JOURNAL OF Evolutionary Biology

doi: 10.1111/j.1420-9101.2012.02467.x

Multilocus analysis of nucleotide variation in *Drosophila madeirensis*, an endemic species of the Laurisilva forest in Madeira

M. KHADEM*, A. MUNTÉ†, R. CAMACHO*, M. AGUADÉ‡§ & C. SEGARRA‡§

*ISOPlexis & Life Sciences, Universidade da Madeira, Funchal, Portugal
†Parc Científic de Barcelona, Barcelona, Spain
‡Facultat de Biologia, Departament de Genètica, Universitat de Barcelona, Barcelona, Spain
§Institut de Recerca de la Biodiversitat (IrBio), Universitat de Barcelona, Barcelona, Spain

Keywords:

DNA divergence;
DNA polymorphism;
introgression;
molecular evolution;
speciation.

Abstract

Drosophila madeirensis is an endemic species of Madeira that inhabits the island Laurisilva forest. Nucleotide variation in *D. madeirensis* is analysed in six genomic regions and compared to that previously reported for the same regions in *Drosophila subobscura*, an abundant species in the Palearctic region that is closely related to *D. madeirensis*. The gene regions analysed are distributed along the O_3 inversion. The O_3 arrangement is monomorphic in *D. madeirensis*, and it was present in ancestral populations of *D. subobscura* but went extinct in this species after the origin of the derived O_{ST} and O_{3+4} arrangements. Levels of nucleotide polymorphism in *D. madeirensis* are similar to those present in the O_{ST} and O_{3+4} arrangements of *D. subobscura*, and the frequency spectrum is skewed towards rare variants. Purifying selection against deleterious nonsynonymous mutations is less effective in *D. madeirensis*. Although *D. madeirensis* and *D. subobscura* coexist at present in Madeira, no clear evidence of introgression was detected in the studied regions.

Introduction

Drosophila species have been used as a model system in evolutionary studies both in the laboratory and in nature (reviewed in Powell, 1997). Standing genetic variation has been characterized at different levels in natural populations of several *Drosophila* species. However, the cosmopolitan species *Drosophila melanogaster* and *Drosophila simulans* are best characterized at the molecular level. Since the work of Kreitman (1983) in *D. melanogaster*, a large number of population studies performed by direct DNA sequencing have been published. These studies, which at first were based on a single locus and later used multilocus approaches, are now being improved in genome-wide surveys (Begun *et al.*, 2007; Sackton *et al.*, 2009). In contrast to the rather large number of surveys in cosmopolitan and other widely distributed species, molecular population genetics studies in island-endemic *Drosophila* species are still scarce. Multilocus analyses have been reported in three island species – *Drosophila mauritiana*, *Drosophila sechellia* and *Drosophila santomea* – that together with *D. melanogaster*, *D. simulans* and *Drosophila yakuba* constitute the *melanogaster* group of the *Sophophora* subgenus. *Drosophila mauritiana* (endemic at the Mauritius Island) and *D. sechellia* (endemic at the Seychelles archipelago) are closely related to *D. simulans*, whereas *D. santomea* (endemic at the São Tomé Island) is a close relative of *D. yakuba*. Comparison of the level of nucleotide variation between each endemic species and its close widely distributed relative has yielded contradictory results. A strong reduction of variation has been described in *D. sechellia* relative to *D. simulans* (Kliman *et al.*, 2000; Legrand *et al.*, 2009). In contrast, *D. mauritiana* exhibits only slightly lower nucleotide diversity than *D. simulans* (Kliman *et al.*, 2000; McDermott & Kliman, 2008). *Drosophila santomea* also exhibits a lower level of variation than *D. yakuba* (Llopart *et al.*, 2005), with an up to two-fold reduction of silent nucleotide diversity for different X-linked loci (Bachtrog *et al.*, 2006). These contradictory results would indicate that not only the speciation process itself and the reduced distribution area but other factors

Correspondence: Carmen Segarra, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.
Tel.: 34 93 4021848; fax: 34 93 4034420;
e-mail: csegarra@ub.edu

might affect the level of variation in endemic species in relation to their close relatives.

Drosophila madeirensis is an endemic species of the Madeira Island (Portugal) located 32.75°N and 16.89°W in the Atlantic Ocean. The Madeira archipelago has a volcanic origin that has been dated to some 5–6 Mya (Galopin de Carvalho & Brandão, 1991). Natural populations of *D. madeirensis* inhabit the Laurisilva forest, which is a relict of the Pliocene subtropical forest. Nowadays, the Laurisilva forest from Madeira represents the largest surviving area of this vegetation type and it is included in the World Heritage List (IUCN 1999). A unique suite of plants and animals, including many endemic species such as *D. madeirensis*, are associated with the different Laurisilva forests. *Drosophila madeirensis* is thus a specialized species with a restricted distribution area. These characteristics might have affected its effective population size that can be expected to be relatively low.

Drosophila madeirensis is closely related to *Drosophila subobscura*, a widespread and abundant Palearctic species (Krimbas, 1993) that has recently colonized the American continent (Prevosti *et al.*, 1988). Natural populations of *D. subobscura* have been extensively surveyed in the last 50 years, and the adaptive character of the species' chromosomal polymorphism is well established. Chromosomal inversions present latitudinal clines in the same direction both in Europe and in America (Prevosti *et al.*, 1988), and their frequencies have undergone temporal changes in Europe that are consistent with a response to global warming (Balanyà *et al.*, 2004). *Drosophila subobscura* and *D. madeirensis* diverged very recently (some 0.6–1.0 Mya; Ramos-Onsins *et al.*, 1998) and can be considered a model species pair to study how differences in effective size and habitat preferences (specialized vs. more generalist species) can affect the level and pattern of nucleotide variation. Indeed, under strict neutrality, endemic species with small effective sizes are expected to harbour low levels of nucleotide variation (Kimura, 1983). However, the small effective size might also cause a relaxation of selection that would imply, according to the nearly neutral model, a longer sojourn time of mildly deleterious mutations and thus an increase in the level of variation (Ohta, 1992).

The species pair *D. madeirensis* – *D. subobscura* is also a good system to analyse the speciation process. Under the biological species concept, species are isolated systems that do not interbreed nor exchange genetic information with other systems (Dobzhansky, 1937; Mayr, 1942). The mechanisms involved in the isolation process, their relative importance and the temporal order in which they are established are important aspects in studies of speciation (Coyne & Orr, 2004). Nevertheless, genes can cross the species barriers when reproductive isolation is incomplete and the production of nonsterile hybrids is possible. Introgression is a common process in incipient species with overlapping distribution areas and has often been detected from variation at the mitochondrial and/or the nuclear genomes in closely related Drosophila species (Wang *et al.*, 1997; Llopart *et al.*, 2005; Bachtrog *et al.*, 2006; Rand *et al.*, 2006). The level of gene flow between not completely isolated species can also differ across the nuclear genome. Indeed, gene regions with restricted gene flow, as reflected by a low number of shared polymorphisms and a high number of fixed differences, can coexist with gene regions that exhibit extensive gene flow that results in a high number of shared polymorphisms and no fixed differences between species (Wang *et al.*, 1997). Some studies have revealed a good correspondence between genomic regions associated with reproductive isolation and regions with restricted gene flow (Ting *et al.*, 2000; Machado *et al.*, 2002). Hence, natural selection may prevent gene flow in those genes involved in sexual isolation and, thus, it contributes to the reinforcement of reproductive isolation.

Gene flow between incipient species can also be prevented in certain genomic regions by the presence of chromosomal inversions. Classical chromosomal speciation models based on the reduction of fitness in heterokaryotypes (White, 1978) have been recently extended to speciation models that focus on the suppression of recombination between different arrangements as a means to preclude gene flow in key genomic regions associated with species-specific adaptations or reproductive isolation (Noor *et al.*, 2001, 2007; Rieseberg, 2001; Navarro & Barton, 2003; Brown *et al.*, 2004).

Reproductive isolation between *D. madeirensis* and *D. subobscura* is not complete because fertile hybrids are obtained in laboratory conditions (Khadem & Krimbas, 1991; Papaceit *et al.*, 1991; Rego *et al.*, 2007). At present, both species coexist in the Madeira Island and thus they might exchange genes in nature via interspecific hybridization. However, *D. madeirensis* and *D. subobscura* differ by chromosomal inversions, either fixed between them or segregating as polymorphic in *D. subobscura*. These inversions might have prevented interspecific genetic exchange in the regions covered by the inversions in the interspecific hybrids (Khadem *et al.*, 2011).

