples; S_{x1}, exclusive polymorphisms sites in population 1; S_{x2}, exclusive polymorphisms sites in population 2; *K*, average number of nucleotide differences between arrangements or species; D_{xy}, average number of nucleotide substitutions per site between samples.

5.1.4.7. Genetic differentiation between species

The average number of nucleotide substitutions per site between the two species was similar independently of the arrangement used to compare with *D. madeirensis* (Table 5.21). There was four shared polymorphisms between *D. madeirensis* and *D. subobscura* and 3 fixed differences between them.

5.1.4.8. Patterns of polymorphism

Tajima's (1989) and Fu and Li's (1993) statistics were negative in *D. madeirensis* and in A_2 arrangement of *D. subobscura*, which indicated an excess of singletons variants in these samples (Table 5.22). In contrast, these tests were positive in the A_{st} arrangement. However, none of tests were statistical significant.

Table 5.22.
Tajima's and Fu and Li's statistics for P125 region.

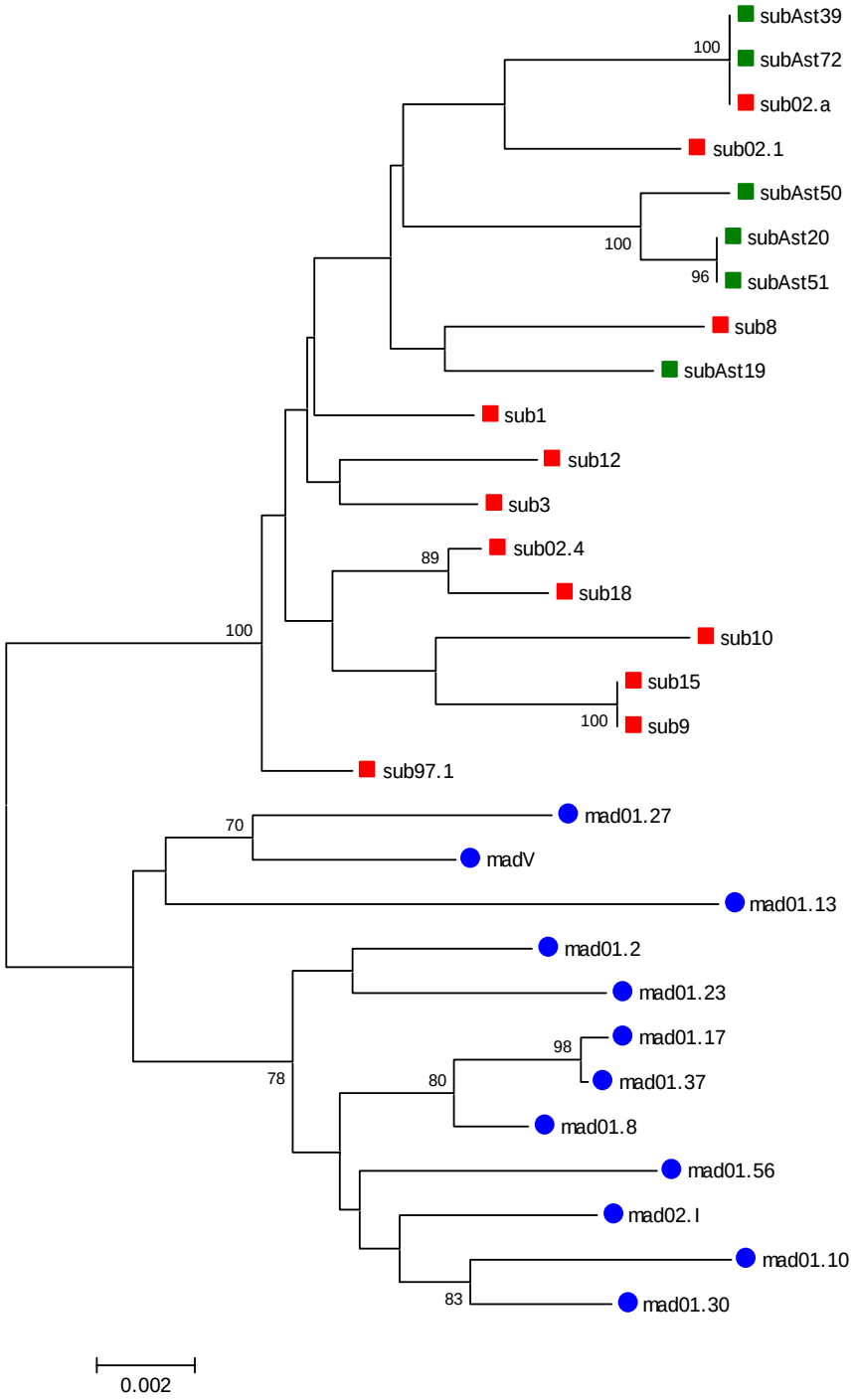
	Tajima's D	*Fu and Li's D*	*Fu and Li's F*	*Fu and Li's D**	*Fu and Li's F**
***D. subobscura* A_{st}**	0.5971	1.0081	1.0925	0.5961	0.6521
***D. subobscura* A_2**	-0.7584	-0.8139	-1.0047	-0.6322	-0.7588
***D. subobscura* Total**	-0.4373	-0.1558	-0.3386	-0.0579	-0.1964
D. madeirensis	-0.4021	-1.2299	-1.2061	-0.8679	-0.8499

5.1.4.9. Gene genealogy

The gene genealogy reconstructed from variation at P125 is shown in Figure 5.12. All *D. madeirensis* lines grouped together in a single cluster. In *D. subobscura* a partial clustering of lines according to their gene arrangement was also detected. The A_{st} lines grouped together in a single cluster, although this cluster also included three A_2 lines (*sub02.a*, *sub02.1* and *sub8*). However, these three lines presented gene conversion tracts from A_{st}, which may explain their clustering with the A_{st} lines.

Figure 5.12.

Gene genealogy reconstructed from nucleotide variation in P125 region.



The gene genealogy was obtained by the neighbour-joining method (Saitou and Nei 1987) using, as genetic distance, the number of substitutions per site, according to Kimura 2-parameter method (Kimura 1980). Bootstrap confidence level (from 1000 replications) above 70% are shown below the branches. The horizontal bar at the bottom represents the distance scale of the branches.

5.1.5. P275 region

5.1.5.1. Studied region

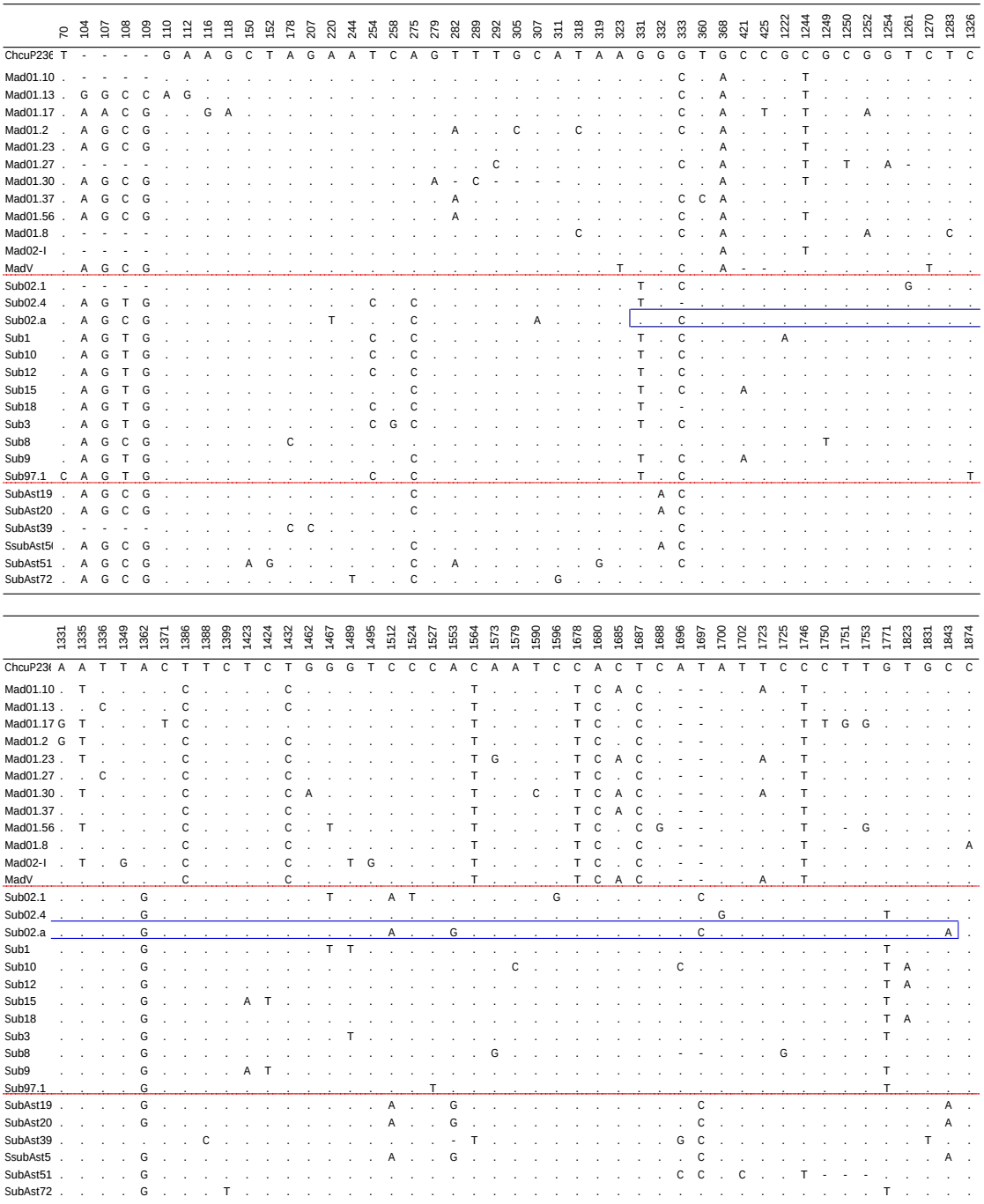
P275 phage was digested with *EcoRI*. Four fragments were obtained (of about 1.8, 3, 5 and 6 kb). These fragments were subcloned and their ends were sequenced using T7 and SK universal primers. BLAST algorithm was used to search for conserved sequences in the *D. melanogaster* genome (release 3.1). Multiple hits with high similarity were found in a 14 kb fragment, with some transposons. A 1.5 kb region was chosen for further analysis avoiding possible tranposons.

5.1.5.2. Nucleotide polymorphism

The multiple alignment of the P275 region in 12 *D. madeirensis* lines and 18 *D. subobscura* lines (12 with the A_2 arrangement and 6 A_{st}) included a total of 1922 sites. This number dropped to 974 when sites with alignment gaps were excluded. All sites were silent as P275 corresponded to a non coding region. A total of 71 polymorphic sites (40 singletons) were detected in the data set and was inferred that the minimum number of mutations was 71. Figure 5.13 shows nucleotide variation of P275 region at polymorphic sites. Sites with alignment gaps were completely excluded from all analysis.

Figure 5.13.

Polymorphic sites in P275 region.



The multiple alignment is given relative to the reference sequence (*chcu* strain) of *D. subobscura*. The number above each site indicates its position in the multiple alignment. The dot (.) indicates the same nucleotide as in the reference sequence, and the dash (-) a deleted nucleotide. The blue box indicate the conversion tracts identified between the A_2 and A_{st} arrangements of *D. subobscura*.

A general description of nucleotide polymorphism at P275 is shown in Table 5.23. The number of polymorphic sites was higher in *D. subobscura* than in *D. madeirensis*, despite their different sample size. When samples with the same number of lines were compared, the number of polymorphic sites was higher in *D. madeirensis* (12 lines) than in the A_2 arrangement of *D. subobscura* (12 lines). The percentage of singletons was higher in *D. madeirensis* than in *D. subobscura* (74% and 66%, respectively). In *D. subobscura*, the A_{st} arrangement presented a higher percentage of singletons (83%) than the A_2 (60%). In P275 region, each line of *D. madeirensis* represented a single haplotype. In *D. subobscura*, 4 different haplotypes were detected in the 6 A_{st} lines and 10 different haplotypes in the 12 A_2 lines, thus haplotype diversity in this species was lower than one.

Table 5.23.
Nucleotide polymorphism in P275 region.

	n	*len*	*len*^{w}	*S*	*mut*	*sin*	*%sin*	*inf*	*hap*	*H*
***D. subobscura* A_{st}**	6	1139	974	18	18	15	83	3	4	0.800
***D. subobscura* A_2**	12	1908	974	25	25	15	60	10	10	0.970
***D. subobscura* Total**	18	1916	974	38	38	25	66	13	14	0.967
D. madeirensis	12	1112	974	31	31	23	74	8	12	1.000

n, sample size or number of sequences; *len*, total number of sites; *len*^{w}, number of sites excluding gaps; *S*, number of polymorphic sites; *mut*, number of mutations; *sin*, number of singletons; *%sin*, percentage of singletons; *inf*, number of parsimony informative sites; *hap*, number of haplotypes; *H*, haplotype diversity.

5.1.5.3. Nucleotide variation

The different estimates of nucleotide variation at P275 region are shown in Table 5.24. Nucleotide variation was similar between *D. madeirensis* and *D. subobscura* according to π or θ estimates. In the latter species, the level of nucleotide variation estimated either by π or θ was also similar between both arrangements.

Table 5.24.
Estimates of nucleotide variation in P275 region.

	n	*S*	*k*	π	θ^{seq}	θ^{sit}
***D. subobscura* A_{st}**	6	18	6.800	0.0069	7.883	0.0080
***D. subobscura* A_2**	12	25	5.970	0.0061	8.278	0.0085
***D. subobscura* Total**	18	38	7.046	0.0072	11.048	0.0113
D. madeirensis	12	31	7.000	0.0071	10.265	0.0105

n, sample size or number of sequences; *S*, number of polymorphic sites; *k*, average number of pairwise nucleotide differences; π, nucleotide diversity; θ^{seq}, nucleotide heterozigosity per sequence inferred from *S*; θ^{sit}, nucleotide heterozigosity per site inferred from *S*.

The levels of nucleotide diversity (π) and nucleotide heterozigosity (θ) are compared in Figure 5.14. The estimates of θ were higher than the π estimates in all samples. This result indicated a general excess of singletons in this region, mainly in *D. subobscura*.

5.1.5.4. Recombination

A minimum of 4 recombination events were detected in the *D. madeirensis* sample and 2 in *D. subobscura*, according to the method proposed by Hudson and Kaplan (1985). In the latter species, 2 recombination events were detected in the A_2 and none in the A_{st} arrangement.

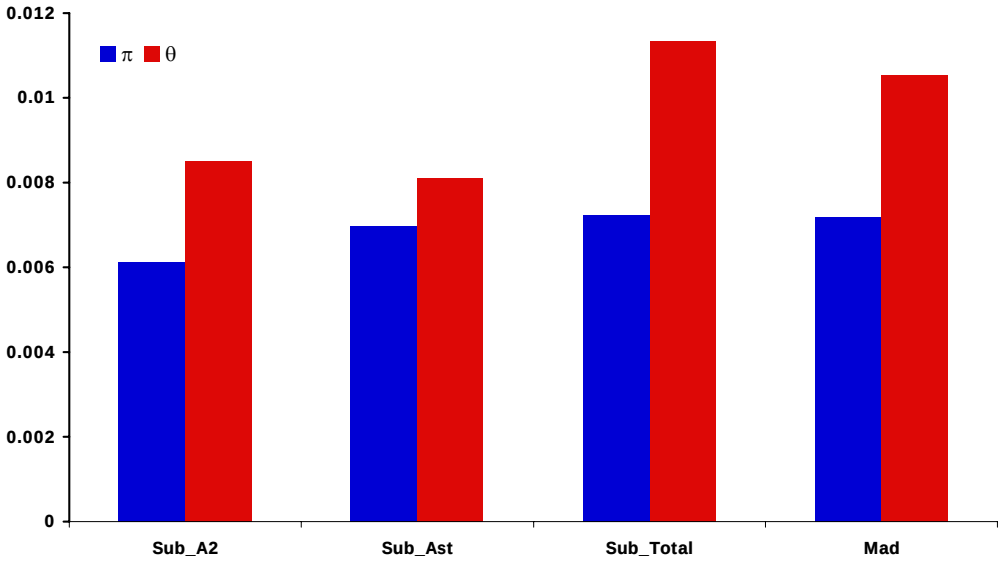
5.1.5.5. Linkage disequilibrium

The results of the linkage disequilibrium analysis are shown in Table 5.25. After applying Bonferroni correction for multiple tests, 29 (4.1%) chi-square tests remained significant for *D. subobscura*. The global level of linkage disequilibrium (*ZnS* and *Za* estimates) was higher in *D. madeirensis* than in *D. subobscura*.

However, the highest level of linkage disequilibrium was detected in the A_{st} arrangement of *D. subobscura*.

Figure 5.14.

Nucleotide diversity (π) and nucleotide heterozigosity per site (θ) estimates.



Sub_A_2, *D. subobscura* lines with A_2 arrangement; Sub_A_{st}, *D. subobscura* lines with A_{st} arrangement; Sub_Total, all *D. subobscura* lines studied (A_2 + A_{st}); Mad, *D. madeirensis* lines; π, nucleotide diversity and θ, nucleotide heterozigosity per site.

Table 5.25.

Linkage disequilibrium analysis for P275 region.

	S_L	*comp*	*Fisher*	*%Fisher*	*chi-square*	*%chi-square*	*ZnS*	*Za*
***D. subobscura* A_{st}**	18	153	0	0	37	24.1	0.3192	0.3129
***D. subobscura* A_2**	25	300	5	1.6	37	12.3	0.1161	0.1083
***D. subobscura* Total**	38	703	12	1.7	68	9.6	0.0819	0.1270
D. madeirensis	31	465	2	0.4	38	8.1	0.0906	0.2070

S_L, number of segregating sites with two variants; *comp*, number of comparisons; *Fisher*, number of significant (P < 0.05) comparisons by the Fisher's test; *%Fisher*, percentage of significant (P < 0.05) comparisons by the Fisher's test; *chi-square*, number of significant (P < 0.05) comparisons by the chi-square test; *%chi-square*, percentage of significant (P < 0.05) comparisons by the chi-square test; *ZnS* (Kelly 1997) and *Za* (Rozas *et al.* 2001) are global measures of linkage disequilibrium.

5.1.5.6. Genetic differentiation between *D. subobscura* arrangements

The average number of nucleotide substitutions per site between the two arrangements was 0.0083 (Table 5.26). Putative significant genetic differentiation between the chromosomal arrangements of *D. subobscura*, was contrasted according to the Hudson *et al.* (1992b) permutation test. The statistical significance of the *Kst** statistic was obtained after 1000 replicates. The observed *Kst** value was 0.083 and resulted statistically significant (P = 0.002). Therefore, the significant genetic differentiation between arrangements, prevented pooling all *D. subobscura* lines in a single sample. Genetic differentiation between arrangements prevailed, despite the presence of 5 shared polymorphisms and the absence of fixed differences between them. According to the hypergeometric distribution the high number of shared polymorphisms cannot be explained by recurrent mutation. So, the high number of shared polymorphisms between the two arrangements was due to genetic exchange between them. Indeed, one gene conversion tract (in A_2) was identified (Figure 5.14) by the algorithm proposed by Betrán *et al.* (1997). This tract was identified in the *sub02.a* line with a length of 728 nucleotides.

Table 5.26.
Genetic differentiation between species and chromosomal arrangements of *D. subobscura*.

	Fixed	*Shared*	S_{x1}	S_{x2}	*K*	*Dxy*
A_2 / A_{st}	0	5	20	13	8.083	0.0083
A_2 / *D. mad*	8	3	22	28	18.674	0.0192
A_{st} / *D. mad*	5	0	18	31	17.167	0.0176

A_2, *D. subobscura* lines with A_2 arrangement; A_{st}, *D. subobscura* lines with A_{st} arrangement; *D. mad*, *D. madeirensis* lines; Fixed, fixed differences between samples; Shared, polymorphic sites segregating for the same two variants in two samples; S_{x1}, exclusive polymorphisms sites in population 1; S_{x2}, exclusive polymorphisms sites in population 2; *K*, average number of nucleotide differences between arrangements or species; D_{xy}, average number of nucleotide substitutions per site between samples.

5.1.5.7. Genetic differentiation between species

The average number of nucleotide substitutions per site between the two species was similar, independently of the arrangement used to compare with *D. madeirensis* (Table 5.26). There were three shared polymorphisms between *D. madeirensis* and *D. subobscura* and five fixed differences between them.

5.1.5.8. Patterns of polymorphism

Tajima's (1989) and Fu and Li's (1993) statistics were negative in all samples, which indicated an excess of singletons variants (Table 5.27). None of the performed tests were statistical significant. However, Tajima's *D* and Fu and Li's *D* statisitcs were significant (*P(D)* = 0.048 and *P(D)* = 0.049, respectively) in *D. subobscura*, after computer simulations under the conservative assumption of no recombination.

Table 5.27.
Tajima's and Fu and Li's statistics for P275 region.

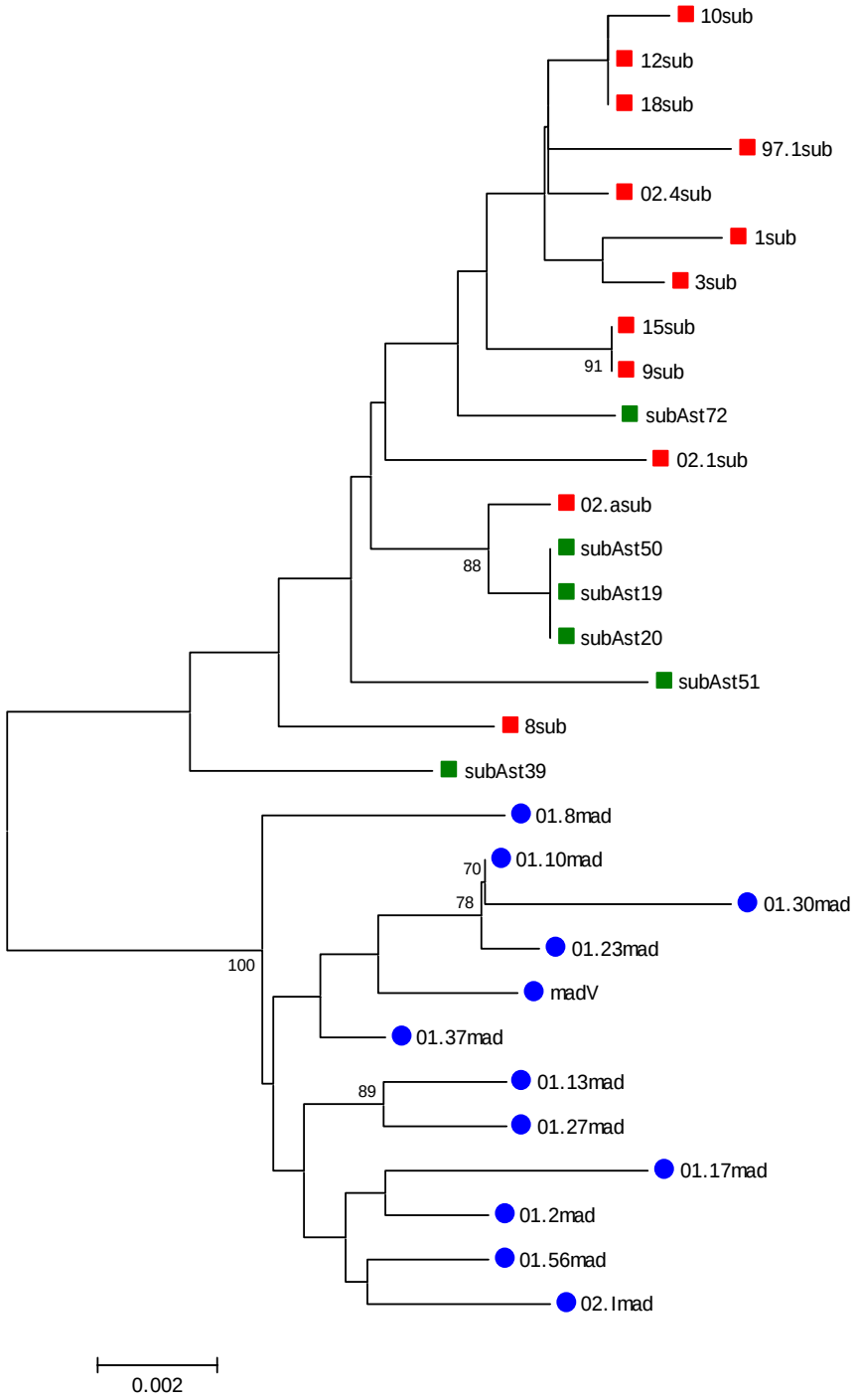
	Tajima's D	*Fu and Li's D*	*Fu and Li's F*	*Fu and Li's D**	*Fu and Li's F**
***D. subobscura* A_{st}**	-0.8545	-0.7811	-0.9643	-0.9885	-1.0448
***D. subobscura* A_2**	-1.2425	-1.5233	-1.7562	-1.0430	-1.2473
***D. subobscura* Total**	-1.4741	-1.9646	-2.2062	-1.8992	-2.0603
D. madeirensis	-1.4332	-2.0122	-2.2478	-1.6864	-1.8457

5.1.5.9. Gene genealogy

The gene genealogy reconstructed from variation at P275 is shown in Figure 5.15. All *D. madeirensis* grouped together in a single cluster. In *D. subobscura* a partial clustering of lines according to their gene arrangement was also detected. The A_{st} lines grouped together in a single cluster, although this cluster also included three A_2 lines (*sub02.1*, *sub02.a*, and *sub8*).