Nucleotide variation was previously surveyed at the *rp49* gene region in a sample of *D. madeirensis* and *D. subobscura* from Madeira (Khadem *et al.*, 2001). No genetic differentiation between continental and island *D. subobscura* populations and similar levels of genetic diversity in *D. subobscura* and *D. madeirensis* were detected. The *rp49* gene region is located in the O chromosome (Muller's element E) of *D. madeirensis*. This species is monomorphic for the O_3 arrangement, which was present in ancestral populations of *D. subobscura* but went extinct in this species after the origin of its derived arrangements O_{ST} and O_{3+4} (Ramos-Onsins *et al.*, 1998). Current natural populations of *D. subobscura* in Madeira are almost monomorphic for the O_{3+4} arrangement (Prevosti, 1972; Larruga *et al.*, 1983). Therefore, interspecific hybrids would be O_{3+4}/O_3 heterokaryotypes and would form an inversion loop including almost

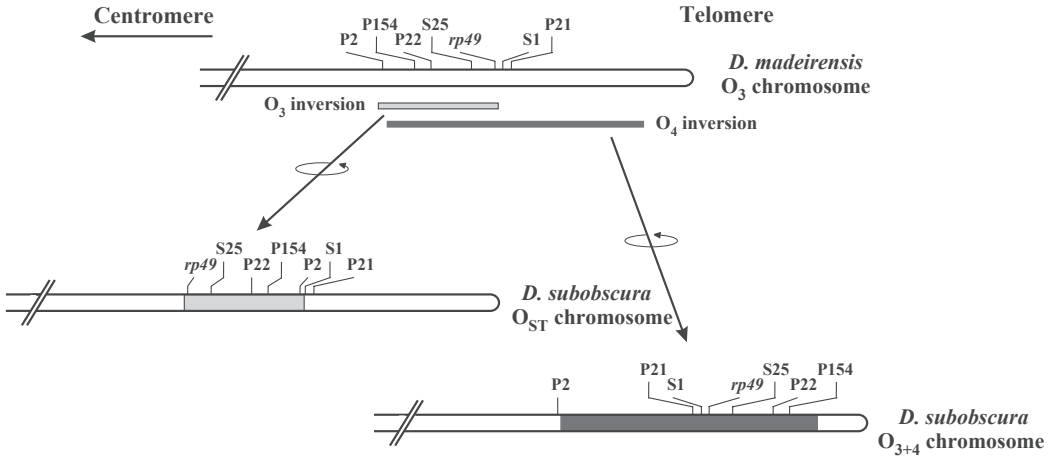

Fig. 1 Distribution of the genomic regions included in this study along the O chromosome of *Drosophila madeirensis* and *Drosophila subobscura* (O_{ST} or O_{3+4}). The effect of the O_3 (shading) and O_4 (solid) inversions on the location of the regions studied is also shown.

completely segment I of the O chromosome (sections 91–99). Here, the previous analysis of nucleotide variation at *rp49* is extended to six new genomic regions in a set of 14 lines of *D. madeirensis*. These six regions, like *rp49*, are associated with the O_3 inversion (Fig. 1), and their variation was previously analysed in a sample of 14 O_{ST} and 14 O_{3+4} chromosomes of *D. subobscura* (Munté *et al.*, 2005).

The main aim of the present multilocus study is to compare the level and pattern of nucleotide variation between an endemic and specialized species (*D. madeirensis*) and a species with a wide distribution in the Paleartic region (*D. subobscura*). The expected strong differences in effective population size between *D. madeirensis* and *D. subobscura* might be reflected in their level and pattern of nucleotide variation. In addition, the present study also aims to shed some light on the speciation process that gave rise to *D. madeirensis*. Indeed, introgression between species in the studied regions is expected to be low as they are associated with the inversion and would, thus, lie in the loop formed in putative interspecific hybrids. Therefore, the level and pattern of genetic variation in these regions would not be affected by introgression, unlike collinear regions, and would better reflect the speciation history of *D. madeirensis* and *D. subobscura*.

The results obtained show that: (1) the level of nucleotide variation in *D. madeirensis* is similar to that present in the O_{ST} and O_{3+4} arrangements of *D. subobscura*, (2) the pattern of variation in *D. madeirensis* is skewed towards polymorphisms with rare variants, (3) purifying selection against nonsynonymous mutations is stronger in *D. subobscura* than in *D. madeirensis* and iv) introgression at loci located along the O_3 inversion is negligible.

Materials and methods

Fly samples

Fourteen highly inbred lines of *D. madeirensis* were established after 12 generations of sib-mating from flies collected in Ribeiro Frio (Madeira, Portugal) in 2000. All lines were homozygous for the O_3 arrangement. The 14 O_{ST} and 14 O_{3+4} *D. subobscura* lines used in some analyses (Rozas *et al.*, 1995; Munté *et al.*, 2005) were obtained from a sample collected in El Pedroso (Galicia, Spain).

DNA extraction and sequencing

A modification of protocol 48 in Ashburner (1989) was used to extract genomic DNA from a single individual of the inbred lines. Six genomic regions (P2, P21, P22, P154, S1 and S25) were PCR-amplified in all lines. The primers and conditions used in the amplification process were the same as described in Munté *et al.* (2005). PCR products were purified with Qiaquick columns (Qiagen, Chatswort, CA, USA) and used as template for sequencing with internal primers designed at increasing intervals of ~300 nucleotides. Both strands were cycle-sequenced using the ABI Prism® BigDye™ Terminators 3.0 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) and run on an ABI Prism 3700 sequencer (Applied Biosystems). Sequences were assembled using the SeqMan program of the DNASTAR Lasergene software package (Burland, 2000), multiply aligned using Clustal W (Thompson *et al.*, 1994), and edited with the MacClade (version 4.0) program (Maddison & Maddison, 2000). Newly reported sequences have been deposited in the EMBL/GenBank Data Libraries under accession numbers HE653145-HE653228.

Data analysis

The level and pattern of nucleotide variation in *D. madeirensis* were analysed in the six genomic regions studied. DNA sequences from the six orthologous regions of *D. subobscura* previously reported in Munté *et al.* (2005) and the *rp49* sequences of *D. subobscura* and *D. madeirensis* (Rozas & Aguadé, 1994; Ramos-Onsins *et al.*, 1998; Khadem *et al.*, 2001; EMBL accession nos. X80076–X80109, Y09708 and J310269–J310306, respectively) were also used in most data analyses. The orthologous *Drosophila pseudoobscura* sequences used for interspecific comparisons were obtained from the Drosophila Genome Project database at the HGSC using the Mega BLAST search engine at NCBI. Multiply aligned sequences of the six newly reported regions were concatenated in the same order as they are in the O_3 arrangement of *D. madeirensis*, and the analysis was performed either for each genomic region independently or for the concatenated data set.

The level of nucleotide polymorphism in *D. madeirensis* was estimated as S (the number of polymorphic sites), π (nucleotide diversity; Nei, 1987) and θ (the heterozygosity per site expected at mutation–drift equilibrium; Watterson, 1975). Sites with insertion/deletion variants (indels) were not considered in the analyses. Interspecific analyses were performed independently between *D. madeirensis* and either O_{ST} or O_{3+4}, given that these gene arrangements are genetically differentiated in *D. subobscura* (Munté *et al.*, 2005). Genetic differentiation between species was estimated as D_{XY} and D_A (the average number of nucleotide substitutions per site and the number of net nucleotide substitutions per site, respectively; Nei, 1987). The statistical significance of genetic differentiation as measured by the K_S* statistic was determined by the permutation test (Hudson *et al.*, 1992a). F_{ST} or the proportion of nucleotide diversity attributed to variation between populations was estimated according to Hudson *et al.* (1992b). This parameter was used to estimate *Nm* (migration index) under the island model of population structure and assuming migration–drift equilibrium (Wright, 1951; Hudson *et al.*, 1992b). The hypergeometric distribution was applied, as described in Rozas & Aguadé (1994), to infer whether the observed number of shared polymorphisms between species (i.e. sites with the same polymorphic variants in *D. madeirensis* and *D. subobscura*) could be explained by recurrent mutations. The algorithm developed by Betran *et al.* (1997) to identify gene conversion tracts was used to detect putative introgressed haplotypes between *D. subobscura* and *D. madeirensis*. The DnaSP v5 (Librado & Rozas, 2009) and SITES (made kindly available by J. Hey at http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm) software packages were used for most DNA polymorphism analyses.

Neutrality tests (Hudson *et al.*, 1987; Tajima, 1989; Fu & Li, 1993) were performed for each region of *D. madeirensis*, and their statistical significance was determined by computer simulations using the coalescent algorithm without recombination. Multilocus neutrality tests were carried out with the HKA program (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm).

Scaled selection coefficients ($\gamma = N_e s$, where N_e is the effective population size and s is the selection coefficient) were estimated for synonymous and nonsynonymous polymorphic sites within *D. madeirensis* and within the *D. subobscura* O_{ST} and O_{3+4} chromosomal arrangements. We used a Poisson random field (PRF) approach (Sawyer & Hartl, 1992) based on the frequency distribution (fd) of polymorphic variants (Hartl *et al.*, 1994; Akashi & Schaeffer, 1997). The PRFML program, made kindly available by S. Sawyer at http://www.math.wustl.edu/~sawyer/, was used to estimate γ, its confidence interval and its associated P-value. The distribution of fitness effects on new nonsynonymous mutations was inferred according to Eyre-Walker & Keightley (2009) using the DFE-alpha server kindly available at http://liberty.cap.ed.ac.uk/~eang33/dfe-alpha-server.html. This maximum-likelihood approach uses information from the site frequency spectra at selected and neutral sites in a population sample. Nonsynonymous and synonymous sites in the coding regions were used as selected and neutral sites, respectively. The method also takes into account demographic events that can alter the frequency spectrum and allows the estimation of the fraction of adaptive amino acid mutations (α) when divergence data are available. Divergence from *D. pseudoobscura* was used with this purpose.

Gene genealogies were reconstructed by the neighbour-joining method (Saitou & Nei, 1987) as implemented in the MEGA v4.0 program (Tamura *et al.*, 2007). Genetic distances were corrected for multiple hits according to Jukes & Cantor (1969). Bootstrap values were obtained after 1000 replicates.

Speciation models

The method developed by Wakeley (1996a,b) was applied to contrast whether the observed pattern of variation was consistent with that expected under the isolation model. The ψ-test (Wakeley, 1996a) based on the variance of the number of nucleotide differences between species was performed with a C program kindly provided by J. Wakeley. The numbers of exclusive mutations (S_{X1}, S_{X2}), fixed differences between groups (S_F) and shared polymorphisms between groups (S_S) were used to estimate the population parameters θ_1 (O_{ST} or O_{3+4}), θ_2 (*D. madeirensis*), θ_A (ancestral species) and τ ($\tau = 2\mu t$, where μ is the mutation rate per sequence and t the time since the split) for each genomic region and for the concatenated data set as described in Wakeley & Hey (1997). In addition, the fit of the data to the isolation model was contrasted by a multilocus approach according to the test statistics χ^2 (Kliman *et al.*, 2000) and *wh*

(Wakeley & Hey, 1997; Wang *et al.*, 1997) as implemented in the WH program (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm). This model assumes that an ancestral population split into two descendent populations with constant effective sizes.

The isolation with migration model, as implemented in the IM program, was also used to contrast whether the data could be explained by isolation with some degree of directional gene flow between species (Nielsen & Wakeley, 2001; Hey & Nielsen, 2004). The IM program (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm) applies Markov chain Monte Carlo (MCMC) simulations to estimate the posterior probability distribution of model parameters from multiple unlinked neutral loci data sets assuming no recombination within loci. Thus, only the largest fragment for each region with no evidence for recombination according to the four gamete test (Hudson & Kaplan, 1985) was considered in this analysis. The basic procedure to obtain maximum-likelihood estimates and confidence intervals of model parameters was to start with a burn-in period of 100 000 steps and proceed for more than ten million steps after the burn-in period. We applied the Hasegawa–Kishino–Yano substitution model (HKY; Hasegawa *et al.*, 1985) and Metropolis coupling with six coupled Markov chains and the two-step increment model. Inheritance scalars were assigned as constants ($h = 1$, for autosomal loci). The IM program was run under the changing population size model that allows changes in the population size of the descendent populations. Sequence divergence from *D. pseudoobscura* in the studied regions (7.4×10^{-9} changes per site per year assuming a divergence time of 8.1 Myr; Ramos-Onsins *et al.*, 1998) was used to infer the mutation rate per locus per year. Time estimates in mutation units (t) were converted to years ($T = t/\mu$) according to the geometric mean of the mutation rate per locus per year (μ).

Results

Nucleotide diversity and neutrality tests

Six genomic regions in segment I of the O chromosome (Fig. 1) were sequenced in 14 *D. madeirensis* lines. Information about the studied regions is available in supplementary Table S1 (Supporting Information). The concatenated data set including the six newly reported regions had 11 255 sites after excluding alignment gaps. A detailed description of nucleotide variation is shown in supplementary Fig. S1 (Supporting Information). Table 1 shows a summary of nucleotide variation estimates in the different regions of the O chromosome of *D. madeirensis*. Silent nucleotide diversity (π_{sil}) ranges from 0.006 at regions S25 and P21 to 0.018 at region P154. These estimates are within the range observed in the orthologous

Table 1 Nucleotide polymorphism and neutrality tests in *Drosophila madeirensis*.

	P2	P154	P22	S25	*rp49*	S1	P21	Concatenated†
n	14	14	14	14	22	14	14	14
Hd	1	1	1	0.989	0.996	0.978	0.989	1
# sites	1453	2159	1926	2142	1525	1174	2401	11 255
# silent sites	1299.17	787.23	1657.73	1879.93	1219.17	1082.40	1539.19	8245.64
S	46	69	61	58	59	38	49	321
Singletons	30	42	32	40	22	15	32	191
π_{total}	0.007	0.008	0.008	0.006	0.008	0.009	0.004	0.007
π_{sil}	0.008	0.018	0.009	0.006	0.010	0.010	0.006	0.009
θ_{sil}	0.012	0.023	0.011	0.009	0.014	0.011	0.009	0.012
π_{syn}	0.016	0.023	0.011	0.026	0.013	0.054	0.009	0.019
π_a	0.000	0.002	0.001	0.010	0.000	0.000	0.001	0.002
$\pi_{noncoding}$	0.008	0.012	0.009	0.005	0.010	0.009	0.006	0.008
Tajima's *D*	−1.194	−0.994	−0.914	−1.244	−0.955	−0.421	−1.488	−1.060
P-value	*0.122*	*0.158*	*0.166*	*0.096*	*0.174*	*0.362*	*0.058*	*0.131*
Tajima's $D_{noncoding}$	−1.271	−1.905**	−0.888	−1.685*	−1.006	−0.537	−1.596*	−1.329
P-value	*0.093*	*0.010*	*0.192*	*0.030*	*0.152*	*0.329*	*0.046*	*0.074*
Fu and Li's *D*‡	−3.008**	−1.256	−0.852	−2.234**	−0.336	−0.420	−1.947*	−1.716
P-value	*0.003*	*0.172*	*0.215*	*0.007*	*0.415*	*0.337*	*0.048*	*0.055*
Fu and Li's *F*‡	−2.986**	−1.438	−1.053	−2.573**	−0.605	−0.497	−2.202*	−1.834
P-value	*0.002*	*0.142*	*0.222*	*0.010*	*0.322*	*0.364*	*0.038*	*0.072*

n, Number of DNA sequences in the sample; Hd, haplotype diversity; # sites, number of sites after excluding alignment gaps; S, number of polymorphic sites; π_{total}, nucleotide diversity in all sites; π_{sil}, nucleotide diversity in silent sites (noncoding and synonymous sites); θ_{sil}, heterozygosity per silent site based on the number of silent segregating sites; π_{syn}, nucleotide diversity at synonymous sites; π_a, nucleotide diversity at nonsynonymous sites; $\pi_{noncoding}$, nucleotide diversity at noncoding sites.

*, $0.01 < P < 0.05$; **, $P \leq 0.01$ (two-tailed test).

†*rp49* was not included in the concatenated data set as this region was sequenced in a different set of *D. madeirensis* lines.

‡Fu and Li tests based on the *D* and *F* statistics were performed using *D. pseudoobscura* as the outgroup.

Table 2 Silent nucleotide polymorphism (π_{sil}) in *Drosophila madeirensis* and *Drosophila subobscura*.

	D. madeirensis	*D. subobscura* (O_{ST})	*D. subobscura* (O_{3+4})	*D. subobscura* (O_{ST} + O_{3+4})
P2	0.0081	0.0069	0.0088	0.0125
P154	0.0179	0.0129	0.0099	0.0235
P22	0.0091	0.0052	0.0129	0.0139
S25	0.0063	0.0073	0.0100	0.0118
S1	0.0100	0.0041	0.0081	0.0108
P21	0.0063	0.0061	0.0099	0.0125
Concatenated*	0.0087	0.0068	0.0100	0.0134
rp49	0.0101	0.0080	0.0101	0.0139

*Concatenated data set including all regions, except *rp49* (see Table 1).

regions of *D. subobscura*, either when the O_{ST} and O_{3+4} arrangements are analysed independently or when they are jointly analysed (Table 2). Therefore, no strong reduction of nucleotide variability was observed in the endemic species *D. madeirensis* relative to *D. subobscura*.