Figure 5.15.

Gene genealogy reconstructed from nucleotide variation in P275 region.



The gene genealogy was obtained by the neighbour-joining method (Saitou and Nei 1987) using, as genetic distance, the number of substitutions per site, according to Kimura 2-parameter method (Kimura 1980). Bootstrap confidence level (from 1000 replications) above 70% are shown below the branches. The horizontal bar at the bottom represents the distance scale of the branches.

5.2. *Drosophila subobscura* species analysis

5.2.1. Total sample

5.2.1.1. Nucleotide variation

Estimates of nucleotide diversity in *D. subobscura* (18 lines) in each of the five studied regions and in the concatenated data set are shown in Table 5.28. Silent nucleotide diversity (π_{sil}) ranges from 0.0055 (*Sxl*) to 0.0117 (P125).

Table 5.28.
Estimates of nucleotide variation in *D. subobscura*.

	n	*S*	π_{tot}	π_{sil}	θ_{tot}	θ_{sil}
P236	18	50	0.0093	0.0093	0.0094	0.0094
P150	18	39	0.0055	0.0060	0.0089	0.0099
Sxl	18	38	0.0049	0.0055	0.0067	0.0077
P125	18	65	0.0117	0.0117	0.0129	0.0129
P275	18	38	0.0072	0.0072	0.0113	0.0113
Concatenated data	18	50	0.0078	0.0081	0.0097	0.0102

n, sample size or number of sequences; *S*, number of polymorphic sites; π_{tot}, nucleotide diversity in all sites; π_{sil}, silent nucleotide diversity; θ_{tot}, heterozygosity per site based on the number of segregating sites; θ_{sil}, silent heterozygosity per site based on the number of segregating sites.

5.2.1.2. Patterns of polymorphism

Tajima's *D* (1989) and Fu and Li's *D* (1993) statistics were negative in all regions, indicating an excess of singletons variants in *D. subobscura*, although only Fu and

Li's *D* statistic was statistically significant in P150 region. However, these tests also turned out to be significant (0.01 < *P* < 0.05) in the P275 and *Sxl* regions, after computer simulations under the conservative assumption of no recombination. This trend detected toward negative values of these statistics was further analyzed using the multilocus test based on the mean value of the Tajima's *D* statistic ($\overline{D}$) after 10.000 computer simulations. The empirical *D* -value averaged across the five studied regions ($\overline{D}$ = -0.934) was not significantly lower (*P* = 0.166, one-tailed test) than the average *D*-value obtained from computer simulations. A similar result was obtained for the multilocus test based on Fu and Li's *D* statistic ($\overline{D}$ = -1.133, *P* = 0.122).

The HKA test (Hudson *et al.* 1987) was also performed to compare the level of polymorphism in *D. subobscura* and of divergence between *D. subobscura* and *D. madeirensis* among regions. No decoupling between polymorphism and divergence was detected in any of the 10 pairwise comparisons performed. A similar result was obtained by the HKA multilocus test, in which all regions were jointly analyzed (χ^2 = 1.828, 8 df, *P* = 0.985).

5.2.2.Comparison between chromosomal arrangements of *D. subobscura*

5.2.2.1. Nucleotide variation

Estimates of nucleotide diversity (π_{tot} and π_{sil}) in the A_{st} and the A_2 chromosomal arrangements of *D. subobscura* were very similar in each of the five studied regions and in the concatenated data set (Table 5.29).

The level of silent nucleotide variation in the five studied regions was analyzed in regard to the physical distance of each region to the A_2 inversion nearest breakpoint.

With this purpose, π_{sil} estimates were corrected by K_{sil} (silent divergence) between *D. subobscura* and *D. madeirensis*. The nearest region to an inversion breakpoint (P236) was the region with the lowest level of variation in both arrangements (Figure 5.16). The low level of variation relative to divergence at P236 was, however, not significant. In fact, none of the HKA tests (Hudson *et al.* 1987) performed between P236 and the other studied regions were significant either in A_2 or A_{st}. A similar result was obtained in pairwise comparison between P150, *Sxl*, P125 and P275 regions. The HKA multilocus tests performed within arrangements were also not significant (χ^2= 1.65, 8 df, P = 0.98 in A_2 and χ^2 = 1.25, 8 df, P = 0.99 in A_{st}). Therefore, no heterogeneity in the ratio between polymorphism and divergence among the different regions was detected.

Table 5.29.

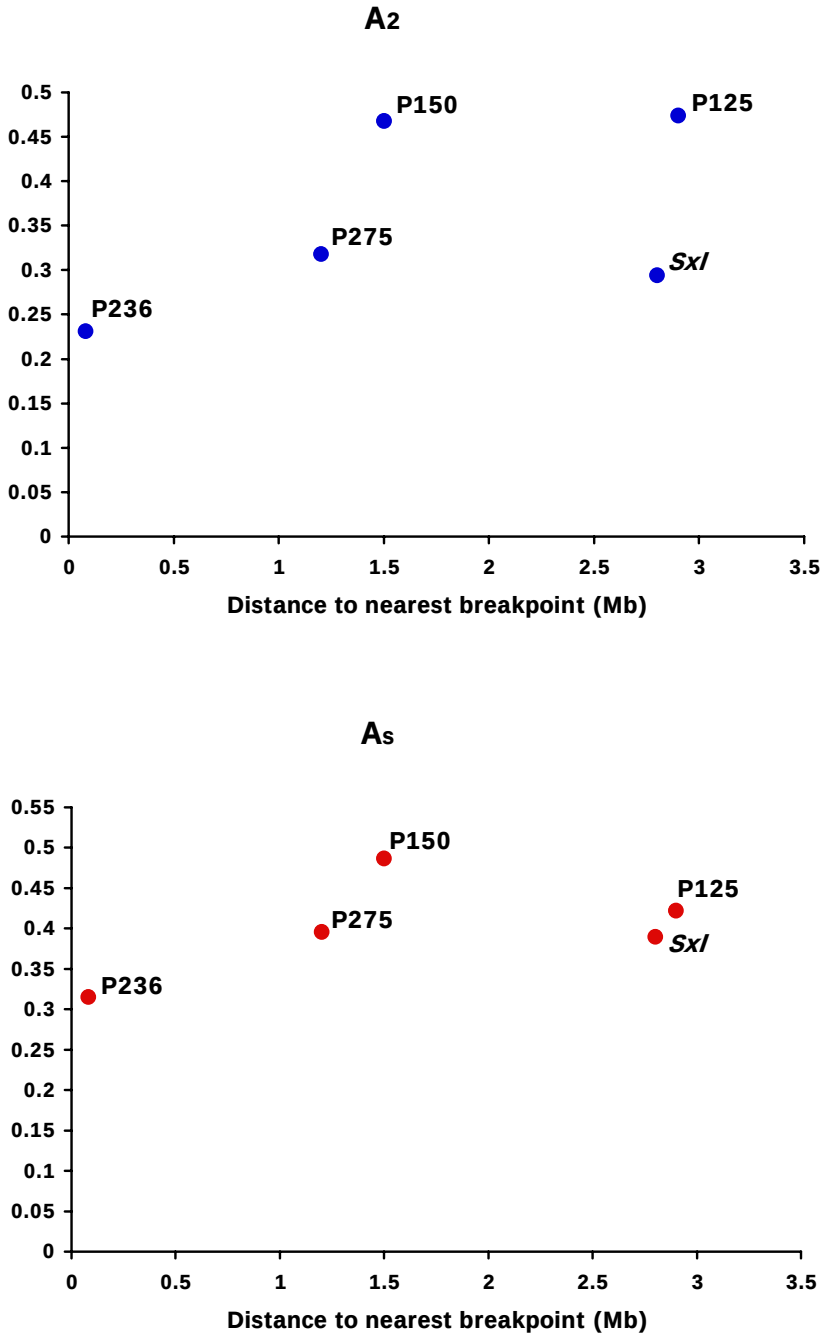
Estimates of nucleotide variation within the chromosomal arrangements of *D. subobscura*.

		n	*S*	π_{tot}	π_{sil}	θ_{sil}	*Ksil*
P236	**A_{st}**	6	28	0.0076	0.0076	0.008	0.024
	A_2	12	37	0.006	0.006	0.008	0.026
P150	**A_{st}**	6	18	0.0053	0.0063	0.0072	0.012
	A_2	12	26	0.005	0.0054	0.0074	0.011
Sxl	**A_{st}**	6	21	0.0056	0.0063	0.0065	0.016
	A_2	12	28	0.0042	0.0047	0.0063	0.015
P125	**A_{st}**	6	31	0.0105	0.0105	0.0093	0.024
	A_2	12	57	0.0109	0.0109	0.0129	0.023
P275	**A_{st}**	6	18	0.0069	0.0069	0.008	0.017
	A_2	12	25	0.0061	0.0061	0.0085	0.019
Concatenated data	**A_{st}**	6	116	0.0072	0.0079	0.0082	0.019
	A_2	12	173	0.0065	0.0067	0.0085	0.019

n, sample size or number of sequences; *S*, number of polymorphic sites; π_{tot}, nucleotide diversity in all sites; π_{sil}, silent nucleotide diversity; θ_{sil}, silent heterozygosity per site based on the number of segregating sites; K_{sil}, silent divergence per site between *D. subobscura* and *D. madeirensis*.

Figure 5.16.

π_{sil}/K_{sil} ratio in each region versus its distance to the A_2 inversion nearest breakpoint in each chromosomal arrangement



Silent divergence per site (K_{sil}) was estimated between *D. madeirensis* and each one of the *D. subobscura* arrangements. The distance between each region and the A_2 inversion nearest breakpoint was estimated assuming a homogenous DNA distribution along the A_2 inversion (7.1 Mb).

5.2.2.2. Genetic differentiation between arrangements

A significant genetic differentiation between the A_{st} and A_2 arrangements was detected in each of the five studied regions, and in the concatenated data set. Therefore, both arrangements were genetically differentiated despite the lack of fixed differences between them and the presence of 60 shared polymorphisms. Estimates of genetic differentiation (D_{xy}) between the A_2 and A_{st} arrangements range from 0.0054 (*Sxl*) to 0.0127 (P236 and P125) (Table 5.30). When D_{xy} was corrected by the level of variation in each region (D_a), *Sxl* still exhibited the lowest level of differentiation and P236 the highest.

Table 5.30.
Genetic differentiation between arrangements of *D. subobscura*.

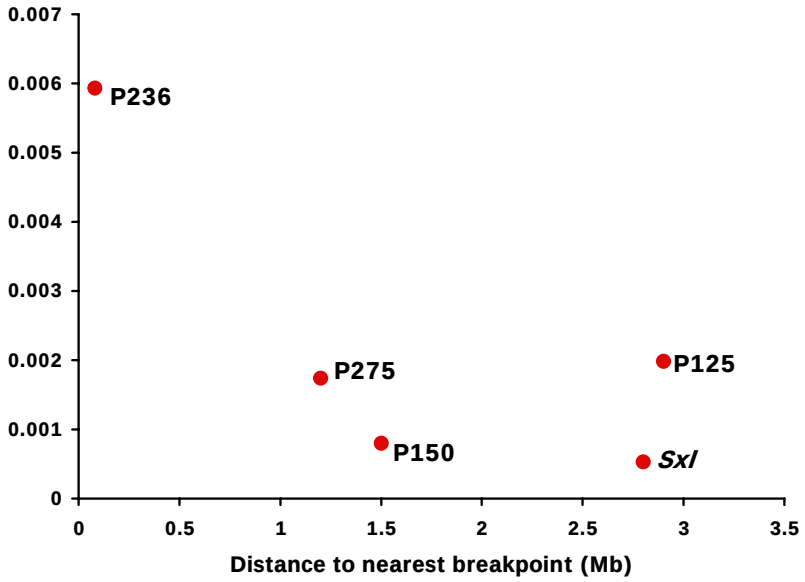
	Fixed	*Shared*	S_{x1}	S_{x2}	D_{xy}	D_a	$P(K_{st})$
P236	0	15	22	13	0.0127	0.0059	0.001
P150	0	7	22	15	0.0071	0.0008	0.002
Sxl	0	11	17	11	0.0054	0.0005	0.020
P125	0	24	34	8	0.0127	0.0019	0.030
P275	0	5	20	13	0.0083	0.0017	0.002
Concatenated data	0	60	114	58	0.0091	0.0022	0.000

Fixed, fixed differences among arrangements; *Shared*, polymorphic sites segregating for the same two variants; S_{x1}, sites polymorphic in A_2 and monomorphic in A_{st}; S_{x2}, sites polymorphic in A_{st} and monomorphic in A_2; D_{xy}, average number of nucleotide differences per site between arrangements; D_a, net number of nucleotide substitutions per site between arrangements; $P(K_{st})$, genetic diferentiation test statistic probability.

Genetic differentiation between arrangements was also analyzed in relation to the distance of each region to the A_2 inversion nearest breakpoint (Figure 5.17). The region with the strongest net genetic differentiation (P236) was also the closest region to an inversion breakpoint. Genetic differentiation declined with distance in the other regions, except in P125.

Figure 5.17.

Net genetic differentiation between arrangements in each region versus its distance to the A_2 inversion nearest breakpoint.



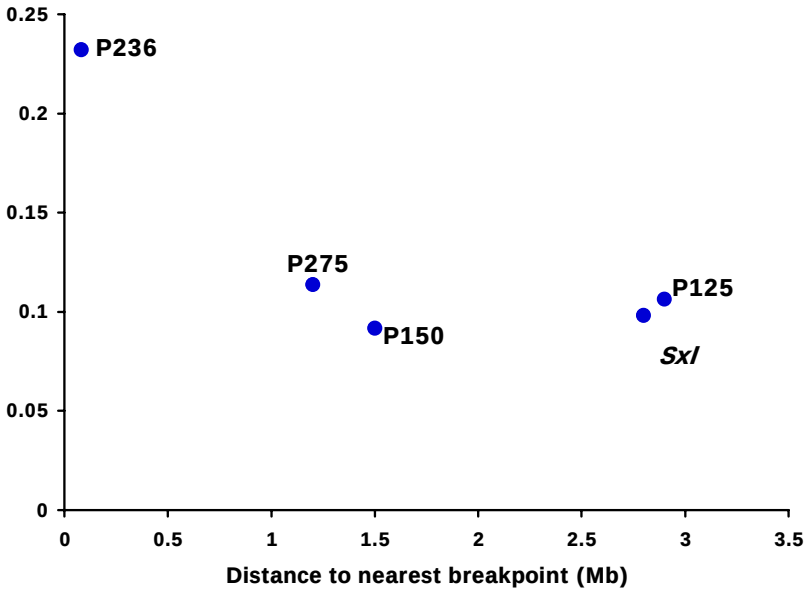
D_a, net number of nucleotide substitutions per site between arrangements. The distance between each region and the nearest A_2 inversion breakpoint was estimated assuming a homogenous DNA distribution along the A_2 inversion (7.1 Mb).

5.2.2.3. Linkage disequilibrium

Association between chromosomal arrangements (A_2 and A_{st}) and the variants present at informative sites was also analyzed using the average r^2 as a measure of linkage disequilibrium (Figure 5.18). The closest region to an inversion breakpoint (P236) showed the strongest association between chromosomal arrangement and nucleotide variation. This association decreased considerably in the other four regions. The extent of association in these regions was, however, very similar, and thus it was not related to its distance to the nearest inversion breakpoint that ranges from 1.5 Mb to 2.9 Mb.

Figure 5.18.

Association between chromosomal arrangements and variants at informative polymorphic sites in each region (average r^2) versus its distance to the nearest inversion breakpoint



The distance between each region and the A_2 inversion nearest breakpoint was estimated assuming a homogenous DNA distribution along the A_2 inversion (7.1 Mb).

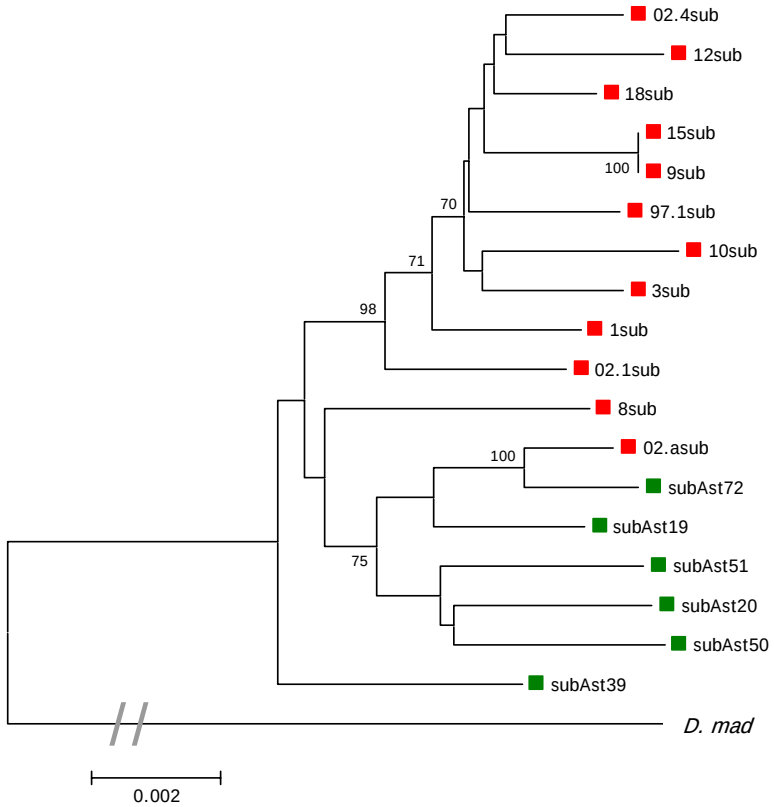
5.2.2.4. Patterns of polymorphism

Tajima's *D* (1989) and Fu and Li's *D* (1993) statistics were negative in the five studied regions in A_2, but only in four in A_{st}. These data were also analyzed by a multilocus test based on the mean value of Tajima's test statistic ($\overline{D}$). In A_{st}, the empirical *D* -value averaged across the five studied regions ($\overline{D}$ = -0.3248) was not significantly lower (one-tailed test) than the average *D* -value obtained from the computer simulations (*P* = 0.25). However, the multilocus test was statistically significant for the A_2 arrangement ($\overline{D}$ = -1.0824; *P* = 0.005). This result indicated a significant excess of low frequency variants, mostly singletons, in this arrangement. A similar result was obtained for the multilocus test based on Fu and Li's *D* statistic: $\overline{D}$ = -1.2962, *P* = 0.004 in A_2, and $\overline{D}$ = -0.5267, *P* = 0.16 in A_{st}.

5.2.2.5. Gene genealogy

The gene genealogy reconstructed from the total variation in the concatenated data set is shown in Figure 5.19. There was a partial clustering of lines according to their chromosomal arrangement. Most A_2 lines grouped together in a single cluster with a rather high bootstraap support. However, two A_2 lines (*8sub* and *02.asub*) clustered with the A_{st} lines. Gene conversion tracts were identified in three of the five studied regions in these two lines. Therefore, the detected genetic exchange between both arrangements may explain the clustering of *8sub* and *02.asub* with the A_{st} lines.

Figure 5.19.
Gene genealogy reconstructed from the total nucleotide variation in *D. subobscura* lines.



The gene genealogy was obtained by the neighbour-joining method (Saitou and Nei 1987) using, as genetic distance, the number of substitutions per site, according to Kimura 2-parameter method (Kimura 1980). Bootstrap confidence level (from 1000 replications) above 70% are shown below the branches. The horizontal bar at the bottom represents the distance scale of the branches. *D. madeirensis* was used as outgroup.

5.3. *Drosophila madeirensis* species analysis

5.3.1. Total sample

5.3.1.1. Nucleotide variation

Estimates of nucleotide diversity in *D. madeirensis* (12 lines) in the five studied regions and in the concatenated data set are shown in Table 5.31. Silent nucleotide diversity (π_{sil}) ranges from 0.0065 in P150 region to 0.0142 in P125.

Table 5.31.
Estimates of nucleotide variation in *D. madeirensis*.

	n	*S*	π_{tot}	π_{sil}	θ_{tot}	θ_{sil}
P236	12	55	0.0078	0.0078	0.0119	0.0119
P150	12	23	0.0054	0.0065	0.0060	0.0071
Sxl	12	55	0.0080	0.0089	0.0111	0.0124
P125	12	69	0.0142	0.0142	0.0156	0.0156
P275	12	31	0.0071	0.0071	0.0105	0.0105
Concatenated data	12	233	0.0087	0.0092	0.0112	0.0118

n, sample size or number of sequences; *S*, number of polymorphic sites; π_{tot}, nucleotide diversity in all sites; π_{sil}, silent nucleotide diversity; θ_{tot}, heterozygosity per site based on the number of segregating sites; θ_{sil}, silent heterozygosity per site based on the number of segregating sites.

5.3.1.2. Patterns of polymorphism

Tajima's *D* (1989) and Fu and Li's *D* (1993) statistics were negative in the five studied regions, although only Fu and Li's *D* statistic was statistically significant in P236 region. Also, Tajima's *D* statistic was statistically significant ($0.01 < P < 0.05$) in

the P236 region, after coalescent simulations under the conservative assumption of no recombination. Data were also analyzed using the multilocus test based on the mean value of Tajima's *D* statistic ($\overline{D}$). The empirical *D* -value averaged across the five regions studied ($\overline{D}$ = -1.030) was significantly lower (*P* = 0.008, one-tailed test) than the average *D*-value obtained from computer simulations. A similar result was obtained for the multilocus test based on Fu and Li's *D* statistic. Therefore, an overall significant excess of low-frequency variants, mainly singletons, was detected in *D. madeirensis*.

The HKA test was performed to contrast levels of polymorphism in *D. madeirensis* and of divergence between *D. madeirensis* and *D. subobscura* among regions. No decoupling between polymorphism and divergence was detected in any of the 10 pairwise comparisons performed. A similar result was obtained in the HKA multilocus test (χ^2 = 1.828, 8 df, *P* = 0.985).

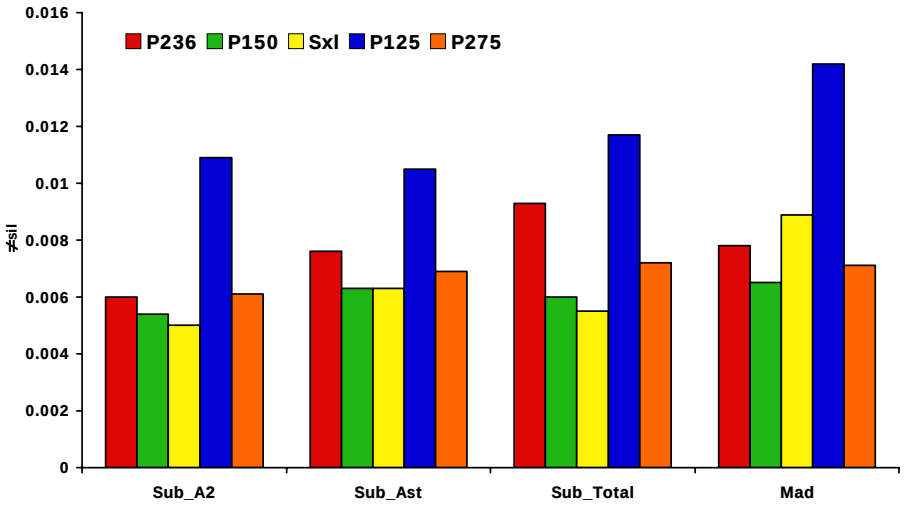
5.4. Comparison between *D. subobscura* and *D. madeirensis*

5.4.1. Species comparison

5.4.1.1. Nucleotide variation

Generally, silent nucleotide diversity (π_{sil}) was higher in *D. madeirensis* than in *D. subobscura* across the five studied regions (Figure 5.20). The only exception was P236, where silent nucleotide diversity was higher in *D. subobscura* (total sample). The highest level of silent nucleotide diversity was detected at P125 both in *D. madeirensis* and in *D. subobscura*. In the latter species, this result holds for the total sample or when each arrangement was analyzed independently. In contrast, the region with the lowest level of silent nucleotide diversity differed between both species: in *D. subobscura* (total sample) was the *Sxl* region, but in *D. madeirensis* was P150.

Figure 5.20.
Silent nucleotide diversity across the studied genomic regions.



π_{sil}, nucleotide diversity in silent sites; Sub_A_2, *D. subobscura* lines with A_2 arrangement; Sub_A_{st}, *D. subobscura* lines with A_{st} arrangement; Sub_Total, all *D. subobscura* lines studied (A_2 + A_{st}) and Mad, *D. madeirensis* lines.

The levels of nucleotide diversity in the concatenated data set are shown in Table 5.32. According to the highest level of polymorphism detected in most regions in *D. madeirensis*, the level of nucleotide diversity in the concatenated data set was also higher in this species than in any *D. subobscura* sample.

Table 5.32.
Nucleotide variation estimates in the concatenated data set.

	π_{tot}	π_{sil}	θ_{tot}	θ_{sil}
***D. subobscura* A_{st}**	0.0072	0.0079	0.0074	0.0082
***D. subobscura* A_2**	0.0065	0.0067	0.0084	0.0085
***D. subobscura* Total**	0.0078	0.0081	0.0098	0.0102
D. madeirensis	0.0087	0.0092	0.0112	0.0118

π_{tot}, nucleotide diversity in all sites; π_{sil}, nucleotide diversity in silent sites; θ_{tot}, heterozygosity per site based on segregating sites; θ_{sil}, heterozygosity per silent site based on segregating sites.

5.4.1.2. Patterns of polymorphism

The empirical Tajima's *D* – value averaged across the five studied regions was negative in all samples and the multilocus test based on this statistic resulted significant in *D. madeirensis* and in the A_2 arrangement of *D. subobscura* (Table 5.33). The same result was detected with Fu and Li's *D* statistic, although in this case the multilocus test was also significant for the *D. subobscura* total sample.

5.4.1.3. Shared variation and sequence divergence

The number of shared polymorphisms and fixed differences between both species is shown in Table 5.34. In the three performed comparisons (*D. madeirensis* – *D. subobscura*, *D. madeirensis* – A_2 arrangement of *D. subobscura*, and *D.*

madeirensis – A_{st} arrangement of *D. subobscura*), the number of fixed differences was higher than the number of shared polymorphisms in most of the studied regions, and in the concatenated data set. The highest difference between these numbers was found in the comparison between *D. madeirensis* and A_{st} arrangement of *D. subobscura*, with 37 fixed differences and only 4 shared polymorphisms.

Table 5.33.
Multilocus tests for Tajima's *D* and Fu and Li's *D*.

	Tajima's D	*Fu and Li's D*
***D. subobscura* A_{st}**	-0.324	-0.526
***D. subobscura* A_2**	-1.082**	-1.296**
***D. subobscura* Total**	-0.325	-1.374**
D. madeirensis	-1.037**	-1.644***

Significance levels of multilocus *D* (one-tailed test*): *** P < 0.001*; *** P < 0.01*).

The number of exclusive polymorphisms in *D. madeirensis* and in *D. subobscura* (total sample) was similar (Table 5.34). Nevertheless, this result may be misleading as the number of studied lines in *D. subobscura* was higher than in *D. madeirensis*. This difference in the number of lines may contributed to an increase in the number of exclusive polymorphisms in *D. subobscura* relative to *D. madeirensis*. Indeed, when *D. madeirensis* (12 lines) was compared with the *D. subobscura* A_2 sample (12 lines), the number of exclusive polymorphisms was higher in *D. madeirensis*, which was consistent with the higher level of nucleotide variation in this species relative to *D. subobscura*.

In the concatenated data set, the level of nucleotide divergence between *D. madeirensis* and *D. subobscura* was similar in the three comparisons performed with the different *D. subobscura* samples (Table 5.35). Across the five studied regions, the

highest level of divergence was found at P236 and the lowest at P150. Accordingly, the highest estimate of N_m was found in the latter region.

Table 5.34.
Genetic differentiation between species.

		Fixed	*Shared*	S_{x1}	S_{x2}
***D. mad*/*D. sub* Tot**	**P236**	15	2	53	49
	P150	3	3	20	36
	Sxl	7	0	59	42
	P125	3	4	65	62
	P275	5	3	28	35
	Concatenated data	33	12	221	220
***D. mad*/*D. sub* A_2**	**P236**	17	1	54	37
	P150	3	3	20	23
	Sxl	8	0	59	31
	P125	3	4	65	54
	P275	8	3	28	22
	Concatenated data	39	11	222	163
***D. mad*/*D. sub* A_{st}**	**P236**	16	2	53	27
	P150	3	0	23	18
	Sxl	7	0	60	22
	P125	6	2	67	30
	P275	5	0	31	18
	Concatenated data	37	4	229	114

D. sub Tot, all *D. subobscura* lines studied (A_2 + A_{st}); *D. mad*, *D. madeirensis* lines; *D. sub* A_2, *D. subobscura* lines with A_2 arrangement; *D. sub* A_{st}, *D. subobscura* lines with A_{st} arrangement; Fixed, fixed differences between samples; Shared, polymorphic sites segregating for the same two variants in two samples; S_{x1}, exclusive polymorphisms sites in population 1; S_{x2}, exclusive polymorphisms sites in population 2.

5.4.1.4. Linkage disequilibrium tests of gene flow

A linkage disequilibrium test was carried out, based in the difference between the linkage disequilibrium found in the shared polymorphisms and the linkage disequilibrium found in which one species showed a shared polymorphism and in the

other showed an exclusive polymorphism. It was only calculated for the P125 region, because was the only region of the five studied with at least four pairs of sites in each of the categories needed to perform the test. In the comparison between *D. subobscura* and *D. madeirensis*, the observed values of *x* were negative for both, and the simulated values were fairly larger than the observed ones. This pattern could indicate that no gene flow happened after divergence between both species, at least for the P125 region (Table 5.36).

Table 5.35.

Population migration rates and sequence divergence estimates between species.

		D_{xy}	D_a	N_m	F_{ST}
***D. mad/D. sub* Tot**	**P236**	0.0254	0.0168	0.17	0.66
	P150	0.0101	0.0046	0.39	0.45
	Sxl	0.0144	0.0078	0.3	0.52
	P125	0.0237	0.0107	0.4	0.45
	P275	0.0187	0.0115	0.18	0.65
	Concatenated data	0.0187	0.0104	0.26	0.55
***D. mad/D. sub* A_2**	**P236**	0.0262	0.0191	0.12	0.73
	P150	0.0099	0.0047	0.37	0.47
	Sxl	0.0142	0.0073	0.28	0.54
	P125	0.0231	0.0105	0.4	0.45
	P275	0.0192	0.0126	0.22	0.59
	Concatenated data	0.0187	0.0111	0.23	0.59
***D. mad/D. sub* A_{st}**	**P236**	0.0243	0.0165	0.16	0.67
	P150	0.0106	0.0052	0.35	0.49
	Sxl	0.0144	0.0071	0.3	0.52
	P125	0.0249	0.0125	0.33	0.5
	P275	0.0176	0.0105	0.21	0.61
	Concatenated data	0.0186	0.0106	0.25	0.57

D. sub Tot, all *D. subobscura* lines studied (A_2 + A_{st}); *D. mad*, *D. madeirensis* lines; . *D. sub* A_2, *D. subobscura* lines with A_2 arrangement; *D. sub* A_{st}, *D. subobscura* lines with A_{st} arrangement; D_{XY}, average number of nucleotide differences per site between species; D_a, number of net nucleotide substitutions per site between species; N_M and F_{ST} are gene flow estimates.

Table 5.36.
Linkage disequilibrium tests of gene flow.

	D. madeirensis			*D. subobscura*		
	Observed	**Simulated**	***P-value***	**Observed**	**Simulated**	***P-value***
P125	-0,034	0,667	0,922	-0,045	0,681	0,863

P-value, estimated probability of observing a simulated value higher than the observed value of *x*.

5.4.1.5. Parameters estimates and isolation model fitting

The multilocus data were fitted to the Wakeley and Hey Isolation Model (Wakeley and Hey 1997, Wang *et al.* 1997). This model assumes divergence in isolation without subsequent gene flow, thus being a simple null model of allopatric speciation. Seeing as the WH model makes quantitative predictions about the patterns of nucleotide diversity across multiple loci, empirical data obtained from recently diverged taxa can, in principle, assess the critical assumption of divergence without gene flow and provide estimates of population-size changes at the time of speciation. The test was performed using the counts of four types of polymorphic base positions: polymorphisms that were exclusive to *D. madeirensis*, polymorphisms that were exclusive to *D. subobscura*, polymorphisms shared by both species, and polymorphisms that appeared as fixed differences between the two species. Also, the population recombination parameter γ for each region was used (Table 5.37). The test was performed for the three possible comparisons: *D. madeirensis* – *D. subobscura*; *D. madeirensis* – *D. subobscura* A_2 arrangement and *D. madeirensis* – *D. subobscura* A_{st} arrangement.

The WH test parameteres estimates are shown in Table 5.38. According to these, *D. madeirensis* presented a higher effective population size than *D. subobscura* samples. In the comparison *D. madeirensis* – *D. subobscura* A_2

arrangement, the ancestral population size should be higher than the *D. subobscura* one, and close to the magnitude of *D. madeirensis* population size.

Table 5.37.

Population recombination parameters estimates.

	P236		P150		SxI		P125		P275	
	γ	γ/θ	γ	γ/θ	γ	γ/θ	γ	γ/θ	γ	γ/θ
***D. subobscura* A_{st}**	0.0018	0.172	0.0177	2.4006	0.0373	7.3177	0	0	0	0
***D. subobscura* A_2**	0.0091	0.9688	0.0189	2.5744	0.0428	6.9741	0.0396	2.9631	0.0026	0.2751
***D. subobscura* Total**	0.0093	0.7793	0.0141	1.4396	0.3355	4.835	0.0343	2.5907	0.0024	0.196
D. madeirensis	0.0318	2.3222	0.0213	0.3304	0.0641	5.7672	0.0718	4.2967	0.0231	1.7915

γ, estimate of the population recombination rate $4Nc$, where c is the recombination rate per generation per base pair (Hey and Wakeley 1997). The ratio of recombination rate per base pair to neutral mutation per base pair was estimated dividing γ by θ.

Table 5.38.

WH parameters estimates and isolation model fitting.

	θ_1	θ_2	θ_A	τ	T	$P\chi^2$	P_{wwh}
***D. mad - D.sub* Tot**	111.03 41.2-455.9	92.54 37.0-230.2	87.18 3.2-299.5	47.61	0.4289	0.983	0.667
***D. mad - D.sub* A_2**	105.4 12.5-578.2	72.63 10.0-235.8	96.54 3.4-329.5	42.71	0.4	0.987	0.652
***D. mad - D.sub* A_{st}**	118.5 60.4-686.0	71.93 736.6-272.3	55.68 2.2-186.8	49.36	0.41	0.99	0.494

θ_1, population mutation parameter for species 1; θ_2, population mutation parameter for species 2; θ_A, population mutation for the ancestral species; τ, scaled time parameter; T, estimated time of species divergence in units of $2N_1$ generations, where N_1 is the effective population size of species 1. Bellow the primary parameter estimates, 95% confidence intervals are shown, determined by 10.000 simulations. The P-values for both the χ^2 test and wwh (Wang *et al.* 1997) statistics are the proportions of simulated values higher of equal to the observed values. *D. sub* A_2, *D. subobscura* lines with A_2 arrangement; *D. sub* A_{st}, *D. subobscura* lines with A_{st} arrangement; *D. sub* Tot, all *D. subobscura* lines studied (A_2 + A_{st}) and *D. mad*, *D. madeirensis* lines.

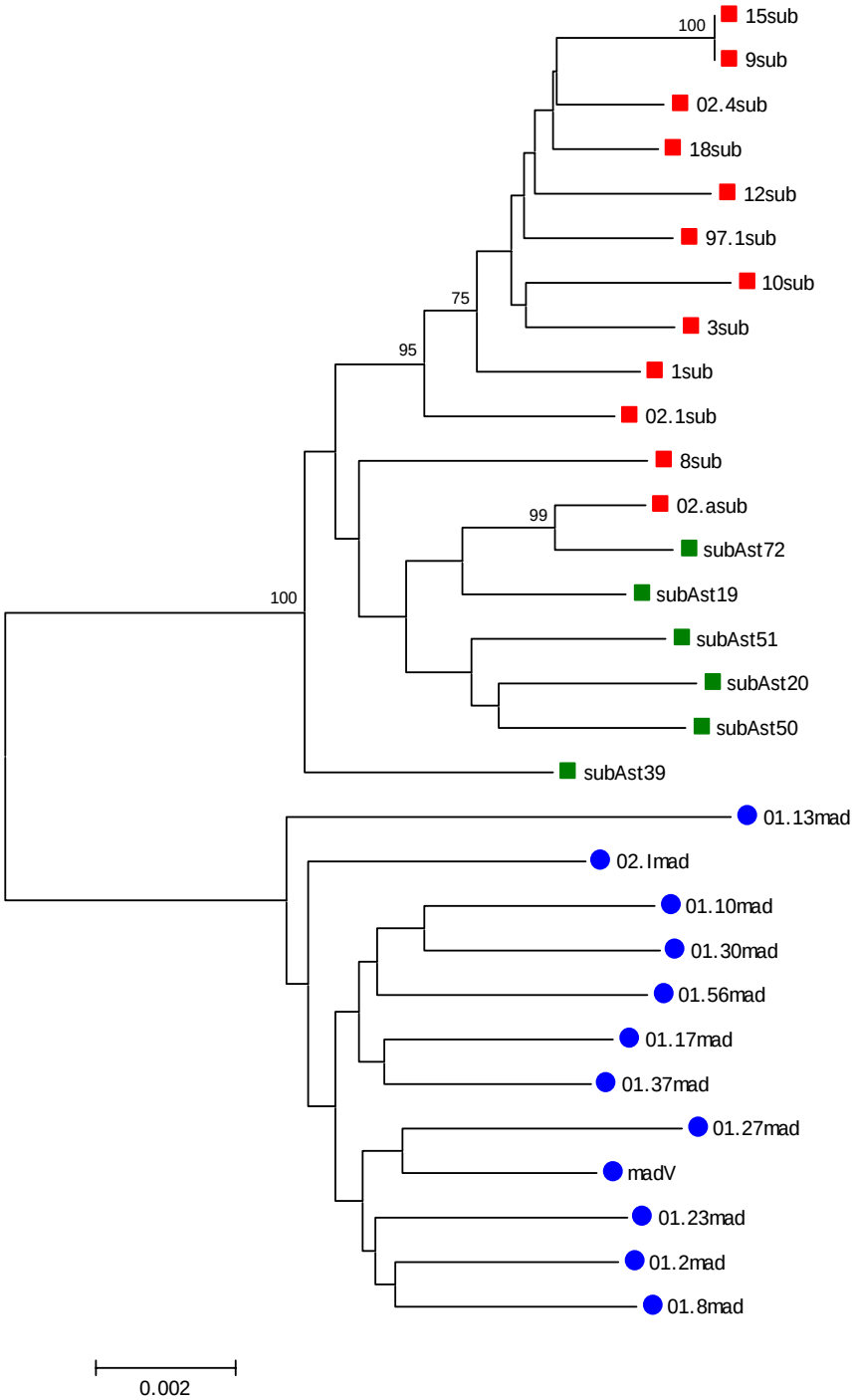
Coalescent simulations implementing the estimated recombination levels did not reject the simple isolation model, as neither the wwh (Wang *et al.* 1997) nor the χ^2 test statistics approach significance in any pairwise species comparison. The homogeneity of nucleotide variation among the studied regions, is consistent with a possible history of divergence without gene flow. Thus, a isolation model of divergence without gene flow adjusts to our multilocus data, and it is a possible explanation to the divergence between *D. madeirensis* and *D. subobscura*.

5.4.1.6. Gene genealogy

The gene genealogy reconstructed from the nucleotide variation in the concatenated data set in the two species is shown in Figure 5.21. There was a clear separation between the lines of the two species, and within *D. subobscura* a partial clustering of the lines according to their chromosomal arrangement.

Figure 5.21.

Gene genealogy reconstructed from the total nucleotide variation in both studied species.



The gene genealogy was obtained by the neighbour-joining method (Saitou and Nei 1987) using, as genetic distance, the number of substitutions per site, according to Kimura 2-parameter method (Kimura 1980). Bootstrap confidence level (from 1000 replications) above 70% are shown below the branches. The horizontal bar at the bottom represents the distance scale of the branches.

6. DISCUSSION

6.1. Nucleotide variation in *D. madeirensis*

Levels of nucleotide polymorphism in the five gene regions studied here were compared to those previously reported for the *rp49* gene region of *D. madeirensis* (Khadem *et al.* 2001). At *rp49*, a silent nucleotide diversity of 0.0096 was detected; this estimate is very similar to the estimate obtained in the present study (0.0092) for the concatenated data set (Table 5.31).

If nucleotide polymorphism is entirely or predominantly neutral, then the level of nucleotide polymorphism is expected to increase linearly with the increase in the effective population size. X-linked genes have a lower effective population size than autosomal genes in *Drosophila*. Therefore, sex-linked genes are expected to exhibit a lower level of neutral variation than autosomal genes. This prediction was not

confirmed in the present study for *D. madeirensis*. When levels of variation in the five X-linked regions were corrected according to neutral expectations (correction factor 4/3), the obtained estimate (0.0123) was higher than the autosomal estimate. To date, the results obtained when levels of variation between the X-chromosome and the autosomes are compared depend on the population and species sampled. In African populations of *D. melanogaster* and *D. simulans*, which are thought to represent ancestral populations for these two species, X-linked diversity appears to be equal to or higher than autosomal diversity (Irvin *et al.* 1998; Begun and Whitley 2000; Andolfatto 2001; Kauer *et al.* 2002; Sheldahl *et al.* 2003). Outside of Africa, however, X-linked diversity may be reduced relative to autosomal diversity (Irvin *et al.* 1998; Begun and Whitley 2000; Andolfatto 2001; Kauer *et al.* 2002; Sheldahl *et al.* 2003; Mousset and Derome 2004). In *D. madeirensis*, more gene regions, mainly from autosomes, need to be studied to further contrast whether corrected levels of nucleotide variation regions are higher in X-linked than in autosomal genes, as present data seem to suggest.

Endemic species inhabiting rather small islands are expected to have a lower effective population size than closely related species with a worldwide distribution. Therefore, under the strict neutral model (Kimura 1983), a lower level of nucleotide variation would be expected in endemic insular species than in mainland species. In fact, comparisons between island endemic species and closely related mainland lineages offer an ideal test of the effect of population size on the rate and pattern of evolution (Ohta 1972b; Llopart and Aguadé 1999; Johnson and Seger 2001; Perez *et al.* 2003; Morton *et al.* 2004). Endemic species are likely to have undergone both a severe population bottleneck during the initial colonization of the island and subsequently, a long-term reduced population size due to the range restriction. The expected reduction in the level of nucleotide variation in endemic species was not detected in the *rp49* region, when *D. madeirensis* and the closely related mainland species *D. subobscura* were compared (Khadem *et al.* 2001). The present study also

failed to detect any difference in the level of nucleotide variation between *D. subobscura* and *D. madeirensis* which might be related to the expected difference in the effective size of both species. However, the population of *D. subobscura* analyzed in the present study was sampled in Madeira Island. Thus, it can be argued that this population is not representative of the mainland populations of *D. subobscura*. Although this possibility cannot be completely discarded, it seems very unlikely. Indeed, natural populations of *D. subobscura* from Madeira and Europe are not genetically differentiated according to data in allozymes, mtDNA (Pinto *et al.* 1997) and at *rp49* (Khadem *et al.* 2001). In addition, levels of variation in the five gene regions in populations of *D. subobscura* from Madeira Island are very similar to those previously reported for other gene regions in mainland populations of the species (Cirera and Aguadé 1998; Rozas *et al.* 1999; Navarro-Sabaté *et al.* 1999a, 1999b; Munté *et al.* 2000, 2005). Therefore, *a priori*, there is no reason to suppose that the Madeira populations of *D. subobscura* present a depletion of variation relative to mainland populations.