When comparing different estimates of nucleotide variation in *D. madeirensis* (Table 1), silent nucleotide diversity (π_{sil}) was lower than silent heterozygosity (θ_{sil}) in all regions, which indicates an excess of variants segregating at low frequency in the species. Among silent sites, nucleotide diversity was always lower at noncoding than at synonymous sites. This result suggests a higher level of constraint at noncoding relative to synonymous sites, as previously reported in *Drosophila* (Andolfatto, 2005). The lowest estimates of nucleotide variation were detected at nonsynonymous sites, as expected by the action of purifying selection acting on amino acid replacements.

Neutrality test statistics based on the frequency spectrum (Tajima, 1989; Fu & Li, 1993) yielded negative values in all regions, indicating an excess of singleton polymorphisms in *D. madeirensis* (Table 1). Tajima's *D* test did not indicate a significant departure from neutral expectations in any region, except at S25 and P21 where the test was marginally significant ($0.05 < P < 0.1$). In these two regions, as well as at P154, the test was significant ($P < 0.05$) when considering only noncoding sites. Nevertheless, Fu and Li's tests based on the *D* or *F* statistics were significant at P2, S25 and P21, using *D. pseudoobscura* (Table 1) or *D. subobscura* (results not shown), as the outgroup. Multilocus Tajima's *D* and Fu and Li's *D* tests (two-tailed tests) showed that the average value of both statistics across regions differed significantly from zero ($\bar{D}_{\text{Tajima}} = -1.0075$, $P = 0.002$; $\bar{D}_{\text{Fu \& Li}} = -1.4288$, $P = 0.002$). These results indicate an overall highly significant excess of singletons in *D. madeirensis*. Indeed, the performed single region tests are rather conservative because the confidence intervals of the test statistics were inferred under the assumption of no recombination.

HKA tests (Hudson *et al.*, 1987) based on silent variation were performed between pairs of regions using *D. subobscura* O_{ST}, *D. subobscura* O_{3+4} or *D. pseudoobscura* as the outgroup. None of the tests yielded a significant result. Therefore, no decoupling between the level of divergence between species and the level of polymorphism within *D. madeirensis* was detected in any region (results not shown). In addition, a multilocus HKA test based on variation in *D. madeirensis* and using *D. pseudoobscura* as the outgroup yielded also a nonsignificant result ($X^2 = 3.47$; d.f. = 6, $P = 0.74$).

Selection coefficient estimates

Interspecific differences in population size might affect the effectiveness of selection acting on new mutations. Estimates of the scaled selection coefficients (γ) in the coding regions are summarized in Table 3. Purifying selection against nonsynonymous mutations is much stronger in *D. subobscura* than in *D. madeirensis*. Indeed, γ estimates for nonsynonymous mutations differ significantly from zero in both the O_{ST} and O_{3+4} arrangements of *D. subobscura*, but not in *D. madeirensis*. In addition, Tajima's *D* statistic for nonsynonymous variants is negative in both species but only significant in *D. subobscura*. These results would indicate that purifying selection is more effective in eliminating nonsynonymous mutations in *D. subobscura* than in *D. madeirensis*. The distribution of fitness effects of deleterious nonsynonymous mutations also supports differences in the strength of purifying selection between species (Fig. 2). *Drosophila madeirensis* presents a higher fraction of nearly neutral and weakly selected mutations and a smaller fraction of mutations under strong purifying selection than *D. subobscura* (O_{3+4}). In addition, the fraction of adaptive nonsynonymous substitutions is lower in *D. madeirensis* than in *D. subobscura* (O_{3+4}) ($\alpha = 41\%$ and $\alpha = 59\%$, respectively). These results are also consistent with a less effective selection and thus with a lower effective size of *D. madeirensis* relative to *D. subobscura*. However, the use of synonymous sites as the neutral class in the maximum-likelihood approach (Eyre-Walker & Keightley, 2009) to infer both the distribution of fitness effects of deleterious nonsynonymous mutations and the fraction of adaptive nonsynonymous substitutions might be questioned. Indeed, γ estimates in Table 3 indicate purifying selection on synonymous mutations in *D. subobscura* (O_{3+4}), which can be related to the maintenance of codon bias as the mutations from preferred to

Table 3 Tajima's *D* and scaled selection coefficient (γ) estimates according to the frequency distribution (fd) method.

Lineage	Polymorphic changes†	Number of polymorphic sites	Tajima's *D*‡	γ	C.I. 95%§	*P* (γ = 0)
Drosophila madeirensis	Synonymous	53	−0.03	−0.336	±1.77	0.732
	P → U	23	−0.36	−1.119	±2.15	0.383
	Nonsynonymous	23	−0.79	−1.834	±1.85	0.087
Drosophila subobscura (O_{3+4})	Synonymous	73	−0.85	−2.225	±1.15	5.4E-04
	P → U	40	−1.03	−2.713	±1.66	1.4E-03
	Nonsynonymous	28	−1.52*	−4.584	±3.13	1.3E-04
D. subobscura (O_{ST})	Synonymous	43	−0.04	−0.199	±2.05	0.856
	P → U	17	−0.21	−0.749	±2.75	0.641
	Nonsynonymous	20	−1.81*	−10.8	±10.55	3.2E-06

†P → U, polymorphic changes from preferred to unpreferred codons inferred according to the *Drosophila melanogaster* codon preferences table (Akashi, 1995). Only polymorphic synonymous changes that could be unambiguously polarized using the sequence of *Drosophila pseudoobscura* as the outgroup were considered.
‡*$P < 0.05$.
§Confidence interval at the 95% level.

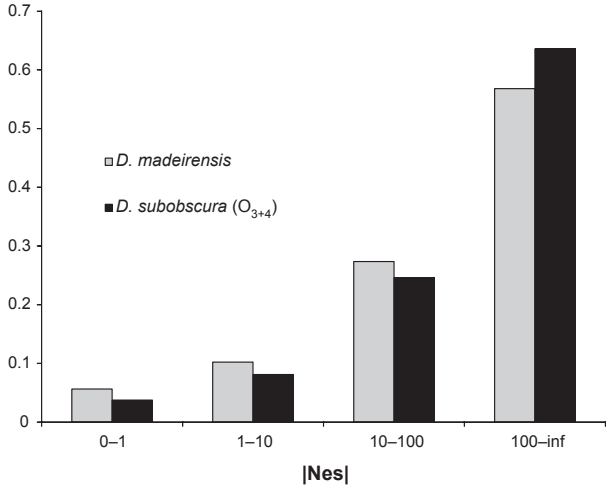

Fig. 2 Distribution of fitness effects of nonsynonymous deleterious mutations in *Drosophila madeirensis* and *Drosophila subobscura* (O_{3+4}).

Table 4 Genetic differentiation and gene flow between *Drosophila subobscura* (O_{ST} or O_{3+4}) and *Drosophila madeirensis*.

Gene region	Data sets	D_{XY}	D_A	*Nm*	F_{ST}
P2*	O_{ST} – *D. madeirensis*	0.0173	0.0094	0.17	0.60
	O_{3+4} – *D. madeirensis*	0.0206	0.0119	0.15	0.62
P154	O_{ST} – *D. madeirensis*	0.0153	0.0087	0.19	0.57
	O_{3+4} – *D. madeirensis*	0.0195	0.0129	0.13	0.66
P22	O_{ST} – *D. madeirensis*	0.0149	0.0085	0.19	0.57
	O_{3+4} – *D. madeirensis*	0.0202	0.0110	0.21	0.54
S25	O_{ST} – *D. madeirensis*	0.0178	0.0116	0.13	0.65
	O_{3+4} – *D. madeirensis*	0.0200	0.0119	0.17	0.59
rp49	O_{ST} – *D. madeirensis*	0.0130	0.0062	0.28	0.47
	O_{3+4} – *D. madeirensis*	0.0169	0.0093	0.21	0.55
S1	O_{ST} – *D. madeirensis*	0.0139	0.0076	0.21	0.55
	O_{3+4} – *D. madeirensis*	0.0158	0.0078	0.26	0.49
P21	O_{ST} – *D. madeirensis*	0.0144	0.0104	0.10	0.72
	O_{3+4} – *D. madeirensis*	0.0161	0.0109	0.12	0.80
Concatenated†	O_{ST} – *D. madeirensis*	0.0152	0.0095	0.15	0.63
	O_{3+4} – *D. madeirensis*	0.0183	0.0112	0.16	0.61

*Lines M27 and M54 of *D. madeirensis* were excluded from the analysis of the P2 region because they lacked the sequences from nucleotide 1900 to 2720 (M27) and from 1680 to 2720 (M54) available in *D. subobscura* and the other *D. madeirensis* lines.
†Concatenated data set including all regions, except *rp49* (see Table 1).

unpreferred codons present the same pattern (Akashi, 1995). On the other hand, it is interesting to note that the distribution of fitness effects differs substantially in both arrangements of *D. subobscura*. Indeed, 98% of mutations fall into the class $10 < |N_e s| < 100$ in *D. subobscura* (O_{ST}).