The neutral theory of molecular evolution predicts that the expected level of polymorphism at mutation-drift equilibrium for a neutral locus is proportional to the effective population size. Other factors can, of course, affect the level of nucleotide polymorphism, including population structure (Cherry and Wakeley 2003), population bottlenecks (O'brien 1994), natural selection (Maynard-Smith and Haigh 1974; Charlesworth *et al.* 1993), life cycle (Caballero and Hill 1992), and mating systems (Amos and Harwood 1998). Thus, how could the high level of nucleotide variation in *D. madeirensis* be explained? Taking into account the limited size of Madeira Island (730 km^2) and the destruction of the natural habitat of this species (the Laurisilva forest) which occurred during the last 400 years (Doria 1945; Frutuoso 1979) it is not easy to explain the rather high level of variation in *D. madeirensis*. Indeed, a lack of genetic diversity is typically considered as evidence for a small or declining, potentially endangered population (Amos and Balmford 2001; Spielman *et al.* 2004).

Moreover, high levels of diversity and low values of linkage disequilibrium may be characteristic features of ancestral populations, whereas low levels of diversity and high amounts of linkage disequilibrium are expected in recently established populations (Jorde *et al.* 2001). Therefore, the level and pattern of variation in *D. madeirensis* is not consistent with those expected in recently established populations.

Present data are consistent with a speciation model with isolation and without migration. Therefore, *D. madeirensis* likely originated in allopatry after Madeira Island was colonized by ancestral *D. subobscura* populations. The isolation model was used to infer estimates of the effective size of the ancestral and *D. madeirensis* populations (Table 5.38). These estimates are rather high and could indicate that the speciation process which originated *D. madeirensis* did not imply a strong bottleneck. The negative and significant Tajima's and Fu and Li's statistics would, however, favour the bottleneck argument, as an excess of low frequency polymorphisms is expected after a strong bottleneck. Population expansions can also cause the same pattern of variation; however, it is not likely that current populations of *D. madeirensis* are in expansion.

Other *Drosophila* species pairs including an endemic and a mainland species (*D. simulans – D. mauritiana*, *D. simulans – D. sechellia* and *D. subobscura – D. guanche*) have also been studied in order to compare the levels of nucleotide variation (Hey and Kliman 1993; Kliman and Hey 1993; Perez *et al.* 2003). The level of allozyme variation is lower in *D. mauritiana* and *D. sechellia* than in the cosmopolitan closely related species (Morton *et al.* 2004). However, only *D. sechellia* exhibits the expected reduction in nucleotide variation relative to *D. simulans*, which can also be due to a strong bottleneck associated with the origin of *D. sechellia*. Levels of nucleotide variation in *D. mauritiana* and *D. guanche* are not significantly reduced relative to *D. simulans* and *D. subobscura*, respectively. Therefore, present

data in *D. madeirensis* can not be considered unusual and may indicate that the expected relationship between effective size and neutral variation is somehow masked by other factors.

The effective population size may not be the unique determinant of genetic variation. In fact, the relationship between population size and genetic variation has been a subject of discussion in past decades (Wright 1931, 1932; Fisher 1958; Kimura 1983; Ohta 1992; Gillespie 1999, 2000a, 2000b, 2001). Gillespie (2000a) proposed the pseudohitchhiking model to explain the lack of correlation between effective size and nucleotide polymorphism. This model is based on linked selection and focuses on the effect that adaptive substitutions at one locus can have in the level of neutral variation at a linked neutral locus. In fact, this level of variation is quite independent on population size after recurrent hitchhiking events. It was suggested that the stochastic effects of linked selection (genetic draft) can be more important than genetic drift, when the population size is greater than about 10^4 and there is no recombination between the selected and the neutral locus (Gillespie 2000a). The pseudohitchhiking model has some important implications: levels of polymorphism would be insensitive to population size; the levels of variation should be relatively constant between species; the frequency spectrum of alleles should be skewed from the neutral spectrum in a direction which leads to negative values of Tajima's *D*, and finally, genetic variation should be proportional to levels of recombination. Recently, Bazin *et al.* (2006) reported a summary study which seems to be consistent with the pseudohitchhiking model. The authors compared allozymic, nuclear DNA and mtDNA diversity in a wide range of animal taxa. In contrast to nuclear diversity, mtDNA nucleotide variation was found to be quite uniform among taxa despite the expected strong differences in their effective sizes. This result supports the putative role of linked selection on neutral variation, as the lack of recombination in the mtDNA would cause that the effects of linked selection were more important in the mitochondrial than in the nuclear genome.

Present data for *D. madeirensis* could be explained by the pseudohitchhiking model, assuming that the effective size of this species is high enough for the effects of linked selection to dominate those of genetic drift. In this case, the expected difference in effective population size between *D. madeirensis* and *D. subobscura* would not be reflected in the level of nucleotide variation. In addition, linked selection is expected to have a stronger effect in *D. subobscura* than in *D. madeirensis*. In contrast to the latter species, which is monomorphic at the chromosomal level, *D. subobscura* has a very rich chromosomal polymorphism. The main genetic effect of paracentric inversions is the suppression of recombination in heterokaryotypes within the inverted segment. The five studied regions are located within the A_2 inversion, and the maintenance in natural populations of the A_{st} and A_2 arrangements could cause a reduction in the recombination in the inverted segment. Indeed, strong linkage disequilibrium between these arrangements and variants at nucleotide polymorphic sites was detected all over the A_2 inversion, indicating that nucleotide variation is highly structured in the inverted segment (see section 5.2.2). In addition, the location of the studied markers is affected by other chromosomal inversions which overlap with inversion 2 and form the arrangements A_{2+4} and A_{2+3}. Therefore, the presence of these arrangements in *D. subobscura* might cause a global reduction of recombination in the studied gene regions and thus, enhance the effects of linked selection in this species. The level of nucleotide variation in the concatenated data set was 0.0081 and 0.0092 in *D. subobscura* and *D. madeirensis* respectively (Table 5.28 and Table 5.31). This slightly lower level of nucleotide variation in *D. subobscura*, despite its much higher effective size, may be consistent with the pseudohitchhiking model, due to a stronger effect of linked selection in this species than in *D. madeirensis*.

6.2. Nucleotide variation in *D. subobscura*

Levels of nucleotide variation in the studied regions were lower than those in the only region studied in *D. subobscura* populations from Madeira. In this region, *rp49*, a silent nucleotide diversity of 0.0102 was detected (Khadem *et al.* 2001), while in the concatenated data set of the present study this value was 0.0081 (Table 5.28). If the latter value was corrected by 4/3 (due to the smaller effective size of X chromosome), then both nucleotide variation levels were quite similar (the corrected silent nucleotide variation was 0.0108 for the concatenated data set).

The previous study which compares the level of nucleotide variation in Madeira and mainland populations of *D. subobscura* did not find a genetic difference between populations (Khadem *et al.* 2001). This pattern of nucleotide variation between insular and mainland populations, together with available data from the mtDNA (Pinto *et al.* 1997), and a previous study in the *rp49* region (Khadem *et al.* 1998), does not support the hypothesis that populations of *D. subobscura* from Madeira have been isolated from mainland populations for a long period of time. Therefore, current populations of *D. subobscura* in Madeira would be the result of a quite recent and massive colonization of the island by continental *D. subobscura* populations.

The levels of silent nucleotide variation in the X chromosome were slightly lower in the present studied regions when compared to those reported for *Pgd*, *RpII215* and the *yellow* in the A_{st} arrangement (Martín-Campos 1998; Llopart 1999; Munté *et al.* 2001), with the exception of the P125 region. This latter region presented a rather high polymorphism in both chromosomal arrangements, only exceeded by that found in the *RpII215* region. The level of silent nucleotide diversity from the concatenated data of the regions reported here (0.0092), is the same that the average level previously reported for other gene regions of the X chromosome (0.0092).

When available data on the level of nucleotide variation was compared between autosomal and X-linked regions in *D. subobscura*, a reduced polymorphism was detected in the X-chromosome (Table 6.1). However, if the levels of nucleotide diversity in X-linked regions were corrected by 4/3 (due to the smaller effective size of X-linked regions), nucleotide variation was quite similar in the X-linked and in the autosomal regions.

Table 6.1.
Nucleotide variation in *D. subobscura*.

	π_{total}	π_{silent}
Total Regions	0.0084	0.0109
Autosomal Regions	0.0083	0.0115
X-linked Regions	0.0085 (0.0064)	0.010 (0.0075)

Average nucleotide variation (estimated by π for autosomic regions and by $^{4/3}\pi$ for X-linked regions) was calculated using data from: *Xdh* (Comerón 1997); *Acp70A* (Cirera and Aguadé 1998); *rp49* (Rozas *et al.* 1999); *rp49* O_{3+4} (Khadem *et al.* 2001); *Acph-1* (Navarro-Sabaté *et al.* 1999a); *S25*, *P22*, *P154*, *P2*, *S1*, *P21* (Munté *et al.* 2005) *Pgd* (Martín-Campos 1998); *RpII215* (Llopart 1999); *yellow* (Munté *et al.* 2001); *P236*, *P150*, *Sxl*, *P125*, *P275* (present study). The uncorrected estimates of nucleotide variation for X-linked regions are in parentheses.

6.3. Comparison of nucleotide variation levels among species and taxa

The level of silent variation in the concatenated sample with the five X-linked gene regions reported here are 0.0081 and 0.0092 in *D. subobscura* and *D. madeirensis*, respectively (Table 5.28 and Table 5.31). This level of variation can be compared with the level present in other *Drosophila* species such as *D. melanogaster* and *D. simulans*, which are the *Drosophila* species best characterized from this point of view. The average level of nucleotide polymorphism for X-linked gene regions is

0.0190 (0.0106 for X-linked and autosomal non-coding regions, Andolfatto 2005) in *D. melanogaster* and 0.0234 in *D. simulans* (Andolfatto 2001). Therefore, there is approximately a two-fold difference in the level of nucleotide polymorphism between the species of the *melanogaster* group and the species of the *obscura* group analyzed here. This difference may be due to a higher effective size of *D. melanogaster* and *D. simulans* relative to *D. subobscura*. However, the rich chromosomal polymorphism of *D. subobscura* may also contribute to its lower level of nucleotide variation (see section 6.1).

Levels of variation in the *Drosophila* species are considerably higher than those reported in humans, where θ estimates at silent sites are about 0.001 (Clark *et al.* 1998; Harris and Hey 1999; Nachman and Crowell 2000; Przeworski *et al.* 2000; Zwick *et al.* 2000; Frisse *et al.* 2001; Payseur and Nachman 2002). The same argument holds when levels of polymorphism in *Drosophila* and *Caenorhabditis elegans* are compared. This latter species is characterized by rather low levels of silent nucleotide diversity with estimates of about 0.002 (Cutter 2006) and thus, much lower than those reported for *Drosophila*. In a general way, plants like *Arabidopsis thaliana*, *A. lyrata* ssp. *petraea* or *Zea mays* ssp. *parviglumis* (Wright and Gaut 2004) present higher levels of nucleotide variation than *D. madeirensis* or *D. subobscura*. In contrast, the *Drosophila* species present higher levels of polymorphism than the wild rice *Oryza rufipogon* (Miyashita *et al.* 2005).

6.4. Inversion effects

The recombination rates vary dramatically across the genome. Several studies in *Drosophila* (Aguadé *et al.* 1989; Begun and Aquadro 1992; Aguadé and Langley 1994; Aquadro *et al*. 1994; Stephan 1994; Pritchard and Schaeffer 1997) have shown that levels of DNA sequence variation are dramatically reduced in regions of

extremely low recombination. It has been further demonstrated that this pattern extends to many organisms (Nachman 1997; Baudry *et al.* 2001; Nachman 2001; Sundstrom *et al.* 2004). It is also well established that recombination rates are strongly influenced by chromosomal inversions. Indeed, the main genetic effect of inversions is the supression of recombination within the inverted segment in heterokaryotypes (Sturtevant 1917; Roberts 1976). This genetic effect also has important evolutionary consequences. Dozhansky (1947, 1954, 1970) argued that inversions represent sets of coadapted alleles, favoured by natural selection under particular conditions and thus, that inversion polymorphism is adaptive. Further development of these ideas led to propose that for a single population in a constant environment, fitness interactions between loci (epistasis) will generally favour the evolution of decreased recombination (Feldman *et al.* 1997). When populations are connected by migration, selection can favour loosely linked modifiers which decrease recombination between loci involved in local adaptation. This kind of selection can act even in the absence of epistasis, suggesting that inversions could be established by a similar mechanism (Charlesworth and Charlesworth 1979; Pylkov *et al.* 1998; Lenormand and Otto 2000). Recently, Kirkpatrick and Barton (2005) proposed that inversions can spread by suppressing recombination between loci implied in local adaptation. According to this mechanism, neither drift nor epistasis are needed. Therefore, this local mechanism may apply to a broad range of genetic and demographic situations.

Inversions reduce recombination within the inverted region because crossovers are partly inhibited by the inversion loop (Roberts 1976; Coyne *et al.* 1991, 1993). In addition, when crossovers do take place, they most often give rise to unbalanced meiotic products (Sturtevant and Beadle 1936; Roberts 1976). Despite the reduction of recombination, genetic exchange inside the inversion loop is not completely suppressed, since viable gametes may arise by double crossing over inside the inversion loop (Sturtevant and Beadle 1936; Spurway and Philip 1952; Novitski and

Braver 1954; Levine 1956), and by gene conversion (Chovnick 1973; Rozas and Aguadé 1993, 1994; Popadic and Anderson 1995). The gene conversion rate would be uniformly distributed along the inversion loop, except at the breakpoints themselves, whereas the contribution of double crossovers to genetic exchange would be considerably higher in the central part of the inversion loop (Navarro *et al.* 1997).

Munté *et al.* (2005) analyzed nucleotide variation in eight gene regions distributed along the O_3 inversion. The results indicated a strong genetic differentiation between inverted and non-inverted chromosomes in all regions along the inversion, and showed no evidence for the higher genetic exchange between arrangements expected in the central part of the inversion loop. The detected rather homogeneous distribution of genetic exchange along the inversion indicated either that no double crossovers were produced inside the inversion loop or, alternatively, if they occur, that selection has acted against the recombinant chromosomes. The authors also suggested that the distance of the studied regions to the nearest inversion breakpoint was high enough for gene conversion to have contributed to the recovery of variation in all regions. Indeed, gene conversion tracts were detected in most of the studied regions and the polymorphism-to-divergence ratio was not reduced in the gene regions located closer to the inversion breakpoints. The most plausible explanation for the strong genetic differentiation detected in all the studied regions would be that selection has acted against the recombinant chromosomes produced by double crossover events between the chromosomal arrangements. Alternatively, it could be argued that the O_3 inversion was not large enough (3.5 Mb) for the occurrence of double crossover events.

The occurrence of double crossovers over evolutionary time inside an inversion loop may depend on the length and age of the inversion. Navarro *et al.* (1997)

suggested that in *Drosophila* double crossover is unlikely only in short inversions (<20 cM). The O_3 inversion has been estimated to be 27.4 cM in length. However, genetic exchange between inversions has not been effective in eroding genetic differentiation even in the central part of the inversion. As stated before, this result would indicate that the length of the O_3 inversion is still too small for double crossover to occur within the inversion loop. In addition, the O_3 inversion is always found with the overlapping O_4 inversion forming the O_{3+4} arrangement. The complex double inversion loop present in O_{st}/O_{3+4} heterozygotes would also help prevent double crossover events. Alternatively, it could be argued that double crossover does occur, but that selection acts against the recombinant chromosomes.

The selective explanation would be further reinforced if the strong genetic differentiation detected along the O_3 inversion is also present along other inversions larger than O_3 which form single inversion loops in heterokaryotypes. The A_2 inversion has an estimated length of 41.3 cM and forms a single inversion loop in A_{st}/A_2 heterokaryotypes. Therefore, it seems *a priori* reasonably to assume that double crossovers inside the inversion loop are likely and thus, double crossovers can contribute to the genetic exchange between arrangements. Therefore, the gene regions analyzed in this study are distributed along the A_2 inversion and nucleotide variation has been analyzed in A_{st} and A_2 chromosomes. One of the main aims of this study is to address the following question: Is the A_2 inversion large enough to allow the occurrence of double crossovers which reduce the genetic differentiation between arrangements in the central part of the inversion loop? In addition, how the presence of the inversion affects the level and pattern of nucleotide variation was evaluated. No clear relationship between the level of silent variation (corrected by silent divergence) and the distance to the nearest inversion breakpoint was detected in any gene arrangement: A_{st} and A_2 (Figure 5.16). Only P236 showed a slight reduction in variation. This result can be explained taking into account that this region is very close to an inversion breakpoint. At inversion breakpoints genetic

exchange between arrangements by gene conversion is highly prevented due to mechanical problems in the synapsis of the homologous chromosomes. Therefore, after the origin and establishment of an inversion, the only factor which contributes to the recovery of variation at the breakpoints is mutation. In the other gene regions, nucleotide variation was slightly higher, as both mutation and gene conversion contributed to the recovery of variation. The detection of gene conversion tracts in some of these regions support this argument. However, as revealed by the HKA test, the level of nucleotide variation does not differ significantly among the studied regions, even when P236 is considered.

A strong genetic differentiation was detected in all the studied regions regardless of their distance to the nearest breakpoint (Figure 5.17). Once again, P236 exhibits the higher differentiation likely due to its close proximity to a breakpoint. Linkage disequilibrium also showed the same pattern (Figure 5.18). Therefore, there is no evidence of a higher genetic exchange between arrangements in the central part of the inversion. This result indicates that double crossover does not contribute significantly to homogenize nucleotide variation between arrangements. As the A_2 inversion seems to be large enough for double crossovers to occur, it has to be inferred that selection acts against the recombinant chromosomes. Consequently, present results would support the adaptive character of chromosomal polymorphism and its role in maintaining blocks of coadapted genes together. In addition, the strong genetic differentiation between arrangements detected along the A_2 (present study) and the O_3 inversions (Munté *et al.* 2005) would indicate that the genome of this species is highly structured, if extensive to other arrangements that segregate as polymorphic in *D. subobscura*.

6.5. Inversions origin, ancestry and phylogeny

Levels of nucleotide variation in gene regions associated with inversions may provide valuable information about the origin and establishment of chromosomal inversions. Assuming that naturally occurring inversions have a unique origin, gene regions associated with inversions should be depleted of nucleotide variation soon after the origin of the inversion. If selection contributes to the increase the frequency of new inversion until it reaches its equilibrium, the lack of variation may prevail even after the inversion achieves a quite high frequency. Thereafter, variation in the new established inversion may be introduced by mutation and by genetic exchange with the preexisting arrangement. Genetic exchange by gene conversion is expected to be uniformly distributed along the inversion except very close to the inversion breakpoints where it would be negligible. In contrast, genetic exchange by double crossover would be higher in the central part of the inversion.

The A_{st} and A_2 arrangements of *D. subobscura* differ by a single inversion located in the central part of the chromosome (inversion 2 with breakpoints in 8C/D and 12C/D). The gene order in *D. madeirensis* in the chromosome sections affected by inversion 2 corresponds to the gene order in A_{st} (Papaceit and Prevosti 1991), and thus indicates that A_{st} is most likely the ancestral arrangement (at least for sections 8 to 12) from which A_2 derived. In addition, the gene order in *D. guanche* in the central part of the chromosome is related to A_{st} by a single step (inversion 10C/D-13A/B, Moltó *et al.* 1987). The higher level of variation in A_{st} relative to A_2 detected in this study would also support this argument. Therefore, the standard arrangement, at least for the chromosomal sections 8 to 12, had to predate the *D. madeirensis* split. A_2 would be a younger arrangement which likely originated in North Africa, where it reaches the highest frequencies. The age of the A_2 arrangement can be inferred from the level nucleotide variation detected in the present study.

Approaches to estimate the age of chromosomal inversions have focused mainly on the level of variation accumulated within the inversion since its origin assuming that it has not reached mutation-drift equilibrium (Rozas *et al.* 1999), and alternatively, on the level of net genetic differentiation between inverted and non inverted chromosomes (Hasson and Eanes 1996, Kovacevic and Schaeffer 2000). The first approach takes into account that according to the expansion model (Rogers and Harpending 1992; Rogers 1995), the level of silent nucleotide diversity (π_{sil}) is equal to $2\mu t$, where μ is the neutral mutation rate and t the time since the expansion. This approach assumes that all variation within the inversion has originated by mutation. For this reason, two estimates of π_{sil} have been considered (Table 6.2). One estimate corresponds to silent nucleotide variation in the concatenated data set. However, this may be an overestimate of the silent variation originated by mutation, since gene conversion tracts were identified in some regions. The alternative estimate is based on silent nucleotide variation at *Sex-lethal* and P150 which are the only studied regions in which no gene conversion tracts were detected. Two estimates of the neutral mutation rate inferred from the rate of silent substitutions were used: 15.3 x 10^{-9} and 7.03 x 10^{-9} silent substitutions per site and per year. These rates are based on the silent divergence between *D. madeirensis* and *D. subobscura* estimated in the present study for the concatenated sample (0.0197) and in a divergence time between these species of 0.63 and 1.4 Myr, respectively. These divergence times correspond to those inferred from the study of Ramos-Onsins *et al.* (1998) assuming that the *obscura* group radiation occurred 30 Mya (Trockmorton 1975) or 17.7 Mya (Tamura *et al.* 2004). The most conservative estimate for the age of the A_2 arrangement is 160.000 years.

The age of the A_2 inversion can also be inferred from net divergence between arrangements which was estimated to be 0.0022 (Table 5.30). Net divergence between *D. subobscura* and *D. madeirensis* is 0.0104 (Table 5.35). Therefore, the estimated age of the A_2 inversion is about 133.000 years assuming a divergence time

between *D. madeirensis* and *D. subobscura* of 0.63 Myr. This estimate although slightly lower, is in agreement with the previous estimate.

Table 6.2.
Age of *D. subobscura* A_2 chromosomal arrangement.

	***D. subobscura* A_2**	
	π_{sil}^{a} **(0.0067)**	π_{sil}^{b} **(0.0049)**
T_a	218	160
T_b	476	348

π_{sil}^{a}, silent nucleotide diversity estimated from the total concatenated data set; π_{sil}^{b}, silent nucleotide diversity estimated from the P150 and *Sxl* regions; T_a, time in thousands years for $\mu_a = 15.3 \times 10^{-9}$; and T_b, time in thousands years for $\mu_b = 7.03 \times 10^{-9}$.

6.6. Speciation process

The multilocus data reported in this study can also shed some light on the speciation process which originated *D. madeirensis*. Indeed, the multilocus approach allows us to contrast whether the level and pattern of variation in different loci is consistent with a common historical model for all of them. Demographic forces (e.g., population splitting, population size changes and migration) are expected to affect the whole genome more homogenously than selective forces which are expected to be more specific to particular loci.

Incipient species are expected to share genetic variation present in the ancestral species, unless at least one of them experiences a strong bottleneck. The shared polymorphisms may persist for a long period of time particularly at those loci which are not responsible for adaptive changes or not directly implied in the speciation process (Kliman *et al.* 2000). Recurrent mutation is another explanation for the

presence of shared polymorphisms between species. However, according to the data obtained in the present study, the number of shared polymorphisms is too large to be explained by recurrent mutations.

The data reported here are consistent with a model of speciation without migration (Table 5.38). Therefore, it seems reasonable to assume that Madeira Island was colonized by ancestral populations of *D. subobscura* which gradually diverged in allopatry originating *D. madeirensis*. Current populations of *D. subobscura* in Madeira would thus be the result of a posterior colonization process, when some kind of reproductive isolation was already established between both species. This scenario was already proposed according to variation at *rp49* (Khadem *et al.* 2001), and is further supported by present data.