Genetic differentiation

The O_{ST} and O_{3+4} arrangements of *D. subobscura* are genetically differentiated ($P < 0.001$) in the different gene regions distributed along the O_3 inversion here studied (Munté *et al.*, 2005). The permutation test (Hudson *et al.*, 1992a) also indicated a significant genetic differentiation between *D. madeirensis* and *D. subobscura* (either O_{ST} or O_{3+4}) when regions are independently analysed and for the concatenated data set (K_S* = 4.11 and K_S* = 4.33 in the *D. madeirensis* vs. O_{ST} and *D. madeirensis* vs. O_{3+4} comparison, respectively; $P < 0.001$ in both cases). The D_{XY} and D_A estimates were used to infer the level of genetic differentiation between species (Table 4). In each region and in the concatenated data set, the level of genetic differentiation between *D. madeirensis* and O_{3+4} was higher than between *D. madeirensis* and O_{ST}.

As expected from the significant genetic differentiation between species, estimates of gene flow (*Nm*) were rather

low (Table 4). The observed number of shared polymorphisms can be explained by recurrent mutation in most genomic regions and in the concatenated data set (Table 5), which is also consistent with restricted gene flow between species. Only at the P154 (O_{ST} – *D. madeirensis*) and *rp49* (O_{3+4} – *D. madeirensis*) regions, the number of shared polymorphisms was slightly higher than expected. Indeed, the maximum number of shared polymorphisms expected from recurrent mutation was three at P154 and four at *rp49*, which is lower than the four and five shared polymorphisms observed in each comparison, respectively. In addition, no introgressed haplotypes between *D. subobscura* and *D. madeirensis* were detected in five of the seven regions studied, and in the remaining two regions only very short tracts were identified (17 nt long at P154 and 15 nt long at P22). Therefore, it can be inferred that in the regions studied, neither introgression nor the presence of ancestral polymorphisms have greatly contributed to increase the number of shared polymorphisms between species.

Gene genealogy

The gene genealogy of the *D. subobscura* and *D. madeirensis* lines was inferred from total variation in the concatenated data set. Fig. 3 shows the neighbour-joining tree reconstructed using *D. pseudoobscura* as the outgroup. Lines group with a high bootstrap support in three well-defined clusters: *D. madeirensis*, *D. subobscura* O_{ST} and *D. subobscura* O_{3+4}. This result is consistent with the significant genetic differentiation detected between *D. madeirensis* and the *D. subobscura* O_{ST} and O_{3+4} arrangements. It is also consistent with a restricted genetic exchange between the O_{ST} and O_{3+4} arrangements of *D. subobscura*. However, the branching order of the three main clusters detected (*D. madeirensis*, O_{ST} and O_{3+4}) is not well resolved in the inferred genealogy as the bootstrap support of the corresponding nodes is low.

Gene genealogies reconstructed from each genomic region independently also support the presence of the three main clusters with rather high bootstrap support (Fig. S2; Supporting Information). However, the branching pattern of the *D. madeirensis*, O_{ST} and O_{3+4} clusters differs from that inferred from the concatenated data set in some cases. Indeed, only *rp49* and P22 rendered the same topology as in Fig. 2 relative to the three main clusters. According to nucleotide variation at S1 and S25, the *D. madeirensis* lines cluster with the O_{3+4} lines. In contrast, the O_{ST} and O_{3+4} *D. subobscura* lines form a monophyletic group in the gene genealogies reconstructed from variation at P21, P2 and P154. Nevertheless, bootstrap values supporting the branching order of the three main clusters are also low in the gene genealogies based on individual genome regions.

Speciation model tests

The ψ-test statistic (Wakeley, 1996b) was used to contrast whether the data were consistent with the isolation model of speciation, in which an ancestral population splits into two descendent populations without migration. The isolation model was not rejected for any of the

Table 5 Population parameter estimates according to the isolation model (Wakeley & Hey, 1997).

Gene region	Data sets	S_F	S_s†	S_{X1}	S_{X2}	ψ	*P*-value	θ_1	θ_2	θ_A	τ
P2‡	O_{ST} – *Drosophila madeirensis*	7	3 ns	35	58	0.27	0.997	11.93	18.40	17.23	10.29
	O_{3+4} – *D. madeirensis*	7	4 ns	50	51	0.25	1.000	17.26	15.17	18.94	10.94
P154	O_{ST} – *D. madeirensis*	9	4*	36	67	0.28	0.987	13.09	20.20	18.91	11.29
	O_{3+4} – *D. madeirensis*	18	4 ns	51	67	0.25	0.998	21.15	18.59	23.22	13.41
P22	O_{ST} – *D. madeirensis*	8	1 ns	37	54	0.23	1.000	11.29	17.41	16.30	9.74
	O_{3+4} – *D. madeirensis*	10	2 ns	66	51	0.23	1.000	19.49	17.13	21.39	12.35
S25	O_{ST} – *D. madeirensis*	17	1 ns	59	49	0.23	1.000	14.79	22.81	21.35	12.75
	O_{3+4} – *D. madeirensis*	13	2 ns	92	58	0.24	0.999	24.93	21.91	27.36	15.80
rp49	O_{ST} – *D. madeirensis*	1	4 ns	42	52	0.26	0.999	10.25	15.81	14.80	8.84
	O_{3+4} – *D. madeirensis*	7	5*	50	47	0.24	1.000	15.04	13.22	16.51	9.53
S1	O_{ST} – *D. madeirensis*	5	2 ns	14	36	0.27	0.999	6.43	9.92	9.29	5.55
	O_{3+4} – *D. madeirensis*	2	3	28	35	0.28	0.997	10.27	9.03	11.28	6.51
P21	O_{ST} – *D. madeirensis*	19	0 ns	30	48	0.23	1.000	10.95	16.89	15.81	9.44
	O_{3+4} – *D. madeirensis*	19	1 ns	58	45	0.19	1.000	18.58	16.33	20.40	11.78
Concatenated§	O_{ST} – *D. madeirensis*	63	9 ns	195	296	0.11	1.000	78.72	121.45	113.69	67.91
	O_{3+4} – *D. madeirensis*	67	13 ns	328	293	0.10	1.000	126.73	111.38	139.09	80.33

S_F, fixed differences; S_S, shared polymorphic sites; S_{X1}, exclusive polymorphic sites in *D. madeirensis*; S_{X2}, exclusive polymorphic sites in O_{ST} or O_{3+4}; ψ statistic (Wakeley, 1996a,b); *P*-value, $P[\psi \geq \psi_{observed}]$ after 1000 simulations. θ_1 (*Drosophila subobscura* O_{ST} or O_{3+4}), θ_2 (*D. madeirensis*), θ_A (ancestral population) and τ estimates obtained according to the isolation model with the WH program.

†Recurrent mutation test based on the hypergeometric distribution. *$P < 0.05$; ns, not significant.

‡Lines M27 and M54 of *D. madeirensis* were excluded from the analysis in this region (see Table 2).

§Concatenated data set including all regions, except *rp49* (see Table 1).