D. madeirensis and *D. subobscura* split some 0.6 or 1.0 Mya according to divergence at *rp49* (Ramos-Onsins *et al.* 1998). These species are rather similar morphologically (Monclús 1984) and their reproductive isolation is not complete, as fertile and viable hybrids are obtained in the laboratory in some interspecific crosses (Khadem and Krimbas 1991a, 1991b, Papaceit *et al.* 1991). Therefore, it can be argued that after the second colonization event genetic introgression was likely. However, genetic introgression is not expected to affect the genome uniformly (Ting *et al.* 2000). Genetic introgression is expected to be highly prevented in gene regions directly implied in the reproductive isolation. In contrast, introgression might be likely, and more important than previously thought, in regions not involved in the speciation process. The test developed by Wang *et al.* (1997), and applied to present data contrasts whether different loci present very different histories reflected by no gene flow or large amounts of recent gene flow. Present results indicate that natural selection has not acted differentially among the studied loci, as in contrast to the results previously reported, for instance in *D. pseudoobscura* and close relatives

(Wang *et al.* 1997). Therefore, no significant differential introgression between the studied loci was detected, which indicates that data are consistent with an isolation model without migration.

According to the isolation model without migration, shared polymorphisms would be ancestral polymorphisms. However, the number of shared polymorphisms is higher between *D. madeirensis* and the A_2 lines of *D. subobscura* than between *D. madeirensis* and the A_{st} lines. This is an unexpected result, as the A_{st} is the ancestral arrangement, and thus it would be expected that most ancestral polymorphisms would be shared between *D. madeirensis* and this arrangement. The higher number of shared polymorphisms in A_2 than in A_{st} might be due to the different number of lines studied for each arrangement. However, the number of fixed differences is similar in both arrangements, and 6.4 % of the polymorphic sites in A_2 are shared with *D. madeirensis*, but only 3.5 % of the Ast polymorphic sites present the same variants in *D. madeirensis*. Thus, the corrected number of shared polymorphism is still higher in A_2 than in A_{st}.

Shared polymorphisms can also be the result of introgression. Therefore, it is possible that introgression occurred after the second colonization of Madeira Island by *D. subobscura*. The high frequency of the A_2 arrangement in Madeira would explain why the number of shared polymorphisms is higher in this arrangement than in A_{st}. Nevertheless, putative hybrids between both species may differ by inversion 2, which would strongly prevent the genetic exchange between the species within the inverted segment. Indeed, the results obtained in *D. subobscura* indicate that genetic exchange between arrangements is highly restricted, as reflected by the strong genetic differentiation between them along the A_2 inversion. However, the detected gene conversion tracts in the studied regions would indicate that some genetic exchange between arrangements does occur.

7. CONCLUSIONS

1. *D. madeirensis* harbors a quite high level of nucleotide polymorphism. In fact, nucleotide variation in this species is similar to that present in other *Drosophila* species. Therefore, *D. madeirensis* can not be considered an endangered species from a genetic point of view. Although the progressive destruction of its natural habitat threat the survival of the species, genetic drift and inbreeding have not had until now a strong effect in eroding the genetic variability of the species.

2. The expected strong differences in effective size between *D. madeirensis* and *D. subobscura* are not reflected in their level of nucleotide variation. Therefore, present data are not consistent with the strictly neutral model, and support other population genetic models, as the pseudohithiking model, that

predicts that levels of nucleotide polymorphism can be insensitive to the effective size.

3. Levels of polymorphism in the studied X-linked regions in *D. madeirensis* are similar to those previously reported in the *rp49* autosomic gene region. Therefore, an excess of nucleotide polymorphism is detected in *D. madeirensis* when nucleotide diversity in X-linked genes is corrected according to the smaller effective size expected for X-linked relative to autosomic loci. More data on nucleotide variation in autosomic gene regions of *D. madeirensis* are needed to corroborate this result.

4. Levels of nucleotide polymorphisms in natural population of *D. subobscura* from Madeira are similar to those previously reported for other gene regions in continental populations. In addition, the levels of nucleotide variation in *D. subobscura* are quite similar between autosomic and X-linked genes when nucleotide diversity at X-linked loci is corrected by effective size.

5. The pattern of variation in *D. madeirensis* shows a significant excess of low frequency variants at polymorphic sites in all gene regions studied as reflected by the tests proposed by Tajima and Fu and Li. This result would be expected in a species under an expansion process. However, the ecological conditions of *D. madeirensis* make this explanation unlikely. The statistics proposed by Tajima and Fu and Li are also negative in *D. subobscura*, however, in this species they are significant only within the A_2 arrangement. This result might indicate that the A_2 arrangement is not at mutation-drift equilibrium and that this arrangement is still in the transient phase of recovering nucleotide variation after in its origin.

6. A significant genetic differentiation between the A_{st} and the A_2 arrangements is detected in each of the studied gene regions, regardless to its distance to the nearest inversion breakpoints. The significant genetic differentiation is also reflected by the strong linkage disequilibrium between chromosomal arrangement and the variants at polymorphic sites. Genetic differentiation and linkage disequilibrium are larger at P236 than in the other regions. This result might indicate that in the P236 gene region that is located very close to an inversion breakpoint, genetic exchange by gene conversion between arrangements is somewhat prevented by mechanical problems in the synapsis of the homologous chromosomes.

7. No relationship is detected between the distance to the nearest breakpoint and the genetic differentiation (or linkage disequilibrium) at P275, P150, *Sxl* and P125. Therefore, double crossovers do not contribute significantly to the genetic exchange between arrangements in the central part of the inversion loop formed in A_{st}/A_2 arrangements. As it seems that the A_2 inversion is large enough to support double crossover, this result would indicate that natural selection acts against the recombinant chromosomes produced by double crossover events.

8. The A_2 inversion results to be 160.000 years old, when the age is estimated in based on the level of nucleotide diversity and according to the expansion model, and 133.000 when this age is estimated according to the net genetic differentiation between arrangements.

9. The simple speciation model without migration is consistent with the obtained data. Therefore, the origin of *D. madeirensis* can be explained by a gradual process of speciation in allopatry after the Madeira Island was colonized by

ancestral populations of *D. subobscura*. Current population of *D. subobscura* in Madeira would thus be the result of a second colonization event when *D. subobscura* and *D. madeirensis* were already reproductively isolated.

10. The number of shared polymorphisms between *D. madeirensis* and *D. subobscura* can not be explained by recurrent mutation. However, the number of shared polymorphisms is higher in the A_2 than in the A_{st} arrangement, which seems to preclude that they are ancestral polymorphisms. Introgression between *D. madeirensis* and *D. subobscura* in Madeira might explain the relative high number of shared polymorphisms in A_2. Nevertheless, introgression between *D. madeirensis* and A_2 *D. subobscura* would be somewhat hindered by the presence of the inversion loop formed in heterokaryotypes.

8. REFERENCES

AGUADÉ, M., N.T. MIYASHITA and C.H. LANGLEY, 1989. Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122**: 607-615.

AGUADÉ, M. and C.H. LANGLEY, 1994. Polymorphism and divergence in regions of low recombination in *Drosophila*, pp. 67-76 in *Non-neutral Evolution. Theories and Molecular Data*, edited by B. GOLDING, Chapman and Hall, New York.

AMOS, W. and J. HARWOOD, 1998. Factors affecting levels of genetic diversity in natural populations. Philos. Trans. R. Soc. London B Biol. Sci. **353**: 177-186.

AMOS, W. and A. BALMFORD, 2001. When does conservation genetics matter? Heredity **87**: 527-265.

ANDERSON, W.W., 1989. Selection in natural and experimental populations of *Drosophila pseudoobscura*. Genome **31**: 239-245.

ANDERSSON, M., 1994. *Sexual Selection*, Princeton University Press, Princeton, New Jersey.

ANDERSON, A.R., J.E. COLLINGE, A.A. HOFFMANN, M. MCKECHNIE and S.W. MCKECHNIE, 2003. Thermal tolerance trade-offs associated with the right arm of chromosome 3 and marked by the *hsr-omega* gene in *Drosophila melanogaster*. Heredity **90**: 195-202.

ANDJELKOVIC, M., V. SAVKOVIC and P. KALAJDZIC, 2003. Inversion polymorphism in *Drosophila subobscura* from two different habitats from the mountain of Goc. Hereditas **138**: 241-243.

ANDOLFATTO, P., 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *D. simulans*. Mol. Biol. Evol. **18**: 279-290.

ANDOLFATTO, P. and M. PRZEWORSKI, 2001. Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. Genetics **158**: 657-665.

ANDOLFATTO, P., F. DEPAULIS and A. NAVARRO, 2001. Inversion polymorphisms and nucleotide variability in *Drosophila*. Genet. Res. **77**: 1-8.

ANDOLFATTO P., 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature **437**:1149-50.

AQUADRO, C.F., 1992. Why is the genome variable? Insights from *Drosophila*. Trends Genet. **8**: 355-362.

AQUADRO, C.F. and D. J. BEGUN, 1993. Evidence for an implication of genetic hitchhiking in the *Drosophila* genome, pp. 159-178 *in Mechanisms of Molecular Evolution*, edited by N. TAKAHATA and A.G. CLARK, Sinauer Associates, Sunderland, MA.

AQUADRO, C.F., D.J. BEGUN and E.C. KINDAHL, 1994. Selection, recombination and DNA polymorphism in *Drosophila*, pp. 46-56 in *Non-neutral Evolution. Theories and Molecular Data*, edited by B. GOLDING, Chapman and Hall, New York.

AQUADRO, C.F., 1997. Insights into the evolutionary process from patterns of DNA sequence variability. Curr. Opin. Genet. Dev. **7**: 835-840.

ASHBURNER, M., 1989. *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, New York.

AVERY, P.J., 1984. The population genetics of haplo-diploids and X-linked genes. Gent. Res. **44**: 321-341.

AYALA, F.J., J.R. POWELL and M.L. TRACEY, 1972. Enzyme variability in the *Drosophila willistoni* group. V. Genic variability in natural populations of *Drosophila equinoxialis*. Genet. Res. **20**: 19-42.

BACHMANN, L. E. MULLER, M.L. CARIOU and D. SPERLICH, 1992. Cloning and characterization of *KM190*, a specific satellite DNA family of *Drosophila kitumensis* and *D. microlabis*. Gene **120**: 267-269.

BALANYA, J., L. SERRA, G.W. GILCHIST, R.B. HUEY, M. PASCUAL, F. MESTRES and E. SOLE, 2003. Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: an evolutionary time series. Evolution **57**: 1837-1845.

BARRIO, E., A. LATORRE and F.J. AYALA, 1992. Phylogenetic reconstruction of the *Drosophila obscura* group on the basis of mitochondrial DNA. Mol. Biol. Evol. **9**: 621-635.

BARRIO, E., A. LATORRE and A. MOYA, 1994. Phylogeny of the *Drosophila obscura* species group deduced from mitochondrial DNA sequences. J. Mol. Evol. **39**: 478-488.

BARRIO, E. and F.J. AYALA, 1997. Evolution of the *Drosophila obscura* species group inferred from the *Gpdh* and *Sod* genes. Mol. Phylogenet. Evol. **7**: 70-93.

BAUDRY, E., C. KERDELHUE, H. INNAN and W. STEPHAN, 2001. Species and recombination effects on DNA variability in the tomato genus. Genetics **158**: 1725-1735.

BAZIN, E., S. GLÉMIN and N. GALTIER, 2006. Population size does not influence mitochondrial genetic diversity in animals. Science **312**: 570-572.

BECKENBACH, A.T. and A. PREVOSTI, 1986. Colonization of North America by the European species *Drosophila subobscura* and *D. ambigua*. The Amer. Midland Naturalist **115**: 10-18.

BECKENBACH, A.T., Y.W. WEI and H. LIU, 1993. Relationships in the *Drosophila obscura* group, inferred from mitochondrial cytochrome oxides-II sequences. Mol. Biol. Evol. **10**: 619-634.

BEGON, M., 1978. Population densities in *Drosophila obscura* Fallen and *D. subobscura* Collin. Ecol. Ent. **3**: 1-12.

BEGON, M., C.B. KRIMBAS and M. LOUKAS, 1980. The genetics of *Drosophila subobscura* populations. XV. Effective size of a natural population estimated by three independent methods. Heredity **45**: 335-350.

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

BEGUN, D.J. and C.F. AQUADRO, 1993. African and North American populations of *Drosophila melanogaster* are very different at DNA level. Nature **365**: 548-550.

BEGUN, D.J. and P. WHITLEY, 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **97**: 5960-5965.

BERNARDI, G., P. SORDINO and D. A. POWERS, 1993. Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. Proc. Natl. Acad. Sci. USA **90**: 9271-9274.

BERRY, A.J., J.W. AJIOKA and M. KREITMAN, 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics **129**: 1111-1117.

BÉTRAN, E., J. ROZAS, A. NAVARRO and A. BARBADILLA, 1997. The estimation of the number and the lenght distribution of gene conversion tracts ffrom populational DNA sequence data. Genetics **146**: 88-991.

BÉTRAN, E., M. SANTOS and A. RUIZ, 1998. Antagonistic pleiotropic effect of second-chromosomes inversions on body size and early life history traits in *Drosophila buzzatii*. Evolution **52**: 144-154.

BIRKY, C.W.Jr. and J.B. WALSH, 1988. Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA **85**: 6414-6418.

BITNER-MATHE, B., A.A. PEIXOTO and L.B. KLACZKO, 1995. Morphological variation in a natural population of *Drosophila mediopunctata*: altitudinal cline, temporal changes and influence of chromosome inversions. Heredity **75**: 54-61.

BONDINAS, G.P., M.G. LOUKAS, G.N. GOULIELMOS and D. SPERLICH, 2001. The actin loci in the genus *Drosophila*: establishment of chromosomal homologies among five paleartic species of the *Drosophila obscura* group by in situ hybridization. Chromosoma **110**: 441-450.

BONDINAS, G.P., M.G. LOUKAS, G.N. GOULIELMOS and D. SPERLICH, 2002. The actin loci in the genus *Drosophila*: establishment of chromosomal homologies among five neartic species of the *Drosophila obscura* group by in situ hybridization. Chromosoma **111**: 256-266.

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.

BREHM, A. and C.B. KRIMBAS, 1990a. The phylogeny of 9 species of the *Drosophila obscura* group inferred by the banding homologies of chromosomal regions. 2. Element E. Hereditas **113**: 157-168.

BREHM, A. and C.B. KRIMBAS, 1990b. Evolution of the *obscura* group *Drosophila* species. III. Phylogenetic relationships in the *subobscura* cluster based on homologies of chromosome A. Heredity **65**: 269-275.

BREHM, A., C.B. KRIMBAS, J. SOURDIS and M.L. CARIOU, 1991. The phylogeny of 9 species of the *Drosophila obscura* group inferred by the banding homologies of chromosomal regions. 1. Element B. Genome **34**: 464-471.

BREHM, A. and C.B. KRIMBAS, 1992. The phylogeny of 9 species of the *Drosophila obscura* group inferred by the banding homologies of chromosomal regions. 3. Element D. Genome **35**: 1075-1085.

BREHM, A. and C.B. KRIMBAS, 1993. The phylogeny of 9 species of the *Drosophila obscura* group inferred by the banding homologies of chromosomal regions. 4. Element C. Heredity **70**: 214-220.

BRNCIC, D., A. PREVOSTI, M. BUDNIK, M. MONCLUS and J. OCANA, 1981. Colonization of *Drosophila subobscura* in Chile. I. First population and cytogenetic studies. Genetica **56**: 3-9.

BROCKETT, M.M., H. ALAVI and W.W. ANDERSON, 1996. Relative effects of female fecundity and male mating success on fertility selection in *Drosophila pseudoobscura*. Proc. Natl. Acad. Sci. USA **93**: 3080-3082.

BROOKE, B.D., R.H. HUNT, F. CHANDRE, P. CARNEVALE and M. COETZEE, 2002. Stable chromosomal inversion polymorphisms and insecticide resistence in the malaria vector mosquito *Anopheles gambiae* (Diptera: Culicidae). J. Med. Entomol. **39**: 568-573.

BUGGE, M., G. BRUUN-PETERSEN, K. BRODUM-NIELSEN, U. FRIEDRICH, J. HANSEN, G. JENSEN, P.K. JENSEN, U. KRISTOFFERSSON, C. LUNDSTEEN, E. NIEBUHR, K.R. RASMUSSEN, K. RASMUSSEN and N. TOMMERUP, 2000. Disease associated balanced chromosome rearrangements: a resource for large-scale genotype-phenotype dileneation in man. J. Med. Genet. **16**: 171-187.

BULMER, M., 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics **129**: 897-907.

BURLA, H., 1951. Systeematik, Verbreitung und Oekologie der *Drosophila*-Arten der Schweiz. Rev. SuisseZool. **58**: 23-175.

BURLA, H., 1954. Distinction between four species of the *melanogaster* group, *Drosophila seguyi*, *D. montium*, *D. kikkawai n. sp*, and *D. auraria* (Drosophilidae: Diptera). Rev. Bras. Biol. **14**: 41-54.

BURTON, R. S. and B. N. LEE, 1994. Nuclear and mitochondrial gene genealogies and allozyne polymorphism across a major phylogeographic break in the copepod *Tigriopus californicus*. Proc. Natl. Acad. Sci. USA **91**: 5197-5201.

BUZZATI-TRAVERSO, A. and R.E. SCOSSIROLI, 1955. The *obscura* group of the genus *Drosophila*. Adv. Genet. **7**: 675-689.

CABALLERO, A. and W.G. HILL, 1992. Effects of partial inbreeding on fixation rates and variation of mutant genes. Genetics **131**: 493-507.

CABALLERO, A., 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics **139**: 1007-1011.

CABOT, E. L., A. W. DAVIS, N. A. JOHNSON and C.-I. WU, 1994. Genetics of reproductive isolation in the *Drosophila simulans* clade: complex epistasis underlying hybrid male sterility. Genetics **137**: 175-189.

CÁCERES, M., J.M. RANZ, A. BarbadiLLA, M. LONG and A. RUIZ, 1999. Generation of a widespread *Drosophila* inversion by a transposable element. Science **285**: 415-418.

CÁCERES, M., M. PUIG and A. RUIZ, 2001. Molecular characterization of two natural hotspots in the *Drosophila buzzatii* genome induced by transposon insertions. Genome Research **11**: 1353-1364.

CALBOLI, F.C.F., W.J. KENNINGTON and L. PARTRIDGE, 2003a. QTL mapping reveals a striking coincidence in the positions of genomic regions associated with adaptive variation in body size in parallel clines of *Drosophila melanogaster* in different continents. Evolution **57**: 2653-2658.

CALBOLI, F.C.F., G.W. GILCHRIST and L. PARTRIDGE, 2003b. Different cell size and cell number contribution in two newly estabilished and one ancient body size cline of *Drosophila subobscura*. Evolution **57**: 566-573.

CARIOU, M.L., D. LACHAISE, L. TSACAS, J. SOURDIS, C.B. KRIMBAS and M. ASHBURNER, 1988. New African species in the *Drosophila obscura* species group: genetic variation, differentiation and evolution. Heredity **61**: 73-84.

CASALS, F., M. CÁCERES and A. RUIZ, 2003. The foldback-like transposon *Galileo* is involved in the generation of two different natural chromosomal inversions of *Drosophila buzzatii*. Mol. Biol. Evol. **20**: 674-685.

CHARLESWORTH, D. and B. CHARLESWORTH, 1979. Selection and recombination in clines. Genetics **91**: 581-589.

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

CHARLESWORTH, B., 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63**: 213-227.

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

CHARLESWORTH, B., 1996. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. **68**: 131-149.

CHERRY, J.L. and J. WAKELEY, 2003. A difussion approximation for selection and drift in a subdivided population. Genetics **163**: 421-428.

CHOVNICK, A., 1973. Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. Genetics **75**: 123-131.

CIRERA, S., J.M. MARTÍN-CAMPOS, C. SEGARRA and M. AGUADÉ, 1995. Molecular characterization of the breakpoints of an inversion fixed between *Drosophila melanogaster* and *D. subobscura*. Genetics **139**: 321-326.

CIRERA, S. and M. AGUADÉ, 1998. Molecular evolution of a duplication: the sex-peptide (*Acp70A*) gene region of *Drosophila subobscura* and *Drosophila madeirensis*. Mol. Biol. Evol. **15**: 988-996.

CLARK, A.G., K.M. WEISS, D.A. NICKERSON, S.L. TAYLOR, A. BUCHANAN, J. STENGARD, V. SALOMAA, E. VARTIANIEN, M. PEROLA, E. BOERWINKLE *et al.*, 1998. Haplotype structure and population genetics inferences from nucleotide-sequence variation in human lipoprotein lipase. Am. J. Hum. Genet. **63**: 595-612.

CLARKE, B. C., M. S. JOHNSON and J. MURRAY, 1996. Clines in the genetic distance between two species of island land snails: how " molecular leakage" can mislead us about speciation. Philos. Trans. Soc. Lond. Ser. B **351**: 773-784.,

COLLIN, J.E., 1936. Note *Drosophila subobscura* sp.n. male, female. Addedum p. 60 to Gordon (1936). J. Genetics **33**: 25-60.

COLUZZI, M., A. SABATINI, A. DELLA TORRE, M.A. DI DECO and V. PETRARCA, 2002. A polytene chromosome analysis of the *Anopheles gambiae* species complex. Science **298**: 1451-1418.

COMERÓN, J.M., 1997. Estudi de la variabilitat nucleotídica a *Drosophila*: regió *Xdh* a *Drosophila subobscura*. Tesi doctoral. Universitat de Barcelona, Spain.

COYNE, J. A., 1984. Genetic basis of male sterility in hybrids between two closely related species of Drosophila. Proc. Natl. Acad. Sci. USA **81**: 4444-4447.

COYNE, J. A., 1985. The genetic basis of Haldane's rule. Nature **314**: 736-738.

COYNE, J. A. and B. CHARLESWORTH, 1989. Genetic analysis of X-linked factor causing sterility between three sibling species of Drosophila. Heredity **62**: 97-106.

COYNE, J.A., J. RUX and J.R. DAVID, 1991. Genetics of morphological differences and hybrid sterility between *Drosophila sechellia* and its relatives. Genet. Res. **57**: 113-122.

COYNE, J.A., W. MEYERS, A. CRITTENDEN and P. SNIEWOWSKI, 1993. The fertility effects of pericentric inversions in *Drosophila melanogaster*. Genetics **134**: 487-496.

COYNE, J.A., A.P. GRITTENDEN and K. MAH, 1994. Genetics of a pheromonal difference contributing to reproductive isolation in Drosophila. Science **265**: 1461-1464.

COYNE, J.A. and B. CHARLESWORTH, 1997. Genetics of a pheromonal difference affecting sexual isolation between *Drosophila mauritiana* and *D. sechellia*. Genetics **145**: 1015-1030.

COYNE, J.A. and H.A. ORR, 1998. The evolutionary genetics of speciation. Phil. Philos. Trans. Soc. Lond. Ser. B **353**: 287-305.

CROW, J. and N.E. MORTON, 1954. Measurement of gene frequency drift in small populations. Evolution **9**: 202-214.

CUTTER, A.D., 2006. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. Genetics **172**: 171-184.

DEAN, M.D., K.J. BALLARD, A. GLASS and J.W. BALLARD, 2003. Influence of two Wolbachia strains on population structure of East African *Drosophila simulans*. Genetics **165**: 1959-1969.

DELLA TORRE, A., L. MERZAGORA, J. R. POWELL and M. COLUZZI, 1997. Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. Genetics **146**: 239-244.

DEPAULIS, F.L., L. BRAZIER and M. VEUILLE, 1999. Selective sweep at the *Drosophila melanogaster suppressor of hairless* locus and its association with the *In(2L)t* inversion polymorphism. Genetics **152**: 1017-1024.

DESALLE, R., 1992. The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol. Phylogenet. Evol. **1**: 31-40.

DEUTZ-TERLOW, P.P., M. LOSEKOOT, R. OLMER, W.C. PIENEMAN, S. de VRIES-V D WEERD, E. BRIET and E. BAKER, 1995. Inversions in the factor VIII gene: improvement of carrier detection and prenatal diagnosis in Dutch haemophilia A families. J. Med. Genet. **32**: 296-300.

DICKERSON, R.E., 1971. The structure of cytochrome *c* and the rates of molecular evolution. J. Mol. Evol. **1**: 26-45.

DOBZHANSKY, T., 1936. Studies of hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics **21**: 113-135.

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

DOBZHANSKY, T., 1947. Genetics of natural populations. XIV. A response of certain gene arrangements in the third chromosome of *Drosophila pseudoobscura* to natural selection. Genetics **32**: 142-160.

DOBZHANSKY, T., 1948. Genetics of natural populations. XVIII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographic regions. Genetics **33**: 588-602.

DOBZHANSKY, T., 1954. Evolution as a creative process. Proc. 9th Int. Congr. Genetics 435-449.

DOBZHANSKY, T., 1970. *Genetics of the Evolutionary Process*, Columbia University Press, New York.

DOBZHANSKY, T. and J.R. POWELL, 1975. *Drosophila pseudoobscura* and its American relatives, *Drosophila persimilis* and *Drosophila miranda*. pp. 537-587 in *Handbook of Genetics. Vol. 3: Invertebrates of Genetic Interest*, edited by R.C. KING, Plenum, New York.

DORIA, A.A., 1945. *Estudos de história dos descobrimentos. O problema do descobrimento da Madeira*, Guimarães, Portugal.

DVORAK, J., M.C. LUO and Z.L. YANG, 1998. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. Genetics **148**: 423-434.

ENDLER, J. A., 1977. *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey.

FANARA, J.J., E. HASSON and C. RODRIGUEZ, 1997. The effect of polymorphic inversions on body size in two natural populations of *Drosophila buzzatii* from Argentina. Hereditas **126**: 233-237.

FAY, J. C. and C.-I WU, 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial and nuclear DNA variation. Mol. Biol. Evol. **16**: 1003-1005.

FELDMAN, M.W., S.P. OTTO and F.B. CHRISTIANSEN, 1997. Population geneic perspectives on the evolution of recombination. Ann. Rev. Genetics **30**: 261-295.

FELSENSTEIN, J., 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. Evolution **35**: 124-138.

FERNANDEZ IRIARTE, P. and E. HASSON, 2000. The role of the use of different host plants in the maintenance of the inversion polymorphism in the cactophilic *Drosophila buzzatii*. Evolution **54**: 1295-1302.

FERNANDEZ IRIARTE, P., F.M. NORRY and E. HASSON, 2003. Chromosome inversions effect body size and shape in different breeding resources in *Drosophila buzzatii*. Heredity **91**: 51-59.

FISHER, R.A., 1958. *The Genetical Theory of Natural Selection*, Dover Publications, New York.

FREESE, E., 1962. On the evolution of base composition at DNA. J. Theor. Biol. **3**: 82-101.

FRISCHAUF, A.M., H. LEHRACH, A. POUSTKA and N. MURRAY, 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol Bio. **170**: 827-842.

FRISSE, L., R.R. HUDSON, A. BARTOSZEWICZ, J.D. WALL, J. DONFACK and A. DIRIENZO, 2001. Gene conversion and different population histories nay explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. **69**: 831-843.

FRUTOS, R. DE and A. AGUILAR, 1978. The lack of association between the distinct interchromosomal arrangements of *Drosophila subobscura*. Drosophila Information Service **53**: 136-137.

FRUTOS, R. DE, A. LATORRE and L. PASCUAL, 1987a. Patterns of puffing activity and chromosomal polymorphism in *Drosophila subobscura*. IV. Position effect at the boundaries of the E_{12} inversion. Genetica **75**: 11-22.

FRUTOS, R. DE, A. LATORRE and L. PASCUAL, 1987b. Chromosomal polymorphism and gene expression in *Drosophila subobscura*. Genet. Iberica **39**: 307-314.

FRUTUOSO, G., 1979. *Livro segundo das saudades da Terra*, Ponta Delgada, Portugal.

FRYDENBERG, J., A.A. HOFFMANN and V. LOESCHCKE, 2003. DNA sequence variation and latitudinal associations in *hsp23*, *hsp26* and *hsp27* from natural populations of *Drosophila melanogaster*. Mol. Ecol. **12**: 2025-2032.

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

GILLESPIE, J.H., 1984. Molecular evolution over the mutational lanscape. Evolution **38**: 1116-1129.

GILLESPIE, J.H., 1991. The causes of molecular evolution, Oxford University Press, New York.

GILLESPIE, J.H., 1997. Junk ain't what junk does: neutral alleles in a selected context. Gene **205**: 291-299.

GILLESPIE, J.H., 1999. The role of population size in molecular evolution. Theor. Pop. Biol. **55**: 145-156.

GILLESPIE, J.H., 2000a. Genetic drift in an infinite population: the pseudohitchhiking model. Genetics **155**: 909-919.

GILLESPIE, J.H., 2000b. The neutral theory in an infinite population. Gene **261**: 11-18.

GILLESPIE, J.H., 2001. Is the population size of a species relevant to its evolution? Evolution **55**: 2161-2169.

GOCKEL, J., W.J. KENNINGTON, A.A. HOFFMANN, D.B. GOLDSTEIN and L. PARTRIDGE, 2001. Nonclinality of molecular variation implicates selection in maintaining a morphological cline of *Drosophila melanogaster*. Genetics **158**: 319-323.

GOCKEL, J., W.J. KENNINGTON, D.B. GOLDSTEIN and L. PARTRIDGE, 2002. Quantitative genetic analysis of natural variation in body size in *Drosophila melanogaster*. Heredity **89**: 145-153.

GODDARD, K., A. CACCONE and J.R. POWELL, 1990. Evolutionary implications of DNA divergence in the *Drosophila obscura* group. Evolution **44**: 1656-1670.

GONZALEZ, A.M., M. HERNANDEZ, A. VOLZ, J. PESTANO, J.M. LARRUGA, D. SPERLICH and V.M. CABRERA, 1990. Mitochondrial DNA evolution in the *obscura* species subgroup of *Drosophila*. J. Mol. Evol. **31**: 122-131.

GORDON, C., 1936. The frequency of heterozygosis in free-living populations of *Drosophila melanogaster* and *Drosophila subobscura*. J. Genetics **33**: 25-60.

GOSTELI, M., 1990. Chromosomal polymorphism in natural populations of *Drosophila subobscura* near Zurich, Switzerland: a contribution to long-term comparisons. Genetica **81**: 199-204.

GOSTELI, M., 1991. Differential flight activity among karyotypes: daily and weather-induced changes in chromosomal inversion polymorphism in natural populations of *Drosophila subobscura*. Genetica **84**: 129-136.

GREEN, M.M., 1980. Transposable elements in *Drosophila* and other Diptera. Annu. Rev. Genet. **14**: 109-120.

GRELL, R., 1962. A new model for secondary nondisjunction: the role of distributive pairing. Genetics **47**: 1737-1754.

GRIMALDI, D.A., 1990. A phylogenetic, revised classification of the genera in the Drosophilidae (Diptera). Bull. Am. Mus. Nat. Hist. **197**: 1-139.

GROSSMAN, A.I., 1967. Distribution of breaks limiting autosomal inversions in natural populations of *Drosophila melanogaster*. Sov. Genet. **3**: 31.

HALLAS, R., M. SCHIFFER and A.A. HOFFMANN, 2002. Clinal variation in *Drosophila serrata* for stress resistence and body size. Genet. Res. **79**: 141-148.

HANAHAN, D., 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**: 557.

HARE, M. P. and J. C. AVISE, 1998. Population structure in the american oyster as inferred by nuclear gene genealogies. Mol. Phylogenet. Evol. **15**: 119-128.

HARRIS, E.E. and J. HEY, 1999. X chromosome evidence for ancient human histories. Proc. Natl. Acad. Sci. USA **96**: 3320-3324.

HASSON, E. and W.F. EANES, 1996. Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. Genetics **144**: 1565-1575.

HEED W. B. and P.M. O'GRADY, 2000. *Drosophila maya*, a new Neotropical member of the *Drosophila obscura* species group. J. New York Entomol. Soc. **108**: 97-104.

HEY, J., 1991. The structure of genealogies and the distribution of fixed differences between DNA sequence samples from natural populations. Genetic*s* **128**: 831-840.

HEY, J. and R. M. KLIMAN, 1993. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10**: 804-822.

HEY, J., 1994. Bridging phylogenetics and population genetics with gene tree models, pp. 435 – 447 in *Molecular Ecology and Evolution: Approaches and Applications*, edited by B. SCHIERWATER, B. STREIT, G. P. WAGNER and R. DESALLE, Birkhauser Verlag, Basel, Switzerland.

HEY, J. and J. WAKELEY, 1997. A coalescent estimator of the population recombination rate. Genetics **145**: 833-846.

HEY, J. and R. NIELSEN, 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics **167**: 747-760.

HILL, W.G. and A. ROBERTSON, 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38**: 226-231.

HILTON, H. and J. HEY, 1997. A multilocus view of speciation in the *Drosophila virilis* group reveals complex histories and taxonomic conflicts. Genet. Res. **70**: 185-846.

HOFFMANN, A.A., A.R. ANDERSON and R. HALLAS, 2002. Opposing clines for high and low temperature resistence in *Drosophila melanogaster*. Ecol. Lett. **5**: 614-618.

HUDSON, R.R. and N.L. KAPLAN, 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111**: 147-164.

HUDSON, R. R., 1987. Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50**: 245-250.

HUDSON, R. R., M. KREITMAN and M. AGUADÉ, 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**: 153-159.

HUDSON, R.R. and N.L. KAPLAN, 1988. The coalescent process in models with selection and recombination. Genetics **120**: 831-840.

HUDSON, R. R., 1990. Gene genealogies and the coalescent process, pp. 1-44 in *Oxford Surveys in Evolutionary Biology*, edited by P. H. HARVEY and L. PARTRIDGE, Oxford University Press, New York.

HUDSON, R.R., M. SLATKIN and W.P. MADDISON, 1992a. Estimation of levels of gene flow from DNA sequence data. Genetics **132**: 583-589.

HUDSON, R.R., D.D. BOOS and N.L. KAPLAN, 1992b. A statistical test for detecting population subdivision. Mol. Biol. Evol. **9**: 138-151.

HUDSON, R.R., 1994. How can levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? Proc. Natl. Acad. Sci. USA **91**: 6815-6818.

HUDSON, R.R. and N.L. KAPLAN, 1994. Gene trees with background selection, pp. 140-153 *in Mechanisms of Molecular Evolution*, edited by N. TAKAHATA and A.G. CLARK, Japan Scientific Society, Tokyo.

HUDSON, R.R. and N.L. KAPLAN, 1995a. Deleterious background selection with recombination. Genetics **141**: 1605-1617.

HUDSON, R.R. and N.L. KAPLAN, 1995b. The coalescent process and background selection. Philos. Trans. R. Soc. Lond. Ser. B **349**: 19-23.

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.

IKEMURA, T., 1981. Correlation between the abundance of *Escherichia coli* transfer rnas and the occurrence of the respective codons in its protein genes. J. Mol. Biol. **146**: 1-21.

IRVIN, S.D., K.A. WETTERSTRAND, C.M. HUTTER and C.F. AQUADRO, 1998. Genetic variation and differentiation at microsatellite loci in *Drosophila simulans*: evidence for founder effects in new world populations. Genetics **150**: 777-790.

JIANG, C.X., P. W. CHEE, X. DRAYE, P.L. MORRELL, C.W. SMITH and A.H. PATERSON, 2000. Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). Evolution **54**: 798-814.

JOHNSON, K.P. and J. SEGER, 2001. Elevated rates of nonsynonymous substitution in island birds. Mol. Biol. Evol. **18**: 874-881.

JONES, C.D., 1998. The genetic basis of *Drosophila sechellia*'s resistence to a host plant toxin. Genetics **149**: 1899-1908.

JORDE, L.B., W.S. WATKINS and M.J. BAMSHAD, 2001. Population genomics: a bridge from evolutionary history to genetic medicine. Hum. Mol. Genet. **10**: 2199-2207.

JUNGEN, H., 1968a. Inversionspolymorphismus in tunesischen populationen von *Drosophila subobscura* Collin. Archiv. Julius Klaus-Stiftung Vererbungsforschung Socialanthropologie Rassenhygiene **43**: 3-55.

JUNGEN, H., 1968b. "Sex ratio" in naturlichen populationen von *Drosophila subobscura*. Archiv. Julius Klaus-Stiftung Vererbungsforschung Socialanthropologie Rassenhygiene **43**: 52-57.

KAMPING, A. and W. VAN DELDEN, 1999. The role of fertility restoration in the maintenance of the inversion *In(2L)t* polymorphism in *Drosophila melanogaster*. Heredity **83**: 460-468.

KAPLAN, N.L., T. DARDEN and R.R. HUDSON, 1988. The coalescent process in models with selection. Genetics **120**: 819-829.

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

KAROTAM, J., T.M. BOYCE and J.G. OAKESHOTT, 1993. Nucleotide variation at the hypervariable *esterase 6* isozyme locus of *Drosophila simulans*. Mol. Biol. Evol. **12**: 113-122.

KAUER, M., B. ZANGERL, D. DIERENGER and C. SCHLOTTERER, 2002. Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *D. melanogaster*. Genetics **160**: 247-256.

KELLY, J.K., 1997. A test of neutrality based on interlocus associations. Genetics **146**: 1197-1206.

KELLY, J.K. and M.J. WADE, 2000. Molecular evolution near a two-locus blanced polymorphism. J. Theor. Biol. **204**: 83-101.

KHADEM, M. and C. B. KRIMBAS, 1991a. Studies of the species barrier between *Drosophila subobscura* and *D. madeirensis*. I. The genetics of male hybrid sterility. Heredity **67**: 157-165.

KHADEM, M. and C. B. KRIMBAS, 1991b. Studies of the species barrier between *Drosophila subobscura* and *D. madeirensis*. II. Genetic analysis of developmental incompabilities in hybrids. Hereditas **114**: 189-195.

KHADEM, M. and C. B. KRIMBAS, 1993. Studies of the species barrier between *Drosophila subobscura* and *D. madeirensis*. III. How universal are the rules of speciation? Heredity **70**: 353-361.

KHADEM, M. and C. B. KRIMBAS, 1997. Studies of the species barrier between *Drosophila subobscura* and *D. madeirensis*. IV. A genetic dissection of the X chromosome for speciation genes. J. Evol. Biol. **10**: 909-920.

KHADEM, M., J. ROZAS, C. SEGARRA, A. BREHM and M. AGUADÉ, 1998. Tracing the colonization of Madeira and the Canary Islands by *Drosophila subobscura* through the study of the *rp49* gene region. J. Evol. Biol. **11**: 439-452.

KHADEM, M., J. ROZAS, C. SEGARRA and M. AGUADÉ, 2001. DNA variation at the *rp49* gene region in *Drosophila madeirensis* and *D. subobscura* from Madeira: inferences about the origin of an insular endemic species. J. Evol. Biol. **14**: 379-387.

KILGER, C. and K. SCHMID, 1994. Rapid characterization of bacterial by microwace treatment and PCR. Trends in Genetics **10**: 149.

KIMURA, M. and J.F. CROW, 1964. The number of alleles that can be maintained in a finite population. Genetics **49**: 725-738.

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

KIMURA, M., 1969. The rate of molecular evolution considered from the standpoint of population genetics. Proc. Natl. Acad. Sci. USA **63**: 1181-1188.

KIMURA, M., 1971. Theoretical foundation of population genetics at the molecular level. Theor. Pop. Biol. **2**: 174-208.

KIMURA, M. and T. OHTA, 1974. On some principles governing molecular evolution. Proc. Natl. Acad. Sci. USA **71**: 2848-2852.

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

KIMURA, M., 1981. Possibility of extensive neutral evolution under stabilizing selection with special reference to nonrandom usage of synonymous codons. Proc. Natl. Acad. Sci. USA **78**: 5773-5777.

KIMURA, M., 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK.

KING, J.L. and T.H. JUKES, 1969. Non-Darwinian evolution. Science **164**: 788-798.

KIRKPATRICK, M. and N. BARTON, 2005. Chromosome inversions, local adaptation and speciation. Genetics **173**: 419-434.

KITTEL, M. and D. SPERLICH, 1989. Restriction fragment and allozyme polymorphism in a wild population of *Drosophila subobscura* from tubingen/Germany. Abstracts 11th European *Drosophila* Research Conference (Marseille) 36.

KLIMAN, R.M. and J. HEY, 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10**: 1239-1258.

KLIMAN, R. M., P. ANDOLFATTO, J.A. COYNE, F. DEPAULIS, M. KREITMAN, A.J. BERRY, J. MCCARTER, J. WAKELEY and J. HEY, 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics **156**: 1913-1931.

KOVACEVIC, M. and S.W: SCHAEFFER, 2000. Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. Genetics **156**: 155-172.

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

KREITMAN, M. and R.R. HUDSON, 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127**: 565-582.

KREITMAN, M. and M.L. WAYNE, 1994. Organization of genetic variation at molecular level: lessons from *Drosophila*, pp. 157-183 in *Molecular ecology and evolution: approaches and implications*, edited by D. SCHIERWATER, B. STREIT, G.P. WAGNER and R. DE SALLE, Biirkhauser, Basel, Switzerland.

KRIMBAS, C.B., 1964. The genetics of *Drosophila subobscura* populations. II. Inversion polymorhism in a population from Holland. Z. Vererbunslehre **95**: 125-128.

KRIMBAS, C.B., 1967. The genetics of *Drosophila subobscura* populations. III. Inversion polymorphism and climatic factors. Mol. Gen. Genet. **99**: 133-152.

KRIMBAS, C.B. and M. LOUKAS, 1980. The inversion polymorphism of *Drosophila subobscura*. Evol. Biol. **12**: 163-234.

KRIMBAS, C.B. and M. LOUKAS, 1984. Evolution of the *obscura* group *Drosophila* species. I. Salivary chromosomes and quantitative characters in *D. subobscura* and two closely related species. Heredity **53**: 469-482.

KRIMBAS, C.B. and J.R. POWELL (eds), 1992. *Drosophila Inversion Polymorphism*, CRC Press, Boca Raton, Florida.

KRIMBAS, C.B., 1993. *Drosophila subobscura*. *Biology, Genetics and Inversion Polymorphism*, Verlag Kovac, Hamburg.

KUMAR, S., K. TAMURA and M. NEI, 2004. Mega 3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics **5**: 150-163.

KUNZE-MUHL, E. and E. MULLER, 1958. Weitere untersuchungen uber die chromosomale struktur und naturalichen strukturtypen von *Drosophila subobscura*. Chromosoma **9**: 559-570.

KWIATOWSKI, J., D. SKARECKY, K. BAILEY and F.J. AYALA, 1994. Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu,Zn *Sod* gene. J. Mol. Evol. **38**: 443-454.

KWIATOWSKI, J., M. KRAWCZYK, M. JAWORSKI, D. SKARECKY and F.J. AYALA, 1997. Erratic evolution of glycerol-3-phosphate dehydrogenase in *Drosophila*, *Chymomyza*, and *Ceratitis*. J. Mol. Evol. **44**: 9-22.

KWIATOWKSI, J. and F.J. AYALA, 1999. Phylogeny of *Drosophila* and related genera: conflict between molecular and anatomical analysis. Mol. Phylogenet. Evol. **13**: 319-328.

LAAYOUNI, H., E. HASSON, M. SANTOS and A. FONTDEVILA, 2003. The evolutionary history of *Drosophila buzzatii*: XXXV. Inversion polymorphism and nucleotide variability in different regions of the second chromosome. Mol. Biol. Evol. **20**: 931-944.

LACHAISE, D., M.L. CARIOU, J.R. DAVID, F. LEMEUNIER, L. TSACAS and M. ASHBURNER, 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22**: 159-225.

LAKOVAARA, S. and A. SAURA, 1982. Evolution and speciation in the *Drosophila obscura* group, pp. 1-59 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER, H.L. CARSON and J.M. THOMPSON, Academic Press, New York.

LANGLEY, C.H. and C.F. AQUADRO, 1987. Restriction-map variation in natural populations of *Drosophila melanogaster*: white-locus region. Mol. Biol. Evol. **4**: 651-663.

LANGLEY, C.H., 1990. The molecular population genetics of *Drosophila*, pp. 75-91 in *Population Biology of Genes and Molecules*, edited by N. TAKAHATA and J.F. CROW, Baifukan, Tokyo.

LANGLEY, C., B. LAZZARO, W. PHILIPS, E. HEIKKINEN and J. BRAVERMAN, 2000. Linkage disequilibria and the site frequency spectra in the *su(s)* and *su(w^a)* regions of the *Drosopjila melanogaster* X chromosome. Genetics **156**: 1837-1852.

LATORRE, A., E. BARRIO, A. MOYA and F.J. AYALA, 1988. Mitochondrial DNA evolution in the *Drosophila obscura* group. Mol. Biol. Evol. **5**: 717-728.

LEMEUNIER, F., J.R. DAVID, L. TSACAS and M. ASHBURNER, 1986. The *melanogaster* species group, pp. 147-256 in *The Genetics and Biology of Drosophila* Vol. 3a, edited by M. ASHBURNER, J.M. THOMPSON and H.L. CARSON, Academic Press, London.

LENORMAND, T. and S.P.OTTO, 2000. The evolution of recombination in a heterogeneous environment. Genetics **156**: 423-438.

LEVINE, R.P., 1956. Crossing over and inversions in coadapted systems. Am. Nat. **90**: 41-45.

LEWONTIN, R.C. and K. KOJIMA, 1960. The evolutionary dynamics of complex polymorphisms. Evolution **14**: 458-472.

LEWONTIN, R.C., 1964. The interaction of selection and linkage I. General considerations: heterotic models. Genetics **49**: 49-67.

LEWONTIN, R.C., 1974. *The genetics basis of evolutionary change*, Columbia University Press, New York.

LEWONTIN, R.C., 1981. The scientific work of Th. Dobzhansky, pp. 93-155 in *Dobzhansky's genetics of natural populations*, edited by R.C. LEWONTIN, J.A. MOORE, W.B. PROVINE and B. WALLACE, Columbia University Press, New York.

LI, W.-H., T. GOJOBORI and M. NEI, 1981. Pseudogenes as a paradigm of neutral evolution. Nature **292**: 237-239.

LI, W.-H., 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. **24**: 337-345.

LI, Y.J., Y. SATTA and N. TAKAHATA, 1999. Paleo-demography of the *Drosophila melanogaster* subgroup: application of the maximum likelihood method. Genes Genet. Syst. **74**: 117-127.

LINDSEY, D.L. and G. ZIMM, 1992. The genome of *Drosophila melanogaster*. Academic Press, San Diego.

LLOPART, A., 1999. Evolución de la región *RpII215* en el grupo *obscura* de *Drosophila* y comparación con el grupo *melanogaster*. Análisis a nivel nucleotídico. Tesi doctoral. Universitat de Barcelona, Spain.