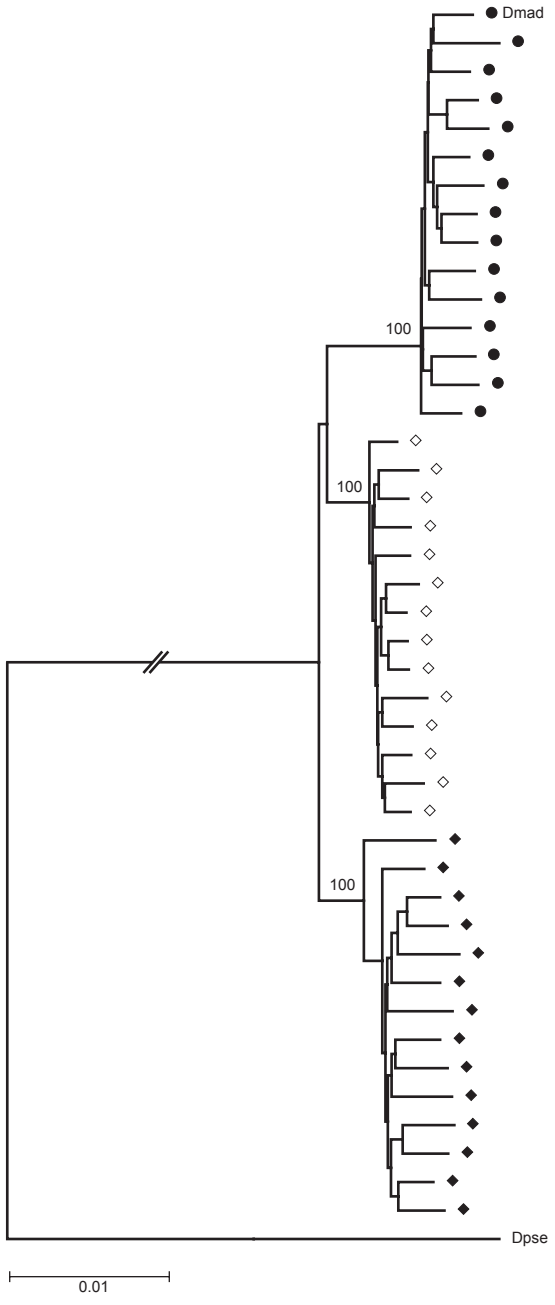

Fig. 3 Neighbour-joining tree reconstructed from total nucleotide variation in the concatenated data set of the six genomic regions studied with the complete deletion option. Bootstrap values were obtained after 1000 replicates and values higher than 95% are shown. Distances were obtained with the Jukes & Cantor (1969) correction and assuming uniform rates of substitution among sites. Open and black diamonds refer to O_{ST} and O_{3+4} *Drosophila subobscura* lines, respectively, and black circles to *Drosophila madeirensis* lines. Dpse refers to the *Drosophila pseudoobscura* line used as the outgroup and Dmad to a *D. madeirensis* line previously sequenced (Munté *et al.*, 2005). The scale bar represents the corrected number of substitutions per site.

Table 6 Multilocus analysis according to the isolation model.

Data set	*wh*	*P*-value	χ^2	*P*-value
O_{ST} – *Drosophila madeirensis*	22	0.239	42.0	0.344
O_{3+4} – *D. madeirensis*	21	0.523	32.4	0.654

wh, Wang *et al.* (1997) multilocus test statistic. χ^2, chi-square multilocus test statistic (Kliman *et al.*, 2000).

studied regions or for the concatenated data set (Table 5). The multilocus tests based on both the *wh* (Wang *et al.*, 1997) and the χ^2 (Kliman *et al.*, 2000) test statistics also failed to reject the isolation model of speciation (Table 6). The population parameters θ_1, θ_2, θ_A and τ were estimated as proposed by Wakeley & Hey (1997) under the isolation model in comparisons between *D. madeirensis* and *D. subobscura* (Table 5). Estimates of θ ($\theta = 4N\mu$) in each region and in the concatenated data set are lower in *D. madeirensis* than in the *D. subobscura* O_{3+4} arrangement but higher than in O_{ST}. Assuming a similar mutation rate across species, this result would indicate that *D. madeirensis* has an effective population size intermediate between those of O_{3+4} and O_{ST}. Estimates of τ in Table 5 would indicate that the divergence time between *D. madeirensis* (with the ancestral O_3 arrangement) and O_{3+4} is higher than between *D. madeirensis* and O_{ST}.

The isolation model without migration was also supported by the Markov chain Monte Carlo approach implemented in the IM program. Indeed, the marginal posterior probability distribution of the migration rate revealed a peak at the lower limit of resolution in comparisons between *D. madeirensis* and *D. subobscura* (O_{ST} or O_{3+4}). Maximum-likelihood estimates of different population parameters inferred by the IM program are summarized in Table 7. The ratio θ_1/θ_2 in both comparisons is quite similar, although slightly higher, to that inferred from the concatenated data set in Table 5 (0.65 vs. 0.71 for *D. madeirensis* – O_{ST} and 1.14 vs. 1.37 for *D. madeirensis* – O_{3+4}), which also supports that the effective size of *D. madeirensis* is lower relative to O_{3+4} but higher relative to O_{ST}. However, maximum-likelihood estimates of θ_A relative to either θ_1 or θ_2 are much lower than estimates in Table 5, which may be a consequence of the different assumptions in the IM changing population size model (as implemented in the IM program) and the isolation model (WH program). The IM changing population size model was also used to obtain maximum-likelihood estimates of the elapsed time since the split of *D. madeirensis* and either O_{ST} or O_{3+4} (Table 7). Although the marginal posterior probability distribution of the *t* parameter is partly overlapping in both comparisons, the results also suggest that the divergence time between *D. madeirensis* (with the ancestral O_3 arrangement) and O_{3+4} is higher than between *D. madeirensis* and O_{ST}.

Table 7 Maximum-likelihood estimates according to the IM changing population size model.

Comparison	θ_1	θ_2	θ_1/θ_2	θ_A	t	t (years)*
O_{ST} – *Drosophila madeirensis*						
Maximum-likelihood estimates	27.30	38.59	0.71	0.03	2.88	742 446
90% HPD†	16.46–56.21	23.61–60.22		0.03–0.72	2.26–3.51	582 614–904 856
O_{3+4} – *D. madeirensis*						
Maximum-likelihood estimates	66.33	48.52	1.37	0.03	3.19	764 029
90% HPD†	39.96–67.40	29.77–67.40		0.03–0.64	2.55–3.90	610 744–934 079

θ_1 (*Drosophila subobscura* O_{ST} or O_{3+4}), θ_2 (*D. madeirensis*), θ_A (ancestral population) and t estimates obtained according to the IM program.
*t (years) were estimated assuming a mutation rate per locus per year of 3.88×10^{-6} (O_{ST} – *D. madeirensis*) and 4.17×10^{-6} (O_{3+4} – *D. madeirensis*).
‡HPD, highest posterior density interval.

Discussion

Natural populations of *D. madeirensis* are restricted to the Laurisilva forest of the Madeira Island. Island-endemic species are expected to have a low effective population size and thus reduced genome-wide variation as compared to abundant species with a wide distribution range. However, levels of nucleotide diversity in *D. madeirensis* (π_{sil} = 0.0087 for the concatenated data) are within the range than those in *D. subobscura*, not only for the genomic regions here analysed (π_{sil} = 0.0080 in O_{ST} and π_{sil} = 0.0100 in O_{3+4}, Table 2), but also for different *D. subobscura* X-linked regions (π_{sil} = 0.0072; Nóbrega *et al.*, 2008). The first intron of the X-linked *RpII215* gene also presents similar levels of nucleotide variation in both species (Llopart & Aguadé, 2000; Khadem, unpublished data). In addition, silent nucleotide variation in *D. madeirensis* is only slightly lower than in the continental species *D. pseudoobscura* (π_{sil} = 0.0105 for different genes of the second chromosome that is not affected by polymorphic inversions; Hamblin & Aquadro, 1999). When compared to other endemic species, silent nucleotide diversity in *D. madeirensis* is much higher than in *D. sechellia* (π_{sil} = 0.0011; Legrand *et al.*, 2009) and also higher than in the single gene so far analysed of *Drosophila guanche* (π_{sil} = 0.0059; Pérez *et al.*, 2003), an endemic species of the Canary Islands and also closely related to *D. subobscura*. Therefore, it would seem that the historical effective size of *D. madeirensis* has not been extremely low, although the highly significant excess of singletons indicates that variation is not at mutation–drift equilibrium.

According to the nearly neutral model of molecular evolution, slightly deleterious mutations may segregate in populations at low frequencies when the effective population size is low (Ohta, 1992). In small populations, selection can be counteracted by genetic drift causing a longer sojourn time in the population of mildly deleterious mutations. The scaled selection coefficients estimated for nonsynonymous mutations and Tajima's D tests applied to this kind of mutations indicate that purifying selection is less effective in *D. madeirensis* than in *D. subobscura* (Table 3). In addition, the distribution of fitness effects of nonsynonymous mutations is slightly shifted towards more weakly selected mutations (i.e. $0 < |N_e s| < 100$) in *D. madeirensis* relative to *D. subobscura* (O_{3+4}). Selection against nonsynonymous mutations