LLOPART, A. and M. AGUADÉ, 1999. Synonymous rates at the *RpII215* gene of *Drosophila*: variation among species and across the coding region. Genetics **152**: 269-280.

LLOPART, A. and M. AGUADÉ, 2000. Nulceotide polymorphism at the *RpII215* gene in *Drosophila subobscura*. Weak selection on synonymous mutations. Genetics **155**: 1245-1252.

LOUKAS, M., C.B. KRIMBAS and Y. VERGINI, 1979. The genetics of *Drosophila subobscura* populations. IX. Studies on linkage disequilibrium in four natural populations. Genetics **93**: 497-523.

LOUKAS, M. and C.B. KRIMBAS, 1980. The genetics of *Drosophila subobscura* populations. XVI. Heterokaryosis and heterozygosity. Genetica **54**: 69-74.

MACHADO, C.A., R.M. KLIMAN, J.A. MARKERT and J. HEY, 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. Mol. Biol. Evol. **19**: 472-488.

MADDISON, D.R. and W.P. MADDISON, 2002. MacClade: Analysis of phylogeny and character evolution, version 4.05. Sinauer Associates. Sunderland, MA.

MAINX, F., T. KOSKE and E. SMITAL, 1953. Untersuchugen uber die chromosomale struktur europaischer vertreter der *Drosophila obscura* gruppe. Z. F. Indukt. Abstamm. Vererbungslehre **85**: 354-372.

MARTÍN-CAMPOS, J., 1998. Análisis de la variabilidad a nivel nucleotídico en la región *6-Pgd* en *Drosophila subobscura*. Tesi doctoral. Universitat de Barcelona, Spain.

MARTIN-CAMPOS, J., J. COMERON, N. MIYASHITA and M. AGUADÉ, 1992. Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *D. melanogaster*. Genetics **130**: 805-816.

MATHER, W.B., 1954. The genus *Drosophila* (Diptera) in Eastern Queensland. I. Taxonomy. Aust. J. Zool. **4**: 76-89.

MATZKIN, L., T. MERRIT, C-T ZHU and W. EANES, 2005. The structure and population genetics of the breakpoints associated with the cosmopolitan chromosomal inversion *In(3R)Payne* in *D. melanogaster*. Genetics **170**: 1143-1152.

MAYNARD SMITH , J., 1966. Sympatric speciation. Am. Nat. **100**: 637-650.

MAYNARD SMITH, J. and J. HAIGH, 1974. The hicthhiking effect of a favorable gene. Genet. Res. Cambridge **23**: 23-35.

McColl, G. and S.W. McKechnie, 1999. The *Drosophila* heat shock *hsr-omega* gene: an allele frequency cline detected by quantitative PCR. Mol. Biol. Evol. **16**: 1568-1574.

McVean, G.A.T. and J. Vieira, 2001. Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. Genetics **157**: 245-257.

Mestres, F. and L. Serra, 1991. Lethal allelism in *Drosophila subobscura*: difficulties in the estimation of certain population parameters. Z. Zool. System. Evolut/forschung **29**: 264-279.

Mestres, F., L. Serra and F.J. Ayala, 1995. Colonization of the Americas by *Drosophila subobscura*: lethal-gene allelism and association with chromosomal arrangements. Genetics **140**: 1297-1305.

Mestres, F., J. Sanz and L. Serra, 1998. Chromosomal structure and recombination between inversions in *Drosophila subobscura*. Hereditas **128**: 105-113.

Miyashita, N. and C.H. Langley, 1988. Molecular and phenotypic variation at the *white* locus region in *Drosophila melanogaster*. Genetics **120**: 199-202.

Miyashita, N.T., K. Yoshida and T. Ishii, 2005. DNA variation in the metallothionein genes in wild rice *Oryza rufipogon*: relationship between DNA sequence polymorphism, codon bias and gene expression. Genes Genet. Syst. **80**: 173-183.

Miyata, T. and T. Yasunaga, 1981. Rapidly evolving mouse α-globin related pseudogene and its evolutionary history. Proc. Natl. Acad. Sci. USA **78**: 450-453.

Molto, M.D. and M.J. Martinez-Sebastian, 1986. Gene arrangements in polytene chromosomes of *Drosophila guanche* differing from standard arrangements of *D. subobscura*. Drosophila Information Service **63**: 97-98.

Molto, M.D., L. Pascual and R. de Frutos, 1987. Puff activity after heat shock in two species of the *Drosophila obscura* group. Experientia **43**: 1225-1227.

Monclús, M., 1953. Variacion geografica de los peines tarsales de los machos de *D. subobscura*. Genet. Iberica **5**: 101-114.

Monclús, M. and A. Prevosti, 1971. The relationship between mating speed and wing length in *Drosophila subobscura*. Evolution **25**: 214-217.

Monclús, M., 1984. Drosophilidae of Madeira, with the description of *Drosophila madeirensis* n.sp. Z. Zool. Syst. Evolutionsforsch **22**: 94-103.

Moriyama, E.N. and J.R. Powell, 1996. Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13**: 261-277.

Morton, R.A., M. Choudhary, M.L. Cariou and R.S. Singh, 2004. A reanalysis of protein polymorphism in *Drosophila melanogaster, D. simulans, D. sechellia* and *D. mauritiana*: effects of population size and selection. Genetica **120**: 101-114.

Mousset, S. and N. Derome, 2004. Molecular polymorphism in *Drosophila melanogaster* and *D. simulans*: what have we learned from recent studies? Genetica **120**: 79-86.

MULLER, H.J., 1940. Bearings of the Drosophila work on systematics, pp.185-268 in *The New Systematics*, edited by J. HUXLEY, Clarendon Press, Oxford, U.K.

MUNTÉ, A., M. AGUADÉ and C. SEGARRA, 2000. Nucleotide variation at the *yellow* gene region is not reduced in *Drosophila subobscura*: a study in relation to chromosomal polymorphism. Mol. Biol. Evol. **17**: 1942-1955.

MUNTÉ, A., M. AGUADÉ and C. SEGARRA, 2001. Changes in the recombinational environment affect divergence in the *yellow* gene of *Drosophila*. Mol. Biol. Evol. **18**: 1045-1056.

MUNTÉ, A., J. ROZAS, M. AGUADÉ and C. SEGARRA, 2005. Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. Genetics **169**: 1573-1581.

NACHMAN, M.W. and G. CHURCHILL, 1996. Heterogeneity in rates of recombination across the mouse genome. Genetics **142**: 537-548.

NACHMAN, M.W., 1997. Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics **147**: 1303-1316.

NACHMAN, M.W., V.L. BAUER, S.L.CROWELL and C.F. AQUADRO, 1998. DNA variability and recombination rates at X-linked loci in humans. Genetics **150**: 1133-1141.

NACHMAN, M.W. and S.L. CROWELL, 2000. Contrasting evolutionary histories of two introns of the duchenne muscular dystrophy gene, *Dmd*, in humans. Genetics **155**: 1855-1864.

NACHMAN, M.W., 2001. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. **17**: 481-485.

NAVARRO, A. and A. RUIZ, 1997. On the fertility effects of pericentric inversions. Genetics **147**: 931-933.

NAVARRO, A., E. BETRAN, A. BARBADILLA and A. RUIZ, 1997. Recombination and gene flux caused by gene conversion and crossing over in inversion heteroksryotypes. Genetics **146**: 695-709.

NAVARRO, A., A. BARBADILLA and A. RUIZ, 2000. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in *Drosophila*. Genetics **155**: 685-698.

NAVARRO-SABATÉ, A., M. AGUADÉ and C. SEGARRA, 1999a. The relationship between allozyme and chromosomal polymorphism inferred from nucleotide variation at the *Acph-1* gene region of *Drosophila subobscura*. Genetics **153**: 871-889.

NAVARRO-SABATÉ, A., M. AGUADÉ and C. SEGARRA, 1999b. The *acid phosphatase-1* gene region in the *Drosophila* species of the *subobscura* cluster. Hereditas **130**: 65-75.

NAVARRO-SABATÉ, A., M. AGUADÉ and C. SEGARRA, 2003. Excess of nonsynonymous polymorphism at *Acph-1* in different gene arrangements of *Drosophila subobscura*. Mol. Biol. Evol. **20**: 1833-1843.

NEI, M., T. MARUYAMA and R. CHAKRABORTY, 1975. The bottleneck effect and genetic variability in populations. Evolution **29**: 1-10.

NEI, M., 1987. *Molecular Evolutionary Genetics*. Columbia University Press, NY.

NOOR, M.A., K.L. GRAMS, L.A. BERTUCCI and J. REILAND, 2001a. Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA **98**: 12084-12088.

NOOR, M.A., K.L. GRAMS, L.A. BERTUCCI, Y. ALMENDAREZ, J. REILAND and K.R. SMITH, 2001b. The genetics of reproductive isolation and the potential for gene exchange between *Drosophila pseudoobscura* and *D. persimilis* via backcross hybrid males. Evolution **55**: 512-521.

NORRY, F.M., J.J. FANARA, E. HASSON and C. RODRIGUEZ, 1995. An adaptive chromosomal polymorphism affecting size-related traits, and longevity selection in a natural population of *Drosophila buzzatii*. Genetica **96**: 285-291.

NOVITSKI, E. and G. BRAVER, 1954. An analysis of crossing over within an heterozygous inversion in *Drosophila melanogaster*. Genetics **39**: 197-209.

O'BRIEN, S.J., 1994. A role for molecular genetics in biological conservation. Proc. Natl. Acad. Sci. USA **91**: 5748-5755.

O'GRADY, P.M., 1999. Reevaluation of phylogeny in the *Drosophila obscura* species group. Mol. Phylogenet. Evol. **12**: 124-139.

O'GRADY, P.M. and M.G. KIDWELL, 2002. Phylogeny of the subgenus Sophophora (Diptera: drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. Mol. Phylo. Evol. **22**: 442-453.

ORENGO, D.J. and A. PREVOSTI, 2002. Relationship between chromosomal polymorphism and wing size in a natural population of *Drosophila subobscura*. Genetica **115**: 311-318.

ORR, H.A., 1987. Genetics of male and female sterility in hybrids of *Drosophila pseudoobscura* and *D. persimilis*. Genetics **116**: 555-563.

ORR, H.A., 1989. Genetic of sterility in hybrids between two subspecies of Drosophila. Evolution **43**: 180-189.

ORR, H.A., L.D. MADDEN, J.A. COYNE, R. GOODWIN and R.S. HAWLEY, 1997. The developmental genetics of hybrid inviablity: a mitotic defect in Drosophila hybrids. Genetics **145**: 1031-1040.

OHTA, T., 1972a. Evolutionary rate of cistrons and DNA divergence. J. Mol. Evol. **1**: 150-157.

OHTA, T., 1972b. Population size and rate of evolution. J. Mol. Evol. **1**: 305-314.

OHTa, T., 1973. Slightly deleterious mutant substitutions in evolution. Nature **246**: 96-98.

OHTA, T., 1974. Mutational pressure as the main cause of evolution and polymorphisms. Nature **252**: 351-354.

OHTA, T., 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23**: 263-286.

OHTA, T., 1993. An examination of the generation-time effect on molecular evolution. Proc. Natl. Acad. Sci. USA **90**: 10676-10680.

OHTA, T., 1994. Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics **136**: 1331-1337.

OHTA, T., 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40**: 56-63.

PAGE, R.D.M., 2000. TreeView v1.6.6. Web http://taxonomy.zoology.gla.ac.uk

PAPACEIT, M. and A. PREVOSTI, 1989. Differences in chromosome A arrangement between *Drosophila madeirensis* and *D. subobscura*. Experientia **45**: 310-312.

PAPACEIT, M. and A. PREVOSTI, 1991. A photographic map of *Drosophila madeirensis* polytene chromosomes. J. Hered. **82**: 471-478.

PAPACEIT, M., J. SAN ANTONIO and A. PREVOSTI, 1991. Genetic analysis of extra sex combs in the hybrids between *Drosophila subobscura* and *D. madeirensis*. Genetica **84**: 107-114.

PAYSEUR, B.A. and M.W. NACHMAN, 2002. Natural selection at linked sites in humans. Gene **300**: 31-42.

PEIXOTO, A.A. and L.B. KLACZKO, 1991. Linkage disequilibrium analysis of chromosomal inversion polymorphisms of *Drosophila*. Genetics **129**: 773-777.

PÉLANDAKIS, M. and M. SOLIGNAC, 1993. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. J. Mol. Evol. **37**: 525-543.

PENALVA, L.O.F., H. SAKAMOTO, A. NAVARRO-Sabaté, E. SAKASHITA, B. GRANADINO, C. SEGARRA and L. SÁNCHEZ, 1996. Regulation of the gene *Sex-lethal*: a comparative analysis of *Drosophila melanogaster* and *D. subobscura*. Genetics **144**: 1653-1664.

PEREZ, D.E., C.-I. WU, N.A. JOHNSON and M.-L. WU, 1993. Genetics of reproductive isolation in the *Drosophila simulans* clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, *Odysseus (Ods)*. Genetics **134**: 261-275.

PEREZ, J.A., A. MUNTÉ, J. ROZAS, C. SEGARRA and M. AGUADÉ, 2003. Nucleotide polymorphism in the *RpII215* gene region of the insular species *Drosophila guanche*: reduced efficacy of weak selection on synonymous variation. Mol. Biol. Evol. **20**: 1867-1875.

PETTENATI, M.J., P.N. RAO, M.C. PHELAN, F. GRASS, K.W. RAO, P. COSPER, A.J. CARROLL, F. ELDER, J.L. SMITH, M.D. HIGGINS *et al.*, 1995. Paracentric inversions in humans: a review of 446 paracentric inversions with presentation of 120 new cases. A. J. Med. Genet. **16**: 171-187.

PHILIP, U., J.M. RENDEL, H. SPURWAY and J.B.S. HALDANE, 1944. Genetics and karyology of *Drosophila subobscura*. Nature **154**: 260-262.

PINTO, F.M., A. BREHM, M. HERNÁNDEZ, J.M. LARRUGA, A.M. GONZÁLEZ and V.M. CABRERA, 1997. Population genetic structure and colonization sequence of *Drosophila subobscura* in the Canaries and Madeira Atlantic Islands as inferred by autosomal, sex-linked and mtDNA traits. J. Hered. **88**: 108-114.

POMINI, F.P., 1940. Contributi alla conoscenza delle *Drosophila* (Diptera, Acalyptera) europee. I. Descrizione di alcune specie rideribili al grouppo *obscura*. Boll. Ist. Ent. Univ. Bologna **12**: 145-164.

POPADIC, A. and W. ANDERSON, 1995. Evidence of gene conversion in the Amylase multigene family of *Drosophila pseudoobscura*. Mol. Biol. Evol. **12**: 564-572.

POWELL, J.R., 1997. *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.

PRAKASH, S. and R.C. LEWONTIN, 1968. A molecular approach to the study of genetic heterozigosity in natural populations. III. Direct evidence of coadaptation in gene arrangements of *Drosophila*. Proc. Natl. Acad. Sci. USA **59**: 398-405.

PRESGRAVES, D.C., L. BALAGOPALAN, S.M. ABMAYR and H.A. ORR, 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. Nature **423**: 715-719.

PREVOSTI, A., 1954. Variacion geografica de varios caracteres cuantitativos en poblaciones catalanas de *Drosophila subobscura*. Genet. Iberica **6**: 33-68.

PREVOSTI, A., 1955. Variacion geografica de caracteros cuantitativos en poblaciones britanicas de *Drosophila subobscura*. Genet. Iberica **7**: 3-44.

PREVOSTI, A., 1967. Inversion heterozygosity and selection for wing length in *Drosophila subobscura*. Genet. Research **10**: 81-93.

PREVOSTI, A., G. RIBO, L. SERRA, M. AGUADÉ, J. BALANYA, M. MONCLUS and F. MESTRES, 1988. Colonization of America by *Drosophila subobscura*: Experiments in natural populations that support the adaptive role of chromosomal-inversion polymorphism. Proc. Natl. Acad. Sci. USA **85**: 5597-5600.

PRICE, C.S., 1997. Conspecific sperm precedence in Drosophila. Nature **388**: 663-666.

PRICE, C.S., K.A. DYER and J.A. COYNE, 1999. Sperm competition between Drosophila males involve both displacement and incapacitation. Nature **400**: 449-452.

PRITCHARD, J.K. and S.W. SCHAEFFER, 1997. Polymorphism and divergence at a *Drosophila* pseudogene locus. Genetics **147**: 199-208.

PRZEWORSKI, M., R.R. HUDSON and A. DI RIENZO, 2000. Adjusting the focus on human variation. Trends Genet. **16**: 296-302.

PYLKOV, K.V., L.A. ZHIVOTOVSKY and M.W. FELDMAN, 1998. Migration versus mutation in the evolution of recombination under multilocus selection. Genet. Res. **71**: 247-256.

QUINTANA, A. and A. PREVOSTI, 1991. Genetic and environmental factors in the resistence of *Drosophila subobscura* adults to high temperature shock. III. Chromosomal-inversion and enzymatic polymorphism variation in lines selected for heat shock resistence. Genetica **84**: 165-170.

RAMOS-ONSINS, S., C. SEGARRA, J. ROZAS and M. AGUADÉ, 1998. Molecular and chromosomal phylogeny in the *obscura* of *Drosophila* inferred from sequences of the *rp49* gene region. Mol. Phylogenet. Evol. **9**: 33-41.

RAMOS-ONSINIS, S., B.E. STRANGER, T. MITCHELL-OLDS and M. AGUADÉ, 2004. Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. Genetics **166**: 373-388.

REMSEN, J. and R. DESALLE, 1998. Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. Mol. Phylogenet. Evol. **9**: 225-235.

REMSEN, J. and P. O'GRADY, 2002. Phylogeny of Drosophilidae (Diptera: Drosophilidae), with comments on combined analysis and character support. Mol. Phylogenet. Evol. **24**: 249-264.

RICE, W.R. and E. E. HOSTERT, 1993. Laboratory experiments on speciation: what have we learned in forty years? Evolution **47**: 1637-1653.

RICHARDS, S., Y. LIU, B.R. BETTENCOURT, P. HRADECKY, S. LETOVSKY, R. NIELSEN, K. THORNTON, M.J. HUBISZ, R. CHEN, R.P. MEISEL, O. COURONNE, S. HUA, M.A. SMITH, P. ZHANG, J. LIU, H.J. BUSSEMAKER, M.F. VAN BATENBURG, S.L. HOWELLS, S.E. SCHERER, E. SODERGREN, B.B. MATTHEWS, M.A. CROSBY, A.J. SCHROEDER, D. ORTIZ-BARRIENTOS, C.M. RIVES, M.L. METZKER, D.M. MUZNY, G. SCOTT, D. STEFFEN, D.A. WHEELER, K.C. WORLEY, P. HAVLAK, K.J. DURBIN, A. EGAN, R. GILL, J. HUME, M.B. MORGAN, G. MINER, C. HAMILTON, Y. HUANG, L. WALDRON, D. VERDUZCO, K.P. CLERC-BLANKENBURG, I. DUBCHAK, M. A.F. NOOR, W. ANDERSON, K.P. WHITE, A.G. CLARK, S.W. SCHAEFFER, W. GELBART, G.M. WEINSTOCK and R.A. GIBBS, 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. Genome Research **15**: 1-18.

RIESEBERG, L.H., J. WHITTON and K. GARDNER, 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics **152**: 713-727.

ROBERTS, P.A., 1976. The genetics of chromosomal aberration, pp. 67-184 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER and E. NOVITSKI, Academic Press, London.

RODRIGUEZ, C., J.J. FANARA and E. HASSON, 1999. Inversion polymorphism, longevity and body size in a natural population of *Drosophila buzzatii*. Evolution **53**: 612-620.

RODRIGUEZ-TRELLES, F., G. ALVAREZ and C. ZAPATA, 1996. Time-series analysis of seasonal changes of the O inversion polymorphism of *Drosophila subobscura*. Genetics **142**: 179-187.

RODRIGUEZ-TRELLES, F., and M. RODRIGUEZ, 1998. Rapid micro-evolution and loss of chromosomal diversity in *Drosophila* in response to climate warming. Evol. Ecol. **12**: 829-838.

ROGERS, A.R. and H.C. HARPENDING, 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9**: 552-569.

ROGERS, A.R., 1995. Genetic evidence for a Pleistocene population explosion. Evolution **49**: 608-615.

ROWLEY, J., 1998. The critical role of chromosome translocations in human leukemias. Annu. Rev. Genet. **32**: 495-519.

ROZAS, J. and M. AGUADÉ, 1990. Evidence of extensive genetic exchange in the *rp49* region among polymorphic chromosome inversions in *Drosophila subobscura*. Genetics **126**: 417-426.

ROZAS, J. and M. AGUADÉ, 1991a. Restriction-map variation in the *rp49* region of *Drosophila subobscura* from the Canary Islands. Mol. Biol. Evol. **8**: 202-211.

ROZAS, J. and M. AGUADÉ, 1991b. Using restriction-map analysis to characterize the colonization process of *Drosophila subobscura* on the American continent. I. *rp49* region. Mol. Biol. Evol. **8**: 447-457.

ROZAS, J. and M. AGUADÉ, 1993. Transfer of genetic information in the *rp49* region of *Drosophila subobscura* between different chromosomal gene arrangements. Proc. Natl. Acad. Sci. USA **90**: 8083-8087.

ROZAS, J. and M. AGUADÉ, 1994. Gene conversion is involved in the transfer of genetic information between naturally ocurring inversions of *Drosophila*. Proc. Natl. Acad. Sci. USA **91**: 11517-11521.

ROZAS, J., C. SEGARRA, C. ZAPATA and G. ALVAREZ, 1995. Nucleotide polymorphism at the *rp49* region of *Drosophila subobscura*: Lack of geographic subdivision within chromosomal arrangements in Europe. J. Evol. Biol. **8:** 355-367.

ROZAS, J., C. SEGARRA, G. RIBÓ and M. AGUADÉ, 1999. Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. Genetics **151**: 189-202.

ROZAS, J., M. GULLAUD, G. BLANDIN and M. AGUADÉ, 2001. DNA variation at the *rp49* region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype strcuture. Genetics **158**: 1147-1155.

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

RUSSO, C.A., N. TAKEZAKI and M. NEI, 1995. Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. **12**: 391-404.

RUTTKAY, H., M. SOLIGNAC and D. SPERLICH, 1992. Nuclear and mitochondrial ribosomal RNA variability in the *obscura* group of *Drosophila*. Genetica **85**: 131-138.

RYCHLIK, W., 1992. Oligo v.4.1. Primer analysis software. Nat. Biosc. Inc. Plymouth.

SAITOU, N. and M. NEI, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406-425.

SAMBROOK, J. and D. RUSSELL., 2001. *Molecular Cloning: A laboratory manual*, 3rd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SANTOS, M., P. FERNADEZ IRIARTE, W. CESPEDES, J. BALANYA, A. FONTDEVILA and L. SERRA, 2004. Swift laboratory thermal evolution of wing shape (but not size) in *Drosophila subobscura* and its relationship with chromosomal inversion polymorphism. J. Evol. Biol. **17**: 841-855.

SCHAEFFER, S.W., C.F. AQUADRO and C.H. LANGLEY, 1988. Restriction map variation in the *Notch* region of *Drosophila melanogaster*. Mol. Biol. Evol. **5**: 30-40.

SCHAEFFER, S.W., P. GOETTING-MINESKY, M. KOVACEVIC, J.R. PEOPLES, J.L. GRAYBILL, J.M. MILLER, K. KIM, J.G. NELSON and W.W. ANDERSON, 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. Proc. Natl. Acad. Sci. USA **100**: 8319-8324.

SEGARRA, C. and M. AGUADÉ, 1992. Molecular organization of the X chromosome in different species of the *obscura* group of *Drosophila*. Genetics **130**: 513-521.

SERRA, L., G. PEGUEROLES and F. MESTRES, 1987. Capacity of dispersal of a colonizing species: *Drosophila subobscura*. Genetica **73**: 223-235.

SHELDAHL, L.A., D.M. WEINRICH and D.M. RAND, 2003. Recombination, dominance and selection on amino acid polymorphism in the Drosophila genome: contrasting patterns on the *X* and fourth chromosomes. Genetics **165**: 1195-1208.

SINGH, S. and B.N SINGH, 2001. *Drosophila bipectinata* species complex. Indian J. Exp. Biol. **39**: 835-844.

SMART, J., 1945. *Drosophila subobscura* Collin: Descriptive note on the species with comments on its nomenclatorial status (Diptera). Proc. R. Ent. Soc. London **14**: 53-56.

SNOOK, R.R., T.A. MARKOW and T.L. KARR, 1994. Functional non-equivalence of sperm in *Drosophila pseudoobscura*. Proc. Natl. Acad. Sci. USA **91**: 11222-11226.

SOKOLOV, N.N. and N.P. DUBININ, 1941. Permanent heterozygosity in *Drosophila*. Drosophila Information Service **15**: 39-40.

SOLE, E., F. MESTRES, J. BALANYA, C. ARENAS and L. SERRA, 2000. colonization of America by *Drosophila subobscura*: spatial and temporal lethal-gene allelism. Hereditas **133**: 75-72.

SOLE, E., J. BALANYA, D. SPERLICH and L. SERRA, 2002. Long-term changes in the chromosomal inversion polymorphism of *Drosophila subobscura*. I. Mediterranean populations from southwestern Europe. Evolution **56**: 830-835.

SPERLICH, D. and H. FEUERBACH-MRAVLAG, 1974. Epistatic gene interaction, crossing over, and linked and unliked inversions in *Drosophila subobscura*. Evolution **28**: 67-75.

SPERLICH, D. and P. PFRIEM, 1986. Chromosomal polymorphism in natural and experimental populations, pp. 257-309 in *The genetics and biology of Drosophila*, edited by M. ASHBURNER, H.L. CARSON and J.N. THOMPSON JR, Academic Press, New York.

SPIELMAN, D., B.W. BROOK and R. FRANKHAM, 2004. Most species are not driven to extinction before genetic factors impact them. Proc. Natl. Acad. Sci. USA **101**: 15261-15264.

SPURWAY, H., 1945. The genetics and cytology of *Drosophila subobscura*. I. Element A. Sex-linked mutants and their standard order. J. Genet. **46**: 268-286.

SPURWAY, H. and U. PHILIP, 1952. Genetics and cytology of *Drosophila subobscura* VII. Abnormal gene arrangements in element A. J. Genet. **51**: 198-215.

STADLER, L.J., 1932. On the genetic nature of induced mutations. Proc. 6th Int. Congr. Genet. **1**: 274-294.

STADLER, T., K. ROSELIUS, and W. STEPHAN, 2005. Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. Evolution **59**: 1268-1279.

STALKER, H.D., 1976. Chromosome studies in wild populations of *Drosophila melanogaster*. Genetics **82**: 323-347.

STEPHAN, W., 1994. Effects of genetic recombination and population subdivision on nucleotide sequence variation in *Drosophila annanasae*, pp. 56-66 in *Non-neutral evolution. Theories and molecular data*, edited by B. GOLDING, Chapman and Hall, London.

STEPHAN, W., 1995. An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorhism data. Mol. Biol. Evol. **12**: 959-962.

STEPHAN, W. and C.H. LANGLEY, 1998. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. Genetics **150**: 1585-1593.

STURTEVANT, A.H., 1917. Genetic factors affecting the strenght of linkage in *Drosophila*. Proc. Natl. Acad. Sci. USA **3**: 555.

STURTEVANT, A.H. and G.W. BEADLE, 1936. The raltions of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. Genetics **21**: 544-604.

STURTEVANT, A.H. and T. DOBZHANSKY, 1936. Inversions in the third chromosome of wild race of *Drosophila pseudoobscura*, and their use in the study of the history of the species. Proc. Natl. Acad. Sci. USA **22**: 448-450.

STURTEVANT, A.H., 1939. On the subdivision of the genus *Drosophila*. Proc. Natl. Acad. Sci. USA **25**: 137-141.

STURTEVANT, A.H., 1942. The classification of the genus *Drosophila* with the description of nine new species. Univ. Texas Publ. **4213**: 5-51.

SUEOKA, N., 1962. On the genetic basis of variation and heterogeneity of DNA base compositions. Proc. Natl. Acad. Sci. USA **48**: 582-592.

SUNDSTROM, H., M.T. WEBSTER and H. ELLEGREN, 2004. Reduced variation on the chicken Z chromosome. Genetics **167**: 377-385.

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.

TAKAHATA, N., K. ISHII and H. MATSUDA, 1975. Effect of temporal fluctuation of selection coefficient on gene frequency in a population. Proc. Natl. Acad. Sci. USA **72**: 45411-4545.

TAKAHATA, N., 1987. On the overdispersed molecular clock. Genetics **116**: 169-179.

TAMURA, K., S. SUBRAMANIAN and S. KUMAR, 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. Mol. Biol. Evol. **21**: 36-44.

TATARENKOV, A., J. KWIATOWSKI, D. SKARECKY, E. BARRIO and F.J. AYALA, 1999. On the evolution of dopa decarboxylase (*Ddc*) and *Drosophila* systematics. J. Mol. Evol. **48**: 445-462.

THOMAS, R.H. and J.A. HUNT, 1993. Phylogenetic relationships in *Drosophila*: A conflict between molecular and morphological data. Mol. Biol. Evol. **10**: 362-374.

THOMPSON, J.D., T.J. GIBBON, F. PLEWNIAK, F. JEANMOUGIN and D.G. HIGGINS, 1997. The ClustalX windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. Nucleic Acids Research **24**: 4876-4882.

THROCKMORTON, L.H., 1975. The phylogeny, ecology, and geography of *Drosophila*, pp. 421-469 in *Handbook of Genetics*, edited by R.C. KING, Plenum Press, New York.

TING, C.-T., S.-C. TSAUR, M.-L. WU and C.-I. WU, 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282**: 1501-1504.

TING, C.-T., S.-C. TSAUR and C.-I. WU, 2000. The phylogeny of closely related species as revelead by the genealogy of a speciation gene, *Odysseus*. Proc. Natl. Acad. Sci. USA **97**: 5313-5316.

TSACAS, L., 1979. Contribution des donnes africaines a la comprehension de la biogeographie et l'evolution de sous-genere *Drosophila* (*Sophophora*) Surtevant (Diptera, Drosophilidae). C. R. Soc. Biogeogr. **48**: 29-51.

TSACAS, L., 1980. Les groupes d'especes du sous-genere *Sophophora* Sturtevant (Diptera, Drosophilidae, *Drosophila*) et le role du fonctionnement des genitalia males dans la definition des taxons supraspecifiques. Bull. Soc. Zool. Fr. **105**: 529-543.

TSACAS, L., M.L. CARIOU and D. LACHAISE, 1985. Le groupe *Drosophila obscura* en Afrique de l'Est, description de trois nouvelles espèces (Diptera: Drosophilidae). Ann. Soc. Entr. Fr. (NS) **21**: 413-424.

TYLER, R.H., H. BRAR, M. SINGH, A. LATORRE, J.L. GRAVES, L.D. MUELLER, M.R. ROSE and F.J. AYALA, 1993. The effect of superoxide dismutase alleles on aging in *Drosophila*. Genetica **91**: 143-149.

VIGNEAULT, G. and E. ZOUROS, 1986. The genetics of asymmetrical male sterility in *Drosophila mojavensis* and *D. arizonensis* hybrids. Interactions between the Y-chromosome and the autosomes. Evolution **40**: 1160-1170.

WAKELEY, J. and J. HEY, 1997. Estimating ancestral population parameters. Genetics **145**: 847-855.

WALLACE, B. and T. DOBZHANSKY, 1946. Experiments on sexual isolation in *Drosophila*. VIII. Influence of light on the mating behavior of *Drosophila subobscura*, *Drosophila persimilis* and *Drosophila pseudoobscura*. Proc. Natl. Acad. Sci. USA **32**: 226-234.

WANG, R.L., J. WAKELEY and J. HEY, 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics **147**: 1091-1106.

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

WEEKS, A., S.W. MCKECHNIE and A.A. HOFFMANN, 2002. Dissecting adaptive clinal variation: markers, inversions and size/stress associations in *Drosophila melanogaster* from a central field population. Ecol. Lett. **5**: 756-763.

WESLEY, C.S. and W.F. EANES, 1994. Isolation and analysis of the breakpoint sequences of chromosome inversion *In(3L)Payne* in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **91**: 3132-3136.

WHEELER, M.R., 1981. The Drosophilidae: A taxonomic overview, pp. 1-97 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER, H.L. CARSON and J.N. THOMPSON JR., Academic Press, New York.

WHEELER, M.R., 1986. Additions to the catalog of the world's Drosophilidae, pp. 395-409 in *The Genetics and Biology of Drosophila*, edited by M. ASHBURNER, H.L. CARSON and J.N. THOMPSON JR., Academic Press, New York.

WIEHE, T.H.E. and W. STEPHAN, 1993. Analysis of a genetic hitchhiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. **10**: 842-854.

WRIGHT, S., 1931. Evolution in Mendelian populations. Genetics **16**: 97-159.

WRIGHT, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the Sixth International Congress in Genetics, vol **1**: 356-366.

WRIGHT, S., 1951. The gentical structure of populations. Ann. Eugenics **15**: 323-354.

WRIGHT, S.I. and B.S. GAUT, 2004. Molecular population genetics and search for adaptive evolution in plants. Mol. Biol. Evol. **22**: 506-519.

WU, C.-I. and M. F. PALOPOLI, 1994. Genetic of post mating reproductive isolation in animals. Ann. Rev. Genetics **28**: 283-308.

YADAV, J. and B.N. SINGH, 2003. Population genetics of *Drosophila ananassae*: inversion polymorphism and body size in Indian geographical populations. J. Zool. Syst. Evol. Res. **41**: 217-226.

ZAPATA, C. and G. ÁLVAREZ, 1987. Gametic disequilibrium in populations of *Drosophila subobscura*: a reviw of experimental evidence. Genét. Ibér. **39**: 593-616.

ZAPATA, C. and G. ÁLVAREZ, 1992. The detection of gametic disequilibrium between allozyme loci in natural populations of *Drosophila*. Evolution **46**: 1900-1917.

ZAPATA, C. and G. ÁLVAREZ, 1993. On the detection of nonrandom associations between DNA polymorphism in natural populations. Mol. Biol. Evol. **10**: 823-841.

ZAPATA, C., G. ALVAREZ, F. RODRIGUEZ-TRELLES and X. MASIDE, 2000. a long-term study on seasonal changes of gametic disequilibrium between allozymes and inversions in *Drosophila subobscura*. Evolution **54**: 1673-1679.

ZIVANOVIC, G., M. ANDJELKOVIC and D. MARINKOVIC, 2000. Genetic load and coadaptation of chromosomal inversions. II. O-chromosme in *Drosophila subobscura* populations. Hereditas **133**: 105-113.

ZIVANOVIC, G. and D. MARINKOVIC, 2003. Viabilities of originally natural O-chromosomal inversion homo and heterokaryotypes in *Drosophila subobscura*. Hereditas **139**: 128-142.

ZOUROS, E., 1981. The chromosomal basis of sexual isolation in two sibling species of *Drosophila*: *D. arizonensis* and *D. mojavensis*. Genetics **97**: 703-718.

ZWICK, M.E., D.J. CUTTER and A. CHAKRAVARTI, 2000. Patterns of genetic variation in mendelian and complex traits. Annu. Rev. Genomics Hum. Genet. **1**: 387-407.

Appendix A

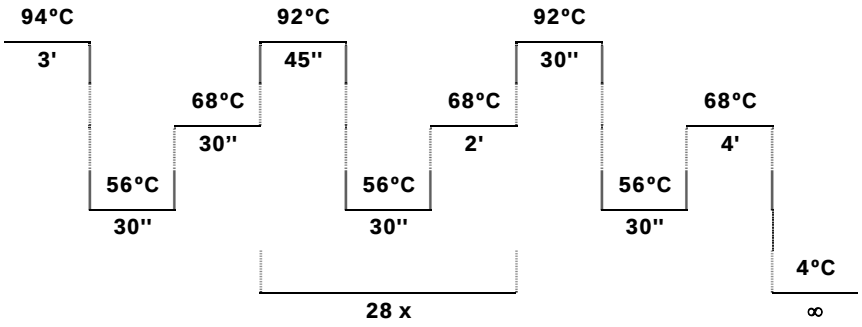
Amplification primers used for the five genomic regions studied

Primer	Sequence (5'→3')	N° nucleotides	Tm	% GC
	P236 region			
U482	CTTCAATGCGGTACACACAG	20	61.8	50
U433	TACAACACAGGAGCACTC	18	54.2	50
U613	GTGTAGCGATGGCAACTCTT	20	61.8	50
	P150 region			
F1	GTGGACACAACAGGCATCAGA	21	66.3	52.4
R1	TGCTACCACAAAAGGGCTTAC	21	62.8	47.6
F2	TCACAGCCCAAACCATACACC	21	66.8	52.4
R2	GTGGTGGGCTGGCTGTGAAAA	21	71.4	57.1
R4	CAAAACCCAACACCAATATGA	21	62.0	38.1
	***Sex-lethal* region**			
SFX	GCAGCGGTGGGCGTGGATTT	20	75.3	65.0
Sxl1484	TGGCTCTGAATAAGGCGTACA	21	64.4	47.6
	P125 region			
U2	CTACTTATTCTGGGCTCATTC	21	57.0	42.8
L1	TAAAAGGCAAACGGCATTCTG	21	65.7	47.6
UmadII	GTATGCTTCTCCCACAGTGT	20	59.2	50.0
L1771	TGGATAAAACGCAGAGATAG	20	57.0	40.0
U249	ATGATGACGCTGATGAAGAC	20	60.3	45.0
U1	TGCTTCTCCCACAGTGTTTCA	21	65.8	47.6
	P275 region			
F1	GTAGGGTGTGTGTGCTTTCAT	21	61.5	47.6
R1	GCCATATTCAATTTACCACAT	21	57.7	33.3
R2	AATCAATGGACAACACGAAGT	21	60.8	33.3

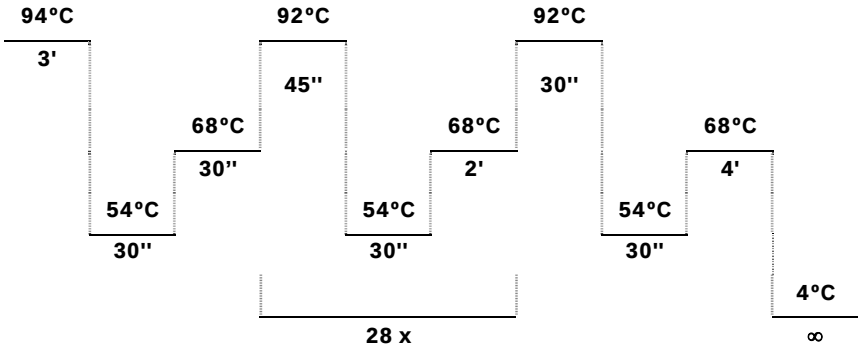
Appendix B

PCR amplification conditions used for the genomic regions studied

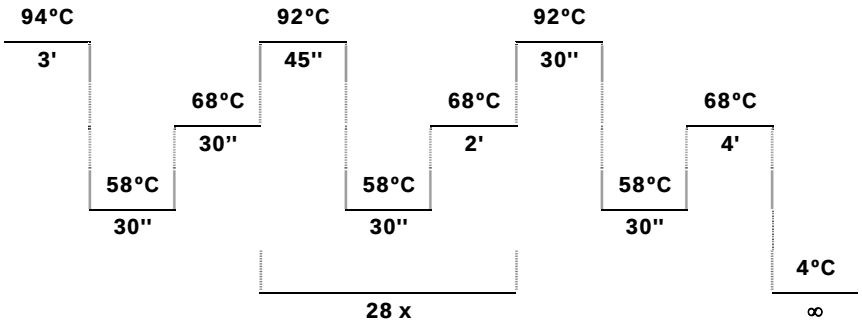
- P236 region



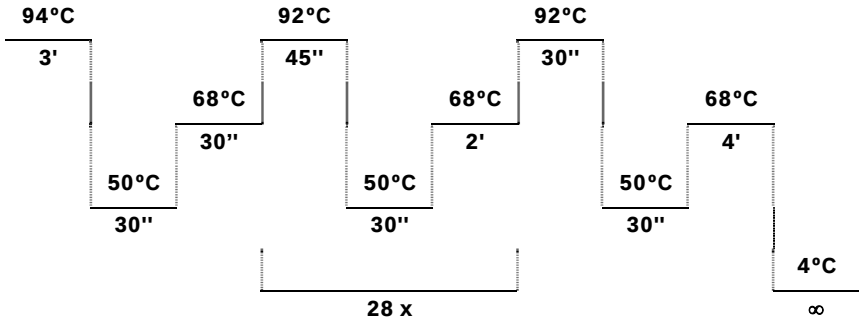
- P150 region



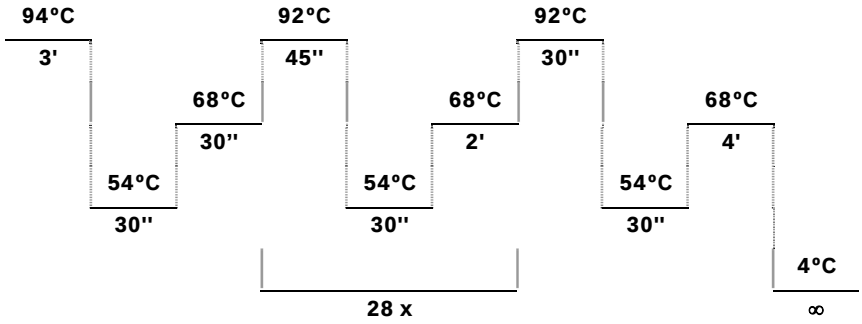
- *Sex-lethal* region



- P125 region



- P275 region



Appendix C

Sequencing primers used for the five genomic regions studied

Primer	Sequence (5'→3')	N° nucleotides	% GC
	P236 region		
U613	GTGTAGCGATGGCAACTCTT	20	45.0
U482	CTTCAATGCGGTACACACAG	**20**	**50.0**
U628	CTCTCGCTCCTTCTATGT	**18**	**50.0**
U433	TACAACACAGGACACTC	**17**	**47.1**
U449	GTTTGTGCCACTCGTTTC	**18**	**50.0**
U404	CGAAACGAGTGGCACAAA	**18**	**50.0**
U624	CTGTTTTCCCCCGCTGTG	**18**	**61.1**
U577	TTTACGGCTCAACTATTT	**18**	**33.3**
	P150 region		
553	GTCTCTCTCCCTCTCTAT	**18**	**50.0**
558	CATACCCCACCCGACATA	**18**	**55.5**
F1	GTGGACACAACAGGCATCAGA	**21**	**52.4**
R1	TGCTACCACAAAAGGGCTTAC	**21**	**47.6**
R4	CAAAACCCAACACCAATATGA	**21**	**38.1**
666R	GACTTGGACTTGGCTCTC	**18**	**55.5**
1281R	GTGATTTTGCCTTTGTTT	**18**	**33.3**
F2	TCACAGCCCAAACCATACACC	**21**	**52.4**
R2	GTGGTGGGCTGGCTGTGAAAA	**21**	**57.1**
	***Sex-lethal* region**		
Sxfw502	AACAACAAACCCTAAACA	**18**	**33.3**
Sxfw1137	ACAGACACGCAAACAGAC	**18**	**50.0**
Ksxl293	CGTTTCATTGCGAGACAG	**18**	**50.0**
Sxl1451	TATCCTGGGGCAAGTAGT	**18**	**50.0**
Sxl142	CGTTTCCGCTTCCGTTTC	**18**	**55.5**
Ksxl591	GATTGAGGTTCGGGTTGA	**18**	**50.0**
SXF	GCAGCGGTGGGCGTGGATTT	**20**	**65.0**
Ksxl343	TAATCAAAAGCGTGTGTC	**18**	**38.9**
Sxl342M	GGACACACGCTTTTGATT	**18**	**44.4**
Sxl307M	AACCAAACCCCAAGACAG	**18**	**50.0**

	P125 region		
556F	CACTGCCCACCACCCATC	**18**	**66.7**
288R	TAATGCCCTCCCCCAAAC	**18**	**55.5**
462	TGAAAAAATGCCAGACAA	**18**	**33.3**
559R	TGCTTCTTCCCACAGTGTT	**19**	**47.3**
374F	AGAGAGAGCGGTAGAACA	**18**	**50.0**
L1	TAAAAGGCAAACGGCATTCTG	**21**	**47.6**
U2	CTACTTATTCTGGGCTCATTC	**21**	**42.8**
X4	TCACCTCCCTTTTTCTTC	**18**	**44.4**
UmadII	GTATGCTTCTCCCACAGTGT	**20**	**50.0**
L1771	TGGATAAAACGCAGAGATAG	**20**	**40.0**
598	AAAAAAAGGCGAAGAAG	**17**	**35.3**
413	CATTTTGCTTGTGGTTTA	**18**	**33.3**
559	GATGGGTGGTGGGCAGTG	**18**	**66.7**
239R	GTATTTATGGCACTGTGT	**18**	**38.9**
377R	TAATGCCCTCCCCCAAAC	**18**	**55.5**
U249	ATGATGACGCTGATGAAGAC	**20**	**45.0**
X2	ACACAGTGCCATAAATAC	**18**	**38.9**
550	GAAGCCCCACAAGGAATG	**18**	**55.5**
LmadII	CCCAGTGAATGCGGATAGTT	**20**	**50.0**
	P275 region		
X1	CGTAAAATGGCTGGAATG	**18**	**44.4**
X2	CCATTCCAGCCATTCTAC	**18**	**50.0**
329F	CGAGTTGTTTTGGTTCAC	**18**	**44.4**
466F	CTCCATTCCAGCCTTCT	**17**	**52.9**
324R	GAGAGCGGCAGAGAGAAA	**18**	**55.5**
469R	AATGGCTGGATGGAGTG	**17**	**52.9**
F2	GGGTGTGTGTGCTTTCATTTG	**21**	**47.6**
F1	GTAGGGTGTGTGTGCTTTCAT	**21**	**47.6**
R1	GCCATATTCAATTTACCACAT	**21**	**33.3**
R2	AATCAATGGCAACACGAAGT	**21**	**33.3**
F3	AATCGGCAAGTTCCAGGTACA	**21**	**47.6**
506	CGGAATGGCGTTTTTAGT	**18**	**44.4**
445	GTGCGGATAACGGTTGTA	**18**	**50.0**
183	GGACAAATAACCGCTAAA	**18**	**38.9**