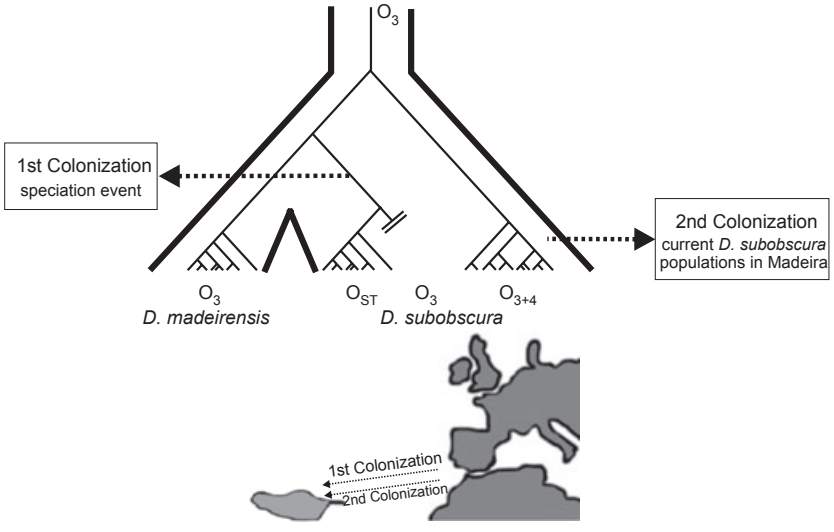

Fig. 4 Schematic representation of relationships among *Drosophila subobscura* chromosomal arrangements and *Drosophila madeirensis*, and the geographical setting of the two colonization events of Madeira by *D. subobscura*. Thin lines represent gene genealogies; thick lines, species boundaries; and dotted arrows, colonization events. The Madeira Island is not drawn to scale.

would seem to have been less effective in *D. madeirensis*, which would be consistent with a stronger effect of genetic drift in this species and thus with the expected difference in effective population size between *D. madeirensis* and *D. subobscura*.

The closely related species *D. madeirensis* and *D. subobscura* are not completely isolated as viable and fertile hybrids are recovered in the laboratory (Krimbas & Loukas, 1984). Although reproductive isolation between species in the laboratory may differ from that in their natural habitats (Counterman & Noor, 2006), hybridization between *D. madeirensis* and *D. subobscura* might be likely in nature at present. However, the present multilocus approach indicates that nucleotide variation in the genomic regions analysed is consistent with the speciation model of isolation without migration, and confirms that the studied regions are especially suitable to trace the history of these species. *Drosophila madeirensis* likely arose in allopatry by a speciation process in Madeira after ancestral O_3 *D. subobscura* populations colonized the island (Fig. 4). Current populations of *D. subobscura* in Madeira with a high frequency of the O_{3+4} arrangement would thus be the result of a second colonization event (Khadem *et al.*, 2001). O_{3+4} appears as a basal cluster in the inferred gene genealogy, although with a very low bootstrap support. Indeed, O_{3+4} is a rather old arrangement according to its standing level of variation (Rozas & Aguadé, 1994; Navarro-Sabaté *et al.*, 1999; Rozas *et al.*, 1999). O_{3+4} might have its origin in Iran or Asia Minor, where it currently reaches its highest frequency (Krimbas & Powell, 1992). After its origin, the new arrangement would have expanded westwards into Europe, where its frequency is lowest in the north. A putative scenario would thus be that O_{3+4} was not present either in Europe or North Africa when the ancestral O_3 *D. subobscura* population colonized Madeira. It is even likely that the origin of O_{ST}, which according to the previously estimated ages of both arrangements (Rozas & Aguadé, 1994; Navarro-Sabaté *et al.*, 1999; Rozas *et al.*, 1999) would be younger than O_{3+4}, occurred in Europe after the first colonization of the Madeira Island.

Reproductive isolation between *D. madeirensis* and *D. subobscura* would have arisen before the second colonization event and likely reinforced when both species coexisted. The lack of introgressed haplotypes in *D. madeirensis* from *D. subobscura* supports that the gene pool of both species has evolved independently in the studied regions. However, interspecific hybridization between *D. madeirensis* and *D. subobscura* in nature after the second colonization event cannot be completely discarded according to the present results. Putative interspecific hybrids would be O_3/O_{3+4} heterokaryotypes. Therefore, the presence of inversion 4 in hybrids would have prevented genetic exchange in the studied regions. Only the analysis of variation at nuclear regions not affected by inversions and thus collinear between *D. madeirensis* and *D. subobscura* might contribute to elucidate the extent of introgression between both species in Madeira. Also, the study of variation at the mtDNA genome might be informative, as in the *D. yakuba* species group, a more extensive gene flow in mitochondrial than in nuclear genes has been detected (Llopart *et al.*, 2005; Bachtrog *et al.*, 2006). However, preliminary data on mtDNA variation in *D. madeirensis* (Khadem, unpublished results) do not support introgression from *D. subobscura* and indicate genetic differentiation between both species at the organelle genome, as previously reported in phylogenetic studies (Barrio *et al.*, 1994; Gleason *et al.*, 1997; Brehm *et al.*, 2001).

Drosophila madeirensis inhabits the Laurisilva forest that, although formerly covered much of Madeira, at present extends over an area of 150 km^2, which represents 20% of the island surface. Therefore, the natural habitat of the species has been progressively reduced in the recent past by human pressure. Average heterozygosities are lower in species with a high extinction risk than in their unthreatened relatives (Spielman *et al.*, 2004). However, present data indicate that the endemic species *D. madeirensis* harbours a rather high level of nucleotide polymorphism and a pattern of variation characterized by a significant excess of low frequency variants. Therefore, evolutionary factors that reduce variation (i.e. genetic drift and inbreeding) have not had until now a strong effect in wearing away the genetic variability of the species. However, ecological factors, such as global warming and the risk of damage of the Laurisilva forest, might threaten the survival of *D. madeirensis*.

Acknowledgments

We thank J. Rozas for critical comments on a previous version of the manuscript and the anonymous reviewers for their constructive comments. We also thank Serveis Científico-Tècnics, Universitat de Barcelona, for automated DNA sequencing facilities. This work was supported by grant POCTI/BSE/43097/2001 by Foundation for Science and Technology (FCT), Portugal, to M.K. and grants BFU2004-2253, BFU2007-63228 from Comisión Interdepartamental de Ciencia y Tecnología, Spain, and 2005SGR-166, 2009SGR-1287 from Comissió Interdepartamental de Recerca i Innovació Tecnològica, Catalonia, Spain, to M.A.

References

Akashi, H. 1995. Inferring weak selection from patterns of polymorphism and divergence at ''silent'' sites in *Drosophila* DNA. *Genetics* **139**: 1067–1076.

Akashi, H. & Schaeffer, S.W. 1997. Natural selection and the frequency distributions of ''silent'' DNA polymorphism in *Drosophila*. *Genetics* **146**: 295–307.

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.

Ashburner, M. 1989. *Drosophila, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bachtrog, D., Thornton, K., Clark, A. & Andolfatto, P. 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution* **60**: 292–302.

Balanyà, J., Solé, E., Oller, J.M., Sperlich, D. & Serra, L. 2004. Long-term changes in the chromosomal inversion polymorphism of Drosophila subobscura. II. European populations. *J. Zool. Syst. Evol. Research* **42**: 191–201.

Barrio, E., Latorre, A. & Moya, A. 1994. Phylogeny of the *Drosophila obscura* species group deduced from mitochondrial DNA sequences. *J. Mol. Evol.* **39**: 478–488.

Begun, D.J., Holloway, A.K., Stevens, K., Hillier, L.W., Poh, Y.-P., Hahn, M.W. *et al.* 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila similans*. *PLoS Biol.* **5**: 2534–2559.

Betran, E., Rozas, J., Navarro, A. & Barbadilla, A. 1997. The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* **146**: 89–99.

Brehm, A., Harris, D.J., Hernandez, M., Cabrera, V.M., Larruga, J.M., Pinto, F.M. *et al.* 2001. Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup. *Insect Mol. Biol.* **10**: 573–578.

Brown, K.M., Burk, L.M., Henagan, L.M. & Noor, M.A.F. 2004. A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. *Evolution* **58**: 1856–1860.

Burland, T.G. 2000. DNASTAR's Lasergene sequence analysis software. *Methods Mol. Biol.* **132**: 71–91.

Counterman, B.A. & Noor, M.A.F. 2006. Multilocus test for introgression between the cactophilic species *Drosophila mojavensis* and *Drosophila arizonae*. *Am. Nat.* **168**: 682–696.

Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Sunderland, MA.

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

Eyre-Walker, A. & Keightley, P.D. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* **26**: 2097–2108.

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

Galopin de Carvalho, A. & Brandão, J. 1991. *Geologia do Archipélago da Madeira*. Museu Nacional de História Natural (Mineralogia e Geologia) da Universidade de Lisboa, Lisbon.

Gleason, J.M., Caccone, A., Moriyama, E.N., White, K.P. & Powell, J.R. 1997. Mitochondrial DNA phylogenies for the *Drosophila obscura* group. *Evolution* **51**: 433–440.

Hamblin, M.T. & Aquadro, C.F. 1999. DNA sequence variation and the recombinational landscape in *Drosophila pseudoobscura*: a study of the second chromosome. *Genetics* **153**: 859–869.

Hartl, D.L., Moriyama, E.N. & Sawyer, S.A. 1994. Selection intensity for codon bias. *Genetics* **138**: 227–234.

Hasegawa, M., Kishino, H. & Yano, T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**: 160–174.

Hey, J. & Nielsen, R. 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.

Hudson, R.R. & Kaplan, N.L. 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., Kreitman, M. & Aguadé, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.

Hudson, R.R., Boos, D.D. & Kaplan, N.L. 1992a. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**: 138–151.

Hudson, R.R., Slatkin, M. & Maddison, W.P. 1992b. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

Jukes, T.H. & Cantor, C.R. 1969. Evolution of protein molecules. In: *Mammalian Protein Metabolism* (H.M. Munro, ed.), pp. 21–120. Academic Press, New York, NY.

Khadem, M. & Krimbas, C.B. 1991. Studies of the species barrier between *Drosophila subobscura* and *D. madeirensis.* I. The genetics of male hybrid sterility. *Heredity* **67**: 157–165.

Khadem, M., Rozas, J., Segarra, C. & Aguadé, M. 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.

Khadem, M., Camacho, R. & Nóbrega, C. 2011. Studies of the species barrier between *Drosophila subobscura* and *D. madeirensis* V: the importance of sex-linked inversion in preserving species identity. *J. Evol. Biol.* **24**: 1263–1273.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Kliman, R.M., Andolfatto, P., Coyne, J.A., Depaulis, F., Kreitman, M., Berry, A.J. *et al.* 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* **156**: 1913–1931.

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

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

Krimbas, C.B. & Loukas, M. 1984. Evolution of the *obscura* group *Drosophila* species. I. Salivary chromosomes and quantitative characters in Drosophila subobscura and two closely related species. *Heredity* **53**: 469–482.

Krimbas, C.B. & Powell, J.R. 1992. *Drosophila Inversion Polymorphism*. CRC Press, Inc., Boca Raton, FL.

Larruga, J.M., Cabrera, V.M., González, A.M. & Gullón, A. 1983. Molecular and chromosomal polymorphism in continental and insular populations from the southwestern range of *Drosophila subobscura*. *Genetica* **60**: 191–205.

Legrand, D., Tenaillon, M.I., Matyot, P., Gerlach, J., Lachaise, D. & Caroiu, M.-L. 2009. Species-wide genetic variation and demographic history of *Drosophila sechellia*, a species lacking population structure. *Genetics* **182**: 1197–1206.

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

Llopart, A. & Aguadé, M. 2000. Nucleotide polymorphism at the *RpII215* gene in *Drosophila subobscura*: weak selection on synonymous mutations. *Genetics* **155**: 1245–1252.

Llopart, A., Lachaise, D. & Coyne, J.A. 2005. Multilocus analysis of introgression between two sympatric sister species of Drosophila: *Drosophila yakuba* and *D. santomea*. *Genetics* **171**: 197–210.

Machado, C.A., Kliman, R.M., Markert, J.A. & Hey, J. 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. & Maddison, W.P. 2000. *MacClade 4: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, MA.

Mayr, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York, NY.

McDermott, S.R. & Kliman, R.M. 2008. Estimation of isolation times of the island species in the *Drosophila simulans* complex from multilocus DNA sequence data. *PLoS ONE* **3**: e2442.

Munté, A., Rozas, J., Aguadé, M. & Segarra, C. 2005. Chromosomal inversion polymorphism leads to extensive genetic structure: a multilocus survey in *Drosophila subobscura*. *Genetics* **169**: 1573–1581.

Navarro, A. & Barton, N.H. 2003. Chromosomal speciation and molecular divergence – accelerated evolution in rearranged chromosomes. *Science* **300**: 321–324.

Navarro-Sabaté, A., Aguadé, M. & Segarra, C. 1999. The relationship between allozyme and chromosomal polymorphism inferred from nucleotide variation at the *Acph-1* gene region of *Drosophila subobscura*. *Genetics* **153**: 871–889.

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

Nielsen, R. & Wakeley, J. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* **158**: 885–896.

Nóbrega, C., Khadem, M., Aguadé, M. & Segarra, C. 2008. Genetic exchange versus genetic differentiation in a medium-sized inversion of *Drosophila*: the A_2/A_{st} arrangements of *Drosophila subobscura*. *Mol. Biol. Evol.* **25**: 1534–1543.

Noor, M.A.F., Grams, K.L., Bertucci, L.A. & Reiland, J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* **98**: 12084–12088.

Noor, M.A.F., Garfield, D.A., Schaeffer, S.W. & Machado, C.A. 2007. Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics* **177**: 1417–1428.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Ann. Rev. Ecol. Syst.* **23**: 263–286.

Papaceit, M., San Antonio, J. & Prevosti, A. 1991. Genetic analysis of extra sex combs in the hybrids between *Drosophila subobscura* and *D. madeirensis*. *Genetica* **84**: 107–114.

Pérez, J.A., Munté, A., Rozas, J., Segarra, C. & Aguadé, M. 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.

Powell, J.R. 1997. *Progress and Prospects in Evolutionary Biology. The Drosophila Model*. Oxford University Press, New York, NY.

Prevosti, A. 1972. Chromosomal polymorphism in *Drosophila subobscura* populations from the Madeira Island. *Genét. Ibér.* **24**: 1–11.

Prevosti, A., Ribó, G., Serra, L., Aguadé, M., Balañà, J., Monclús, M. *et al.* 1988. Colonization of America by Drosophila subobscura: experiment in natural populations that supports the adaptive role of chromosomal inversion polymorphism. *Proc. Natl. Acad. Sci. USA* **85**: 5597–5600.

Ramos-Onsins, S., Segarra, C., Rozas, J. & Aguadé, M. 1998. Molecular and chromosomal phylogeny in the *obscura* group of *Drosophila* inferred from sequences of the *rp49* gene region. *Mol. Phylogenet. Evol.* **9**: 33–41.

Rand, D.M., Fry, A. & Sheldahl, L. 2006. Nuclear-mitochondrial epistasis and *Drosophila* aging: introgression of *Drosophila simulans* mtDNA modifies longevity in *D. melanogaster* nuclear backgrounds. *Genetics* **172**: 329–341.

Rego, C., Santos, M. & Matos, M. 2007. Quantitative genetics of speciation: additive and non-additive genetic differentiation between *Drosophila madeirensis* and *Drosophila subobscura*. *Genetica* **131**: 167–174.

Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**: 351–358.

Rozas, J. & Aguadé, M. 1994. Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **91**: 11517–11521.

Rozas, J., Segarra, C., Zapata, C., Alvarez, G. & Aguadé, M. 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., Segarra, C., Ribó, G. & Aguadé, M. 1999. Molecular population genetics of the *rp49* gene region in different chromosomal inversions of *Drosophila subobscura*. *Genetics* **151**: 189–202.

Sackton, T.B., Kulathinal, R.J., Bergman, C.M., Quinlan, A.R., Dopman, E.B., Carneiro, M. *et al.* 2009. Population genomic inferences from sparse high-throughput sequencing of two populations of *Drosophila melanogaster*. *Genome Biol. and Evol.* **1**: 449–465.

Saitou, N. & Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.

Sawyer, S.A. & Hartl, D.L. 1992. Population genetics of polymorphism and divergence. *Genetics* **132**: 1161–1176.

Spielman, D., Brook, B.W. & Frankham, R. 2004. Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. USA* **101**: 15261–15264.

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

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**: 1596–1599.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.

Ting, C.T., Tsaur, S.C. & Wu, C.I. 2000. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. USA* **97**: 5313–5316.

Wakeley, J. 1996a. The variance of pairwise nucleotide differences in two populations with migration. *Theor. Popul. Biol.* **49**: 39–57.

Wakeley, J. 1996b. Distinguishing migration from isolation using the variance of pairwise differences. *Theor. Popul. Biol.* **49**: 369–386.

Wakeley, J. & Hey, J. 1997. Estimating ancestral population parameters. *Genetics* **145**: 847–855.

Wang, R.L., Wakeley, J. & Hey, J. 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.
White, M. 1978. *Modes of Speciation*. W.H. Freeman & Co, San Francisco, CA.
Wright, S. 1951. The genetical structure of populations. *Ann. Eugen.* **15**: 323–354.

Supporting information

Additional Supporting Information may be found in the online version of this article:
Table S1 *Drosophila madeirensis* O chromosome gene regions.
Figure S1 Polymorphic sites at the six newly reported genomic regions of *Drosophila madeirensis*.
Figure S2 Gene genealogies reconstructed from individual loci (symbols as in Figure 3).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Received 7 November 2011; revised 21 December 2011; accepted 3 January 2012